Health & Place 67 (2021) 102482

Contents lists available at ScienceDirect

Health and Place

journal homepage: http://www.elsevier.com/locate/healthplace

A dual-method approach toward measuring the built environment - sampling optimization, validity, and efficiency of using GIS and virtual auditing

Weimin Li [a,*], Patricia L. Winter [b], Lee-Anne Milburn [a], Pamela E. Padgett [b]

[a] California State Polytechnic University at Pomona, United States

[b] US Forest Service, Pacific Southwest Research Station, Riverside, California, United States

A B S T R A C T

In recent years, GIS and virtual auditing have been widely used to measure the built environment, and each method carries its strengths and weaknesses. To generate higher quality, more cost-effective, and less time-consuming measures, it is necessary to explore dual- or multi-method strategy toward sampling optimization, improvement of measurement, and enhancement of efficiency. To justify the proposed dual-method approach, the study has three major objectives. First, it examines the uncertainties associated with different sample sizes by using GIS to generate scenarios that contrast the validity of measurements to aid sampling optimization in auditing. Second, it compares the validity of GIS measures with those generated through Google Street View Auditing (GSVA) by human raters. Third, it further examines the efficiency of the proposed dual-method approach in comparison to the two individual methods.

Such investigation generates several novel findings. First, the study presents important evidence to support that GIS measures can offer sampling guidance applicable to the GSVA method. It leads to a recommendation of sampling sizes (5%–20%) for cases in settings with a mixture of affluent and disadvantaged neighborhoods. Results further indicate that different communities and certain individual features and characteristics may demand different sampling practices. Second, the study found that while GSVA is trustworthy for most characteristic variables, especially those that required subjective input, GIS provides well-validated measures for certain objective environmental attributes. Furthermore, the study reports that a dual-method approach of GIS and GSVA had a lower financial and time burden than using GSVA alone and is thus recommended as a comprehensive solution for optimal measurement of an objective built environment in mixed urban neighborhoods.

1. Introduction

Features and characteristics of the built environment have been intensively investigated for their impact on people's outdoor physical activities, recreational behaviors, and quality of life (Brownson et al., 2009; Handy et al., 2002; Humpel et al., 2002; Owen et al., 2004; Papas et al., 2007; Sallis, 2009). How to accurately and efficiently measure the built environment thus became a central concern in active living research (Brownson et al., 2009). Attention on objective measures generated by GIS and auditing have increased in the past two decades (Brownson et al., 2009; Frank et al., 2012; Hoehner et al., 2005; Kirtland et al., 2003; Pikora et al., 2002; Sallis, 2009). Each method bears their advantages and disadvantages.

The GIS method often uses archived data to measure the built environment. The reliability, validity, and comparability of GIS measures were reported as mixed in the past due to inaccurate and incomplete data, multiple data sources, inconsistent computation definitions and procedures, and lack of metadata (Brownson et al., 2009). On the one hand, some found that GIS measures present high validity on objective attributes such as land use, public amenities, recreational sites, schools, and commercial stores as well as composite variables such as walkability (Brownson et al., 2009; Hajna et al., 2013; Kirtland et al., 2003; Owen et al., 2004). On the other hand, others found that physical activity resources and facilities were poorly validated by GIS measures (Boone et al., 2008). These inconsistent findings indicated that the validity of GIS measures may vary given difference in the quality of data sources, attributes investigated, types of active living spaces under investigation, etc. Despite these limitations, GIS has continued to rise in popularity (Porter et al., 2004; Torio, 2012) and is often recommended when other methods are not feasible or affordable, for the increasingly improved quality of GIS data in the big data era (Li and Milburn, 2016) and their low financial and labor cost (Brownson et al., 2009).

Auditing uses human raters to examine the built environment through pre-defined protocols or tools. It provides reliable measures for many built environment attributes and is often used as a gold standard to assess the validity of other measures (Brownson et al.,

* Corresponding author.
E-mail addresses: wli@cpp.edu (W. Li), pat.winter@usda.gov (P.L. Winter), lsmilburn@cpp.edu (L.-A. Milburn), ppadgett1460@gmail.com (P.E. Padgett).

https://doi.org/10.1016/j.healthplace.2020.102482
Received 7 May 2020; Received in revised form 5 November 2020; Accepted 11 November 2020
Available online 29 December 2020

2009; Hajna et al., 2013; Hoedl et al., 2010; Hoehner et al., 2005). Comparatively, auditing methods are preferred for measuring physical features that are best assessed through direct observation (e.g., architectural character, landscape maintenance, etc.) or those not available in existing data sources (Brownson et al., 2009). While its high validity and reliability were reported, in-situ auditing is often costly and time-consuming, which has led to the increasing popularity of virtual auditing in recent years. The validity of virtual auditing using tools such as Google Street View, Google Earth, and Bing Map was found consistently moderate to high for land-use, transportation characteristics, walkability, recreational facilities, traffic safety, etc., in comparison to on-site auditing (Ben-Joseph et al., 2013; Charreire et al., 2014; Chiang et al., 2017; Clarke et al., 2010; Gullón et al., 2015; Kelly et al., 2012; Vanwolleghem et al., 2014; Wilson et al., 2012). However low validity was also reported for variables related to social presence and behaviors as well as qualitative environmental characteristics such as detailed street elements, physical decay and disorder, and perceived safety (Charreire et al., 2014). High inter-rater reliability of virtual auditing has been reported for land use, presence of street and building features, walkability, and safety. However, low reliability was also reported for natural features, traffic design, disorder and amenities, aesthetics, pedestrian access, etc. (Bader et al., 2015; Plascak et al., 2020; Rzotkiewicz et al., 2018). The reported cost of virtual auditing is a lot less than that of on-site auditing (Bader et al., 2015; Rzotkiewicz et al., 2018) but remains a burden (see 3.3).

The unique but complementary advantages carried by GIS and virtual auditing present a good possibility for a dual-method approach to capture the benefits of both to measure the built environment. To validate this idea, some critical gaps have to be addressed. These gaps include, first, to provide a valid sampling guidance for virtual auditing in different built environments to reduce high auditing costs. The current literature has started to address sampling issues through the optimization of sample size and sampling methods (Adu-Brimpong et al., 2017; McMillan et al., 2010; Moniruzzaman and Páez, 2012; Ussery et al., 2019). For example, Adu-Brimpong et al. used a scoring method for virtual neighborhood audits and assessed the representativeness of different sample sizes of street segments for walkability measurement in low-income micro-neighborhoods, defined by 12 streets immediately next to a household (Adu-Brimpong et al., 2017). Their results indicated that a threshold of 25% sampling of streets yielded walkability results comparable to higher sampling rates. Similar results were reported previously using on-site audit methods (Cerin et al., 2011; McMillan et al., 2010). However, these prior studies are largely limited to low-income micro-neighborhoods. Evidence is mostly lacking where the definition and sizes of neighborhoods being examined are largely different in many outdoor recreation and active living studies (Cohen et al., 2007; Sister et al., 2010). The second gap is related to the disparity in measurement validity involving the application of different methods for various environmental characteristics. This calls for further evidence that may support the use of a dual-method approach in a single study. In other words, which neighborhood characteristics can be reliably measured by GIS, and which of them still have to rely on virtual auditing? What accounts for the major variations between measures derived from GIS and virtual auditing? Third, while GIS is well-recognized as less time-consuming and the efficiency of virtual auditing as compared to on-site auditing has been well-reported, a direct comparison is lacking between the efficiency of GIS and virtual auditing measures (Bader et al., 2015; Brownson et al., 2009; Rzotkiewicz et al., 2018), let alone the true efficiency of a dual-method approach in a single study.

To fill in these gaps and validate the feasibility of a dual-method approach, this study focuses on three major tasks. (1) It examines the uncertainties associated with different sample sizes to aid better auditing sampling. In an ideal world, virtual auditing measures can be generated for all variables and all street segments in the study area for this purpose. A GIS method instead was used here for sampling optimization to allow generating scenarios that contrast the validity of measurements associated with different sample sizes given the cost-prohibitive nature of GSVA and the high sampling validity of GIS measures to substitute GSVA-based sampling found in this study (see section 3.1.2, 3.2.1, and 4.1 for full report). (2) It validates GIS measures of the built environment with Google Street View Auditing (GSVA) measures as a gold standard and provides evidence of the validity for a dual-method approach. (3) It further documents how efficient the proposed dual-method approach is in measuring the built environment in comparison to the two individual methods used separately and thus further proves its feasibility.

