

Assessing Walking and Cycling Environments in the Streets of Madrid: Comparing On-Field and Virtual Audits

Pedro Gullón, Hannah M. Badland, Silvia Alfayate,
Usama Bilal, Francisco Escobar, Alba Cebrecos, Julia Diez,
and Manuel Franco

ABSTRACT *Audit tools are useful for exploring the urban environment and its association with physical activity. Virtual auditing options are becoming increasingly available potentially reducing the resources needed to conduct these assessments. Only a few studies have explored the use of virtual audit tools. Our objective is to test if the Madrid Systematic Pedestrian and Cycling Environment Scan (M-SPACES) discriminates between areas with different urban forms and to validate virtual street auditing using M-SPACES. Three areas (N=500 street segments) were selected for variation in population density. M-SPACES was used to audit street segments physically and virtually (Google Street View) by two researchers in 2013–2014. For both physical and virtual audits, all analyzed features score significantly different by area ($p<0.05$). Most of the features showed substantial ($ICC=0.6–0.8$) or almost perfect ($ICC\geq 0.8$) agreement between virtual and physical audits, especially neighborhood permeability walking infrastructure, traffic safety, streetscape aesthetics, and destinations. Intra-rater agreement was generally acceptable ($ICC>0.6$). Inter-rater agreement was generally poor ($ICC<0.4$). Virtual auditing provides a valid and feasible way of measuring residential urban environments. Comprehensive auditor training may be needed to guarantee good inter-rater agreement.*

KEYWORDS *Physical activity, Validation studies, Urban environment, Omnidirectional image, Virtual image*

Abbreviations: CVD – Cardiovascular diseases; GSV – Google Street View; SPACES – Systematic Pedestrian and Cycling Environmental Scan; NZ-SPACES – New Zealand Systematic Pedestrian and Cycling Environmental Scan; M-SPACES – Madrid Systematic Pedestrian and Cycling Environmental Scan; HHH – Heart Healthy Hoods; ICC – Intraclass correlation coefficient

Gullón, Alfayate, Bilal, Escobar, Cebrecos, Diez, and Franco are with the Social and Cardiovascular Epidemiology Research Group, School of Medicine, University of Alcalá, Alcalá de Henares, Madrid 28871, Spain; Gullón is with the Escuela Nacional de Sanidad, Instituto de Salud Carlos III, Madrid, 28029, Spain; Badland is with the Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, VIC, Australia; Bilal and Franco are with the Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA; Escobar is with the Department of Geology, Geography and Environment, University of Alcalá, Alcalá de Henares, Madrid 28871, Spain; Franco is with the Social and Cardiovascular Epidemiology Research Group, School of Medicine, University of Alcalá, Alcalá de Henares, Madrid 28871, Spain.

Correspondence: Manuel Franco, Social and Cardiovascular Epidemiology Research Group, School of Medicine, University of Alcalá, Alcalá de Henares, Madrid 28871, Spain. (E-mail: mfranco@uah.es)

Electronic supplementary material The online version of this article (doi:10.1007/s11524-015-9982-z) contains supplementary material, which is available to authorized users.

BACKGROUND

Cardiovascular diseases (CVD) are the leading cause of death in Europe¹ and their social, medical and economic burden will likely increase over the coming decades.² One of the major risk factors for CVD is physical inactivity.³ Walking and cycling, as a mean of active transportation and commuting⁴, are the main determinants of physical activity levels when considering population health approaches.^{5, 6} Worldwide, it is estimated that the failure to spend 15–30 min a day briskly walking is responsible for 6–10 % of the burden of major noncommunicable diseases.⁷

In order to tackle this problem, researchers, health professionals, and policy-makers have identified a range of population approaches to support physical activity accumulation throughout the day. Population approaches for disease prevention attempts to shift the whole distribution of a population for a given risk factor, so the effect of the prevention is not limited to high-risk individuals.⁸ This approach focuses on contextual factors, such as physical and social environments, and is often conceptualized within a socioecological framework.⁹ Physical activity behaviors are increasingly studied within socioecological frameworks, with one of the possible areas for investigation being the urban environment.^{10–14}

The development of reliable tools and methods to audit the physical qualities of the urban environment that likely impact physical activity remains an ongoing challenge for public health researchers. More than 30 field audit instruments have been developed in recent years,¹⁵ which can be further classified into subjective (questionnaires and interviews) and objective tools (direct observation with checklists, official dataset, etc.).^{11, 16, 17} The use of subjective or objective tools depends on the objective of the study and the availability of research resources.¹⁶

Direct observation is one type of objective measure, whereby trained observers undertake audits or checklists, which assess different aspects of the urban environment. These can be simple measures of one characteristic of the environment, such as the distance to specific destinations; information about general environment (e.g., density of buildings or traffic);¹¹ or assess multiple items that represent different aspects of the urban environment. The study of specific urban domains for walking (e.g., streets quality, pedestrians' safety) and cycling (e.g., cycle lanes, cycle storage) has been one of the main focuses when developing audit tools to study the relationship between urban environment and physical activity¹⁶.

For logistical reasons, especially for studies in large geographic areas, there has been increasing interest for virtually measuring attributes of the urban environment thought to be associated with physical activity. Furthermore, many of these measures lend themselves well to using open-access mapping technologies, such as Google Earth, Google Street View, or Microsoft Visual Earth.¹⁸ Compared with physical audits, virtual audits may provide a faster, easier, cheaper, safer, and more reliable method to assess the urban environment.¹⁸ To date, Google Street View is the most available and accessible form of omnidirectional imagery, providing coverage for most European, USA, and Australasian urban areas.

