


METHODOLOGY

Open Access



Adapting the SPOTLIGHT Virtual Audit Tool to assess food and activity environments relevant for adolescents: a validity and reliability study

Oddbjørn Klomsten Andersen^{1*} , Siobhan A. O'Halloran², Elin Kolle¹, Nanna Lien², Jeroen Lakerveld^{3,4}, Onyebuchi A. Arah^{5,6,7} and Mekdes K. Gebremariam²

Abstract

Background: Physical inactivity and unhealthy diet are key behavioral determinants underlying obesity. The neighborhood environment represents an important arena for modifying these behaviors, and hence reliable and valid tools to measure it are needed. Most existing virtual audit tools have been designed to assess either food or activity environments deemed relevant for adults. Thus, there is a need for a tool that combines the assessment of food and activity environments, and which focuses on aspects of the environment relevant for youth.

Objective: The aims of the present study were: (a) to adapt the SPOTLIGHT Virtual Audit Tool (S-VAT) developed to assess characteristics of the built environment deemed relevant for adults for use in an adolescent population, (b) to assess the tool's inter- and intra-rater reliability, and (c) to assess its criterion validity by comparing the virtual audit to a field audit.

Methods: The tool adaptation was based on literature review and on results of a qualitative survey investigating how adolescents perceived the influence of the environment on dietary and physical activity behaviors. Sixty streets (148 street segments) in six neighborhoods were randomly selected as the study sample. Two raters assessed the inter- and intra-rater reliability and criterion validity, comparing the virtual audit tool to a field audit. The results were presented as percentage agreement and Cohen's kappa (κ).

Results: Intra-rater agreement was found to be moderate to almost perfect ($\kappa = 0.44$ – 0.96) in all categories, except in the category *aesthetics* ($\kappa = 0.40$). Inter-rater agreement between auditors ranged from fair to substantial for all categories ($\kappa = 0.24$ – 0.80). Criterion validity was found to be moderate to almost perfect ($\kappa = 0.56$ – 0.82) for most categories, except *aesthetics* and *grocery stores* ($\kappa = 0.26$ – 0.35).

Conclusion: The adapted version of the S-VAT can be used to provide reliable and valid data on built environment characteristics deemed relevant for physical activity and dietary behavior among adolescents.

Keywords: Built environment, Physical activity, Dietary behavior, Adolescence, Virtual audit

Introduction

The global burden of obesity amongst adolescents has increased more than eightfold from 0.7 to 5.6% in girls and 0.9% to 7.8% in boys, in the past 40 years [1]. Physical inactivity and unhealthy diet are key behavioral

*Correspondence: o.k.andersen@nih.no

¹ Norwegian School of Sport Sciences, Ullevaal Stadion, PO Box 4014, 0806 Oslo, Norway

Full list of author information is available at the end of the article



© The Author(s) 2021. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

determinants underlying obesity [2]. According to ecological frameworks [3, 4], these behaviors are driven by multiple levels of influence, including built environmental factors. Broadly, the built environment can be defined as man-made structures such as neighborhoods, roads, buildings, food sources and recreational facilities in which people live, work, are educated, eat and play [5]. These characteristics are thought to facilitate or impede physical activity (PA) and healthy eating behavior [6–8]. However, very few aspects of the built environment have been consistently associated with these health behaviors among adolescents [9, 10]. Only access to school facilities/resources and access to neighborhood playgrounds and recreational facilities have been found to be associated with PA among adolescents across studies [9, 11]. There are currently no environmental factors considered to be consistently associated with dietary behavior among adolescents [10, 12–14]. The key reasons for these inconsistencies include poor quality of measurement tools, and inconsistencies in the measurement approaches used [11, 13, 15].

In public health research, field audits have traditionally been considered the criterion measurement methodology. A field audit is conducted by systematically walking the streets or parks of a given neighborhood to assess the specific characteristics with an audit tool, often in form of a checklist [16]. The main advantage of field audits is that they can provide highly detailed information about certain aspects of the built environment such as the quality and maintenance of recreational facilities and landscapes, which are rarely incorporated in secondhand data sources, such as Geographical Information Systems (GIS) [16]. However, field audits can be costly and time-consuming as they require the auditor to be physically present at the site [16]. Furthermore, concerns related to traffic and crime can make field audits impractical and unsafe in certain areas [17]. Given these concerns, researchers have opted for the use of virtual street audits.

In a virtual audit, the auditor or researcher make use of online cartographic platforms that also provides 360° photo views, such as Bing Maps or Google Street View (GSV) to walk through a neighborhood virtually [18]. Due to its vast coverage, GSV is usually the preferred platform [19]. Compared to a field audit, a virtual audit is thought to be less costly, less time-consuming and safer since the audit is performed from a personal computer [18]. Furthermore, it allows researchers to audit remote locations without being physically present at the site. Since the images are from a fixed point in time it enables the researcher to go back and validate the audits. However, there are some GSV challenges related to the limited image coverage in rural areas and the temporal validity of the images, which can make it difficult to

measure the dynamic features of the built environment, such as aesthetics [20]. Google is constantly improving their GSV coverage and image quality, and a recent systematic review concluded that GSV thus far has been applied successfully in health-related research [18].

Existing virtual audit tools were mostly designed to assess built environmental characteristics deemed relevant for adults in North American or European settings and usually focus on either PA or dietary behavior [18]. However, PA and dietary behaviors are complex and interrelated [21]. It is therefore recommended to combine the measurements of food and activity environments to gain better insight into the association between the built environment and obesity [22]. The SPOTLIGHT Virtual Audit Tool (S-VAT) was the first tool to combine food and PA environments [23] and was originally designed to assess obesogenic characteristics of the built environment considered relevant for adults. The tool displayed high reliability and validity when it was tested in a Dutch setting [23]. The S-VAT was developed as a part of the large EU-funded SPOTLIGHT project [24], and has been used to investigate the association between obesogenic environments and health behavior among adults in five urban regions in France, Hungary, the United Kingdom, Belgium and the Netherlands [22, 25–27].

An important finding derived from the SPOTLIGHT project was the substantial differences in the neighborhood typologies across countries, despite the authors sampling strategies to minimize heterogeneity in neighborhoods [22]. This indicates that there are substantial inter-country differences in the built environment even within Europe. The finding might be particularly relevant for Nordic countries, as the climate allows for different activities during the winter months (e.g. cross-country skiing). To meet this challenge it has been suggested that existing measurement tools should be complemented with country-specific items to capture aspects of the environment that is geographically dependent [28]. Furthermore, the characteristics of the built environment deemed most relevant for PA and dietary behavior are likely to differ across age groups [9]. While walkability and diversity in land-use seems to be most predictive of PA behavior among adults, the presence of school infrastructure/equipment and specific recreational facilities in the neighborhood seems to be more important among adolescents [9, 29]. Thus, existing tools such as the S-VAT may only include items that reflect a limited range of exposure variability, which might not be generalizable to other settings [22, 28] or to adolescents [9, 29]. Hence, further development, adaptations, and validation seem warranted.

The aims of the present study were: (a) to adapt the S-VAT developed to assess characteristics of the built

environment deemed relevant for adults for use in an adolescent population, (b) to assess the tool's inter- and intra-reliability, and (c) to assess its criterion validity by comparing the virtual audit to a field audit. It is hypothesized that the adapted version of the S-VAT will achieve reliable and valid results, comparable to the original S-VAT tool [23].