2. Methods

2.1. Defining features and characteristics of the built environment

Scholars rely on established or newly developed audit protocols and tools to capture the unique definitions, features, and characteristics of the built environment that they want to measure (Brownson et al., 2009; Kepper et al., 2017; Ramirez et al., 2006; Torio, 2012). For this study, the authors adopted the Active Neighborhood Checklist (ANC) Version 2 (available via URL: https://activelivingresearch.org/active-neighborhood-checklist)(Hoehner et al., 2007), a street-based auditing tool that defines features and characteristics of the built environment, used in many studies including those focused on sampling optimization (Adu-Brimpong et al., 2017). In this study, both the GIS and GSVA methods were used to measure 56 variables, including 50 variables from (A) “land use”, 3 from (B) “public transportation”, and 3 from (D) “quality of the environment” of ANC 2.0 (see Table 1). Another 39 variables of sections (C) “street characteristics”, (D) “environmental quality”, and (E) “place to walk or bicycle” were measured by GSVA but not by GIS due to the lack of publicly available GIS data. Only the 56 variables measured by both methods were reported to keep the paper focused on the three major aims.

2.2. Study areas

Four communities in the City of Los Angeles, Brentwood, Hollywood Hills West, Sun Valley and Wilmington, were chosen for the study (see Fig. 1 and Table 2). Brentwood and Hollywood Hills West are markedly affluent (>70% white; median incomes $108,199 and $112,927 respectively) and the other two are disadvantaged (>90% Hispanic; median incomes $40,627 and $51,290 respectively) (Winter et al., 2019). Two are located along the coast and the other two are located inland. These combinations provide different physical, socioeconomic, and cultural characteristics of the built environment in a metropolitan area. In each of the four communities, two park neighborhoods were generated, yielding eight neighborhoods in total. Each neighborhood was defined through spatially buffering ½ mile around the border of a selected park, as widely used in recreation and active living studies (Cohen et al., 2007; Sister et al., 2010). All eight parks are in similar size with no entrance fee and similar facilities and amenities such as restroom, visitor parking, pedestrian trails, children’s playgrounds, sports and other active recreation, etc. (Winter et al., 2019). The population size of each park neighborhood varies from 2818 in Coldwater Canyon Park neighborhood to 18475 in Banning Park and Museum neighborhood (see Table 2). Park neighborhoods were used here to augment an investigation of park visitor activities and exposure to ozone (Winter et al., 2019). The choice of park neighborhood as a research setting enables comparison to previous studies that used other neighborhood definitions and sizes to assess sampling optimization and measurement (c.f. Adu-Brimpong et al., 2017).

2.3. Data collection, sampling and preprocessing

The authors collected GIS data for the 56 variables in the ANC 2.0

Table 1
Selected Active Neighborhood Checklist Index Variables Accessed by both GIS and GSVA.

i	Questions and Answers on ANC	*i*	Questions and Answers on ANC
	A. What land uses are present?		
	A.1. Are residential and non-residential land uses present?	29	d. Sports/playing field
1	a. All residential	30	e. Basketball/tennis/volleyball court
2	b. Both Residential and non-residential	31	f. Playground
3	c. All non-residential	32	g. Outdoor pool
	A.2. What is the predominant land use?	33	h. Other
4	a. Residential buildings/yards		**A.6. What types of non-residential uses are present?**
5	b. Commercial, institutional, office or industrial building(s)	34	a. None
6	c. School/school yards (elementary, middle, high school)	35	b. Abandoned building or vacant lot
7	d. Parking lots or garages	36	c. Small grocery, convenience store (including in gas station), or pharmacy
8	e. Park with exercise/sport facilities or playground equipment	37	d. supermarket
9	f. Abandoned building or vacant lot	38	e. Food establishment (restaurant, bakery, café, coffee shop, bar)
10	g. Undeveloped land	39	f. Entertainment (e.g., movie theatre, arcade)
11	h. Designated green space (including park with no exercise/ play facilities)	40	g. Library or post office
12	i. Other non-residential	41	h. Bank
	A. 3. What types of residential uses are present?	42	i. Laundry/dry cleaner
13	a. None	43	j. Indoor fitness facility
14	b. Abandoned homes	44	k. School (elementary, middle, high school)
15	c. Single family homes	45	l. College, technical school, or university
16	d. Multi-unit homes (2–4 units)	46	m. High-rise building (>5 stories)
17	e. Apartments or condominiums (>4units, 1-4 stories)	47	n. Big box store (e.g., Walmart, Office Depot, Best Buy)
18	f. Apartments or condominiums (>4 stories)	48	o. Mall
19	g. Apartment over retail	49	p. Strip mall
20	h. Other (retirement home, mobile home, dorms)	50	q. Large office building, warehouse, factory, or industrial building
	A.4. What functioning parking facilities are present?		**B. Is Public Transportation Available?**
21	a. None (no parking allowed on street most or all of the time)		**B.1. Any transit stops (bus, train, or other)?**
22	b. On-street, including angled parking	51	a. No
23	c. Small lot or garage (<30 spaces)	52	b. Yes, one side
24	d. Medium to large lot	53	c. Yes, both sides
25	e. Garage		
	A. 5. What public recreational facilities and equipment are present (including schoolyard if publicly accessible)?		**D. What is the quality of the environment?** **D. 6. Tree shade on the walking are?**
26	a. None	54	a. None or a little
27	b. Park with exercise/sport facilities or playground equipment	55	b. Same
28	c. Off-road walking/biking trail	56	c. A lot

Note: The table only lists 56 ANC2.0 items that are assessed by both GIS and GSVA methods. The rest of the items on ANC 2.0 in the Street Characteristics, Quality of the Environment, and Place to Walk or Bicycle categories was assessed by the GSVA method but is not listed here as the relevant results are not reported in this paper.

using all street segments (N = 2280). Among them, 200 (8.8%) street segments were randomly sampled and audited by raters using the GSVA method. The choice was to meet both the sampling validity threshold (5%–20%) as reported in section 3.1 and project budget limitation covering the costs of human raters to carry out the procedures.

2.3.1. *Street segment defining and preprocessing*

Following guidance of the ANC 2.0, street segments are generally defined by both crossroads and T-shaped intersections. To assemble a street segment population, the authors obtained public 2015 street data from the County of Los Angeles Enterprise GIS (https://egis-lacounty.hub.arcgis.com). The spatial typology of the data was cleaned in ESRI ArcGIS 10.6 so that all line segments were merged into continuous ones unless they were divided by intersections. Segments longer than ¼ mile were then divided into two equal segments until they were equal to or shorter than ¼ mile. Segments less than 10 feet were combined with the neighboring segment.

2.3.2. *Sampling for Google Street View Auditing (GSVA)*

While GIS measures were generated for all street segments in the four communities, 50 random samples in each community (25 within each park neighborhood), for a total of 200 samples, were assessed via the GSVA method. The percentage of segments sampled in different park neighborhoods varied from 5% in the highest density neighborhood (Wilmington R.C.) to 15% in the lowest density neighborhood (Coldwater Canyon Park), which also meet the sampling threshold (5%–20%) recommended in section 3.1. To facilitate the GSVA method, street segments were randomly sampled among street segments accessible by pedestrians excluding freeways and on-and-off ramps. Owing to issues such as unavailability of Google Street View images in certain areas such as undeveloped land and alleyways, some street segments chosen in the first round of sampling were removed and replaced by newly sampled ones from an undrawn pool. A total of 48 out of the 200 original samples were problematic and contributed to a 24% resampling rate (see Table 2). The highest resampling rate occurred in the Coldwater Canyon Park neighborhood, an affluent canyon zone with a lot of undeveloped spaces and private roads lacking GSV data (See Fig. 1). Once the samples were finalized, street segments in GIS format were converted to Google Earth KML format for virtual auditing.

2.3.3. *Auditing the built environment with GSVA*

To ensure the reliability of the audit results, the authors developed a practical step-by-step GSVA protocol to guide and train the raters working through an independent, blind, but identical process within one month. Included in the protocol are how to install and access Google Earth Pro, open segment KML files, navigate the sample segments in a pre-defined order within a neighborhood, and then to audit the street segment from a specified starting point to an ending point on both sides of a street segment. Imagery available on Google Earth Pro was captured during 2014–2015, a time period matching the dates of most GIS data. Every sampled street segment was rated independently by two different raters. No rater would be auditing the same neighborhood continuously to avoid the potential impact of their memories of the neighborhood. The audit results were recorded by raters on a paper version of the ANC V.2.0, which then were input to digital format, using a double-entry process by separate research team members. To check the validity of GIS measures against GSVA measures, the authors resolved unidentical measures by the two different raters, first by checking for a data entry error, and where necessary, a review by a third rater was conducted as a form of arbitration between the two raters' entries. A scoring method similar to the one used by Adu-Brimpong et al.,2017 was adopted to assign a value of 1 to any street segment satisfying a variable criterion, e.g., the presence of "A.1.a All residential" and a value of 0 depicting the absence of each feature.