Experiences of using open-source mapping technologies have emerged mostly in the USA^{15, 19–24}, but also in Canada,^{25, 26} Australia,²⁷ New Zealand²⁸, UK,²⁹ Netherlands,³⁰ and Belgium.³¹ Due to historical reasons, European cities³² have different urban forms compared to cities in North America or Australia, and therefore, the study of the urban environment in Europe has its own challenges. As there are few experiences in Europe where the validity of applying a streetscape

audit tool virtually has been assessed, another European context is warranted. Besides, within the European context, Mediterranean cities are characterized for being more compact in terms of urban sprawl.³³

The objectives of this study are the following: (1) to test if the Madrid Systematic Pedestrian and Cycling Environment Scan (M-SPACES) can differentiate walking and cycling environments across different urban-form areas in Madrid, Spain; (2) to assess the validity of the M-SPACES in a virtual urban setting using Google Street View; and (3) to assess the reliability (intra- and inter-rater) of the M-SPACES audit tool.

METHODS

Setting

This study is part of the Heart Healthy Hoods (HHH) project (<http://hhhproject.eu/>). The HHH project examines the association between the social and physical features of the urban environment in relation to adults' cardiovascular health living in the city of Madrid, Spain.

In 2011, Madrid city had an estimated population of 3,198,645 citizens.³⁴ Madrid is structured in 21 districts, each of which is subdivided into neighborhoods. Neighborhoods are also divided in units of ~1000–1500 residents, called census sections. A total of 36 census sections were selected for this study, providing an estimated population of 49,260 residents.³⁴ Three areas, each comprising 12 census sections, were selected in Madrid based on variation (low, medium, high) in residential population per square mile and on homogeneity in terms of sociodemographic characteristics (all three areas had average scores in terms of education, immigration and aging).³⁴ In order to obtain clusters of 12 census sections, the Kulldorff' spatial scan statistic software was used (Fig. 1).³⁵ These 36 census sections included a total of 500 street segments defined as the street line from one intersection to another. These areas were located in the districts of Carabanchel, Ciudad Lineal, and Chamartín, respectively. Population density was used as a proxy of different urban form; also, population density has been positively related with walking in a recent systematic review.³⁶

Measurement

Development of M-SPACES The M-SPACES is an observational audit tool of urban attributes associated with walking and cycling along a street network. The original tool was developed by Australian researchers to represent physical environments that may promote or inhibit walking or cycling.³⁷ Using stakeholder interviews and a Delphi study, four factors were identified that likely support physical activity behavior within the neighborhood environment: function, safety, aesthetics, and destinations.³⁷ Each factor consisted of different elements, which were further reduced to items. Within the same Delphi group, depending on the importance for supporting neighborhood walking and cycling, weights were applied.

As part of the URBAN study,³⁸ the SPACES tool was further refined by adjusting the item weights for the New Zealand context.²⁸ From the NZ-SPACES tool, small adjustments were made for the M-SPACES measure to be applied to Madrid, as some features of the audit tool could not be differentiated in this urban context. Small adjustments were: summing the item weights for “negotiation of footpath”

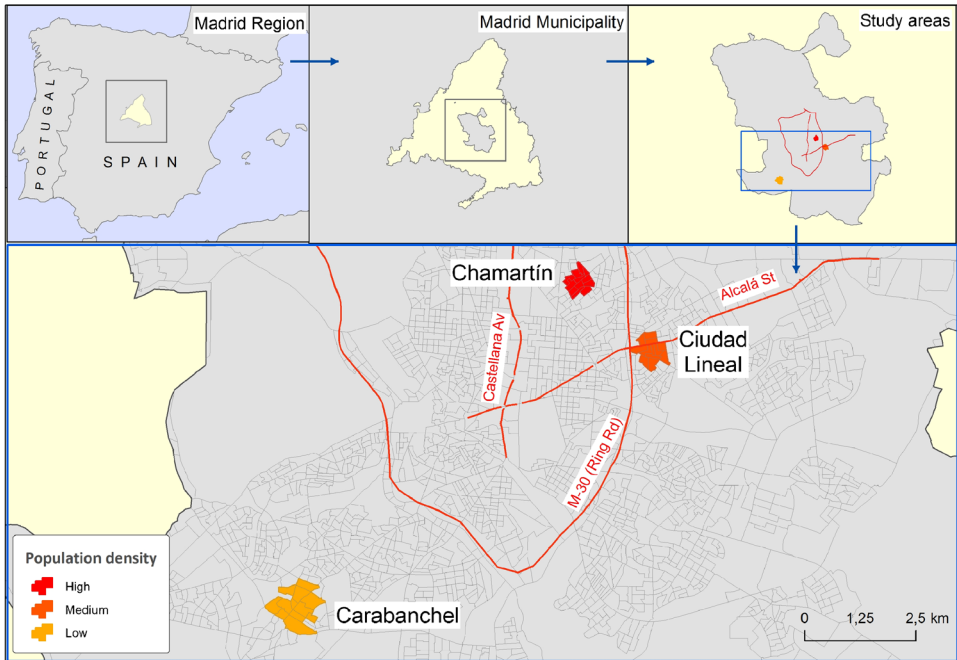


FIG. 1 Madrid city with the study areas selected.

and “type of footpath” into a single item called “type of footpath”; similarly, we aggregated “footpath smoothness” and “footpath smoothness/condition” into a variable called “footpath smoothness.” We also modified the “Destinations” item by adding the number of destinations present in the street segment, as most of Madrid’s segments presented many destinations, and we considered important to discriminate the number of destinations. These adjustments did not affect the final weights of the elements and factors. Final items, elements, factors and their weights are shown in Table 1. M-SPACES audit tool can be found in the online [supplementary material](#).