Methods

The tool selected for the adaptation process was the S-VAT [23]. The S-VAT utilizes GSV in Google Earth (GE) and has 42 items (two duplicates) divided in eight categories: six walking related items (e.g. presence of sidewalk), eight cycling related items (e.g. presence of bicycle lane), two public transport items (e.g. presence of bus/tram stop), nine aesthetics items (e.g. presence of litter/graffiti), three land use-mix (e.g. type of residential buildings), five grocery store items (e.g. presence of supermarket), six food outlet items (e.g. presence of fast food restaurants), and three recreational facility items (e.g. presence of outdoor recreational facilities) [23]. We chose to keep the eight categories as they represent the main components of the built environment thought to influence health behavior [11], and to include or modify a number of items depending on relevance for adolescents, aged 12–17 years old.

Adaptation process

The adaptation of the S-VAT tool was performed in a three-step process. First, a literature review of existing field and virtual audit tools, and environmental correlates and determinants of PA and dietary behavior was conducted to identify built environment aspects specific to adolescents, aged 12–17 years old [9, 10, 23, 29–33]. Emerging topics derived from the literature review were the importance of access to school facilities and specific PA infrastructure/equipment in the neighborhood [9, 29]. In addition, the attractiveness of the recreational facilities and parks was reported to be relevant for PA behavior and public open space visitation [29]. No additional environmental determinants were found for dietary behavior [10]. Thereafter, to identify built environment aspects specific to the geographical context, results from a qualitative survey which investigated how adolescents aged 13–14 years old from Oslo perceived the influence of the environment on dietary and PA behaviors were used (submitted work). In concordance with the first step, adolescents reported that certain recreational facilities, specifically soccer fields, were important for PA behavior. In addition, the presence of a forest was reported to be an important arena for cross-country skiing, which is the second largest sport in Norway [34]. No food outlets or stores were mentioned by the adolescents outside the

already existing items in the S-VAT tool. Finally, items from the original S-VAT tool were amended accordingly.

Two duplicates and the items *public bicycle renting facilities* and *wine liquor store* were removed. The reasons for removal was either because the item was repetitive, or the item was regarded as irrelevant for adolescents or for the Norwegian context. Based on the results from the literature review and the qualitative survey, we chose to add a substantial number of items in the category *recreational facilities*. This included seven common outdoor recreational facilities and three indoor recreational facilities. The quality of the outdoor recreational facilities was added in separate items. In addition, the items *school*, *presence of forest*, *youth club*, *bakery* were added to the tool. The items *convenience store/grocery store* and *Cafe/Bar* was split into four individual categories, and *take away restaurants* was disaggregated into *take away* (pizza, kebab, hamburger) and *take away—others* (Indian, Asian food etc.). As more than one type of *traffic calming device* and *type of residential building* can be observed in each segment, these items were separated into individual categories. A complete list of added or modified items and their scoring/rating is available in Additional file 1.

The final tool included a total of 73 items (9 walking related items, 6 cycling related items, 2 public transport items, 13 aesthetics items, 11 land-use mix items, 6 grocery store items, 7 food outlets and 19 recreational facility items). The complete list of items is available in Additional files 2 and 3. In line with Bethlehem and colleagues [23], we adapted their standard operational procedure (SOP). The SOP is an instruction manual with detailed descriptions on how to rate each individual item. It was developed to minimize heterogeneity between audits and inconsistent scoring. For more detail, see the original S-VAT study [23]. The modifications made to the original SOP are available in Additional file 1.

The first and second authors conducted the audit using the adapted tool. From this point on, they are referred to as auditor 1 and auditor 2. Auditor 1 was a Norwegian researcher with good familiarization of the study area, while auditor 2 was an Australian researcher who was less familiar with the study area. Both auditors were trained in the use of the tool. The auditor training was conducted using the adapted SOP. Prior and during the training, the auditors were in contact with the project leader (JL) of the original S-VAT study, who gave instructions on how to rate the original items. The training was conducted by assessing a total of 36 random street segments in Oslo with various residential density (RD) and socioeconomic position (SEP). In accordance with Krieger et al. [35] the term SEP is used instead of socioeconomic status (SES), as SES does not clearly distinguish between actual resources and status. Inter-rater agreement was assessed

mid-way and after completion of all the street segments. Any disagreements and misunderstandings related to the tool and its concepts were solved by discussions and further refinements of the adapted SOP to maximize homogeneity between the raters.

Selection of study area and data collection

Sixty streets in six neighborhoods (ten in each neighborhood) in Oslo, Norway with varying RD and SEP were randomly selected as the study sample. The capital city of Norway was selected as it has a high ethnic, socioeconomic and environmental variability compared to other cities/regions [36] and the food and activity environments are likely to be more diverse compared to other cities in Norway. Oslo consists of 96 administratively defined neighborhoods. We stratified the neighborhoods in six groups based on RD and neighborhood SEP to maximize exposure variability. Residential density was obtained from Statistics Norway [37] and defined as number of inhabitants per square kilometer (inhabitants/km²) and stratified into tertiles (high, medium, low). Socioeconomic position was obtained from the Oslo municipality website [38], defined by the percent of the population (>16 years) with higher education (minimum bachelor's degree) and stratified into two groups based on median education (high/low). One neighborhood from each stratum (high SEP/low RD, high SEP/medium RD, high SEP/low RD, low SEP/high RD, low SEP/medium RD, low SEP/low RD) was randomly selected. All streets in the selected neighborhoods were identified through GE and Google Maps (GM) and ten streets from each neighborhood were randomly selected as the final study sample. The number of streets selected represented $\geq 25\%$ of all streets identified in each respective neighborhood which has previously been suggested to be sufficient to assess the neighborhood built environment exposure [39]. All randomization was conducted using an online random number generator [40].

In line with previous studies [23, 41], the streets were further divided into street segments to ensure comparability. A street segment was defined as the length between two four-way intersections, with a minimum length of 50 m and no longer than 300 m. With the absence of a four-way intersection, an arbitrary cut-off was used at approximately 300 m. If a street crossed neighborhood boundaries, the street was assessed for the entire length, or at a maximum length of 300 m outside neighborhood borders. The streets and street segments were measured and drawn in GE using the *add path* function. This allowed for the street segments to be clearly visible in the GSV mode. Data were recorded in a Microsoft Excel spreadsheet and the data validation function was applied to mitigate typographical errors (Fig. 1). Food

outlets and recreational facilities were pinpointed with x and y coordinates to enable the possibility to add them as a layer in a Geographical Information System (GIS) tool such as ArcGIS Pro. Both sides of each street segment were audited. A feature was recorded as present if it was available on minimum one side of the road (e.g. *sidewalk present*), unless otherwise stated (e.g. *sidewalk present on both side of the road*). If the quality of a feature varied across a street segment (e.g. *quality of residential gardens*), an overall quality score was given (e.g. *does residential gardens generally look trimmed and clean*). In streets with GSV images available in both driving directions, the most recent image was used to audit both sides of the road. If the timestamps were identical, both images were consulted. Times taken to complete the virtual and field audit was recorded by auditor 1. Cost analysis was conducted based on salary and expenses related to transportation (car rental, parking, fuel and bus tickets) for the field audit.

Inter-rater reliability

The first virtual audit was conducted in the period from 9th to 16th July 2019. During the first audit, each auditor assessed 60 streets, in the six neighborhoods with varying RD and SEP independently. Inter-rater reliability was then assessed by comparing the results from the first audit of the two auditors.

