Journal of Physical Activity and Health, 2004, 1, 189-207

Reliability of Two Instruments for Auditing the Environment for Physical Activity

Ross C. Brownson, Christine M. Hoehner, Laura K. Brennan, Rebeka A. Cook, Michael B. Elliott, and Kathleen M. McMullen

Purpose: To understand the relationships between street-scale environments and rates of physical activity, it is crucial to develop reliable methods of measurement. Community audits are commonly used to test the walkability and bikability of environments, yet few have been tested for reliability. *Methods:* Audit tools were collected from the peer-reviewed literature, the Internet, and via experts from a variety of backgrounds. Two versions of an audit instrument were created: an "analytic" (with Likert scale and ordinal response choices) and a "checklist" (with dichotomous response choices) audit tool. Audits were conducted in St Louis, Missouri for 147 street segments, representing both higher income and lower income neighborhoods. The same segments were re-audited to assess inter-rater reliability. *Results:* Characteristics of the physical environment varied considerably across lower and higher income segments. For example, in the checklist audit, physical disorder was present for 67 segments in lower income segments, compared with 0 segments in higher income segments. Among 8 questions from each audit tool designed to broadly capture environmental attributes, most had moderate to poor agreement. Most of the transportation and land-use items demonstrated high (substantial or perfect) agreement, while the aesthetics and social environment items showed reliability in the moderate to fair range. *Conclusions:* A community audit tool can be relatively easy and quick to administer and, for many domains, is reliable. Our audit tools appear particularly well suited for capturing elements in the transportation and land-use environments.

Key Words: ■

The positive health benefits of regular physical activity are now well established.[1] Despite these known benefits, over one quarter of the American population remains completely inactive and U.S. trends in leisure time activity showed little improvement from 1990 to 1998.[2]

The physical environment is important in providing cues and opportunities for activity,[3] and certain environmental attributes (e.g., presence of sidewalks,

The authors are with the Prevention Research Center in the School of Public Health at Saint Louis University, St. Louis, MO 63104.

scenery) are positively associated with rates of physical activity in intervention studies and in large population-based surveys.[4,5] Most studies of the effects of the physical environment have focused on community-scale changes—for example, city-wide urban-design features such as greater population density, land- use mix, or street connectivity.[6-9] Many fewer studies have sought to evaluate the effects of street-scale characteristics, sometimes called "fine grain" features. These are characteristics, policies, or practices occurring in an area of several square miles and might include quality of sidewalks, traffic volume and speed, presence of multiple destinations, or indicators of social or physical disorder (e.g., presence of garbage, graffiti). It appears that automobile trips could be more heavily influenced by community-scale characteristics, whereas walking trips might more likely be influenced by the street-scale environment.[10]

To understand the relationships between street-scale environments and rates of physical activity, it is crucial to develop reliable methods of measurement.[11] In a recent review, Moudon and Lee identified 31 instruments designed to audit the walkability and bikability of environments.[12] These instruments were drawn from a variety of disciplines, including transportation planning, urban design, and public health. Some instruments were comprehensive and sought to measure as many relevant aspects of the environment as possible, whereas others were brief and were designed to be user friendly, capturing selected characteristics quickly. In the review by Moudon and Lee, nearly 200 variables were used to capture environmental factors, and few instruments assessed walking and cycling destinations.[12] Few of these instruments have undergone rigorous testing for reliability or validity.

The present study builds on work originating from an expert review to identify evidence-based indicators of activity-friendly communities. In the first phase of this project, an extensive literature review and 3-stage Delphi process identified key indicators of activity-friendly communities. To measure these indicators, two audit tools were developed. In the present study, we tested the interrater reliability of "analytic" and "checklist" audit instruments.

Methods

Audit-Tool Development

Between November 2001 and August 2002, audit tools were collected from the peer-reviewed literature, the Internet, experts from the fields of transportation and health, and advocacy groups. Thirty-six audit tools were identified and compiled into a database, with descriptive fields for the mode of administration (trained research assistants, community audit, self-administered survey), number of items (range, 2 to 72), scoring style (e.g., point system, Likert scales), topics audited (e.g., attractiveness, comfort, convenience, safety, security), and field of origin (e.g., transportation, health, advocacy). Many of the tools identified were part of a recent review by Moudon and Lee.[12]

Items from existing audit tools were categorized by the community environment indicators from the first phase of this project (i.e., transportation environment, land-use environment, aesthetics, facilities, social environment, travel patterns, promotion, transport/economic, land use/economic, organization/policy). This helped us to select relevant items from these tools and identify where new items were necessary. To permit consistency across the response choices (i.e., 4-point Likert or ordinal scales) and to maximize the fit with the community indicators identified from the first phase, audit items were either newly developed or adapted from instruments developed by Pikora and colleagues,[13] Raudenbusch and Sampson,[14] and Kirtland and colleagues.[15] The audit instrument by Pikora and colleagues was used primarily in guiding the development of audit items related to destinations, sidewalk and bike-lane characteristics, street characteristics, and aesthetics. Audit items related to residential buildings, recreational facilities and amenities, physical disorder, signage, and the social environment were modified from the systematic social-observation tool by Raudenbusch and Sampson. The sidewalk audit by Kirtland and colleagues was used to develop items for assessing sidewalk conditions. When new items were developed (e.g., sidewalk width), efforts were made to comply with existing standards or guidelines (e.g., *Highway Capacity Manual*[16]). A multidisciplinary expert panel assembled during the first phase of the project performed a final review of the audit items, including comprehensiveness, measurement properties, and response scales.