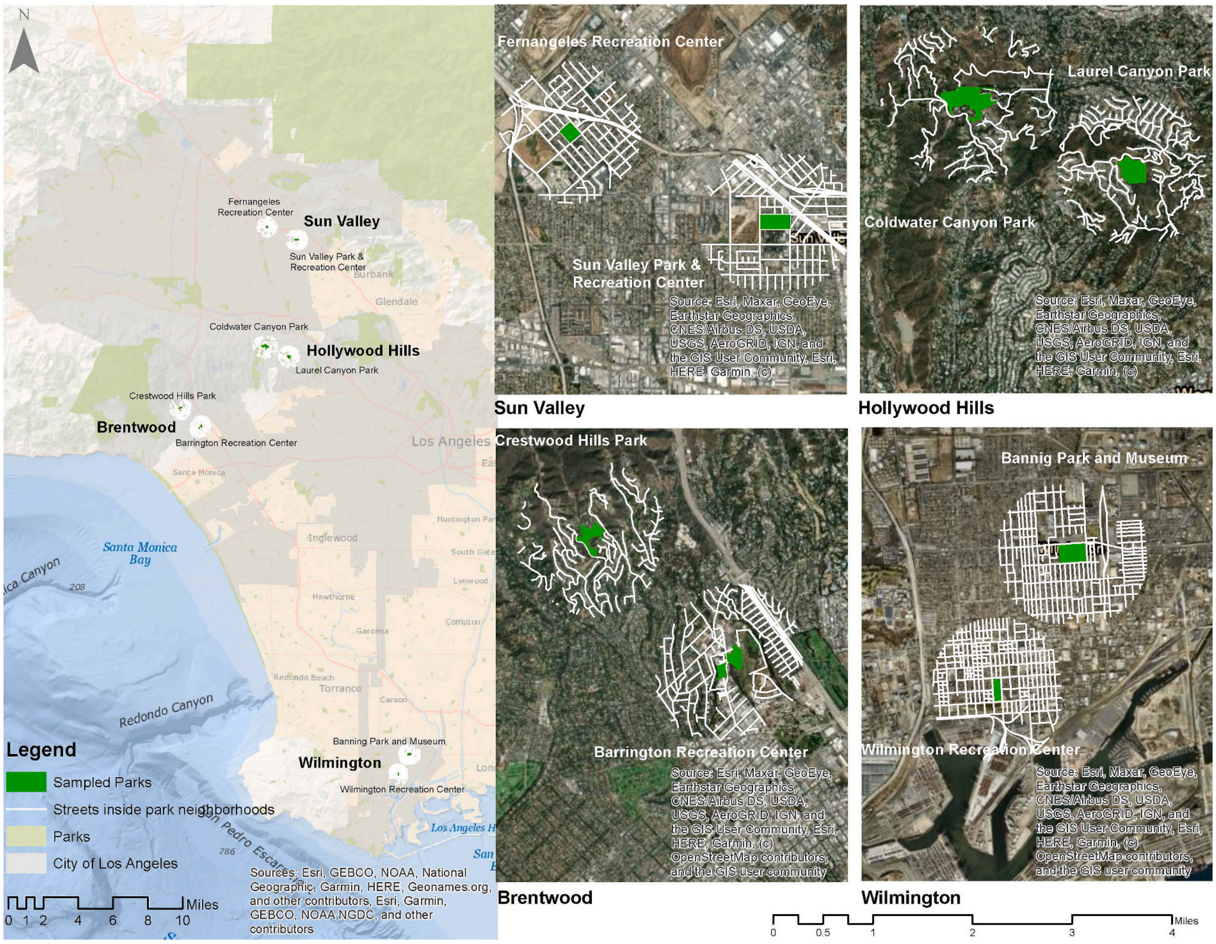

Fig. 1. The eight park neighborhoods in both affluent and disadvantaged communities in the City of Los Angeles.

Table 2
Resampling rate for GSVA measures in studied communities.

Community (Est. 2010 Pop. within Buffer)	Type	Park Neighborhood (Est. 2010 Pop.)	Resampling No./Total Sample	Resampling Rate (%)	Major Issues
Brentwood (14974, 70.8% White)	Affluent, Coastal	Crestwood Hills Park (4161)	6/25	16	GSV not available for private roads and undeveloped area
		Barrington R.C. (10813)	2/25		GSV not available in alleys
Hollywood Hills West (9228, 72.8% White)	Affluent, Inland	Laurel Canyon Park (6410)	1/25	54	GSV not available in hilly undeveloped areas
		Coldwater Canyon Park (2818)	26/25		
Sun Valley (23751, 96.5% Hispanic)	Disadvantaged, Inland	Fernangeles R.C. (8793)	1/25	2	GSV not available
		Sun Valley R.C. (14958)	0/25		n/a
Wilmington (35005, 93.7% Hispanic)	Disadvantaged, Coastal	Banning Park and Museum (18475)	4/25	24	GSV not available in alleys
		Wilmington R.C. (16530)	8/25		
Total			48/200	24	

2.3.4. Data measuring, processing and sampling through GIS modeling

In addition to street data, the authors collected other GIS data to help quantify the 56 selected variables (see Table 1). Most GIS data were obtained from the County of Los Angeles Enterprise GIS and other public data sources dated around 2015 and 2016, with one exception of vegetation cover data which was dated 2006, the most updated of its kind at the time of the study. Variables were measured through geo-processing and other spatial analysis. Table 6 in the Online supplementary further describes the GIS data layers, their date and sources, and analysis methods applied to each category of variables. The same scoring method was used to assign a value (1 or 0) to each variable criterion for any street segment using rule-based modeling rather than by human raters.

Each variable representing one criterion in ANC 2.0 is denoted as i ($i \in I$ (1, 2, 3,…56)). To better facilitate quantitative analyses, the raw measures (1, 0) were further processed. First, the authors computed the chosen rate R of each variable i. For instance, for characteristic variable $i = 4$ (see Table 1), a chosen rate of 40% for R_4 means 40% of all street segments was assigned value 1 for criterion variable 4 (A.2.a "residential buildings/yards" for the question "A.2. What is the predominant land use?). Second, to examine the performance of different sample sizes, the authors applied 21 different sample sizes J to extract samples from the street segment population (N = 2280) measured in GIS. Samples were created randomly from the four communities based on 1%, 5%, 8%, 10%, 15% and 16 additional sample sizes with a 5% increment from 20% to 95% of the total population, gathered as scenarios to contrast the implication of varying sample sizes. Then, the authors computed the difference $Dif_{i,j}$ between the value of each variable $i's$ chosen rate R_{ij} generated by sample j with each variable $i's$ R_{iP} generated from the entire street segment population P (see Equation (1)) across each of the sampling size scenarios.

$$Dif_{i,j} = R_{i,j} - R_{i,P} \tag{1}$$

$I \in [1, 2 \ldots 56]; \quad J \in [1\%, 5\%, 8\%, 10\%, 15\% \ldots 95\%];$

2.4. Assessing sampling optimization, validity, and efficiency of measures

2.4.1. Sampling performance and optimization

A group of statistics tools were deployed to examine the performance of different sample sizes in comparison with that of the population using GIS measures for all variables.

Standard descriptive statistics, e.g., mean, standard deviation and variance, were used to examine the distribution of the $|Dif_{I,j}|$ of the GIS measures for sample size $j (j \in J)$. Absolute value is used here to allow easier comparison on graph. These statistics allowed a comparison between the distribution generated with different individual sample size j and the distribution generated with population P and to identify the optimal sample sizes in representing the entire population P. The higher the value of the mean, standard deviation, and variance of $|Dif_{I,j}|$, the poorer the representation of the population P by sample size j. The 95% Confidence Interval (CI) of each sample's mean of $Dif_{I,j}$ in predicting that of the population ($Dif_{I,j} = 0$) was also calculated based on the mean and standard deviation of $Dif_{I,j}$ of each sample. **Sampling validity** was computed with $Dif_{i,j}$ of all variables i for the 21 sample sizes J according to equation (2) below (Krippendorff, 2004). The higher the value of Sampling Validity a_j of sample size j, the better the representation of the sample size j for the entire population P.

$$\text{Sampling Validity of sample } j(a_j) = 1 - \text{sampling error} = 1 - \frac{\sigma_j}{\sqrt{j}}\sqrt{\frac{P-j}{P-1}} \tag{2}$$

where σ_j is the standard deviation of $Dif_{I,j}$

P - denotes the population size of all the street segments of the GIS measures.

j - denotes the sample sizes of a specific sample set, $j \in [1\%, 5\%, 8\%, 10\%, 15\% \ldots 95\%]$.

In addition to examining sampling optimization for all communities, sampling validity was assessed to compare sample performance between the affluent and disadvantaged communities.