Intra-rater reliability

A second virtual audit of the same streets and street segments was conducted in the time period 30th July–6th of August 2019. Intra-rater reliability was assessed by comparing the results from the first audit to the second audit of the same 60 streets in the same six neighborhoods. To minimize recall bias, during the second audit the streets were audited in the reverse direction a minimum of 14 days after the first audit. Intra-rater reliability was performed for both auditors.

Criterion validity

A field audit of the same streets and street segments was conducted in the time period 20th August–3rd September 2019. Criterion validity was assessed by comparing the results from the first virtual audit to the field audit. In the field audits, the auditor walked the same 60 streets in the same six neighborhoods and systematically assessing the same characteristics of interest with the use of two Apple Pro iPads®. Comparable to the virtual audit, GE was used to identify the streets and street segments, and a Microsoft Excel spreadsheet with data validation was used to record the results. Both auditors conducted the field audit.

Neighborhood		Cars form obstacles on road	Traffic calming devices	Bus/tram stop	Railway/ underground station	Green and/or water area visable	Maintenance of green areas
Street 1	Segment A	Yes	No	No	No	Yes	Well maintained
	Segment B	Yes	No	No	No	Yes	Well maintained
Street 2	Segment A	Yes	No	No	No	Yes	Not well maintained
	Segment B	Yes	No	No	No	Yes	Not well maintained
Street 3	Segment A	No	Yes	Yes	No	Yes	Well maintained
	Segment B	No	No	No	No	Yes	Well maintained
	Segment C	No	Yes	No	No	Yes	Well maintained
	Segment D	No	Yes	No	No	Yes	Well maintained
	Segment E	No	Yes	No	No	Yes	Well maintained
Street 4	Segment A	Yes	Yes	No	No	Yes	Well maintained
	Segment B	Yes	Yes	No	No	Yes	Well maintained
Street 5	Segment A	No	Yes	Yes	No	Yes	Well maintained
	Segment B	No	No	No	No	Yes	Well maintained
	Segment C	No	No	No	Yes	Yes	Well maintained
	Segment D	Yes	Yes	Yes	No	Yes	Well maintained
	Segment E	Yes	Yes	No	No	Yes	Well maintained
	Segment F	Yes	Yes	No	No	Yes	Well maintained
	Segment G	Yes	Yes	No	No	Yes	Well maintained
	Segment H	Yes	No	No	Yes	Yes	Not well maintained
Street 6	Segment A	No	No	Yes	No	Yes	Well maintained
	Segment B	No	Yes	No	No	Yes	Not well maintained
	Segment C	No	Yes	Yes	No	Yes	Well maintained
Street 7	Segment A	No	Yes	No	No	Yes	Well maintained
	Segment B	Yes	Yes	No	No	Yes	Not well maintained
	Segment C	No	No	No	No	Yes	Not well maintained
	Segment D	No	Yes	No	No	Yes	Not well maintained
Street 8	Segment A	No	No	No	No	Yes	Well maintained
	Segment B	No	No	No	No	Yes	Well maintained
Street 9	Segment A	Yes	No	No	No	No	Not present
Street 10	Segment A	Yes	Yes	No	No	Yes	Well maintained
	Segment B	Yes	Yes	No	No	No	Not present
	Segment C	Yes	No	No	No	No	Not present

Fig. 1 Example data extraction sheet

Statistical analyses

Cohens’ kappa (κ) and weighted kappa were used to determine levels of agreement in dichotomous and categorical variables, respectively. In accordance with Landis and Koch [42], the following cut-off values were applied to determine agreement; <0.2 poor agreement, 0.21–0.40 fair agreement, 0.41–0.60 moderate agreement, 0.61–0.80 substantial agreement, 0.81–1.00 almost perfect agreement. Landis and Koch kappa cut-offs are commonly used in built environment research [20, 23, 32, 43–45]. As Cohen’s kappa is a relative measure and a low kappa can be observed despite high agreement, we also reported percent agreement [46, 47]. Kappa values were not reported in items with zero

observations. To determine homogeneity, the asymptomatic McNemar- and Stuart-Maxwell-test were performed on dichotomous and categorical variables, respectively [48–50]. An alpha value of 0.05 was used to determine statistical significance. All statistical analyses were performed with Stata/SE 16.0 (StataCorp LLC) statistical software.

Results

A total of 181 streets were identified in the six neighborhoods. Thirty-four streets were excluded due to lack of GSV imaging. Of the remaining 147 streets, 60 streets were randomly selected and included as the final sample. The audits were completed during the summer months

(July–August 2019). The GSV images were on average 5 years and 3 months old (± 3 years and 6 months) when the first virtual audit was completed (07.2019). The oldest images were from 2009 and the most recent from 2017. All the GSV images were from the months May–October. Average time taken to complete a virtual and a field audit of a neighborhood was 181 min and 40 s, and 214 min and 10 s, respectively. This included time taken to walk between streets in the field audit. The field audit was associated with an additional cost of 4818 NOK (511 USD) in salary and transportation expenses per auditor.

Due to few observations, the categories *take away (pizza, kebab, hamburger)* and *take away others (Indian, Asian etc.)* were merged into the category *take away restaurants* in the analysis. For similar reasons, the items *volleyball court, tennis court, basketball courts skate/BMX park, outdoor fitness facilities* and *other sports fields* were merged into the category *other outdoor facilities*. Descriptive statistics for each neighborhood are presented in Table 1. The number of observations per individual item for both auditors in each neighborhood is provided in Additional files 2 and 3.

Inter-rater reliability

Mean level of agreement between auditors ranged from fair to substantial for all categories ($\kappa = 0.24$ – 0.80). Highest agreement was found for *walking related items, public transport, and food outlets* ($\kappa = 0.62$ – 0.80) while the lowest agreement was found for *aesthetics* ($\kappa = 0.24$). Inter-rater reliability results are presented in Table 2.

Intra-rater reliability

Intra-rater agreement was found to be substantial to almost perfect ($\kappa = 0.62$ – 0.96) in all categories for auditor 1. For auditor 2 this was moderate to substantial for all categories ($\kappa = 0.44$ – 0.63), except *aesthetics* where fair agreement was found ($\kappa = 0.40$). Intra-rater reliability results for both auditors are presented in Table 2.

Criterion validity

Moderate to almost perfect agreement ($\kappa = 0.56$ – 0.82) was found for all categories except *aesthetics* ($\kappa = 0.30$) by auditor 1. Auditor 2 found moderate to substantial agreement for all categories ($\kappa = 0.50$ – 0.72), except *aesthetics* and *grocery stores* ($\kappa = 0.26$ – 0.35). Criterion validity results for both auditors are presented in Table 2.

Test for homogeneity

Tests for homogeneity were conducted for inter-reliability and criterion validity. Auditor 1 was more likely to rate *condition of sidewalks* as good ($p \leq 0.05$), and more likely to observe *green and water areas, graffiti* and *trees*, compared to auditor 2 ($p \leq 0.05$). Finally, higher percentage of *commercial and industrial buildings* were more likely to be observed by auditor 2 ($p \leq 0.05$). Regarding criterion validity, both auditors were more likely to report poorer *quality of sidewalk* in the field audit compared to the virtual audit ($p \leq 0.05$). Furthermore, both auditors reported more *traffic calming devices* in the field audit compared to the virtual audit ($p \leq 0.05$). Finally, *more apartment buildings > 5 floors* were observed in the virtual audit compared to the field audit ($p \leq 0.05$) by both auditors.