Two versions of an audit instrument were created: an analytic and a checklist audit tool. The items of the analytic audit tool contained Likert-scale and ordinal response choices, designed to capture variation across street segments. The checklist audit tool contained items similar to the analytic audit tool but with dichotomous response choices. It was intended for use by lay community members and public health practitioners to obtain a quick and simple snapshot of an area, whereas the analytic audit was primarily intended for researchers. The instruments consisted of 6 major domains—transportation environment, land-use environment, recreational facilities, physical disorder (aesthetics), signage, and social environment. Both audit tools included 24 individual questions (some with multiple parts). The audit tools are available from http://prc.slu.edu

Selection of Study Areas

Testing the reliability of the audit tools was part of a larger study designed to assess the association of the indicators of activity-friendly communities with physical activity behavior. Higher and lower income study areas were selected among census tracts in both St Louis, MO (representing a "low walkable" city) and Savannah, GA (representing a "high walkable" city). Census-tract selection was based on the following 2000 US Census Bureau data: number of households, percentage of the population below poverty in 1999 (lowest and highest decile), and area per square mile. In addition, for each higher and lower income area, we sought adjacent census tracts with diverse land uses. The study of interrater reliability was conducted only in the St Louis study areas, composed of 1 higher income (n = 2981 households, 1.0 square miles, and 4.5% of the population below poverty level) and 3 lower income (n = 2565 total households, 1.3 square miles, and a weighted average of 56.5% of the population below the poverty level) census tracts. Maps labeled with street segment identification numbers and street names for each census tract were generated using ArcView 8.2 Geographic Information Systems (GIS) software. In St Louis, the average street segment (i.e., the length of the road between consecutive intersections) was 147 m. The length of the average street segment in Savannah was 128 m.

Training

A comprehensive protocol for audit-tool data collection was developed to address general auditing concerns (e.g., safety), audit methods (e.g., identifying and labeling street segments), and individual audit item and response-scale operational definitions. Six observers participated in a 2-d training session. The 1st day covered the audit protocol, including a comprehensive review of the audit tool item by item. The 2nd day of training familiarized the auditors with the electronic handheld data-collection devices, including software used to collect and download data, processes for labeling files, and general information regarding the devices themselves. The audit protocol is available from the first author on request.

Data Collection

The initial audits of the St Louis segments took place during daylight hours between April 1 and May 13, 2003, for the initial audit and between June 3 and June 19, 2003, for the reaudits. The audits were carried out in pairs, with each observer assigned either the analytic or the checklist audit tool. The same partners were paired throughout the audit period. A member from the community accompanied auditors in lower income study areas. During the audit, clarification of operational definitions for specific street-segment characteristics was permitted within pairs. To promote standardization across audit pairs, the entire audit team convened for periodic consultation about modified operational definitions. However, such discussions were not permitted between pairs during the reaudit period. During the initial audit, 475 (193 higher income, 282 lower income) street segments were audited. From these, a random sample of 150 (75 higher income, 75 lower income) street segments were reaudited by different observer pairs using the same instruments. The mean administration time was 10.6 min for the St Louis audits.

Data Editing and Analysis

The audit data were downloaded from the handheld units and converted to a .dbf format using Pathfinder Office software. Data on individual segments were combined into 6 files that were subsequently reviewed for missing data and for miscoded segments (i.e., invalid or missing segment name, duplicate segment names, or segments existing in 1 audit format but not the other). Possible problem areas were highlighted, and the files were returned to the auditors for correction. Auditors kept detailed notes of problems while conducting the audits and were able to draw on these while correcting the audits. The corrected audit files were again reviewed to ensure that analytic audit and checklist audit matches existed. Text responses were recoded into numeric responses (i.e., *yes* = 1, *no* = 0), and the files were transferred to SPSS® 11.0.1 for analysis. The 150 reliability checklist and analytic audits were matched to the original checklist and analytic audits by segment ID number, yielding a data set of 147 audit matches for analysis in the current study. The 3 nonmatching segments resulted from combinations of short segments that were not used by both auditing teams.