M-SPACES Training Two field researchers underwent familiarization training, where practice physical and virtual audits occurred using the M-SPACES. Both researchers piloted the tool with supervision for 2 h in Ciudad Lineal. Training audits were completed following the SPACES protocol outlined in the manual.³⁹

On Field and Virtual Measurements Two researchers conducted physical and virtual audits of the 36 census sections previously identified using the M-SPACES audit tool (Fig. 2). Between February and May 2014 the field researchers completed the M-SPACES physical audit by walking together along both sides of the street segments. Virtual audits of the streetscapes were completed using Google Street View software. Images of the 36 census sections were recorded by Google between May 2008 and February 2014 (Fig. 3); 152 segments (30.4 %) were recorded before 2010. The order of measurement of the three areas was not randomized; starting with the medium density area (Ciudad Lineal) followed by the high (Chamartín) and the low (Carabanchel) density areas. Within each area, we randomly selected half of the census sections for physical audit first, followed by the virtual audit; the

TABLE 1 M-SPACES tool: factors, elements, and items for walking and cycling

Factor	Element	Element weight	Item	Item weight			
Walking	Function	1.00	Type of path	0.25			
			Path smoothness	0.20			
			Path material	0.10			
			Slope	0.20			
			Continuity	0.20			
			Curb type	0.05			
			Neighborhood permeability	1.00	Other routes available	0.50	
					Neighborhood legibility	0.50	
			Walking infrastructure	0.33	Seats	0.50	
					Trees/verandas	0.50	
			Safety	Street (lanes)	0.33	Number of lanes	1.00
						Fixed traffic controls	0.33
				Path safety	0.66	Traffic control devices present	1.00
						Path location	0.30
Traffic safety	0.66	Fixed obstacles on path		0.10			
		Street lights		0.20			
		Surveillance		0.20			
		Graffiti/vandalism		0.20			
		Crossing type		0.50			
		Crossing aids		0.30			
Aesthetics	Streetscape aesthetics	0.66	Visible driveways	0.20			
			Trees	0.25			
			Gardens maintained	0.25			
			Verges maintained	0.25			
			Cleanliness	0.25			
			Views	1.00			
View aesthetics	0.66	Attractiveness	0.50				
		Subjective walking assessment	0.33				
Destinations	Land use mix	1.00	Physical difficulty	0.50			
			Number of destinations present	Out of 10			
Cycling	Function	1.00	Path type	0.30			
			On-road cycle lane	0.30			
			Slope	0.20			
			Road condition	0.10			
			Curb type	0.10			
			Neighborhood permeability	1.00	Other routes available	0.50	
					Neighborhood legibility	0.50	
			Cycling infrastructure	0.33	Cycle storage	1.00	
					Streets (lanes)	0.66	
			Safety	Fixed traffic controls	0.66	Number of lanes	1.00
						Traffic control devices present	1.00
			Traffic safety	0.33	Crossing type	0.50	
					Crossing aids	0.30	
					Visible driveways	0.20	

TABLE 1 *Continued*

Factor	Element	Element weight	Item	Item weight
Aesthetics	Streetscape aesthetics	0.33	Trees	0.25
			Gardens maintained	0.25
			Verges maintained	0.25
			Cleanliness	0.25
	View aesthetics	0.33	Views	1.00
Destinations	Land use mix	1.00	Subjective cycling assessment	0.50
			Physical difficulty	0.50
			Number of destinations present	Out of 10

remaining half were audited in the reverse order to reduce the effect of taking the same measures of the same streetscape.

Time taken to complete the audit for each street segment was recorded for both virtual and on-field measurements. Also, time travel to each of the areas by public transportation was calculated using Google Maps software. To do so, Instituto de Salud Carlos III (Avenida Monforte de Lemos 5, 28029, Madrid, Spain) was used as the departure location for all areas, and its closest point of the area as the arrival location. Monday 8:00 A.M. was chosen as the departure hour.

Inter-Rater and Intra-Rater Reliability Inter-rater agreement was assessed between the two researchers for both physical and virtual audits. In order to measure intra-rater reliability, physical and virtual audit data collected in April 2013 by one of the researchers were compared for the median-density area (Ciudad Lineal). The

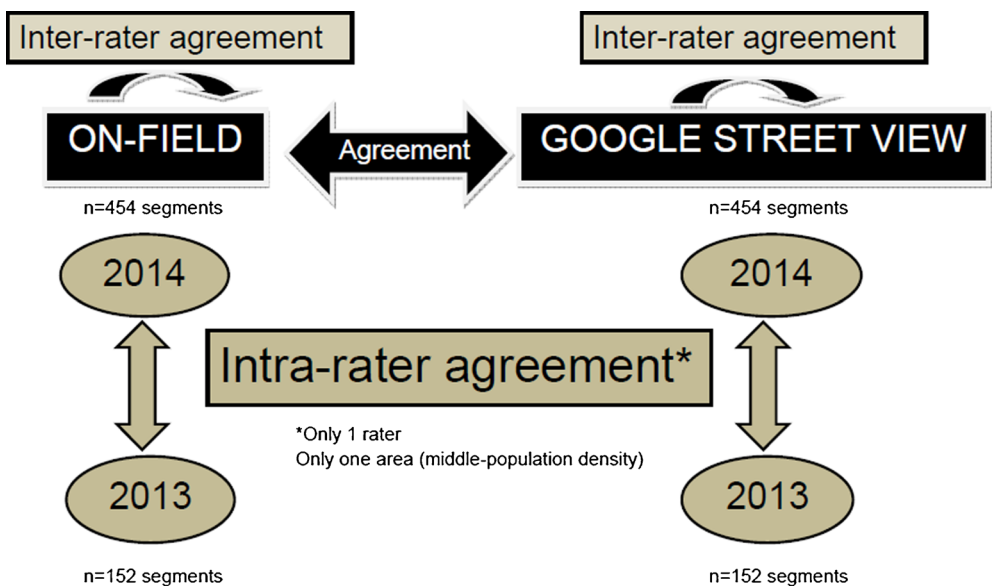


FIG. 2 Description of the M-SPACES assessment by two raters across the study areas.



FIG. 3 Picture of a street feature as it was observed in physical (*left*) and virtual audit (*right*).

adjustments to the M-SPACES audit tool (i.e., grouping and refining items) were made after this first assessment; therefore “Destinations” was excluded for the intra-rater agreement analysis.