Discussion

The main aims of the study were to adapt the S-VAT to assess the food and activity environments relevant for adolescents and to investigate the reliability and validity of the tool. In accordance with the hypothesis, the adapted S-VAT was found to provide reliable and valid data for most built environmental characteristics deemed relevant for PA and dietary behaviors in adolescents. The results are in agreement with previous research suggesting that virtual audit tools generally have adequate reliability and validity and are useful for assessing neighborhood characteristics [17, 18, 51–54]. Slightly lower levels of agreement were observed by auditor 2. Limited exposure variability was observed for the presence of *cycling related items, grocery stores, food outlets and*

Table 1 Neighborhood characteristics

Neighborhood	High socioeconomic position			Low socioeconomic position		
	HRD	MRD	LRD	HRD	MRD	LRD
RD (inhabitants/km ²)	7140	3784	3566	18,913	4694	1443
SEP (% higher education ≥ 16 years)	61	63	65	53	29	28
Number of street segments	23	31	28	16	23	27
Avg. street segment length (m)	216 (75)	231 (71)	198 (86)	178 (59)	241 (61)	260 (62)
Distance audited (m)	4969	7388	5741	2849	5537	7029
Avg. GSV picture date (mm.yyyy)	03.2016	03.2012	08.2016	05.2016	01.2013	09.2012

RD residential density, SEP socioeconomic position, HRD high RD, MRD medium RD, LRD low RD, GSV Google Street View

Table 2 Inter- and intra-rater reliability and criterion validity for all items

Category	Inter-rater reliability		Intra-rater reliability		Intra-rater reliability		Intra-rater reliability		Criterion validity		Criterion validity	
	% Agreement	Kappa	% Agreement	Kappa	% Agreement	Kappa	% Agreement	Kappa	% Agreement	Kappa	% Agreement	Kappa
Walking related items ^a	90	0.69	96	0.80	89	0.63	92	0.72	88	0.63	88	0.63
Type of street (4 categories)	84	0.69	90	0.83	89	0.77	82	0.67	86	0.70	86	0.70
Presence of sidewalks (yes/no)	93	0.84	99	0.97	92	0.81	97	0.92	93	0.83	93	0.83
Sidewalk on both sides (yes/no)	85	0.70	96	0.92	83	0.65	95	0.89	87	0.74	87	0.74
Pedestrian crossing available (yes/no)	86	0.72	93	0.86	87	0.73	91	0.82	87	0.72	87	0.72
Zebra-path (yes/no)	93	0.85	97	0.93	90	0.76	90	0.79	78	0.49	78	0.49
Traffic lights (yes/no)	95	0.88	97	0.93	94	0.87	95	0.89	92	0.83	92	0.83
Over/under pass (yes/no)	95	0.77	98	0.90	92	0.56	95	0.77	96	0.81	96	0.81
Presence of streetlights (yes/no)	95	0	99	0	95	-0.01	99	0	93	-0.03	93	-0.03
Cars form obstacles on the road (yes/no)	87	0.74	95	0.89	77	0.54	85	0.70	81	0.60	81	0.60
Cycling related items ^a	95	0.48	99	0.96	95	0.47	94	0.77	92	0.50	92	0.50
Presence of bicycle lane (yes/no)	98	-0.01	100	1	98	-0.01	100	1	96	0.27	96	0.27
Speed limit	93	0.90	100	1	97	0.96	81	0.70	85	0.85	85	0.85
Type of bicycle lane (yes/no)	100	N/A	100	N/A	100	N/A	100	N/A	100	N/A	100	N/A
Traffic calming devices (yes/no)	78	0.55	94	0.88	75	0.47	80	0.61	68	0.38	68	0.38
Condition of bicycle lane (3. cat.)	100	N/A	100	N/A	100	N/A	100	N/A	100	N/A	100	N/A
Obstacles present bicycle lane (yes/no)	100	N/A	100	N/A	100	N/A	100	N/A	100	N/A	100	N/A
Public transport ^a	97	0.80	99	0.91	94	0.59	96	0.82	94	0.72	94	0.72
Bus/tram stop (yes/no)	94	0.81	98	0.93	90	0.68	93	0.75	91	0.74	91	0.74
Railway/underground station (yes/no)	99	0.79	99	0.89	97	0.49	99	0.89	97	0.70	97	0.70
Aesthetics ^a	81	0.24	91	0.62	87	0.40	84	0.30	82	0.26	82	0.26
Green/water area available (yes/no)	81	0.38	95	0.67	79	0.38	84	0.31	81	0.44	81	0.44
Park (yes/no)	99	-0.01	98	0.39	96	0.55	98	0.39	95	0.40	95	0.40
Condition of park (3. cat.)	100	N/A	100	N/A	100	1	100	N/A	100	0	100	0
Residential gardens visible (yes/no)	90	0.74	95	0.86	84	0.58	92	0.78	87	0.62	87	0.62
Rating of residential gardens (well maintained/not well maintained)	90	-0.03	99	0.80	98	-0.01	95	-0.02	93	-0.02	93	-0.02
Rating of most residential buildings (good/poor)	99	0	100	1	100	N/A	98	N/A	100	N/A	100	N/A
Open vacant area/parking lot available (yes/no)	73	0.46	88	0.74	77	0.52	83	0.64	75	0.49	75	0.49
Maintenance of green areas (well maintained/not well maintained)	45	0.06	82	0.58	80	0.26	47	-0.15	40	0.06	40	0.06
Condition of sidewalks (3. cat.)	55	0.12	80	0.34	67	0.19	68	0.15	48	0.01	48	0.01
Presence of graffiti (yes/no)	54	0.15	78	0.55	79	0.13	64	0.29	76	0.28	76	0.28
Presence of litter (yes/no)	86	0.07	83	0.41	95	-0.02	78	0.11	97	-0.02	97	-0.02
Presence of trees (yes/no)	88	0.21	97	0.43	85	0.23	95	0.31	89	0.23	89	0.23

Table 2 (continued)

Category	Inter-rater reliability		Intra-rater reliability Rater 1		Intra-rater reliability Rater 2		Criterion validity Rater 1		Criterion validity Rater 2	
	% Agreement	Kappa	% Agreement	Kappa	% Agreement	Kappa	% Agreement	Kappa	% Agreement	Kappa
	Presence of forest (yes/no)	91	0.69	91	0.71	88	0.52	87	0.52	90
Land use-mix ^a	92	0.55	95	0.70	92	0.54	94	0.65	92	0.57
Presence of residential buildings (yes/no)	93	0.79	99	0.96	87	0.66	93	0.80	90	0.70
Detached/semidetached homes (yes/no)	90	0.80	97	0.95	92	0.83	95	0.89	89	0.77
Terraced homes (yes/no)	85	0.44	83	0.56	90	0.50	84	0.54	86	0.37
Apartment buildings ≤ 5 stories (yes/no)	81	0.61	92	0.84	78	0.55	89	0.77	80	0.60
Apartment buildings > 5 stories (yes/no)	96	0.76	97	0.86	92	0.56	97	0.78	96	0.80
Apartment above shops (yes/no)	97	0.65	99	0.80	96	0.49	97	0.65	95	0.52
% of commercial buildings (5 categories)	83	0.29	95	0.39	83	0.46	93	0.38	83	0.47
% of industrial buildings (5 categories)	92	0.18	93	0.31	100	0.80	94	0.29	99	0.80
School (yes/no)	95	0.61	95	0.47	95	0.53	96	0.55	95	0.69
Youth clubs (yes/no)	100	N/A	100	N/A	100	N/A	100	N/A	100	N/A
Shopping mall (yes/no)	96	0.32	99	0.85	97	0	99	0.85	97	0
Grocery stores ^a	99	0.55	99	0.65	99	0.44	99	0.56	98	0.35
Supermarket (number)	99	0.96	95	0.74	98	0.77	93	0.56	92	0.43
Local food shop (number)	99	0	99	0	100	N/A	99	0.66	99	0
Bakery (number)	100	1	99	0	100	1	99	-0.01	99	0.66
Street food market (number)	100	N/A	100	N/A	100	N/A	100	N/A	100	N/A
Small grocery store (number)	99	0	98	-0.01	99	0	97	-0.01	99	0
Convenience store (number)	99	0.80	99	0.66	97	-0.01	100	1	99	0.66
Food outlets ^a	100	0.62	100	0.92	100	0.51	100	0.74	99	0.56
Restaurants (number)	99	0.57	100	1	99	0.80	99	0.80	99	0.80
Fast food outlets (number)	99	0	100	N/A	99	0	100	N/A	99	0
Take away restaurants (number)	100	0.89	100	1	100	0.75	99	0.66	97	0.59
On street vendors (number)	100	N/A	100	N/A	100	N/A	100	N/A	100	N/A
Café (number)	99	0.66	99	0.66	99	-0.01	99	0.49	98	0.39
Bar/Pub (number)	100	1	100	1	100	1	100	1	100	1
Recreational facilities ^a	81	0.26	94	0.71	94	0.48	87	0.60	98	0.59
Playground (number)	90	0.50	91	0.66	96	0.65	92	0.65	95	0.64
Condition playground (3. cat.)	43	0	91	0.61	83	0	43	-0.17	100	N/A
Soccer field (number)	97	0.69	98	0.71	99	0.85	95	0.59	96	0.55
Condition soccer field (3. cat.)	43	0	91	0.61	83	0	88	0.60	100	1