One sample *t*-test was used to examine whether the distributions of $R_{I,J}$, the chosen rate of the 56 neighborhood characteristic variables I, of the different sample sets J, were in normal distributions. For each characteristic variable i, one sample t-test evaluated whether there is a significant difference between the mean of its value $R_{i,j}$ returned from each of the 21 sample sets and the mean of $R_{i,P}$ (the tested value) returned from the entire population. For our 2-tailed test, if $p > .05$, the difference between the sample estimated mean was not significantly different from the population mean. Each variable was also tested against different low-end sampling sizes in the one sample t-test to identify the lowest acceptable sampling size for the variable in order to significantly represent the entire population. Sampling results are reported in 3.1.

2.4.2. Validity and reliability of GIS and GSVA measures

To address the study's second aim, the authors assess the inter-approach validity of GIS measures against GSVA and the inter-rater reliability of GSVA as audited by different raters. For this purpose, the authors use Kappa's Coefficient k (see equation (3)) as the major tool supplemented by observed agreement p_0 where valid k is not available. For each of the 56 selected variables i, Kappa coefficient measures the percentage of values in the main diagonal of a crosstab table as observed agreement p_0 (%) (see also equations (5) and (6)), where the two measurement approaches agree on the return value (1 or 0), and then adjusts these values by the amount of hypothetical probability of chance agreement p_e that could be expected due to chance alone. p_e is calculated with equation (4).

$$k = \frac{p_o - p_e}{1 - p_e}, k \in [0, 1] \tag{3}$$

$$p_e = \frac{1}{N^2}\sum_c n_{c1}n_{c2}, c \in C \tag{4}$$

where $N = 200$ is the total number of sampled street segments to be measured by the two methods. C is the mutually exclusively option of answers to each question. There are two situations of C. The first situation includes questions, i.e., A.1., B.1, and D.6, with three mutually exclusive answers. For A.1 as an example, there are three mutually exclusively answers $A.1.a$, $A.1.b$, and $A.1.c$. n_{c1} refers to the number of times category $c \in [1, 2, 3]$, where 1 means A.1.a and so forth, is chosen as measured by method 1 and n_{c2} refers to the number of times category c is chosen as measured by method 2. For this situation, only three p_0 and k were calculated for 9 variables (A.1.a, A.1.b, A.1.c; B.1.a, B.1.b, B.1.c; and D.6.a, D.6.b, D.6.c). In the second situation, each of the rest of 47 answer options is treated as individual variable i to generate mutually exclusive answers (1 and 0) and measured separately as answers to these questions are not necessarily mutually exclusive if processed together. In such situations, $c \in [1, 0]$ where 1 means that a certain answer option is chosen and 0 means that it is not chosen. n_{c1} refers to the number of times category $c \in [1, 0]$ measured by method 1 and n_{c2} refers to the number of times category c is measured by method 2 for each answer to each question. These remaining variables generate 47 p_0 and k, resulting in a total 50 p_0 and k along with an additional three from questions A.1, B.1, and D.6.

As a supplementary assessment, observed agreement p_o was used to assess both the inter-approach validity of GIS and GSVA measures and the inter-rater reliability of the GSVA measures where invalid Kappa's coefficient was reported due to constant results generated by one or both methods or one or both raters. p_o is calculated with equations (5) and (6) for each of the 200 sampled street segments j. Where L is the total number of characteristic variables measured by both of the methods. l_{aj} is the total number of characteristic variables for which the two methods or the two raters generated the same measured results for street segment j.

$$p_{oj} = \frac{l_{aj}}{L}\Big|(L = 56) \tag{5}$$

$$p_{oJ} = \frac{\sum_{j=1}^{200} p_{oj}}{L^* N}\Bigg|(\mathrm{N} = 200) \tag{6}$$

p_{oJ} is the average observed agreement for all variables among the 200 sampled street segments. Validity and reliability results as assessed by both k and p_o are reported in 3.2.

2.4.3. Efficiency of GIS and GSVA measures

Targeting the third aim of the study, the authors documented the total person-hours used for each major task of the two methods. The GSVA method involves data preparation, GIS modeling to prepare sampled street segments, rating protocol development, rater training, rater auditing, data input and correction, etc., while the GIS method mainly involves data preparation, data integration, and geospatial modeling to assign measurement values to the variables (see Table 3).

Table 3
Cost comparison between GSVA and GIS methods.

TASK ITEMS	GSVA Hours*person	Note	GIS Hour*person	Note
GIS Modeling	20	Sampling method development, sampling segment preparation, data generation sexporting for google earth rating, etc.	80	Geospatial analysis and modeling on ANC criteria, data output for statistics analysis, etc.
Data preparation	10	Data Collection, Integration, preprocessing, etc.	20	Data Collection, Integration, preprocessing, etc.
Rating Protocol Development and Supervision	50	Development of Rating Protocols that raters can strictly follow. Provision of instruction, guidance and supervision to raters.	n/a	
Rater Training	15	Practically training raters to follow protocols properly to undertake the auditing process and document findings	n/a	
Rater Auditing	400	Auditing process carried out by the raters	n/a	
Data Input and Correction	100	Inputing data recorded in hard copy forms to digital copies and identifying records of disagreement, and conducting another round of auditing to correct the audit results	n/a	
Total Hours	395		100	
Total Segments	200		2280	
Total Variables	95		56	
Total variables*segments	19000		127680	
Segment-Variables/Hour	32		1277	

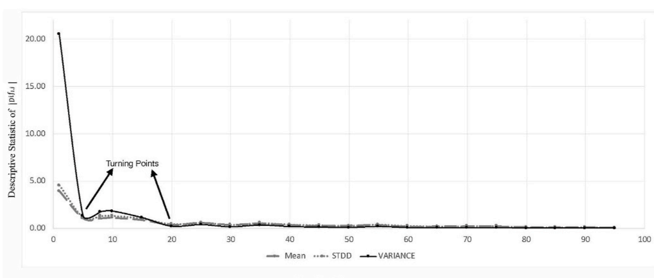

Fig. 2. The relation between sample size and descriptive statistics (Mean, Standard Deviation, and Variance) of the absolute value of the difference between sample and population.

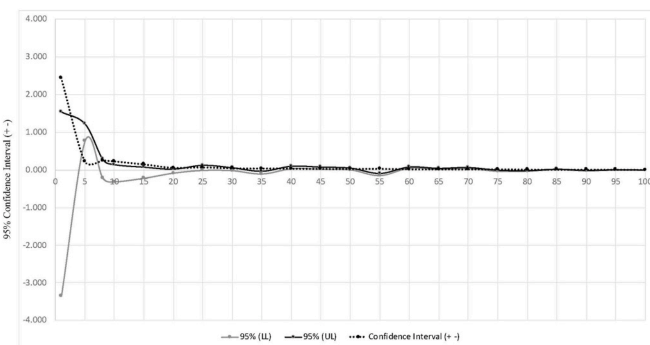

Fig. 3. The relation between sample size and Confidence Interval (+-) of the mean of the difference between sample and population.

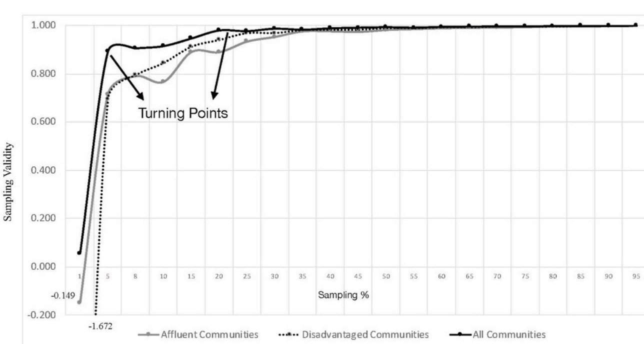

Fig. 4. The relation between sample size and Sampling Validity (Krippendorff, 2004) of the DIFFERENCE between sample and population.

3. Results

3.1. Sampling performance and optimization

3.1.1. Sampling validity for all variables of GIS measures

Descriptive statistics and sampling validity a_j were used to examine the $\left|Dif_{i,j}\right|$ of all variables based on GIS measures. As shown in Fig. 2, the mean, standard deviation, and variance of $\left|Dif_{i,j}\right|$ show a clear L shape declining trend as the sample size increases from 1% to 95% of the total population (N = 2280). Two turning points were found at a 5% (N = 114) sampling size and 20% (N = 456) sampling size, both resulted in significant drops of the mean, standard deviation, and variance of $\left|Dif_{i,j}\right|$, representing significant increases of validity. However, increasing the sampling sizes beyond the 20% level (N = 456) did not substantively improve the representativeness of each sample when compared to the population P. Confidence Interval (CI) was used to examine$Dif_{i,j}$. Its mean and 95% CI increasingly decrease to 0 (the mean of $Dif_{i,j}$ of the population) from both direction as the sample size increases to 100% (see Fig. 3). CI drops below ±0.5 at 5% (±0.228) and further dropped near ±0.05% at 20% (±0.0557). However, at 5% sampling, the projected mean falls on one side (0.761,1.217) of the x-axis, somewhat deviated from the population mean (0). With an 8% sampling, such deviation (CI: ±0.249, Mean: [-0.213,.285]) largely disappears. As sample size increases, the projected mean gets closer to 0.