Statistical Analysis

In order to test if the M-SPACES scores differed by urban form, we performed an analysis of variance (ANOVA) of mean scores by area. Level of significance was set as $p < 0.05$.

A two-way mixed model intraclass correlation coefficient (ICC) was used to measure the following: (a) agreement between physical and virtual audits, (b) inter-rater agreement between the two researchers, and (c) intra-rater agreement comparing M-SPACES assessments in 2013 and 2014.^{40, 41} The ICC measured the percentage of total variability for a given street segment:

$$ICC = \frac{k \cdot ISS - TSS}{(k-1) \cdot TSS}$$

Where “ k ” refers to the number of street segments, “TSS” total sum of squares, and “ISS” inter-group sum of squares.

Following Landis and Koch classification, the cutoff ranges for ICC values used were as follows: 0.0–0.20 (weak agreement), 0.21–0.40 (poor agreement), 0.41–0.60 (moderate agreement), 0.61–0.80 (substantial agreement), and 0.81–1.00 (almost perfect agreement).⁴² All analyses were conducted using Stata SE version 12.1 (StataCorp., College Station, TX, USA).

Mapping

Geographic information systems (GIS) were adopted with a twofold aim. On the one hand, it constitutes the underlying technology allowing for the integration of all collected information together with contextual information such as street network, land parcel and building polygons. On the other hand, the establishment of a geo-referenced database on the subject will make possible future analysis and modeling. All data layers, both collected and downloaded from official sources, were first projected and referenced to a common system. Absolute differences between on-field and virtual audits of M-SPACES total scores for walking and cycling were then joined to the attribute table of the street segment data sets by means of relational

union in order to represent it in a map. All GIS-related operations were undertaken with ArcGIS software (ESRI, Redlands, CA, USA).

RESULTS

A total of 454 street segments (90.8 %) included in the three study areas were measured both physically and virtually. Of the 46 street segments that were not audited, 20 were excluded as Google Street View imagery was unavailable (i.e., pedestrian streets that were inaccessible to the vehicles where images are captured from). The remaining excluded street segments ($n=26$) were too short to measure, and their characteristics were subsequently absorbed into the closest street segment.

Differentiation of Areas with M-SPACES in Madrid

Data presented in Table 2 show the scores for the M-SPACES factors across the three different residential density areas. Function for walking scores was higher in the medium-density area (Ciudad Lineal), while cycling functionality was higher in the lowest-density area (Carabanchel). Safety factor scores for walking and cycling were higher in the highest density area (Chamartín). Also, Chamartín had the highest scores for destinations to walk and cycle. Destinations yielded the greatest difference between the areas, whereby Chamartín (0.55) almost doubled Carabanchel's mean score for destinations present (0.29). All differences were statistically significant ($p<0.05$).

Google Street View Validity

Physical and virtual audits reported substantial ($ICC>0.60-0.80$) or almost perfect agreement ($ICC>0.80$) for 6/11 elements for walking and 5/10 elements for cycling, and also for total scores (Table 3). Walking infrastructure, traffic safety, streetscape aesthetics, and destinations demonstrated almost perfect agreement ($ICC=0.86, 0.89, 0.80,$ and $0.85,$ respectively). However, walking surface and cycling infrastructure showed poor agreement ($ICC=0.36$ and $0.39,$ respectively), while cycling surface, streets lane, and subjective walking and cycling assessment showed

TABLE 2 Differences in M-SPACES factors between the three study areas

Factors	Mean (SD) ($n=454$ street segments)			p Value ^a
	High-density area (Chamartín), $n=124$ street segments	Median-density area (Ciudad Lineal), $n=152$ street segments	Low-density area (Carabanchel), $n=178$ street segments	
Walking factors				
Function	1.18 (0.28)	1.41 (0.31)	1.35 (0.33)	0.002*
Safety	1.07 (0.22)	1.03 (0.24)	1.03 (0.22)	0.026*
Aesthetics	0.75 (0.25)	0.77 (0.27)	0.91 (0.29)	<0.001*
Destinations	0.55 (0.29)	0.43 (0.27)	0.29 (0.21)	<0.001*
Cycling factors				
Function	0.87 (0.24)	0.93 (0.23)	1.00 (0.29)	<0.001*
Safety	1.05 (0.21)	0.94 (0.25)	1.02 (0.23)	0.005*
Aesthetics	0.44 (0.16)	0.45 (0.18)	0.55 (0.17)	<0.001*
Destinations	0.55 (0.29)	0.43 (0.27)	0.29 (0.21)	<0.001*

^aStatistical significance of the difference in means across the study areas in each factor ($p<0.05$)

TABLE 3 Levels of agreement between elements assessed physical and virtual audits ($n=454$ street segments)

Factor	Element	Physical audit score	Virtual audit score	ICC	95 % CI
Walking					
Function	Walking surface	0.92	0.96	0.36 ^b	0.28–0.44
	Neighborhood permeability	0.56	0.56	0.69 ^d	0.64–0.73
Safety	Walking infrastructure	0.19	0.18	0.86 ^e	0.84–0.89
	Streets (lane)	0.88	0.89	0.41 ^c	0.33–0.48
	Fixed traffic controls	0.43	0.43	0.58 ^c	0.52–0.64
	Path safety	0.35	0.35	0.48 ^c	0.41–0.55
Aesthetics	Traffic safety	0.87	0.87	0.89 ^e	0.87–0.91
	Streetscape aesthetics	0.33	0.36	0.80 ^e	0.77–0.84
	View aesthetics	0.56	0.53	0.75 ^d	0.71–0.79
Destinations	Subjective walking assessment	0.56	0.64	0.55 ^c	0.48–0.61
	Land use mix	0.41	0.41	0.85 ^e	0.83–0.88
Total	Sum of factors weights	4.00	4.06	0.87 ^e	0.85–0.89
Cycling					
Function	Cycling surface	0.65	0.41	0.51 ^c	0.44–0.57
	Neighborhood permeability	0.56	0.56	0.69 ^d	0.64–0.73
Safety	Cycling infrastructure	0.01	0.01	0.39 ^b	0.31–0.47
	Streets (lane)	0.88	0.89	0.40 ^c	0.33–0.48
	Fixed traffic controls	0.43	0.43	0.58 ^c	0.52–0.64
Aesthetics	Traffic safety	0.87	0.87	0.89 ^e	0.87–0.91
	Streetscape aesthetics	0.32	0.36	0.80 ^e	0.77–0.84
	View aesthetics	0.56	0.53	0.75 ^d	0.71–0.79
Destinations	Subjective cycling assessment	0.56	0.64	0.53 ^d	0.46–0.59
	Land use mix	0.42	0.42	0.85 ^e	0.83–0.88
Total	Sum of factors weights	3.29	3.35	0.77 ^d	0.72–0.80