Table 2 (continued)

Category	Inter-rater reliability		Intra-rater reliability Rater 1		Intra-rater reliability Rater 2		Criterion validity Rater 1		Criterion validity Rater 2	
	% Agreement	Kappa	% Agreement	Kappa	% Agreement	Kappa	% Agreement	Kappa	% Agreement	Kappa
Other outdoor sports facilities (number)	98	0.39	99	0.89	100	0.89	95	0.66	97	0.16
Condition other sports facilities (3. cat.)	100	N/A	88	0.67	100	N/A	100	1	100	N/A
Indoor facilities (number)	98	0	99	0.85	100	N/A	99	0.85	100	N/A

N/A not applicable

^a Mean category agreement

recreational facilities, which could explain the variability in kappa values despite almost perfect percent agreement in most items in these categories. Finally, low criterion validity was observed for *aesthetics* by both auditors indicating that the tool might be less suitable to assess this feature.

Compared to the original S-VAT, the most substantial modifications were made in the category *recreational facilities*. The decision to make these modifications were informed by existing literature and the qualitative study conducted prior to the adaptation process (submitted work). The literature suggests that certain types of recreational facilities and adjacent quality, might be of particular importance for PA behavior among adolescents [9, 29]. As the original tool only included two general items for recreational facilities (e.g. presence of *indoor recreational facility* and *presence of outdoor recreational facilities*), it was considered relevant and important to separate the more specific components. Determining which key recreational facilities are thought to be relevant for PA behavior among adolescents can potentially lead to the development of targeted interventions in the built environment [55].

Consistently high reliability and validity was found for the categories *walking related items* and *public transport*. This is in accordance with previous research [23, 56]. Most items were of a dichotomous nature (e.g. present/not present) and items such as *type of road*, *presence of sidewalks* and *pedestrian crossing* are clearly visible and can easily be identified in GSV. Thus, little subjective judgement is likely to occur, and high levels of agreement is expected. Some authors have pointed to the difficulties of identifying sidewalk characteristics due to parked cars blocking the view from the GSV images [57, 58]. However, this was not a major issue in the present study, possibly because the adapted S-VAT does not assess sidewalk width and curb type. Importantly, the *walking related items* and the *public transport items* are likely to be relatively stable over time. This could explain the high criterion validity despite the relatively large temporal discrepancy between when the GSV images were taken and when the field audit was carried out. However, it should be mentioned that both auditors observed significantly more *traffic calming devices* in the field audit compared to the virtual audit. This is probably due to a real difference between the photos and the field audit due to new policies in Oslo.

The lowest overall agreement for any category was observed for *aesthetics*. This finding is not surprising and has been reported by several other researchers [20, 23, 51, 52, 59, 60]. There are multiple issues related to the assessment of *aesthetics* in GSV. Many of the items in the *aesthetics* category are subjected to

smaller nuances and subjective judgment [47, 61, 62]. Auditors can have different perceptions of what constitutes “good”, “fair” and “poor” quality, leading to systematic differences in ratings. However, the level of subjective judgment in this study was mitigated by utilizing an adapted version of the SOP used by Bethlehem et al. [23]. Nevertheless, some systematic differences between auditors did occur. For instance, *graffiti* was more frequently rated as present by auditor 1 than auditor 2, which suggests that the SOP was unsuccessfully standardized for all items.

Other issues are more directly related to GSV as a measurement tool. For instance, *condition of sidewalk* was more likely to be rated as poor in the field audit, compared to the virtual audit by both auditors which could be attributed to annual decay or poor image quality. Furthermore, the temporal validity of the images, the fixed point of the camera and poor image quality can make it challenging to assess items such as litter and graffiti [61]. The GSV images in the present study were on average five years old, which most likely affected the temporal validity of the images, resulting in poor criterion validity for both auditors. Given the potential, inherent within GSV, to derive highly detailed information about aesthetics, it is unfortunate that the adapted version of the S-VAT tool and similar tools show poor reliability and validity for many items in this category [19, 20, 59].

The categories *cycling related items*, *grocery stores*, *food outlets* and *recreational facilities* displayed high percentage agreement, but inconstant kappa values. High percentage agreement despite low kappa values can be observed when there is limited exposure variability. This paradox is referred to as the prevalence problem and has been observed in these categories in previous studies [20, 56]. It is attributed to the different mathematical properties of the statistical tests. While percent agreement is an absolute measure of agreement, the kappa statistics is a relative measure which also accounts for chance agreement. High agreement is expected by chance when there is low variability in the observations. By contrast, lower agreement is expected by chance when there is high variability in the observations [47]. Thus, a discordant pair has a more punitive effect on the kappa value when the variability is low. To bypass this issue, some researchers have reported the prevalence and biased adjusted Kappa (PABAK) [58, 59, 63–65]. However, the PABAK has been criticized as it tends to give disproportionate high Kappa values [66]. It has been proposed that in cases where the prevalence problem occurs and the auditors are well trained and items are of a dichotomous nature, more trust can be placed on percent agreement [47]. Nevertheless, the results in these categories should be interpreted with some caution.

A larger sample of streets could potentially have allowed to determine the actual agreement in the categories *cycling related items*, *grocery stores*, *food outlets* and *recreational facilities* more accurately. However, the adapted version of the S-VAT is a community tool designed to assess a vast range of exposures in residential neighborhoods. Thus, it is reasonable to assume that certain items in these categories indeed occur less frequently in the environment. This finding is supported by the results from the original S-VAT validity and reliability study, that also found low exposure variability in these categories [23]. However, low exposure variability in *cycling related items* was not observed in that study. This finding could be due to cultural differences between Norway and the Netherlands. For example, the Netherlands is especially known for its bicycle infrastructure and cycling popularity [67]. Whereas in Norway, cycling is gradually becoming more common, together with the Norwegian government committing to greater funding to support the use of bicycles, recreationally and as a mode of transport [68]. Nevertheless, when present, these features can have an important impact on adolescents' PA and dietary behaviors [69–71]. Hence, it was considered as an important and appropriate addition to the existing tool.