The reliability of each variable from time 1 to time 2 was assessed using the 1-way model intraclass correlation coefficient (ICC).[17] The ICC is derived using 1-way analysis of variance and represents the proportion of total variation accounted for by the variability between, rather than within, audited street segments. We used the ICC for variables that were continuous or polychotomous ordinal. Agreement for dichotomous variables was assessed using Cohen's kappa statistic K).[18] As a rough guide in interpreting results, we followed the adjectival ratings suggested by Landis and Koch[19] in the following categories: 1.0 to 0.8 (almost perfect agreement), 0.8 to 0.6 (substantial agreement), 0.6 to 0.4 (moderate agreement), 0.4 to 0.2 (fair agreement), 0.2 to 0.0 (poor agreement). Another less stringent criterion includes the percentage of items with observed agreement ≥75%, as used by Pikora and colleagues.[13]

In addition to analyses for individual items, we created summary scores for analytic audit items in the categories of the land-use environment (e.g., sum of all commercial destinations) and recreational facilities (e.g., sum of all types of recreational equipment). We also calculated the mean rating for the entire set of analytic items included in the physical disorder, signage, and social-environment sections of the analytic audit tool. For these mean ratings, weighted means were calculated as follows: *none* (weight = 0), *a little* (weight = 2), *some* (weight = 5), or *a lot* (weight = 9). The weights were based on the protocol definitions for the response choices of each of these items, where *none* = 0, *a little* = 1 to 3, *some* = 4 to 6, and *a lot* = 7 or more. For all items with dichotomous response choices, including all of the checklist audit items, we calculated a summary rating based on whether any of the items was coded as present (e.g., any physical disorder, any residential destination). Interrater reliability was assessed for each of these summary measures in addition to the individual items. All means and frequencies in the tables are based on the 1st administration of the audit instruments.

Results

Eight questions from each audit tool were designed to broadly capture environmental attributes (Table 1). In all tables, the percentages are based on the initial audits. For the checklist audit, only the question regarding integration of residential and commercial land uses showed substantial agreement. The reliability of the other questions ranged from *moderate agreement* ($n = 1$), to *fair* ($n = 5$) or *poor* ($n = 1$) agreement. The reliability of the analytic audit tool items was somewhat higher. In the analytic tool, the availability of alternative modes of transportation and physical disorder questions showed substantial agreement. The reliability of the remaining questions ranged from *moderate agreement* ($n = 2$) to *fair* ($n = 3$) or *poor* ($n = 1$) agreement. In both checklist and analytic audit tools, 4 of 8 items showed observed agreement ≥75%.

We also examined the reliability of each of these 8 broad questions across the segments in the higher and lower income areas; these data are not shown in a table. Some characteristics varied considerably across these areas. For example, in the checklist audit, physical disorder was present for 67 segments in lower income segments, compared with 0 segments in higher income segments. There was also higher availability of alternative modes of transportation in lower income segments ($n = 73$, 97%) than in higher income segments ($n = 41$, 57%). Other characteristics did not vary substantially between the 2 types of segments, such as comfort features and the people visible. In addition, the reliability of these questions varied between higher and lower income segments. From the checklist audit tool, we

Table 1 Reproducibility of Data on Broad Community Indicators, St Louis, 2003 (n = 147, Street Segments)

	Checklist Audit			Analytic Audit					
Overall indicator	Yes, % (n)	Observed agreement	Kappa	None, % (n)	A little, % (n)	Some, % (n)	A lot, % (n)	Observed agreement	ICC[a]
Are alternative transportation modes visible?	77.6 (114)	0.82	0.33	6.8 (10)	74.8 (110)	17.7 (26)	0.7 (1)	0.88	0.77
Are diverse land uses visible?	26.5 (39)	0.88	0.69	74.8 (110)	13.6 (20)	4.8 (7)	6.8 (10)	0.80	0.58
Are public recreational facilities visible?	8.2 (12)	0.90	0.29	92.5 (1[illegible]	2.7 (4)	1.4 (2)	3.4 (5)	0.91	0.29
Is public recreational equipment visible?	8.8 (13)	0.94	0.58	91.2 (134)	5.4 (8)	2.0 (3)	1.4 (2)	0.92	0.46
Are attractive features visible?	24.5 (36)	0.53	0.07	48.3 (71)	40.1 (59)	10.2 (15)	1.4 (2)	0.45	0.19
Are comfort features visible?	58.5 (86)	0.63	0.22	46.9 (69)	44.9 (66)	5.4 (8)	2.7 (4)	0.42	0.26
Is physical disorder visible?	45.6 (67)	0.64	0.25	65.3 (96)	19.0 (28)	6.8 (10)	8.8 (13)	0.68	0.64
Are people visible?	66.0 (97)	0.65	0.24	36.1 (53)	46.3 (68)	9.5 (14)	8.2 (12)	0.46	0.32

[a]Intraclass correlation coefficient derived using 1-way analysis of variance.

noted substantial agreement for availability to public recreational equipment in lower income segments but only fair agreement in higher income segments. Lower income segments showed moderate agreement for availability of comfort features, although higher income segments showed poor agreement for this question in the analytic audit tool.

Among transportation items, the reliability of most questions in the checklist audit tool was relatively high, with 3 questions showing almost perfect agreement, 5 with substantial agreement, 1 with moderate agreement, and 3 with poor agreement (Tables 2a and 2b). The analytic audit tool included parallel items from the checklist tool along with additional items related to sidewalk characteristics. The sidewalk items in the analytic audit tool showed almost perfect agreement (n = 2) or substantial agreement (n = 5). The remainder of transportation questions in the analytic audit tool showed agreement similar to the checklist items, including 3 questions with almost perfect agreement and 3 with substantial agreement. All except 1 item in the checklist audit tool satisfied the less stringent criteria of observed agreement ≥75%.