Sample validity a_j matches well with that of descriptive statistics and CI. As shown in Fig. 4, sampling validity scores for all communities (solid black line) rose from 0.05 to 0.9 as sampling size increased from 1% to 5% and continued to rise from 0.9 to 0.98 as sampling sizes increase to 20%, after which sampling validity reached 1 when sampling rose to 100%. Thus, differences among these sampling scenarios over 20% can be largely omitted. This result further verifies that 5% (N = 114) and 20% (N = 456) are the two major thresholds of sampling in this study resulting in different levels of validity. However, sampling performance changes when different types of communities are examined separately. A higher sampling rate is required in affluent communities to achieve the same level of sampling validity as in disadvantaged communities (see Fig. 4 and Table 7 in the online supplementary). However, the difference largely disappears when the sampling rate exceeds 30% or only 15% if all four communities are affluent or disadvantaged.

3.1.2. Sampling validity of individual variables of GIS measures

One sample *t*-test was applied to examine each variable to contrast the different low-end sampling sizes to identify the lowest acceptable sampling size. Results, as shown in Table 4, indicate that the sample-generated mean values are normally distributed for 71.4% (40 out of 56) variables. With $p > .05$, their measures generated from a minimum 1% sampling size are not significantly different from those returned from the population (see Table 4). Among the 41 variables with valid k for validity or reliability measures (see Fig. 5 in the main body and Table 5 in the online supplementary), those with normal distribution and identical 1% minimum sampling requirement rises to 95.1% (39 out of 41 variables) and can be categorized into three groups. (1) Group 1 are 20 out of the 21 (95.2%, A.3.f as an exception) variables that have GIS measures validated by previous studies (see introduction) and by the 8.8% samples of GSVA measures in this study (see section 3.2.1). They cover major land use types, public service and amenities, recreation, and transportation facilities. They share identical one sample *t*-test results denoting consistency in distribution and sampling requirements. Both high inter-approach validity and similar distribution pattern strongly support the adoption of a universal sampling rate for these variables and indicate that GIS measures can offer sampling guidance for the GSVA methods for these variables. (2) Another 10 out of the 11 variables (D.6.c as an exception) showing disparity between GIS and GSVA measures (see 3.2.1). Their distributions were also in a normal distribution with the same 1% required sampling rate. (3) The last nine variables without a valid inter-approach validity but with high inter-rater reliability (see Table 8) also share the same distribution and 1% minimum required sampling rate. In summary, a total of 19 additional variables are found to share similar distributions as the 20 variables with validated GIS measures, resulting in 95.1% (39 out of 42) variables sharing the same distribution and minimum required sample size. The remaining 15 variables were not used for sampling optimization as they lack valid validity and reliability scores owing to their absence in the study area. The above important evidence supports the validity of using GIS measures for sampling optimization. Please see the implication of such results in section 4.1.

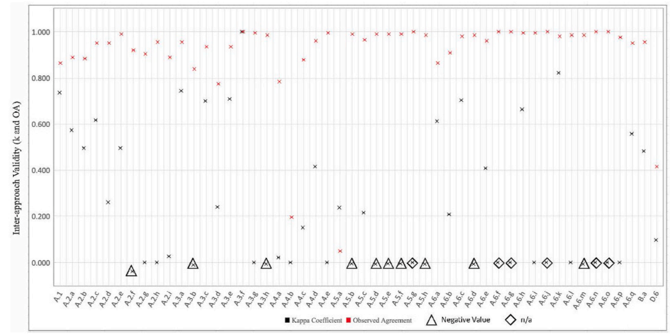

Fig. 5. Inter-approach validity assesment of GIS and GSVA Measures.

3.2. Validity and reliability of GIS and GSVA measures

3.2.1. Inter-approach validity between GSVA and GIS measures

Fig. 5 in the main body and Table 5 in the online supplementary report inter-approach validity results measured by Kappa Coefficient (K) and observed agreements (OA) using GSVA as a gold standard. k values are mixed despite that OAs are generally high among most variables (76% with >0.9 OA, 87.5% with >0.8 OA). Based on their k values, variables are categorized into five groups, i.e., “perfect”, “substantial”, “moderate”, “fair” and “poor” (Landis and Koch, 1977). Variables without valid k are categorized as “No Kappa’s” and those with negative values are noted “<chance”, both of which were largely caused by the dominance of “0” and a lack of “1” in the measurement reported by one or both methods. “Moderate” to “perfect” performances are achieved by variables describing commonly found land uses such as (1) presence of

Table 4
One Sample T-Test (*p*) of 1% Sampling and Sampling Requirement for individual variables.

Variable	p	Minimum Sampling %	Minimum Sampling Size	Variable	p	Minimum Sampling %	Minimum Sampling Size
A.1.a	0.172	1	23	A.5.d	0.145	1	23
A.1.b	0.478	1	23	A.5.e	0.076	1	23
A.1.c	0.078	1	23	A.5.f	0.18	1	23
A.2.a	0.076	1	23	A.5.g	n/a*	1	23
A.2.b	0.046**	5	114	A.5.h	0.219	1	23
A.2.c	n/a*	n/a*	n/a	A.6.a	0.202	1	23
A.2.d***	0.696	1	23	A.6.b***	0.202	1	23
A.2.e	0.101	1	23	A.6.c	0.295	1	23
A.2.f	0.416	1	23	A.6.d****	0.673	1	23
A.2.g****	0.009**	80	1824	A.6.e****	0.625	1	23
A.2.h****	0.011**	75	1710	A.6.f	n/a*	n/a*	n/a
A.2.i***	0.274	1	23	A.6.g	n/a*	n/a*	n/a
A.3.a	0.161	1	23	A.6.h****	0.488	1	23
A.3.b	0.471	1	23	A.6.i	n/a*	n/a*	n/a
A.3.c	0.151	1	23	A.6.j	0.324	1	23
A.3.d***	0.38	1	23	A.6.k	0.401	1	23
A.3.e	0.372	1	23	A.6.l	0.925	1	23
A.3.f	0.002**	65	1482	A.6.m	n/a	n/a*	n/a
A.3.g****	n/a*	n/a*	n/a	A.6.n	n/a	n/a*	n/a
A.3.h	0.147	1	23	A.6.o	0.013**	10	228
A.4.a***	0.602	1	23	A.6.p****	n/a*	n/a*	n/a
A.4.b****	n/a*	n/a*	n/a	A.6.q	0.488	1	23
A.4.c***	0.342	1	23	B.1.a	0.132	1	23
A.4.d	0.279	1	23	B.1.b	0.199	1	23
A.4.e****	0.628	1	23	B.1.c	0.091	1	23
A.5.a***	0.111	1	23	D.6.a***	0.108	1	23
A.5.b	0.106	1	23	D.6.b***	0.284	1	23
A.5.c***	0.663	1	23	D.6.c***	0.038**	20	456

*Variable is invalid for one sample t-test due to lack of variation of values among all street segments. ** variables with p value < .05. *** Variables with “fair” to “poor” inter-approach validity. **** Variables with high GSVA inter-rater reliability.

residential/non-residential (A.1.a, A.1.b, A.1.c and A.2.a), specific types of residential (A.3.a, A.3.c, A.3.e, A.3.f and A.6.a), and specific non-residential usages (A.2.b and A.6.q); (2) schools (A.2.c and A.6.k); (3) local stores (A.6.c and A.6.h) and food establishments (A.6.e); and (4) other highly recognizable public facilities such as parks (A.2.e), medium and large parking lots (A.4.d), and public transportation facilities (B.1.a, B.1.b, and B.1.c). For the above 21 variables, GIS measures provide a good substitution to GSVA measures. Another 11 variables reported with "fair" or "poor" validity scores include those describing (1) rare land uses definitions such as predominant "parking lots or garages" (A.2.d) and "other non-residential" (A.2.i), (2) land uses and facilities hard to discern or define clearly by raters (A.3.d "Multi-unit homes", A.6.b "abandoned building or vacant lot", and D.6.a, D.6.b, and D.6.c. tree shades) and (3) those in lack of accurate GIS data (A.4.a "no parking on street", A.4.c "small lot or garage", A.5.a "non-recreational facilities", A.5.c "off-road walking/biking trails"). Among these 11 variables, GIS may offer more accurate measures for the 5 variables in the group (2) as inter-rater reliabilities were low for these variables and reliable official public or scientific records are available in GIS format. GSVA is found a better choice for the remaining 6 variables. In addition, there are 24 variables without valid inter-approach validity (k). Among them, 9 variables (A.2.g, A.2.h, A.3.g, A.4.b, A.4.e, A.6.d, A.6.,e, A.6.h, and A.6.p.) present high GSVA inter-rater reliability (see Table 8 in the online supplementary) indicating that the method can be comfortably used as a gold standard for these variables. But for the remaining 15 variables, even with high inter-approach OA >.9, due to lack of valid kappa's coefficients for both assessments, the study offers no evidence on whether GSVA measures are reliable among raters or whether GIS measures can substitute GSVA measures.