ICC intraclass correlation coefficient, CI confidence interval

^aWeak agreement (ICC <0.2)

^bPoor agreement (ICC 0.2–0.4)

^cModerate agreement (ICC 0.4–0.6)

^dSubstantial agreement (ICC 0.6–0.8)

^eAlmost perfect agreement (ICC >0.8)

moderate agreement (ICC=0.51, 0.41, 0.55, and 0.53, respectively). Figure 4 shows walking and cycling total scores' difference between physical and virtual audits.

Intra-Rater Reliability

Most elements demonstrated moderate or substantial intra-rater agreement (ICC=0.40–0.80) (Table 4). Walking infrastructure, streets lane, traffic safety, and view aesthetics had almost perfect agreement (ICC>0.80). Intra-rater agreement for the virtual auditing was higher for aesthetics (ICC=0.81) than the intra-rater agreement for aesthetics in the physical audits (ICC=0.45).

Vast differences existed for view aesthetics (physical audit: ICC=0.15; virtual audit: ICC=0.88). Path safety, cycling surface (only for the physical measurement)

ABSOLUTE DIFFERENCE BETWEEN PHYSICAL AND VIRTUAL AUDITS



FIG. 4 Absolute differences between physical and virtual audits for walking (*left*) and cycling (*right*) total M-SPACES score.

and cycling infrastructure had an ICC=0.00, and thus were regarded as being highly unreliable.

Inter-Rater Reliability

In general, inter-rater agreement was low (Table 5). The agreement between the two observers was better for the physical audits when compared with the virtual audits.

TABLE 4 Intra-rater agreement between elements assessed by one rater twice (2013 and 2014) ($n=152$ street segments)

Factor	Element	Physical		Virtual	
		ICC	95 % CI	ICC	95 % CI
Walking					
Function	Walking surface	0.54 ^c	0.43–0.66	0.30 ^b	0.15–0.45
	Neighborhood permeability	0.49 ^c	0.37–0.61	0.48 ^c	0.35–0.60
Safety	Walking infrastructure	0.85 ^e	0.80–0.89	0.82 ^e	0.76–0.90
	Streets (lane)	0.92 ^e	0.89–0.94	0.91 ^e	0.89–0.94
	Fixed traffic controls	0.59 ^c	0.49–0.69	0.74 ^d	0.66–0.81
	Path safety	0.00 ^a	0.00–0.17	0.00 ^a	0.00–0.16
Aesthetics	Traffic safety	0.77 ^d	0.71–0.84	0.80 ^e	0.74–0.86
	Streetscape aesthetics	0.72 ^d	0.65–0.80	0.69 ^d	0.60–0.77
	View aesthetics	0.15 ^a	0.00–0.30	0.88 ^e	0.85–0.92
Destinations	Subjective walking assessment	0.42 ^c	0.29–0.55	0.52 ^c	0.40–0.64
	Land use mix	N.A.	N.A.	N.A.	N.A.
Cycling					
Function	Cycling surface	0.33 ^b	0.18–0.47	0.00 ^f	0.00–0.17
	Neighborhood permeability	0.49 ^c	0.37–0.61	0.48 ^c	0.35–0.60
	Cycling infrastructure	0.00 ^f	0.00–0.00	0.00 ^f	0.00–0.00
Safety	Streets (lane)	0.92 ^e	0.89–0.94	0.91 ^e	0.89–0.94
	Fixed traffic controls	0.59 ^c	0.49–0.69	0.74 ^d	0.66–0.81
	Traffic safety	0.77 ^d	0.71–0.84	0.80 ^e	0.74–0.86
Aesthetics	Streetscape aesthetics	0.72 ^d	0.65–0.80	0.69 ^d	0.60–0.77
	View aesthetics	0.15 ^a	0.00–0.30	0.88 ^e	0.85–0.92
	Subjective cycling assessment	0.41 ^c	0.28–0.54	0.42 ^c	0.29–0.55
Destinations	Land use mix	N.A.	N.A.	N.A.	N.A.

N.A. not assessed, ICC intraclass correlation coefficient, CI confidence interval

^aWeak agreement (ICC <0.2)

^bPoor perfect agreement (ICC 0.2–0.4)

^cModerate agreement (ICC 0.4–0.6)

^dSubstantial agreement (ICC 0.6–0.8)

^eAlmost perfect agreement (ICC >0.8)

^fICC truncated at zero

Both physical and virtual walking and cycling surface elements showed weak inter-rater agreement (ICC<0.20). Apart from virtual cycling infrastructure (e.g., cycle storage) (ICC=0.47), physical and virtual audits of walking and cycling infrastructure showed substantial agreement (ICC=0.60–0.80). Aesthetics and subjective assessments tended to have lower agreement (ICC<0.40) than the streetscape aesthetics. Virtual path safety inter-rater agreement was weak (ICC=0.08), especially when compared with the inter-rater agreement using the physical audit (ICC=0.32).