Additional questions in each audit tool examined the land-use environment and facilities (Tables 3a and 3b). For the checklist audit tool, the majority of land-use questions had either almost perfect (n = 6) or substantial (n = 6) agreement. Similar high reliability was shown for the land-use questions ascertained with the analytic audit tool. Agreement was somewhat lower for questions related to recreational facilities. Using the checklist audit tool, agreement was either almost perfect or substantial for 9 of the 12 items. For the analytic audit tool, 5 of the recreational-facility items were in the almost perfect or substantial category, with the remaining items in the moderate (n = 4), fair (n = 1) or poor (n = 2) agreement categories.

Questions on aesthetics, signage, and the social environment tended to show lower reliabilities than the other domains (Tables 4a and 4b). Most of the questions related to aesthetics showed fair or moderate agreement in the checklist tool (n = 6) and varied from substantial (n = 3) to fair (n = 2) or poor (n = 2) agreement in the analytic instrument. Of the 7 signage items, 5 demonstrated moderate agreement in the checklist tool and either substantial (n = 3) or moderate (n = 2) agreement in the analytic instrument. The items assessing the social-environment tended to have the lowest agreement of any domain. For the checklist tool, the social-environment items showed either fair (n = 3) or poor (n = 6) agreement. In the analytic tool, the social environment items ranged from moderate (n = 2) to fair (n = 5) or poor (n = 2) agreement.

Discussion

To increase population rates of physical activity, it is essential to understand, and intervene on, environmental and policy factors.[4,20,21] To conduct research studies to test these environmental hypotheses, it is essential to improve measurement of environmental variables.[22-24] One innovative approach to obtain data for measuring the community environment is the audit method.[12,13] Development of audit tools with sound measurement properties is still in the early stages.

Only 1 other study, in Perth, Australia, appears to have rigorously tested the reliability of an instrument to assess street-scale characteristics of the physical

Table 2a Reproducibility of Checklist Audit Data on the Transportation Environment, St. Louis, 2003

	% (*n*)	Observed agreement	Kappa
Presence of sidewalks		0.93	0.61
yes	91.8 (135)		
no	8.2 (12)		
Location of sidewalks	—	—	—
Continuity of sidewalks	—	—	—
Sidewalk width	—	—	—
Levelness and condition of sidewalk	—	—	—
Obstructions	—	—	—
Curvilinear curbs	—	—	—
Presence of bus stops or transit stations		0.97	0.93
yes	24.5 (36)		
no	75.5 (111)		
Presence of paths or rails		0.95	0.51
yes	4.1 (6)		
no	95.9 (141)		
General speed limit		0.93	0.77
25	9.5 (14)		
30	5.4 (8)		
35	2.0 (3)		
Special speed limit (schools)		0.99	0.97
20	4.8 (7)		
25	6.8 (10)		
On-street parking		0.90	0.61
yes	16.3 (24)		
no	83.7 (123)		
Street type		0.96	0.90
≥2 narrow lanes	70.1 (103)		
<2 narrow lanes	29.9 (44)		
Connectivity		0.97	–0.01
yes	98.0 (144)		
no	2.0 (3)		
Street-design characteristics (roundabouts)		0.90	0.73
yes	23.8 (35)		
no	76.2 (112)		
Traffic-calming devices (traffic signals)		0.93	-0.04
yes	5.4 (8)		
no	94.6 (139)		
Aggressive drivers		0.71	0.08
yes	14.3 (21)		
no	85.7 (126)		
Crossing aids (traffic island)		0.85	0.64
yes	74.8 (110)		
no	25.2 (37)		
Street lighting		1.0	—
yes	100.0 (147)		
no	0.0 (0)		

Table 2b Reproducibility of Analytic Audit Data on the Transportation Environment, St Louis, 2003

	% (*n*)	Observed agreement	Kappa
Presence of sidewalks		0.93	0.83
none	6.8 (10)		
one side of street	6.1 (9)		
both sides of street	87.1 (128)		
Location of sidewalks		0.78	0.80
adjacent to street or curb	31.3 (46)		
within 2 ft of street	5.4 (8)		
between 2 and 6 ft of street	50.3 (74)		
greater than 6 ft of street	6.1 (9)		
Continuity of sidewalks		0.90	0.83
not continuous	0.7 (1)		
continuous at one end	4.1 (6)		
continuous at both ends	88.4 (130)		
Sidewalk width		0.79	0.76
0–3 ft	8.2 (12)		
>3 to ≤6 ft	65.3 (96)		
>6 ft	19.7 (29)		
Levelness and condition of sidewalk		0.40	0.66
none	35.4 (52)		
a little	21.1 (33)		
some	17.7 (26)		
a lot	17.7 (26)		
Obstructions (cars, trees, etc.)		0.60	0.68
none	53.7 (79)		
a little	17.0 (25)		
some	11.6 (17)		
a lot	8.8 (13)		
Curvilinear curbs (or curb cuts)		0.58	0.77
none	17.7 (26)		
on only one end	10.9 (16)		
on both ends	20.4 (30)		
on both sides and ends	44.9 (66)		
Presence of public transit		0.96	0.90
none	76.2 (112)		
bus stop	23.8 (35)		
other transit stop	0.0 (0)		
multiple forms of transit	0.0 (0)		
Presence of paths or trails		0.97	0.70
none	96.6 (142)		
one side of street	2.0 (3)		
both sides of street	1.4 (2)		