3.2.2. Inter-rater reliability of GSVA measures

High reported inter-rater reliability of GSVA measures support their use as a gold standard for the inter-approach validity assessment. Table 8 in the online supplementary reported .625 averaged of valid Kappa's coefficient (k), varying from moderate (0.4-0.6) to perfect (0.8–1.0). OA is very high for nearly all variables with a mean of 0.9553. The lack of a valid k for some variables was found to be mainly caused by a lack of or very low presence of characteristics. Inter-rater reliability of measures in different communities have nearly identical means of k and OA (see Tables 9 and 10 in online supplementary), indicating no inconsistency. One variable warranting additional study is "D.6 Tree shade on the walking are" with valid but low value for both k and OA.

3.3. Efficiency of GSVA and GIS measures

As shown in Table 3, measuring 56 variables for 200 street segments costs GSVA 350 h (32 segment-variables per hour), and measuring 56 variables for all 2280 street segments costs GIS 100 h (1277 segment-variables per hour), 40 times faster than GSVA. If the 26 variables with high inter-approach validity are measured by GIS, the total time will be reduced to 233.5 h (187.5 by GSVA and 46 by GIS), a 33.3% reduction from the GSVA method alone.

4. Discussion and suggestions

The current study fills in the three gaps identified in the introduction and supports the adoption of a dual-method approach. First, GIS sampling provides a reasonable substitution for GSVA sampling for objective neighborhood features and characteristics (see section 4.1). It offers a low-cost sampling alternative in neighborhoods with different environmental characteristics varying from small to large. Second, it provides evidence that a dual-method approach is feasible for accurate measurement of the built environment. Last but not least, it verifies that a dual-method approach can be more efficient than the GSVA method.

4.1. Sampling optimization and sampling choices

This study provides new evidence upon how to better sample and what is the number of samples needed in order to accurately represent the population of objective environmental characteristics using ½ mile park neighborhood in both affluent and disadvantaged communities. Our findings add evidence to the previous body of work which either focused on sampling optimization in low-income micro-neighborhoods (Adu-Brimpong et al., 2017) or investigated only limited sampling sizes (Moniruzzaman and Páez, 2012). The novel contribution of park neighborhood is an important advancement, given the ½ mile buffer distance is also widely used in the literature on public health, recreation, and active living studies (Cohen et al., 2007; Rigolon et al., 2018). It also provides a new precedent on using GIS measures for sampling optimization as an alternative to virtual audit (Adu-Brimpong et al., 2017; Moniruzzaman and Páez, 2012) and on-site audit (Cerin et al., 2011; McMillan et al., 2010). Further, with sampling findings on individual variables, the study adds to the current literature in which a paucity of information is available on whether and how specific features of the built environment may need to be sampled differently to achieve similar validity.

Specifically, the study presents two pieces of important evidence to support that GIS measures can offers sampling guidance applicable to the GSVA methods. First, the authors found that the distribution and minimum required sample size are largely the same for three groups of variables including (1) the 21 variables with GIS measures highly validated by GSVA measures, (2) the 11 variables with a disparity between GIS measures and GSVA measures, and (3) the 9 variables without valid inter-approach validity but with valid inter-rater reliability. 95.1% of these variables (39/41) with valid validity or reliability scores share the same distribution and sampling requirements. Second, the 8.8% sampling size (very close to the 5% low-end threshold recommended) used for the GSVA measures validated the majority (72.5%) of these variables (40), among which 20 have high validity and 9 have high GSVA reliability). The variables validated are largely consistent with those reported by previous studies. These findings indicate that the disparity between GIS measures and GSVA measures of some variables was largely caused by the usage of different methods but not by the difference in their distribution and the required sample size to represent the population. We, therefore, argue that the GIS measures can be reasonably used to test sampling performance and recommend sampling size for the GSVA method in measuring objective environmental characteristics similar to the ones investigated in the current study.

Our sampling optimization findings, derived from GIS measures and sampling practice based on urban communities in Los Angeles, indicate that proper sampling can generate reliable data to represent an unknown population. The authors found that in a given population N = 2280, one can use 5% (N = 114) as the low-end threshold and 20% (N = 465) as the high-end threshold of sampling to generate a sample that adequately represents the population. Lower than 5% sampling may lead to significant differences between the sample and the population and thus is not recommended. Any sample size higher than 20% would involve further gains in representativeness, however, reductions of uncertainty beyond this level resulted in only modest improvements. While sampling is considered in different types of communities, 5% can be comfortably used as the low-end if all neighborhoods are disadvantaged ones with higher street density, while a slightly higher 8% as the sampling low-end is recommended when all neighborhoods are affluent ones with much lower street density, to achieve a similar level of sampling validity (see Table 7 in online supplementary). These recommendations would be most suited for studies focusing on objective neighborhood characteristics similar to those investigated in the current study with similar street segment population sizes in areas comparable to Los Angeles communities. Studies should apply the above recommendations with caution if they meet one or more of the following conditions. (1) They focus on a more subjectively defined built environment or use an auditing protocol that caters to subjective measures. (2) They investigate geographic areas with physical, cultural, and socioeconomic conditions fundamentally different from the Los Angeles area. (3) They study areas with street segment population sizes in ½ mile park neighborhood outside the ranges presented in this study. When the street segment population is higher than 2280, such low- and high-end thresholds of percentages can be used comfortably and even reduced if the population is much larger and the reported sampling size instead of sampling % may be used as a reference. For populations lower than 2280, the authors recommend proportionally increasing the sample rate to generate similar sampling sizes as reported to derive similarly accurate population estimates.

With regard to individual characteristics outlined, different sampling sizes associated with their varied representation in a neighborhood's population of measures call for further attention. While the authors could use the 1% sampling rate to generate representative data for a majority (71.4%) of examined variables, other variables either called for a much higher sampling rate, e.g., A.6.o Mall, or were not in a normal distribution and thus not subjected to t-tests, e.g., A.2.c Predominant School/school yards (see Table 4). The higher rate was attributable to the fact that these characteristics only existed along very few street segments or were not present in the study area. However, it is impossible to know whether these characteristics are absent unless the entire population is examined. The best way to handle those variables with no or very few cases in the neighborhoods is to collect data for a much larger sample size using GSVA or even for the entire population using GIS if needed data is available. Whether or not these attributes would be rare in other study areas is also not known, where it is possible that in other areas these attributes would be uniquely sparse or abundant, requiring the investigator to determine based on their local knowledge on how to best adjust sampling practice. Using GIS to generate a full population of measures may not produce 100% accurate results for all variables, but it can at least provide evidence for investigators to consider which variables are deserving an investment of more time and resources. Therefore, for future studies involving complicated conditions, an alternative to adopting the 5%–20% sampling rate as a general rule is to adopt a localized dual-method approach. In such an approach, the investigators can collect data for all variables for the entire study area using GIS and then study the sampling validity to assist in choosing the ideal sampling sizes for their auditing effort. The local investigators can then decide whether they apply a universal sampling rate or variable-specific sampling rate for each variable based on their findings. This additional step will also help further verify whether the general sampling recommendation of this study is valid in other geographic areas.

4.2. Strengths and weaknesses of GSVA and GIS measures

4.2.1. Inter-approach validity of GSVA and GIS measures

While the GSVA is used as a gold standard to examine the validity of GIS measures, both methods have limitations depending on the characteristic being assessed. There lacks enough evidence for 15 (26.8%) out of 56 variables due to the absence of characteristics in the study area. Among the other 41(73.2%) variables, GIS measures are found a successful substitution for GSVA measures for 26 (63.4%) of them, which mainly describes (1) land uses that commonly found in all types of communities such as residential (e.g., single-family homes and apartment or condominiums) and non-residential (commercial, institutional, and industrial), and (2) public facilities such as schools, local stores, food establishments, parks, medium and large parking lots, and public transportation. On the one hand, the GIS method is good at measuring characteristic variables with official information from public records, e.g., A.3.d Multi-unit Homes (2–4 units). The number of units may be difficult to assess in a virtual or even on-site audit without entering the units. On the other hand, many variables are difficult to assess using the GIS method simply because such data have not been documented or not documented in the same way in the protocol, e.g., A.4.a and A.4.b on whether parking is allowed on a street. No GIS data about street parking rules were available for the study. This echoes previous findings that GIS measures are a more accurate source of data for characteristics officially documented previously but would not be the case otherwise. In summary, while GIS may afford the ability to successfully assess the full population on some characteristics, GSVA is a more comprehensive and balanced approach that is appropriate for most variables at the expense of higher cost, although without guaranteeing perfect results for all variables. Such findings strongly support using a dual-method approach to maximize the benefit of the two methods to achieve both high validity and efficiency.