Time Taken to Complete the Audits

Overall, the time taken to measure each street segment was faster for on-field auditing (2.45 min per segment) than virtual auditing (2.76 min per segment), and this was consistently shown across the three areas. In Chamartín, virtual auditing took 2.87 min per street segment compared with 2.42 min for the physical assessment. In Ciudad Lineal it took a meantime of 2.84 min virtually and 2.33 min

TABLE 5 Inter-rater agreement between physical and virtual elements by two raters in three study areas ($n=454$ street segments)

Factor	Element	Physical		Virtual	
		ICC	95 % CI	ICC	95 % CI
Walking					
Function	Walking surface	0.02a	0.00–0.12	0.00f	0.00–0.09
	Neighborhood permeability	0.39b	0.31–0.47	0.41c	0.33–0.49
Safety	Walking infrastructure	0.66d	0.60–0.71	0.60d	0.52–0.65
	Streets (lane)	0.79d	0.76–0.83	0.33b	0.25–0.41
	Fixed traffic controls	0.27b	0.18–0.36	0.21b	0.12–0.30
	Path safety	0.32b	0.24–0.40	0.08a	0.00–0.17
Aesthetics	Traffic safety	0.43c	0.36–0.51	0.41c	0.33–0.48
	Streetscape aesthetics	0.61d	0.56–0.67	0.59c	0.52–0.65
	View aesthetics	0.28b	0.19–0.36	0.18a	0.09–0.27
Destinations	Subjective walking assessment	0.30b	0.22–0.39	0.29b	0.21–0.38
	Land use mix	0.67d	0.61–0.72	0.57c	0.51–0.63
Cycling					
Function	Cycling surface	0.10a	0.01–0.19	0.00f	0.00–0.0f
	Neighborhood permeability	0.39b	0.31–0.47	0.41c	0.33–0.49
	Cycling infrastructure	0.75d	0.71–0.79	0.47c	0.40–0.53
Safety	Streets (lane)	0.79d	0.76–0.83	0.33b	0.25–0.41
	Fixed traffic controls	0.27b	0.18–0.36	0.21b	0.12–0.30
	Traffic safety	0.43c	0.36–0.51	0.41c	0.33–0.48
Aesthetics	Streetscape aesthetics	0.61d	0.56–0.67	0.59c	0.52–0.65
	View aesthetics	0.28b	0.19–0.36	0.18a	0.09–0.27
	Subjective cycling assessment	0.44c	0.36–0.51	0.33b	0.25–0.42
Destinations	Land use mix	0.67d	0.61–0.72	0.57c	0.51–0.63

ICC intraclass correlation coefficient, CI confidence interval

aWeak agreement (ICC <0.2)

bPoor agreement (ICC 0.2–0.4)

cModerate agreement (ICC 0.4–0.6)

dSubstantial agreement (ICC 0.6–0.8)

eAlmost perfect agreement (ICC >0.8)

fICC truncated at zero

physically per street segment. In Carabanchel, virtual auditing was faster (2.62 min per street segment), and physical audits were slower (2.57 min per street segment) compared with Carabanchel and Chamartín. Time travel from Instituto de Salud Carlos III to the study areas was estimated as 36 min to Chamartín, 40 min to Ciudad Lineal and 60 min to Carabanchel.

DISCUSSION

These findings showed that M-SPACES is a useful tool to measure walking and cycling environments in the streets segments of a European city like Madrid. SPACES original tool and its variants are based on a conceptual framework which makes it an easy tool to compare different urban forms and has been used and validated in several other settings.^{28, 43, 44} Google Street View was, for most features of the streetscape, a valid instrument to assess physical urban environment using M-SPACES when the physical and virtual audits were compared. Furthermore, most

elements had substantial (neighborhood permeability) or almost perfect agreement (walking infrastructure, traffic safety, streetscape aesthetics, and destinations). Despite that, some features did not reach acceptable agreement when virtual and physical measures were compared. For some of these elements (walking and cycling surface and streets lane), the use of supplementary information via other secondary spatial sources (e.g., council databases) could be used to improve reliability. However, the differences shown for cycling infrastructure (cycle storage) may be a temporal issue, whereby new cycling infrastructure (e.g., bicycle parking) in Madrid were not present when the Google Street View imagery was taken. In summary, our findings are compatible with the work of other studies,⁴⁵ that have found that the more subjective characteristics (e.g., sidewalk conditions) have lower agreement between virtual and physical audits. However, in our findings view aesthetics and streetscape aesthetics showed an acceptable agreement. Furthermore, little availability for some elements (e.g., low prevalence of cycling storage) could explain low levels of agreement with ICC due to low variability.⁴¹

It was not anticipated that the mean time auditing the street segments using Google Street View would be slower than conducting physical audits; however, time taken to go to the areas by public transportation also has to be taken into account. Nevertheless, other studies conducted in Australia, New Zealand, and UK found that virtual audits were faster than physical audits.^{28, 30} A possible explanation may be that the greater complexity of the street structures in Mediterranean and continental Europe cities may eliminate the time advantage of undertaking virtual audits.³² Given this fact, in this study, time spent auditing street segments in the low-density (where the streets segments are expected to be longer) area was similar when measuring the environment physically or virtually. Despite this, virtually assessing urban environments has other advantages as virtual audits: (1) are not affected by daylight restriction or weather changes; (2) require limited resources (computer and internet connection); (3) are less financially costly than physical audits;²⁸ (4) are useful for regional and international comparisons; (5) may be a good alternative when measuring unsafe areas;²¹ and (6) it provides an alternative when visiting the study area physically is not possible. Also, metadata provided by Google Street View allow researchers to match environmental conditions.

Our intra-rater reliability results are comparable with previous studies.^{30, 31} These results, besides demonstrating the intra-reliability of M-SPACES both physically and virtually, showed that in most cases, the urban environment in these three areas did not change between 2013 and 2014. Lower levels of agreement were found for cycling infrastructure and surface, as also path safety, which may be due to low variability of these elements in the study area, as also some new elements (e.g., cycle storage) that were not present in 2013.