(*continued*)

Table 2b (*continued*)

	% (*n*)	Observed agreement	Kappa
General speed limit		0.91	0.73
25	10.2 (15)		
30	6.1 (9)		
35	2.0 (3)		
Special speed limit (schools)		0.97	0.87
20	4.8 (7)		
25	6.8 (10)		
On-street parking		0.89	0.47
yes	87.1 (128)		
no	12.2 (18)		
Street type		0.87	0.92
undivided ≥4 lanes	10.2 (15)		
divided ≥4 lanes	19.7 (29)		
2 marked lanes	12.2 (18)		
No marked lanes	57.8 (85)		
Connectivity		0.83	0.69
segment has unidirectional intersection	0.0 (0)		
segment has 2 directions at intersection(s)	2.7 (4)		
segment has 3–4 directions at intersection(s)	24.5 (36)		
segment has 5+ directions at intersection(s)	72.8 (107)		
Street-design characteristics (roundabouts)		0.88	0.55
none	79.6 (117)		
a few	19.0 (28)		
some	0.0 (0)		
a lot	1.4 (2)		
Traffic-calming devices (traffic signals)		0.90	0.09
none	88.4 (130)		
a few	9.5 (14)		
some	2.0 (3)		
a lot	0.0 (0)		
Aggressive drivers		0.80	–0.09
none	88.4 (130)		
a few	8.8 (13)		
some	2.0 (3)		
a lot	0.7 (1)		
Crossing aids (traffic island)		0.69	0.59
none	30.6 (45)		
a few	55.1 (81)		
some	13.6 (20)		
a lot	0.7 (1)		
Street lighting		0.42	0.35
none	0.0 (0)		
a few	43.5 (64)		
some	40.1 (59)		
a lot	16.3 (24)		

[a]Intraclass correlation coefficient derived using one-way analysis of variance.

Table 3a Reproducibility of Checklist Audit Data on Land-Use Environment, Recreational Facilities, and Amenities, St Louis, 2003

	Visible, % (*n*)	Observed agreement	Kappa
Land-use environment			
family homes, single	48.3 (71)	0.94	0.88
family homes, 2 to 6	36.7 (54)	0.92	0.83
apartment buildings	11.6 (17)	0.92	0.53
All residential destinations	67.3 (99)	0.94	0.86
Fast-food restaurants	3.4 (5)	0.98	0.66
Factory	4.1 (6)	0.98	0.79
All commercial destinations	19.0 (28)	0.91	0.70
Place of worship	14.3 (21)	0.97	0.85
Health or social services	4.1 (6)	0.99	0.83
All schools	6.1 (9)	0.97	0.69
All public or government service destinations	20.4 (30)	0.97	0.89
Parking lot or garage	59.2 (87)	0.63	0.29
Driveway	77.6 (114)	0.84	0.58
Abandoned building or vacant lot	24.5 (36)	0.91	0.74
All other types of destinations	89.1 (131)	0.92	0.68
Open natural spaces	13.6 (20)	0.87	0.36
Recreational facilities			
park	5.4 (8)	0.97	0.69
playground	6.1 (9)	0.97	0.60
sports/playing field, basketball court, tennis court	6.1 (9)	0.99	0.89
All recreational destinations	12.9 (19)	0.98	0.91
Playground equipment	5.4 (8)	0.97	0.65
Complete sports equipment	4.1 (6)	0.97	0.74
Incomplete sports equipment	3.4 (5)	0.97	0.53
All types of recreational equipment	9.5 (14)	0.96	0.75
Sports stands/seating	3.4 (5)	0.98	0.56
Public telephones	3.4 (5)	0.99	0.79
Trash bins	10.2 (15)	0.91	0.47
All types of recreational amenities	15.0 (22)	0.90	0.61

Table 3b Reproducibility of Analytic Audit Data on Land Use Environment, Recreational Facilities, and Amenities, St. Louis, 2003