4.2.2. Inter-rater reliability of GSVA measures

As seen in 3.2.2, reported high inter-rater reliability of GSVA measures (see Tables 8 and 9 in the online supplementary) matches well with findings of previous studies (Brownson et al., 2009; Hoedl et al., 2010). This finding further supports the adoption of the GSVA method as a gold standard to examine the validity of GIS measures for the majority of neighborhood characteristics. However, in the case where sampling is applied, GSVA may contribute a large degree of uncertainty when specific features are rare or hard to define. In such a case, sampling size has to increase or GIS measures for the population should be collected if an official public record is available.

4.2.3. Time and financial cost

Our analysis of the burden of collection uncovered that unit time cost for each characteristic variable is higher by GSVA than by GIS. This further supports the idea of a dual-method approach to achieve similar results at a lower cost. Under such a proposal, available GIS data can be used to generate measures for variables with strong validity records (see Fig. 5 and Table 5 in the online supplementary), paired with GSVA or other methods where desired information is not well measured by GIS or remains undocumented in GIS data. However, in the near future, Artificial Intelligence (AI) technologies may replace human raters in auditing the environment through optimized automated remote sensing techniques and thus greatly reduce the time cost of GSVA.

4.3. Design protocols and definitions measuring the built environment

In this study, a challenge found while establishing measures to be used in both methods was the ambiguity surrounding some of the definitions of characteristics in the ANC V2.0 protocol. For instance, in category D. What is the quality of the environment? where raters assess "6. Tree shade on the walking area" by assigning a rating of "none or a little", "Some" and "A lot". These appear as subjective assessments, difficult to standardize systematically for comparison of methods, or even among raters working independently on assigned segments. Shade cast on a walking area is influenced by time of day, season (for the abundance of leaves), species of tree, and so on. Lower kappa's coefficient and observed agreement between raters and methods are not surprising in such cases. These findings mirror prior findings that the reliability and validity of measures gathered in audits tend to be higher for objective physical attributes than for subjective assessments of neighborhood conditions. Where protocols aim to measure the built environment, care in design, definitions, and training are essential to generate accurate measures.

5. Conclusions

Through investigating sampling optimization using GIS measures in both affluent and disadvantaged communities in Los Angeles, this study recommends a high-end sampling threshold of 20% of street segments and a low-end threshold of 5% for sampling in measuring a mixture of neighborhood environment for a street population at or greater than 2280 segments. If a rather homogeneous neighborhood environment is under investigation, a higher low-end threshold of 8% is recommended for affluent neighborhoods with low street density. Where a characteristic is a rare occurrence, findings demonstrated the likely need to employ a different sampling practice including the use of increased sampling size for such variables or the full segment population assuming their availability in GIS. Second, the study reported that while GSVA measures are trustworthy for most of the characteristic variables, GIS measures are found with high validity for objective characteristics related to commonly found residential and non-residential land use types and public facilities such as schools, local stores, food establishments, parks, medium and large parking lots, and public transportation. Third, the study provided solid evidence that GIS measures come with a lower financial and time cost than that of GSVA and highly valid GIS measures can substitute GSVA measures to reduce cost.

Given this evidence, the authors recommend a dual-method approach to takes advantage of the benefits of both GIS and GSVA for a more comprehensive, accurate, and efficient measurement of the objective built environment in urban communities from disadvantaged to affluent.

6. Limitation and future investigation

A few limitations of the current study are important to note: (1) The use of GIS measure for sampling optimization, while largely validated by both high inter-approach validity and identical sampling requirement of individual variables measuring the objective built environment, can be further examined against studies using GSVA to measure an entire street population to further check the validity of GIS sampling. This is especially needed to develop sampling guidance for measuring rather subjective environmental characteristics which was not fully examined in our study. (2) Temporal mismatch of some of the GIS data, e.g., tree cover (2006), and the GSVA data (2015) and various levels of other uncertainty involved in the two methods may lead to higher inconsistency among the results and impose higher challenges toward applying a dual-method approach in other study areas. (3) The study lacked an examination of test-retest reliability among GIS measures. A future study may consider comparing GIS measures following different conceptual frameworks, definitions, and data modeling protocols. (4) The present GSVA measurement employed an 8.8% sample of all segments, above the recommended low-end threshold of 5% for measurement accuracy, and below the 20% upper-end threshold. Future comparisons of methods may benefit from a lower or higher percent or a full scope of sampling thresholds for auditing practice. (5) The ANC V 2.0 protocol is one of several protocols designed for systematic assessment of a neighborhood. Future researchers may compare multiple audit protocols and offer recommendations on the best fit for varied study objectives. (6) The contrast of both methods focused on only 56 of the 95 neighborhood characteristics assessed in the ANC V 2.0 protocol, owing to the lack of GIS data for the remaining variables. A future multi-method assessment of the ANC V 2.0 protocol might rely on integrating GIS, human virtual auditing, and AI automated auditing on big data documenting the built environment in the forms of high-resolution remote sensing imagery, LiDAR data, and virtual video.

Acknowledgements

This research was conducted jointly by the US Forest Service, Pacific Southwest Research Station, and California State Polytechnic University, Pomona. The research was funded by the US Forest Service and the Research Joint Venture is (PSWRS RJVA 13-JV-11272131-051).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.healthplace.2020.102482.

References

Adu-Brimpong, J., Coffey, N., Ayers, C., Berrigan, D., Yingling, L.R., Thomas, S., Mitchell, V., Ahuja, C., Rivers, J., Hartz, J., Powell-Wiley, T.M., 2017. Optimizing scoring and sampling methods for assessing built neighborhood environment quality in residential areas. Int. J. Environ. Res. Publ. Health 14 (3), 273. https://doi.org/10.3390/ijerph14030273.

Bader, M.D.M., Mooney, S.J., Lee, Y.J., Sheehan, D., Neckerman, K.M., Rundle, A.G., Teitler, J.O., 2015. Development and deployment of the computer assisted neighborhood visual assessment system (CANVAS) to measure health-related neighborhood conditions. Health Place 31, 163–172. https://doi.org/10.1016/j.healthplace.2014.10.012.

Ben-Joseph, E., Lee, J., Cromley, E., Laden, F., Troped, P., 2013. Virtual and actual: relative accuracy of on-site and web-based instruments in auditing the environment for physical activity. Health Place 19, 138–150. https://doi.org/10.1016/j.healthplace.2012.11.001.

Boone, J.E., Gordon-Larsen, P., Stewart, J.D., Popkin, B.M., 2008. Validation of a GIS facilities database: quantification and implications of error. Ann. Epidemiol. 18 (5), 371–377. https://doi.org/10.1016/j.annepidem.2007.11.008.

Brownson, R.C., Hoehner, C.M., Day, K., Forsyth, A., Sallis, J.F., 2009. Measuring the built environment for physical activity: state of the science. Am. J. Prev. Med. 36 (4, Suppl. ment), S99–S123. https://doi.org/10.1016/j.amepre.2009.01.005 e12.

Cerin, E., Chan, K., Macfarlane, D.J., Lee, K., Lai, P., 2011. Objective assessment of walking environments in ultra-dense cities: development and reliability of the Environment in Asia Scan Tool—Hong Kong version (EAST-HK). Health Place 17 (4), 937–945. https://doi.org/10.1016/j.healthplace.2011.04.005.

Charreire, H., Mackenbach, J.D., Ouasti, M., Lakerveld, J., Compernolle, S., Ben-Rebah, M., McKee, M., Brug, J., Rutter, H., Oppert, J.-M., 2014. Using remote sensing to define environmental characteristics related to physical activity and dietary behaviours: a systematic review (the SPOTLIGHT project). Health Place 25, 1.

Chiang, Y.-C., Sullivan, W., Larsen, L., 2017. Measuring neighborhood walkable environments: a comparison of three approaches. Int. J. Environ. Res. Publ. Health 14 (6). https://doi.org/10.3390/ijerph14060593.