Inter-rater agreement between the two observers was low. Only walking infrastructure, streets lane (physical audit), streetscape aesthetics, and destinations have acceptable levels of agreement. One possible explanation for the low agreement observed might be systematic auditor training differences. Even though both researchers received training with the SPACES manual, this manual was not translated and adapted to the Madrid context, which may have resulted in different understanding of the different options to choose in the audit tool. Potentially due to these reasons, our findings are not consistent with the results of Kelly et al., which found high levels of agreement between four auditors when measuring the streetscape with Google Street View.²² In their work, all auditors received systematically 4 h of training; our auditors read the SPACES manual³⁹ and piloted

the tool in Ciudad Lineal for 2 h. Griew et al. found that inter-rater agreement was different between different types of neighborhood (industrial, residential, etc.).²⁹ However, we did not find any differences between inter-rater agreements for our three areas that could help us to understand this low agreement between observers.

Limitations and Strengths

We acknowledge that this study has several limitations. The M-SPACES tool was designed to measure the streetscape, therefore did not assess access to parks, pedestrian- or cycling-only paths, recreation centers, or other facilities, and it is possible that these residential features are associated with walking behavior. Images from Google Street View were taken between 2008 and 2014 (30.4 % before 2010); so there may be some areas for which the images are not updated. As there has been an acceptable agreement for both intra-rater and physical-virtual, we do not believe that timing of the Google Street View pictures was a major problem in this study; however, this may be an important topic to take into account when using virtual assessments to study the relationship between urban environment and health results. Google Street View was the web-based tool used to measure the streets, and when used in conjunction with other measures, such as Google Earth or MS Visual Oblique, may provide additional environmental context; however, we did not explore these additional datasets. Ten percent of the street segments could not be measured by Google Street View, as there were no images, almost all of them being pedestrian streets. It is important to note that pedestrian streets are very relevant to develop walkable neighborhoods in European cities, so it is important to create specific tools to assess pedestrian streets, as also ways to measure it virtually (e.g., not 3D satellite images). Low variability affects ICC agreement values, so the assessment of some aspects of the streets (especially cycling infrastructure and surface characteristics) may need to be completed with other concordance methodologies. Despite this, we regarded Google Street View as being the most accessible and most appropriate web tool for measuring fine-grained streetscape elements.²³ Inter-rater agreement was low, and in the future, we would recommend developing systematic training adapted to the study area. The question for Destinations was adapted between the 2013 and 2014 auditing, and therefore was not included the agreement of this factor in the intra-rater analysis; future work needs to explore the reliability of this factor for the M-SPACES.

The present study also presents several strengths. As far as we know, there are no other studies that have used such a large number of street segments set to test the accuracy of virtual audits. Previous studies have used between 48 and 369 street segments.¹⁸ This is also the first study to test the validity of the virtual assessment of an audit tool that measures streets characteristics that may influence both walking and cycling in Europe, and compared areas based on residential density, a factor potentially related to physical activity. Moreover, to our knowledge, this is the first study to study the agreement between virtual and on-field tools in a Mediterranean context, where urban form patterns differ greatly from the more sprawled North-American or Australasian cities.

CONCLUSIONS

The M-SPACES audit tool is able to discriminate between different population-density areas. Google Street View provided a valid way of measuring most aspects of the residential environment in a European city like Madrid, especially neighborhood

permeability walking infrastructure, traffic safety, streetscape aesthetics, and destinations. However, for some features (e.g., street lane), the audits may need to be completed with other secondary spatial databases. Characteristics of the streets that may inhibit or promote cycling had lower correlation between on-field and virtual audits. Inter-rater agreement was, in general, weak; therefore, intensive observer training and the use of complementary objective techniques may be required. Intra-auditor agreement was substantially better when measuring urban environments virtually.

ACKNOWLEDGMENTS

We would like to thank Carlos Martínez de la Serna, for his collaboration and ideas at the early stages of the study. This study was supported by the European Research Council under the European Union's Seventh Framework Programme (FP/2007–2013)/ERC Starting Grant HeartHealthyHoods Agreement no. 336893. Usama Bilal was supported by a La Caixa Fellowship (2012 edition) and by the Enrique Nájera grant for Young Epidemiologists (10th edition) awarded by the Sociedad Española de Epidemiología and the Escuela Nacional de Sanidad. HB is supported by the NHMRC Centre for Excellence in Healthy Liveable Communities (no. 1061404) and the Australian Prevention Partnership Centre (TAPPC) (the latter is supported by NHMRC, ACT Health, NSW Health, the Australian National Preventive Health Agency (ANPHA), the Hospitals Contribution Fund of Australia (HCF), and the HCF Research Foundation)."

Authors' Contributions. MF, UB, and JD conceived the research idea. PG and HB contributed to the final design. PG and SA conducted data collection. FE and AC organized the database and prepared all the maps. PG conducted the statistical analysis. PG, HB, SA, FE, AC, JD, and MF reviewed and worked in the interpretation of the results. The first draft of the manuscript was prepared by PG and MF and reviewed by all authors. All authors have contributed and approved the final report.

REFERENCES

1. Nichols M, Townsend N, Scarborough P, Rayner M. Cardiovascular disease in Europe: epidemiological update. *Eur Heart J.* 2013; 34(39): 3028–34.
2. Beaglehole R, Bonita R. Global public health: a scorecard. *Lancet.* 2008; 372(9654): 1988–96.
3. WHO. *Global status report on noncommunicable diseases 2010.* Geneva, Switzerland: World Health Organization; 2011.
4. Shephard RJ. Is active commuting the answer to population health? *Sports Med.* 2008; 38(9): 751–8.
5. Lindstrom M. Means of transportation to work and overweight and obesity: a population-based study in southern Sweden. *Prev Med.* 2008; 46(1): 22–8.
6. Hamer M, Chida Y. Walking and primary prevention: a meta-analysis of prospective cohort studies. *Br J Sports Med.* 2008; 42(4): 238–43.
7. Lee IM, Shiroma EJ, Lobelo F, Puska P, Blair SN, Katzmarzyk PT. Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. *Lancet.* 2012; 380(9838): 219–29.