Variable	Mean	Observed agreement	ICC[a]
Land-use environment			
family homes, single	6.1	0.65	0.99
family homes, 2 to 6	1.3	0.78	0.83
apartment buildings	0.25	0.90	0.17
All residential destinations	7.7	0.52	0.99
Fast-food restaurants	0.03	0.98	0.66
Factory	0.05	0.97	0.61
All commercial destinations	0.33	0.85	0.86
Place of worship	0.14	0.99	0.98
Health or social services	0.04	0.99	0.83
All schools	0.05	0.97	0.74
All public or government service destinations	0.25	0.97	0.95
Parking lot or garage	0.48	0.79	0.47
Driveway	3.5	0.42	0.89
Abandoned building or vacant lot	0.56	0.76	0.42
All other types of destinations	4.6	0.29	0.79
Open natural spaces	0.14	0.86	0.26
Recreational facilities			
park	0.05	0.97	0.69
playground	0.04	0.99	0.79
sports/playing field, basketball court, tennis court	0.11	0.95	0.61
All recreational destinations	0.22	0.91	0.87
Playground equipment	0.30	0.95	0.00
Complete sports equipment	0.05	0.95	0.45
Incomplete sports equipment	0.07	0.96	0.39
All types of recreational equipment	0.42	0.90	0.12
Sports stands/seating	4.1 (6)[b]	0.97	0.49
Public telephones	2.7 (4)	0.97	0.49
Trash bins	10.2 (15)	0.90	0.41
All types of recreational amenities	14.7	0.91	0.62

[a]Intraclass correlation coefficient derived using one-way analysis of variance.
[b]Percent visible (*n*) presented for these items with dichotomous responses choices.

Table 4a Reproducibility of Data From the Checklist Audit on Physical Disorder, Signage, and the Social Environment, St Louis, 2003

Variable	Yes, % (*n*)	Observed agreement	Kappa
Physical disorder (aesthetics)			
noise pollution	21.8 (32)	0.83	0.31
whole or broken beer or liquor bottles	38.8 (57)	0.77	0.52
cigarette or cigar butts, discarded cigarette packages	49.7 (73)	0.43	–0.14
garbage, litter, or broken glass	61.2 (90)	0.78	0.54
graffiti	5.4 (8)	0.93	0.38
broken windows	12.2 (18)	0.88	0.50
all types of physical disorder	65.3 (96)	0.71	0.23
Signage			
political message or event	11.6 (17)	0.93	0.58
other pedestrian- or bicyclist-friendly traffic signs	4.8 (7)	0.90	0.16
neighborhood/crime watch	18.4 (27)	0.88	0.57
security warning signs	49.0 (72)	0.76	0.52
no trespassing/beware of dog	17.0 (25)	0.86	0.50
unreadable sign or billboard	4.8 (7)	0.93	–0.04
all types of signs	68.7 (101)	0.82	0.58
Social environment			
children visible	15.0 (22)	0.78	0.17
teenagers or adults visible	63.9 (94)	0.61	0.17
older adults visible	12.9 (19)	0.82	0.18
all people visible	68.7 (101)	0.65	0.20
children engaging in active behaviors	11.6 (17)	0.84	0.16
teenagers or adults engaging in active behaviors	53.1 (78)	0.63	0.25
older adults engaging in active behaviors	8.8 (13)	0.88	0.04
all people engaging in active behaviors	61.2 (90)	0.59	0.17
people talking	15.0 (22)	0.81	0.28

Table 4b Reproducibility of Data From the Analytic Audit on Physical Disorder, Signage, and the Social Environment, St Louis, 2003

Variable	None, % (*n*)	A little, % (*n*)	Some, % (*n*)	A lot, % (*n*)	Observed agreement	ICC[a]
Physical disorder (aesthetics)						
noise pollution	77.6 (114)	11.6 (17)	9.5 (14)	1.4 (2)	0.68	0.14
whole or broken beer or liquor bottles	62.6 (92)	23.1 (34)	7.5 (11)	6.8 (10)	0.64	0.67
cigarette or cigar butts, discarded cigarette packages	51.0 (75)	4.1 (6)	14.3 (21)	30.6 (45)	0.32	0.09
garbage, litter, or broken glass	38.1 (56)	25.9 (38)	10.9 (16)	25.2 (37)	0.62	0.76
graffiti	95.9 (141)	2.7 (4)	1.4 (2)	0.0 (0)	0.90	0.32
broken windows	90.5 (133)	4.1 (6)	2.0 (3)	3.4 (5)	0.81	0.35
all types of physical disorder	mean = 9.1	0.15	0.73			
Signage						
political message or event	90.5 (133)	8.2 (12)	1.4 (2)	0.0 (0)	0.70	0.60
other pedestrian- or bicyclist-friendly traffic signs	93.9 (138)	6.1 (9)	0.0 (0)	0.0 (0)	0.90	0.23
neighborhood/crime watch	82.3 (121)	16.3 (24)	1.4 (2)	0.0 (0)	0.88	0.52
security warning signs	54.4 (80)	38.8 (57)	4.8 (7)	2.0 (3)	0.70	0.66
no trespassing/beware of dog	81.6 (120)	17.0 (25)	1.4 (2)	0.0 (0)	0.82	0.80
unreadable sign or billboard	97.3 (143)	2.7 (4)	0.0 (0)	0.0 (0)	0.95	–0.02
all types of signs	mean = 2.5	0.44	0.64			