Compared to intra-rater reliability, lower levels of agreement were observed for inter-rater agreement. This may be because auditor 2 was non-Norwegian and less familiar with the study area, which could indicate that there are some cultural barriers to achieving high levels of agreement. For instance, there was observed systematic differences in the ratings of quality of sidewalk between raters which could reflect differences in sidewalk quality between Australia and Norway. Furthermore, slightly lower intra-rater agreement was observed in certain items for auditor 2 compared to auditor 1. Language barriers might partially explain this issue, especially when assessing the food and activity environment, as the purpose of a building might not be obvious without being able to translate the language written on the building, leading to more guesswork. Unfortunately, as the auditors did not have any contact during the virtual audits, this was not examined thoroughly. However, discussions after the audits, revealed issues that related to understanding the Norwegian text on certain indoor recreational facilities. Notably, one of the major advantages of the virtual audit tools is the possibility to audit remote locations, thus it is important to be aware of these language translation issues.

Previous research has suggested that differences in seasonal climate can influence levels of agreement [61]. This is an important issue to address in countries with large variation in climate between the summer and winter.

All the GSV images had timestamps from May–October, when the climate in Norway is quite consistent during the warmer months. However, Google has recently updated its GSV images in certain streets included in the present study and although this should theoretically increase criterion validity, some of these images were taken during snowy weather conditions in the winter. This can have major implications on criterion validity as it makes it impossible to assess certain aspects of the environment such as sidewalk characteristics, litter, quality of residential gardens, surface of sports fields, etc. Fortunately, Google has created the option of viewing older images from the same location. However, this increases the temporal discrepancy between the images obtained from GSV and the field audit which can lead to an invalid representation of the actual environment. Nevertheless, Google reports when and where they are going next, which can give researchers some predictability when planning a study in these regions [72].

Strength and weaknesses

The present study included a randomly selected and diverse sample of neighborhoods stratified by SEP and RD, which is likely to be representative of neighborhoods in Oslo. Further validation studies in more rural areas should shed light on the generalizability of the tool's assessments. The qualitative work conducted prior to the tool development ensured that items that are culturally and geographically specific were included. Most modifications made to the S-VAT tool were informed by international studies on environmental correlates and determinants of PA and dietary behavior deemed relevant for adolescents [9, 10, 23, 29–33]. Thus, it is reasonable to assume that the tool is generalizable to cities outside of Norway. Nevertheless, it is recommended that researchers use the present tool in combination with items that are specific for the study region in question [28]. The inclusion of auditors with different nationalities and different degree of familiarization with the study area can be considered a strength as it highlights both the tools robustness and weakness to cultural and language barriers. All reliability and validity assessments were rigorously conducted by both auditors, which further strengthen our results.

The present study has some weaknesses that should be considered. Thirty-four streets in the selected neighborhoods lacked GSV images and were thus not eligible for inclusion. These were mostly private streets leading up to private residences or very small public streets in residential areas where cars are prohibited from driving. While private streets are unlikely to influence walking and cycling behavior of adolescents in general, the small streets in residential areas that were inaccessible

by the Google car can be particularly appealing for walking and cycling behavior.

The results indicate that the SOP was insufficiently standardized for certain items in the *aesthetics* category. It is likely that this contributed to poor inter-rater agreement in this category. However, auditor 1 had substantial intra-rater agreement and still displayed poor criterion validity. Thus, the poor criterion validity in the *aesthetics* category is most likely attributable to the temporal validity of the images and not to unsuccessful standardization of the SOP. This is a common problem reported in most GSV tools [18, 19]. The images in the present study were on average more than 5 years old, and certain images from smaller roads and more rural areas were 10 years old. Previous research has indicated that non-arterial streets are more likely to lack photos or have outdated images, compared to more urban areas [19, 73]. However, this problem could be partially mitigated since rural environments are thought to be more consistent than urban environments [51]. Indeed, the auditor most familiar with the study area achieved moderate to perfect criterion validity for all categories except *aesthetics*. This suggests that there was no significant change in most of the assessed environmental characteristics over time. Thus, the tool appears to be robust when it comes to tackling temporal variability.

Conclusion

Substantial modifications were made to the original S-VAT to make it suitable to measure the characteristics of the built environment deemed relevant for adolescents. The most significant alterations were made in the category *recreational facilities*. The adapted version of the S-VAT can be used to provide reliable and valid data on most neighborhood characteristics deemed relevant for PA and dietary behavior among adolescents. It is recommended that future studies use the present tool in combination with items that are specific for the study region in question. Further validation studies in more rural areas should shed light on the generalizability of the tool's assessments.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12942-021-00258-0>.

Additional file 1: Table S1. Modified or added items to the original S-VAT tool and how to rate them.

Additional file 2: Table S2. Prevalence (%) of all items, across different neighborhood types (based on first auditors result).

Additional file 3: Table S3. Prevalence (%) of all items, across different neighborhood types (based on second auditors result).

Abbreviations

GE: Google Earth; GIS: Geographical Information System; GM: Google Maps; GSV: Google Street View; PA: Physical activity; PABAK: Prevalence and bias adjusted Kappa; RD: Residential density; SEP: Socioeconomic position; SES: Socioeconomic status; SOP: Standard operational procedure; S-VAT: SPOT-LIGHT Virtual Audit Tool.

Acknowledgements

The TACKLE project was funded by the Norwegian Research Council BEDREHELSE program. We would like to thank Hanne Hennig Havalda who conducted the qualitative work that helped inform this study.

Authors' contributions

EK, NL, OA and MG contributed to the conception of this study. JL provided guidance in the use of the original S-VAT tool. OKA and SAO adapted the tool and conducted the audits. OKA performed the statistical analyses and drafted the manuscript and made the greatest contribution to this paper. All authors commented and contributed with important intellectual content and final approval of the version to be published. All authors read and approved the final manuscript.

Funding

Norwegian Research Council BEDREHELSE program (273823/H10).

Availability of data and materials

The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹ Norwegian School of Sport Sciences, Ullevaal Stadion, PO Box 4014, 0806 Oslo, Norway. ² University of Oslo, Oslo, Norway. ³ Department of Epidemiology and Data Science, Amsterdam Public Health Research Institute, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands. ⁴ Upstream Team, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands. ⁵ Department of Epidemiology, Fielding School of Public Health, University of California, Los Angeles (UCLA), Los Angeles, CA, USA. ⁶ Department of Statistics, College of Letters and Science, UCLA, Los Angeles, CA, USA. ⁷ Department of Public Health, Aarhus University, Aarhus, Denmark.

Received: 11 September 2020 Accepted: 7 January 2021

Published online: 18 January 2021

References

1. Abarca-Gómez L, Abdeen ZA, Hamid ZA, Abu-Rmeileh NM, Acosta-Cazares B, Acuin C, Adams RJ, Aekplakorn W, Afsana K, Aguilar-Salinas CA, Agyemang C. Worldwide trends in body-mass index, underweight, overweight, and obesity from 1975 to 2016: a pooled analysis of 2416 population-based measurement studies in 128·9 million children, adolescents, and adults. *Lancet*. 2017;390(10113):2627–42.
2. Bibiloni Mdell M, Pons A, Tur JA. Prevalence of overweight and obesity in adolescents: a systematic review. *ISRN Obes*. 2013;2013:392747.
3. Sallis JF, Cervero RB, Ascher W, Henderson KA, Kraft MK, Kerr J. An ecological approach to creating active living communities. *Annu Rev Public Health*. 2006;27:297–322.
4. Swinburn BA, Sacks G, Hall KD, McPherson K, Finegood DT, Moodie ML, et al. Obesity 1: the global obesity pandemic: shaped by global drivers and local environments. *Lancet*. 2011;378(9793):804–14.
5. Sallis JF, Glanz K. The role of built environments in physical activity, eating, and obesity in childhood. *Future Child*. 2006;16(1):89–108.