Clarke, P., Ailshire, J., Melendez, R., Bader, M., Morenoff, J., 2010. Using Google Earth to conduct a neighborhood audit: reliability of a virtual audit instrument. Health Place 16 (6), 1224–1229. https://doi.org/10.1016/j.healthplace.2010.08.007.

Cohen, D.A., McKenzie, T.L., Sehgal, A., Williamson, S., et al., 2007. Contribution of Public Parks to Physical Activity, vol. 97. American Journal of Public Health, Washington, pp. 509–514, 3.

Frank, L.D., Saelens, B.E., Chapman, J., Sallis, J.F., Kerr, J., Glanz, K., Couch, S.C., Learnihan, V., Zhou, C., Colburn, T., Cain, K.L., 2012. Objective assessment of obesogenic environments in youth: geographic information system methods and spatial findings from the neighborhood impact on kids study. Am. J. Prev. Med. 42 (5), e47–e55. https://doi.org/10.1016/j.amepre.2012.02.006.

Gullón, P., Badland, H.M., Alfayate, S., Bilal, U., Escobar, F., Cebrecos, A., Diez, J., Franco, M., 2015. Assessing walking and cycling environments in the streets of madrid: comparing on-field and virtual audits. J. Urban Health 92 (5), 923–939. https://doi.org/10.1007/s11524-015-9982-z.

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. https://doi.org/10.1016/j.amepre.2013.01.033.

Handy, S.L., Boarnet, M.G., Ewing, R., Killingsworth, R.E., 2002. How the built environment affects physical activity: views from urban planning. Am. J. Prev. Med. 23 (Suppl. 1), 64–73. https://doi.org/10.1016/S0749-3797(02)00475-0, 2.

Hoedl, S., Titze, S., Oja, P., 2010. The bikeability and walkability evaluation table: reliability and application. Am. J. Prev. Med. 39 (5), 457–459. https://doi.org/10.1016/j.amepre.2010.07.005.

Hoehner, C.M., Brennan Ramirez, L.K., Elliott, M.B., Handy, S.L., Brownson, R.C., 2005. Perceived and objective environmental measures and physical activity among urban adults. Am. J. Prev. Med. 28 (Suppl. 2), 105–116. https://doi.org/10.1016/j.amepre.2004.10.023, 2.

Hoehner, C.M., Ramirez, L.K.B., Ivy, A., Handy, S.L., Brownson, R.C., 2007. Active neighborhood checklist: a user-friendly and reliable tool for assessing activity friendliness. Am. J. Health Promot. 21, 534–537.

Humpel, N., Owen, N., Leslie, E., 2002. Environmental factors associated with adults' participation in physical activity: a review. Am. J. Prev. Med. 22 (3), 188–199. https://doi.org/10.1016/S0749-3797(01)00426-3.

Kelly, C.M., Wilson, J.S., Baker, E.A., Miller, D.K., Schootman, M., 2012. Using Google street View to audit the built environment: inter-rater reliability results. Ann. Behav. Med. 45 (S1), 108–112. https://doi.org/10.1007/s12160-012-9419-9.

Kepper, M.M., Sothern, M.S., Theall, K.P., Griffiths, L.A., Scribner, R.A., Tseng, T.-S., Schaettle, P., Cwik, J.M., Felker-Kantor, E., Broyles, S.T., 2017. A reliable, feasible method to observe neighborhoods at high spatial resolution. Am. J. Prev. Med. 52 (Suppl. 1), S20–S30. https://doi.org/10.1016/j.amepre.2016.06.010, 1.

Kirtland, K.A., Porter, D.E., Addy, C.L., Neet, M.J., Williams, J.E., Sharpe, P.A., Neff, L.J., Kimsey Jr., C.D., Ainsworth, B.E., 2003. Environmental measures of physical activity supports: perception versus reality. Am. J. Prev. Med. 24 (4), 323–331. https://doi.org/10.1016/S0749-3797(03)00021-7.

Krippendorff, K., 2004. Content Analysis: an Introduction to its Methodology, second ed. Sage, Thousand Oaks, Calif.

Landis, J.R., Koch, G.G., 1977. The measurement of observer agreement for categorical data. Biometrics 33 (1), 159–174. https://doi.org/10.2307/2529310.

Li, W., Milburn, L.-A.S., 2016. The evolution of geodesign as a design and planning tool. Landsc. Urban Plann. 156, 5–8. https://doi.org/10.1016/j.landurbplan.2016.09.009.

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. Activ. 7 (1), 20. https://doi.org/10.1186/1479-5868-7-20.

Moniruzzaman, Md, Páez, A., 2012. A model-based approach to select case sites for walkability audits. Health Place 18 (6), 1323–1334. https://doi.org/10.1016/j.healthplace.2012.09.013.

Owen, N., Humpel, N., Leslie, E., Bauman, A., Sallis, J.F., 2004. Understanding environmental influences on walking: review and research agenda. Am. J. Prev. Med. 27 (1), 67–76. https://doi.org/10.1016/j.amepre.2004.03.006.

Papas, M.A., Alberg, A.J., Ewing, R., Helzlsouer, K.J., Gary, T.L., Klassen, A.C., 2007. The built environment and obesity. Epidemiol. Rev. 29 (1), 129–143. https://doi.org/10.1093/epirev/mxm009.

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. https://doi.org/10.1016/S0749-3797(02)00498-1.

Plascak, J.J., Rundle, A.G., Babel, R.A., Llanos, A.A.M., LaBelle, C.M., Stroup, A.M., Mooney, S.J., 2020. Drop-and-spin virtual neighborhood auditing: assessing built environment for linkage to health studies. Am. J. Prev. Med. 58 (1), 152–160. https://doi.org/10.1016/j.amepre.2019.08.032.

Porter, D.E., Kirtland, K.A., Williams, J.E., Neet, M.J., Ainsworth, B.E., 2004. Considerations for using a geographic information system to assess environmental supports for physical activity. Prev. Chronic Dis. 1 (4). http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1277960/.

Ramirez, L.K.B., Hoehner, C.M., Brownson, R.C., Cook, R., Oreleans, C.T., Hollander, M., Barker, D., Bors, P., Ewing, R., Killingsworth, R., Petersmarck, K., Schimd, T., Wilkinson, W., 2006. Indicators of activity-friendly communities: an evidence-based consensus process. Am. J. Prev. Med. 31 (6), 515–524. https://doi.org/10.1016/j.amepre.2006.07.026.

Rigolon, A., Browning, M., Jennings, V., 2018. Inequities in the quality of urban park systems: an environmental justice investigation of cities in the United States. Landsc. Urban Plann. 178, 156–169. https://doi.org/10.1016/j.landurbplan.2018.05.026.

Rzotkiewicz, A., Pearson, A.L., Dougherty, B.V., Shortridge, A., Wilson, N., 2018. Systematic review of the use of Google Street View in health research: major themes, strengths, weaknesses and possibilities for future research. Health Place 52, 240–246. https://doi.org/10.1016/j.healthplace.2018.07.001.

Sallis, J.F., 2009. Measuring physical activity environments: a brief history. Am. J. Prev. Med. 36 (4, Suppl. ment), S86–S92. https://doi.org/10.1016/j.amepre.2009.01.002.

Sister, C., Wolch, J., Wilson, J., 2010. Got green? Addressing environmental justice in park provision. Geojournal 75 (3), 229–248. https://doi.org/10.1007/s10708-009-9303-8.

Torio, C.M., 2012. The role of the geographic information systems infrastructure in childhood obesity prevention: perspective from the robert wood johnson foundation. Am. J. Prev. Med. 42 (5), 513–515. https://doi.org/10.1016/j.amepre.2012.02.003.

Ussery, E.N., Omura, J.D., Paul, P., Orr, J., Spoon, C., Geremia, C., Carlson, S.A., 2019. Sampling methodology and reliability of a representative walkability audit. Journal of Transport & Health 12, 75–85. https://doi.org/10.1016/j.jth.2018.11.007.

Vanwolleghem, G., Van Dyck, D., Ducheyne, F., De Bourdeaudhuij, I., Cardon, G., 2014. Assessing the environmental characteristics of cycling routes to school: a study on the reliability and validity of a Google Street View-based audit. Int. J. Health Geogr. 13 (1), 19. https://doi.org/10.1186/1476-072X-13-19.

Wilson, J.S., Kelly, C.M., Schootman, M., Baker, E.A., Banerjee, A., Clennin, M., Miller, D. K., 2012. Assessing the built environment using omnidirectional imagery. Am. J. Prev. Med. 42 (2), 193–199. https://doi.org/10.1016/j.amepre.2011.09.029.

Winter, P.L., Padgett, P.E., Milburn, L.-A., Li, W., 2019. Neighborhood Parks and Recreationists' Exposure to Ozone: A Comparision of Disadvantaged and Affluent Communities in Los Angeles, California. Environmental Management 63 (3), 379–395.