Social environment						
children visible	86.4 (127)	8.8 (13)	0.7 (1)	4.1 (6)	0.78	0.30
teenagers or adults visible	43.5 (64)	43.5 (64)	10.9 (16)	2.0 (3)	0.48	0.33
older adults visible	88.4 (130)	10.9 (16)	0.0 (0)	0.7 (1)	0.82	0.02
all people visible	Mean = 2.5	0.35	0.42			
children engaging in active behaviors	89.8 (132)	5.4 (8)	0.7 (1)	4.1 (6)	0.87	0.30
teenagers or adults engaging in active behaviors	57.1 (84)	35.4 (52)	6.8 (10)	0.7 (1)	0.55	0.36
older adults engaging in active behaviors	91.8 (135)	7.5 (11)	0.0 (0)	0.7 (1)	0.88	0.02
all people engaging in active behaviors	mean = 1.8	0.39	0.41			
people talking	85.0 (125)	9.5 (14)				
2.7 (4)	2.7 (4)	0.78	0.27			

[a]Intraclass correlation coefficient derived using 1-way analysis of variance.

environment.[13] A limitation of the Pikora study was the lack of variation between segments. Because we sampled a range of segments across higher and lower income neighborhoods, our study found considerable variation in the street-scale environment for many variables. Similar to the Australian study, we found that a community audit tool can be relatively easy and quick to administer and, for many domains, is reliable. Our audit tools appear particularly well suited for capturing elements in the transportation and land-use environments. Questions on aesthetics and the social environment showed lower reliability than other domains.

When we tried to assess an entire domain with global questions (e.g., a single question on the transportation environment, as shown in Table 1), reliability was often in the fair to moderate range. This suggests that multiple questions are needed to adequately measure the various domains that are important in facilitating physical activity.

Based on our experience in developing and administering these audit tools, we have identified several limitations and lessons that could be helpful for other researchers and practitioners.

- The checklist audit is easier and simpler to complete and is reasonably reliable, but we are unsure of its ability to fully capture variation or its correlation with behavior.
- The analytic audit might be infeasible for some uses as a result of the more intensive training needs and the need to standardize definitions and categories. For example, we developed decision rules for assigning land-use characteristics present on multiple adjacent segments (e.g., buildings, parking lots) to individual segments to prevent duplication.
- More work is needed to identify and measure elements of the social environment through community audits or other methods (e.g., systematic social observation[25]). The social environment, in contrast to the more static physical environment, presents many challenges for observational assessment corresponding to time of day (e.g., presence of children playing after school hours) and personal safety.
- The community environment might be in a state of change; therefore, audit tools and protocols should provide some inherent flexibility in their representation of the environment (e.g., an apartment building being torn down to leave an abandoned lot behind). Audit training should involve in-depth discussions about perceptions of the environment (e.g., designation of an abandoned lot vs. an open field).
- Our project used handheld devices that are probably too expensive for many projects. However, we asked our community members to complete paper-and-pencil versions of the checklist audit tool.
- In lower income areas, we hired local community members to accompany the auditors. This accomplished 2 purposes: safety for the auditors and a locally known individual to answer questions from residents.

- The auditors' frame of reference probably influenced the low reliability estimates for the aesthetics and physical-disorder items. For the baseline audit, 1 group audited the higher income street segments, and 2 groups audited the lower income segments. Reauditing segments in different areas might have caused auditors to be more or less sensitive to physical disorder (e.g., litter) or aesthetic features, depending on their frame of reference.
- As illustrated in our study, a low reliability coefficient does not necessarily equate with low observed agreement. There are several possible reasons for low ICCs,[26] including actual low reliability, little variation between segments on a given property, and a change in the environment (e.g., for physical disorder) from initial to reaudit. All of these possible explanations should be considered when designing future audit tools.
- Physical-disorder features and the social environment are also likely to change over time, which could explain the low reliability of these items. We attempted to minimize the influence of time by conducting the audits within a close time period to reduce the temporal effects.
- Reliability might differ across other types of neighborhoods (suburban, rural, other big cities), and we lacked the sample size and geographic coverage to address these potential differences.
- Because of the complexity of the street environment and the potential interaction of factors (e.g., aggressive drivers and an absence of traffic signals), the auditor's perception of environment as a whole could influence individual item ratings. This could be of particular relevance for less static features of the environment (e.g., aesthetics, social environment).

In studying the associations between environmental factors and physical activity, we identified a major need to determine the number of variables required to capture key determinants of physical activity behavior.[12] In the next phase of our study, we will determine the association between audit-derived indicators, perceived indicators from telephone interviews, and leisure-time and transportation physical activity behavior, based on the International Physical Activity Questionnaire.[27] To illuminate several issues, our future work will link audit data to perceived indicators derived from telephone interviews with residents of audited street segments; validate audit data against other objective, GIS-derived data; and determine the strength and patterns of the associations between audit-derived environmental data and physical activity behavior.