6. Bauman AE, Reis RS, Sallis JF, Wells JC, Loos RJ, Martin BW. Correlates of physical activity: why are some people physically active and others not? *Lancet*. 2012;380(9838):258–71.
7. Sallis JF, Floyd MF, Rodriguez DA, Saelens BE. Role of built environments in physical activity, obesity, and cardiovascular disease. *Circulation*. 2012;125(5):729–37.
8. O'Donoghue G, Kennedy A, Puggina A, Aleksovska K, Buck C, Burns C, et al. Socio-economic determinants of physical activity across the life course: A "DEterminants of Diet and Physical ACTivity" (DEDIPAC) umbrella literature review. *PLoS ONE*. 2018;13(1):e0190737.
9. Carlin A, Perchoux C, Puggina A, Aleksovska K, Buck C, Burns C, et al. A life course examination of the physical environmental determinants of physical activity behaviour: a "DEterminants of Diet and Physical ACTivity" (DEDIPAC) umbrella systematic literature review. *PLoS ONE*. 2017;12(8):e0182083.
10. Mackenbach JD, Nelissen KGM, Dijkstra SC, Poelman MP, Daams JG, Leijssen JB, et al. A systematic review on socioeconomic differences in the association between the food environment and dietary behaviors. *Nutrients*. 2019;11(9):2215.
11. Nordbø ECA, Nordh H, Raanaas RK, Aamodt G. Promoting activity participation and well-being among children and adolescents: a systematic review of neighborhood built-environment determinants. *JBI Evid Synth*. 2020;18(3):370–458.
12. Williams J, Scarborough P, Matthews A, Cowburn G, Foster C, Roberts N, et al. A systematic review of the influence of the retail food environment around schools on obesity-related outcomes. *Obes Rev*. 2014;15(5):359–74.
13. Engler-Stringer R, Le H, Gerrard A, Muhajarine N. The community and consumer food environment and children's diet: a systematic review. *BMC Public Health*. 2014;14:522.
14. de Vet E, de Ridder DT, de Wit JB. Environmental correlates of physical activity and dietary behaviours among young people: a systematic review of reviews. *Obes Rev*. 2011;12(5):e130–42.
15. Ding D, Gebel K. Built environment, physical activity, and obesity: what have we learned from reviewing the literature? *Health Place*. 2012;18(1):100–5.
16. Brownson RC, Hoehner CM, Day K, Forsyth A, Sallis JF. Measuring the built environment for physical activity: state of the science. *Am J Prev Med*. 2009;36(4 Suppl):S99–123.e12.
17. Aghaabbasi M, Moeinaddini M, Shah M, Asadi-Shekari Z. Addressing issues in the use of Google tools for assessing pedestrian built environments. *J Transp Geogr*. 2018;73:185–98.
18. Rzotkiewicz A, Pearson AL, Dougherty BV, Shortridge A, Wilson N. Systematic review of the use of Google Street View in health research: major themes, strengths, weaknesses and possibilities for future research. *Health Place*. 2018;52:240–6.
19. Charreire H, Mackenbach JD, Ouasti M, Lakerveld J, Compennolle S, Ben-Rebah M, et al. Using remote sensing to define environmental characteristics related to physical activity and dietary behaviours: a systematic review (the SPOTLIGHT project). *Health Place*. 2014;25:1–9.
20. Ben-Joseph E, Lee JS, Cromley EK, Laden F, Troped PJ. Virtual and actual: relative accuracy of on-site and web-based instruments in auditing the environment for physical activity. *Health Place*. 2013;19:138–50.
21. Baranowski T. Why combine diet and physical activity in the same international research society? *Int J Behav Nutr Phys Act*. 2004;1(1):2.
22. Feuillet T, Charreire H, Roda C, Ben Rebah M, Mackenbach JD, Compennolle S, et al. Neighbourhood typology based on virtual audit of environmental obesogenic characteristics. *Obes Rev*. 2016;17(Suppl 1):19–30.
23. Bethlehem JR, Mackenbach JD, Ben-Rebah M, Compennolle S, Glonti K, Bárdos H, et al. The SPOTLIGHT virtual audit tool: a valid and reliable tool to assess obesogenic characteristics of the built environment. *Int J Health Geogr*. 2014;13:52.
24. Lakerveld J, Brug J, Bot S, Teixeira PJ, Rutter H, Woodward E, et al. Sustainable prevention of obesity through integrated strategies: the SPOTLIGHT project's conceptual framework and design. *BMC Public Health*. 2012;12:793.
25. Compennolle S, De Cocker K, Roda C, Oppert JM, Mackenbach JD, Lakerveld J, et al. Physical environmental correlates of domain-specific sedentary behaviours across five European regions (the SPOTLIGHT Project). *PLoS ONE*. 2016;11(10):e0164812.
26. Mertens L, Compennolle S, Deforche B, Mackenbach JD, Lakerveld J, Brug J, et al. Built environmental correlates of cycling for transport across Europe. *Health Place*. 2017;44:35–42.
27. Roda C, Charreire H, Feuillet T, Mackenbach JD, Compennolle S, Glonti K, et al. Mismatch between perceived and objectively measured environmental obesogenic features in European neighbourhoods. *Obes Rev*. 2016;17(Suppl 1):31–41.
28. Ding D, Adams MA, Sallis JF, Norman GJ, Hovell MF, Chambers CD, et al. Perceived neighborhood environment and physical activity in 11 countries: do associations differ by country? *Int J Behav Nutr Phys Act*. 2013;10(1):57.
29. Van Hecke L, Ghekiere A, Veitch J, Van Dyck D, Van Cauwenberg J, Clarys P, et al. Public open space characteristics influencing adolescents' use and physical activity: a systematic literature review of qualitative and quantitative studies. *Health Place*. 2018;51:158–73.
30. Kaczynski AT, Wilhelm Stanis SA, Besenyi GM. Development and testing of a community stakeholder park audit tool. *Am J Prev Med*. 2012;42(3):242–9.
31. Pikora TJ, Giles-Corti B, Knuiaman MW, Bull FC, Jamrozik K, Donovan RJ. Neighborhood environmental factors correlated with walking near home: using SPACES. *Med Sci Sports Exerc*. 2006;38(4):708–14.
32. Badland HM, Opit S, Witten K, Kearns RA, Mavoa S. Can virtual streetscape audits reliably replace physical streetscape audits? *J Urban Health*. 2010;87(6):1007–16.
33. Hoehner CM, Ivy A, Ramirez LKB, Handy S, Brownson RC. Active neighborhood checklist: a user-friendly and reliable tool for assessing activity friendliness. *Am J Health Promot*. 2007;21(6):534–7.
34. Idrettsforbund N. Nøkkeltall - Rapport 2018 Idrettsforbundet.no: Idrettsforbundet; 2018.
35. Krieger N, Williams DR, Moss NE. Measuring social class in US public health research: concepts, methodologies, and guidelines. *Annu Rev Public Health*. 1997;18:341–78.
36. Norwegian Institute of Public Health. Social inequalities in health 2014. <https://www.fhi.no/en/op/hin/groups/social-inequalities/>. Accessed 15 Oct 2019.
37. Statistics Norway. Befolkning. 2020. <https://www.ssb.no/statbank/table/04317/>.
38. Oslo Kommune. Statistikkbanken. 2020. <http://statistikkbanken.oslo.kommune.no/webview/index.jsp?catalog=http%3A%2F%2Fstatistikkbanken.oslo.kommune.no%3A80%2Fobj%2Fcatalog%2Fcatalog48&submode=catalog&mode=documentation&top=yes>.
39. McMillan TE, Cubbin C, Parmenter B, Medina AV, Lee RE. Neighborhood sampling: how many streets must an auditor walk? *Int J Behav Nutr Phys Act*. 2010;7(1):20.
40. Haahr M. RANDOM.ORG: Randomness and Integrity Services Ltd. 1998. <https://www.random.org/>.
41. Griev P, Hillsdon M, Foster C, Coombes E, Jones A, Wilkinson P. Developing and testing a street audit tool using Google Street View to measure environmental supportiveness for physical activity. *Int J Behav Nutr Phys Act*. 2013;10:103.
42. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159–74.
43. Odgers CL, Caspi A, Bates CJ, Sampson RJ, Moffitt TE. Systematic social observation of children's neighborhoods using Google Street View: a reliable and cost-effective method. *J Child Psychol Psychiatry*. 2012;53(10):1009–17.
44. Clarke P, Ailshire J, Melendez R, Bader M, Morenoff J. Using Google Earth to conduct a neighborhood audit: reliability of a virtual audit instrument. *Health Place*. 2010;16(6):1224–9.
45. Gullón P, Badland HM, Alfayate S, Bilal U, Escobar F, Cebrecos A, et al. Assessing walking and cycling environments in the streets of Madrid: comparing on-field and virtual audits. *J Urban Health*. 2015;92(5):923–39.
46. de Vet HCW, Mookink LB, Terwee CB, Hoekstra OS, Knol DL. Clinicians are right not to like Cohen's κ . *BMJ*. 2013;346:f2125.
47. McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med*. 2012;22(3):276–82.
48. Maxwell AE. Comparing the classification of subjects by two independent judges. *Br J Psychiatry*. 1970;116(535):651–5.
49. Fagerland MW, Lydersen S, Laake P. The McNemar test for binary matched-pairs data: mid-p and asymptotic are better than exact conditional. *BMC Med Res Methodol*. 2013;13:91.