8. Rose G. Sick individuals and sick populations. *Int J Epidemiol*. 1985; 14(1): 32–8.
9. Giles-Corti B, Donovan RJ. The relative influence of individual, social and physical environment determinants of physical activity. *Soc Sci Med*. 2002; 54(12): 1793–812.
10. Mozaffarian D, Afshin A, Benowitz NL, et al. Population approaches to improve diet, physical activity, and smoking habits: a scientific statement from the American Heart Association. *Circulation*. 2012; 126(12): 1514–63.
11. Chow CK, Lock K, Teo K, Subramanian SV, McKee M, Yusuf S. Environmental and societal influences acting on cardiovascular risk factors and disease at a population level: a review. *Int J Epidemiol*. 2009; 38(6): 1580–94.
12. Townshend T, Lake AA. Obesogenic urban form: theory, policy and practice. *Health Place*. 2009; 15(4): 909–16.
13. Jackson RJ, Dannenberg AL, Frumkin H. Health and the built environment: 10 years after. *Am J Public Health*. 2013; 103(9): 1542–4.
14. Franco M, Bilal U, Diez-Roux AV. Preventing non-communicable diseases through structural changes in urban environments. *J Epidemiol Community Health*. 2015; 69(6): 509–11.
15. Wilson JS, Kelly CM, Schootman M, et al. Assessing the built environment using omnidirectional imagery. *Am J Prev Med*. 2012; 42(2): 193–9.
16. Diez Roux AV. Neighborhoods and health: where are we and where do we go from here? *Rev Epidemiol Sante Publique*. 2007; 55(1): 13–21.
17. 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. e112.
18. Charreire H, Mackenbach JD, Ouasti 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*. 2013; 25C: 1–9.
19. Vargo J, Stone B, Glanz K. Google walkability: a new tool for local planning and public health research? *J Phys Act Health*. 2012; 9(5): 689–97.
20. 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.
21. Rundle AG, Bader MD, Richards CA, Neckerman KM, Teitler JO. Using Google Street View to audit neighborhood environments. *Am J Prev Med*. 2011; 40(1): 94–100.
22. 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.
23. 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.
24. Kelly C, Wilson JS, Schootman M, Clennin M, Baker EA, Miller DK. The built environment predicts observed physical activity. *Front Public Health*. 2014; 2: 52.
25. Janssen I, Rosu A. Measuring sidewalk distances using Google Earth. *BMC Med Res Methodol*. 2012; 12: 39.
26. Carson V, Janssen I. Neighborhood disorder and screen time among 10–16 year old Canadian youth: a cross-sectional study. *Int J Behav Nutr Phys Act*. 2012; 9: 66.
27. Taylor BT, Fernando P, Bauman AE, Williamson A, Craig JC, Redman S. Measuring the quality of public open space using Google Earth. *Am J Prev Med*. 2011; 40(2): 105–12.
28. 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.
29. Griew P, Hillsdon M, Foster C, Coombes E, Jones A, Wilkinson P. Developing and testing a street audit tool using Google Street View to measure environmental supportiveness for physical activity. *Int J Behav Nutr Phys Act*. 2013; 10: 103.

30. Bethlehem JR, Mackenbach JD, Ben-Rebah M, 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.
31. Vanwolleghem G, Van Dyck D, Ducheyne F, De Bourdeaudhuij I, Cardon G. 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.* 2014; 13(1): 19.
32. Kazepov Y. *Cities of Europe: changing contexts, local arrangement and the challenge to urban cohesion.* Oxford, United Kingdom: Wiley; 2005.
33. Kasanko M, Barredo JI, Lavalle C, et al. Are European cities becoming dispersed?: A comparative analysis of 15 European urban areas. *Landsc Urban Plan.* 2006; 77(1–2): 111–30.
34. Censo de Población y viviendas. *Resultados municipales Madrid. Instituto Nacional de Estadística.* 2011. http://www.ine.es/censos2011_datos/cen11_datos_resultados.htm. Accessed 11/19/2013.
35. Kulldorff M. A spatial scan statistic. *Comm Stat Theor Meth.* 1997; 26(6): 1481–96.
36. McCormack GR, Shiell A. In search of causality: a systematic review of the relationship between the built environment and physical activity among adults. *Int J Behav Nutr Phys Act.* 2011; 8: 125.
37. Pikora T, Giles-Corti B, Bull F, Jamrozik K, Donovan R. Developing a framework for assessment of the environmental determinants of walking and cycling. *Soc Sci Med.* 2003; 56(8): 1693–703.
38. Badland HM, Schofield GM, Witten K, et al. Understanding the Relationship Between Activity and Neighbourhoods (URBAN) study: research design and methodology. *BMC Public Health.* 2009; 9: 224.
39. Pikora T. SPACES instrument: observers manual
40. Bartko JJ. The intraclass correlation coefficient as a measure of reliability. *Psychol Rep.* 1966; 19(1): 3–11.
41. DeVellis RF. *Scale development: theory and applications.* Thousand Oaks, CA: SAGE Publications; 2003.
42. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977; 33(1): 159–74.
43. 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.
44. Jago R, Baranowski T, Zakeri I, Harris M. Observed environmental features and the physical activity of adolescent males. *Am J Prev Med.* 2005; 29(2): 98–104.
45. Curtis JW, Curtis A, Mapes J, Szell AB, Cinderich A. Using Google Street View for systematic observation of the built environment: analysis of spatio-temporal instability of imagery dates. *Int J Health Geogr.* 2013; 12: 53.