References

1. US Dept of Health and Human Services. *Physical Activity and Health. A Report of the Surgeon General*. Atlanta, Ga: US Dept of Health and Human Services, Centers for Disease Control and Prevention; 1996.
2. Centers for Disease Control and Prevention. Physical activity trends—United States, 1990 to 1998. *MMWR*. 2001;50(9):166-169.

3. Giles-Corti B, Donovan RJ. The relative influence of individual, social and physical environment determinants of physical activity. *Soc Sci Med*. 2002;54:1793-1812.
4. Humpel N, Owen N, Leslie E. Environmental factors associated with adults' participation in physical activity. a review. *Am J Prev Med*. 2002;22:188-199.
5. Saelens BE, Sallis JF, Frank LD. Environmental correlates of walking and cycling: findings from the transportation, urban design and planning literatures. *Ann Behav Med*. 2002;25:80-91.
6. Shriver K. Influence of environmental design on pedestrian travel behavior in four Austin neighborhoods. *Transportation Res Rec*. 1578. 1997:64-75.
7. Handy SL. Understanding the link between urban form and nonwork travel behavior. *J Plann Educ Res*. 1996;15:183-198.
8. Cervero R, Gorham R. Commuting in transit versus automobile neighborhoods. *J Amer Plann Assoc*. 1995;61:210-225.
9. Craig CL, Brownson RC, Cragg SE, Dunn AL. Exploring the effect of the environment on physical activity. a study examining walking to work. *Am J Prev Med*. 2002;23(2 suppl 1): 36-43.
10. Handy SL, Boarnet MG, Ewing R, Killingsworth RE. How the built environment affects physical activity: views from urban planning. *Am J Prev Med*. 2002;23(2 suppl):64-73.
11. Baker EA, Brennan LK, Brownson RC, Housemann RA. Measuring the determinants of physical activity in the community: current and future directions. *Res Q Exerc Sport*. 2000;71: 146-158.
12. Moudon AV, Lee C. Walking and bicycling: an evaluation of environmental audit instruments. *Am J Health Promot*. 2003;18:21-37.
13. Pikora T, Bull F, Jamrozik K, Knuiman M, Giles-Corti B, Donovan R. Developing a reliable audit instrument to measure the physical environment for physical activity. *Am J Prev Med*. 2002;23:187.
14. Raudenbusch SW, Sampson RJ. Ecometrics: toward a science of assessing ecological settings, with application to the systematic social observation of neighborhoods. *Sociol Methodol*. 1999;29:1-41.
15. Kirtland KA, Porter DE, Addy CL, et al. Environmental measures of physical activity supports: perception versus reality. *Am J Prev Med*. 2003;24:323-331.
16. Transportation Research Board, Transportation Research Council. *Highway capacity manual*. 4th ed. Washington, DC: National Research Council; 2000.
17. Fisher RA. *Statistical Methods for Research Workers*. 14th ed. New York, NY: Hafner; 1973.
18. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas*. 1960;20:37-46.
19. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159-174.
20. King AC, Jeffery RW, Fridinger F, et al. Environmental and policy approaches to cardiovascular disease prevention through physical activity: issues and opportunities. *Health Educ Q*. 1995;22:499-511.
21. Sallis JF, Bauman A, Pratt M. Environmental and policy interventions to promote physical activity. *Am J Prev Med*. 1998;15:379-397.
22. Sallis JF, Kraft K, Linton LS. How the environment shapes physical activity. a transdisciplinary research agenda (1). *Am J Prev Med*. 2002;22:208.

23. Orleans CT, Kraft MK, Marx JF, McGinnis JM. Why are some neighborhoods active and others not? charting a new course for research on the policy and environmental determinants of physical activity. *Ann Behav Med*. 2003;25(2):77-79.
24. Brownson RC, Chang JJ, Eyler AA, et al. Measuring the environment for friendliness toward physical activity: a comparison of the reliability of three questionnaires. *Am J Public Health*. 2004;94(3):473-483.
25. Sampson RJ, Raudenbush SW. Systematic social observation of public spaces: a new look at disorder in urban neighborhoods. *Am J Sociol*. 1999;105:603-651.
26. Streiner DL, Norman GR. *Health Measurement Scales: A Practical Guide to Their Development and Use*. New York, NY: Oxford University Press; 1995.
27. Craig CL, Marshall AL, Sjostrom M, et al. International physical activity questionnaire: 12-country reliability and validity. *Med Sci Sports Exerc*. 2003;35:1381-1395.

Acknowledgments

This study was funded through The Robert Wood Johnson Foundation contract #046480, including support from the Centers for Disease Control and Prevention contract U48/CCU710806 (Prevention Research Centers Program). Human-subjects approval was obtained from the Saint Louis University Institutional Review Board. The authors are grateful for the assistance of Phil Bors, Reid Ewing, Mark Fenton, Marla Hollander, Richard Killingsworth, Jim Sallis, Tom Schmid, Wendell Taylor, Anne Vernez-Moudon, and Bill Wilkinson for providing information and/or reviewing audit instruments and to Montenia Anderson, Cheryl Kelly, and Brandi Meriwether of the Prevention Research Center at Saint Louis University for assisting in data collection.