50. Stuart A. A test for homogeneity of the marginal distributions in a two-way classification. *Biometrika*. 1955;42(3–4):412–6.
51. Wu Y-T, Nash P, Barnes LE, Minett T, Matthews FE, Jones A, et al. Assessing environmental features related to mental health: a reliability study of visual streetscape images. *BMC Public Health*. 2014;14(1):1094.
52. Less EL, McKee P, Toomey T, Nelson T, Erickson D, Xiong S, et al. Matching study areas using Google Street View: a new application for an emerging technology. *Eval Program Plan*. 2015;53:72–9.
53. Steinmetz-Wood M, Velauthapillai K, O'Brien G, Ross NA. Assessing the micro-scale environment using Google Street View: the virtual systematic tool for evaluating pedestrian streetscapes (virtual-STEPS). *BMC Public Health*. 2019;19(1):1246.
54. Pliakas T, Hawkesworth S, Silverwood RJ, Nanchahal K, Grundy C, Armstrong B, et al. Optimising measurement of health-related characteristics of the built environment: comparing data collected by foot-based street audits, virtual street audits and routine secondary data sources. *Health Place*. 2017;43:75–84.
55. Heath GW, Brownson RC, Kruger J, Miles R, Powell KE, Ramsey LT. The effectiveness of urban design and land use and transport policies and practices to increase physical activity: a systematic review. *J Phys Act Health*. 2006;3(s1):S55–S76.
56. Zhu W, Sun Y, Kurka J, Geremia C, Engelberg JK, Cain K, et al. Reliability between online raters with varying familiarities of a region: micro-scale audit of pedestrian streetscapes (MAPS). *Landsc Urban Plan*. 2017;167:240–8.
57. Chudyk AM, Winters M, Gorman E, McKay HA, Ashe MC. Agreement between virtual and in-the-field environmental audits of assisted living sites. *J Aging Phys Act*. 2014;22(3):414–20.
58. Kelly CM, Wilson JS, Baker EA, Miller DK, Schootman M. Using Google Street View to audit the built environment: inter-rater reliability results. *Ann Behav Med*. 2013;45(Suppl 1):S108–12.
59. Wilson JS, Kelly CM, Schootman M, Baker EA, Banerjee A, Clennin M, et al. Assessing the built environment using omnidirectional imagery. *Am J Prev Med*. 2012;42(2):193–9.
60. Hanibuchi T, Nakaya T, Inoue S. Virtual audits of streetscapes by crowd-workers. *Health Place*. 2019;59:102203.
61. Phillips CB, Engelberg JK, Geremia CM, Zhu W, Kurka JM, Cain KL, et al. Online versus in-person comparison of microscale audit of pedestrian streetscapes (MAPS) assessments: reliability of alternate methods. *Int J Health Geogr*. 2017;16(1):27.
62. Kurka JM, Adams MA, Geremia C, Zhu W, Cain KL, Conway TL, et al. Comparison of field and online observations for measuring land uses using the microscale audit of pedestrian streetscapes (MAPS). *J Transp Health*. 2016;3(3):278–86.
63. Byrt T, Bishop J, Carlin JB. Bias, prevalence and kappa. *J Clin Epidemiol*. 1993;46(5):423–9.
64. Fisher BD, Richardson S, Hosler AS. Reliability test of an established pedestrian environment audit in rural settings. *Am J Health Promot*. 2010;25(2):134–7.
65. Seymour M, Reynolds KD, Wolch J. Reliability of an audit tool for systematic assessment of urban alleyways. *J Phys Act Health*. 2010;7(2):214–23.
66. Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Phys Ther*. 2005;85(3):257–68.
67. Pucher J, Buehler R, Sgm F. Making cycling irresistible: lessons from The Netherlands, Denmark and Germany. *Transp Rev*. 2008;28:495–5281464.
68. Regjeringen. 78,5 millioner kroner til meir sykling og gåing på fylkeskommunale og kommunale vegar i 2019. [Regjeringen.no](http://regjeringen.no) 2019.
69. Kopcakova J, Veselska ZD, Geckova AM, Bucksch J, Nalecz H, Sigmundova D, et al. Is a perceived activity-friendly environment associated with more physical activity and fewer screen-based activities in adolescents? *Int J Environ Res Public Health*. 2017;14(1):39.
70. Markevych I, Smith MP, Jochner S, Standl M, Bruske I, von Berg A, et al. Neighbourhood and physical activity in German adolescents: GINIplus and LISAPLUS. *Environ Res*. 2016;147:284–93.
71. Oreskovic NM, Kuhlthau KA, Romm D, Perrin JM. Built environment and weight disparities among children in high- and low-income towns. *Acad Pediatr*. 2009;9(5):315–21.
72. Google. Where we've been and where we're headed next. Google. <https://www.google.com/intl/en-CA/streetview/explore/>.
73. Mygind L, Bentsen P, Badland H, Edwards N, Hooper P, Villanueva K. Public open space desktop auditing tool—establishing appropriateness for use in Australian regional and urban settings. *Urban For Urban Green*. 2016;20:65–70.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

