

Test–Retest Reliability of the Twin Cities Walking Survey

Ann Forsyth, J. Michael Oakes, and Kathryn H. Schmitz

Background: The Twin Cities Walking Study measured the associations of built environment versus socioeconomic and psychosocial variables with total physical activity and walking for 716 adults. **Methods:** This article reports on the test–retest reliability of the survey portion of the study. To test the reliability of the study measures, 158 respondents completed measures twice within 1 to 4 weeks. Agreement between participants' responses was measured using Pearson r and Spearman ρ , and kappa statistics. **Results:** Demographic questions are highly reliable ($R > .8$). Questions about environmental and social features are typically less reliable (ρ range = 0.42–0.91). Reliability of the International Physical Activity Questionnaire (last 7 days version) was low ($\rho = 0.15$ for total activity). **Conclusions:** Much of the survey has acceptable-to-good reliability. The low test–retest reliability points to potential limitations of using a *single* administration of the IPAQ to characterize *habitual* physical activity. Implications for sound inference are accordingly complicated.

Keywords: physical activity, built environment, NEWS, IPAQ

Overview

Rising obesity levels and high levels of physical inactivity are of immense concern to those working in the field of public health. Recently, there has been great interest in the relative influence of individual factors, the social environment, and the built environment on these 2 health

issues. To this end, the Twin Cities Walking Study (TCWS) measured the associations of built environment versus socioeconomic and psychosocial variables on total physical activity, total walking, and walking for different purposes for 716 adults. Principal outcomes are reported elsewhere.^{1–3}

Among the threats to validity, the effects of measurement error on outcomes and predictors are well known but frequently overlooked.^{4–8} In short, invalid measures amount to studying the wrong question. Unreliable measures decrease precision in estimates and, if differential across study groups, can significantly undermine any comparative inference.

The TCWS survey contains questions taken from a variety of existing, adapted, and new sources. Questions were typically taken from surveys for which there had been some reliability and/or validity testing (see Table 1).^{9,10} The major survey components included the International Physical Activity Questionnaire Long Form (IPAQ-LF),^{11,12} NEWS (Neighborhood Environment Walkability Survey),¹³ the National Household Transportation Survey (NHTS)¹⁴ for demographic variables, and a variety of questions about social and psychological supports for exercise.^{15–32} This combination was partly modeled after the survey developed for the Neighborhood Quality of Life Study soon to be posted online.^{33–35} The entire TCWS survey is available online,⁹ sources are outlined in Table 1, and analyses reported herein correspond to sections of the survey itself.

The survey instrument was compiled to be used by others doing related projects to enhance the potential for cross-study analyses. Apart from the survey, the study used a number of additional measures including a new urban design inventory, new geographic information system (GIS) based measures of the built environment, 7 days of accelerometer readings (MTI Actigraph accelerometer, MTI Inc, Fort Walton Beach, FL, with 1-minute epoch), 7 days of travel diaries, and measured height and weight. These are mentioned for context only and are not included in this reliability analysis.

However, although many questions included in the study had some level of prior testing, the character of those tests was uneven. In addition, we proposed that the reliability of measures might be differentiated by

Forsyth is with the Dept of City and Regional Planning, Cornell University, Ithaca, NY 14853. Oakes is with the Division of Epidemiology, University of Minnesota, Minneapolis, MN 55454. Schmitz is with the Center for Clinical Epidemiology and Biostatistics, University of Pennsylvania, Philadelphia, PA 19104.

Table 1 Survey Question Sources and Prior Reliability Results

Question sources	Prior test–retest reliability?
a. Single question on quality of life (US Dept of Health and Human Services, 1989) ³²	Not reported
b. Neighborhood definition (new item—open-ended question)	No
c. Social cohesion of neighborhood (Sampson et al, 1997) ³¹	No (has validity testing)
d. Social life in neighborhood	
#1–5 Neighborhood Quality of Life Survey (NQLS) ³³ developed and adapted from Parker et al, 2001 ²⁵	Adapted ^a
#6–9 (Sampson et al, 1997) ³¹	No (has validity testing)
e.–m. Neighborhood Environment Walkability Survey (NEWS): Types of residences in neighborhood, stores and facilities, access to services, streets, places for walking and cycling, neighborhood surroundings, safety from traffic, safety from crime, neighborhood satisfaction (Saelens et al, 2003) ¹³	Yes; tested scales ^{13,38}
n.–o. Home environment for exercise, convenient exercise facilities (Sallis et al, 1997) ²⁹	Yes; tested scales. Home environment $R = .89$; convenient facilities $R = .80$ (N = 110)
p. Reasons for moving to neighborhood (adapted from Frank et al, 2001) ¹⁸ (unlettered). International Physical Activity Questionnaire long form for last 7 days (IPAQ-LF; http://www.ipaq.ki.se/)	Adapted ^a Yes (Craig et al, 2003); ¹¹ Telephone version tested with N = 200; survey with several sites and N = 149, 68, 28, 19, and 30
q. Activity check list	
#1–7 (Sallis et al, 1997, 1999) ^{27,28}	Total MET hours per week all activities (exclude other activities) $R = .33$ (N = 100)
r. Stage of change for commuting (everyday commuting activity; Miilunpalo et al, 2000) ²²	No (has validity testing)
s. Stages of readiness for change in vigorous exercise (Marcus et al, 1992a, 1992b) ^{20,21}	Yes; Kappa = 0.78 (n = 20)
t. Self-confidence for vigorous exercise (Sallis et al, 1988) ³⁰	Yes; test–retest for factors (not individual questions) with reliability = 0.68 for the two exercise factors (N = 52)
u. Enjoyment of vigorous exercise (developed by NQLS group www.nqls.org) ³³	No
v. Self-confidence for moderate physical activity (Sallis et al, 1988) ³⁰	Yes; see explanation in part t
w. Enjoyment of moderate physical activity (developed by NQLS group www.nqls.org) ³³	No
x.–y. Benefits of regular physical activity, Barriers to regular physical activity (adapted from Hovell et al, 1989; Calfas et al, 1994) ^{17,19}	Adapted ^a
z. Exercise social support (Sallis et al, 1987) ²⁶	Yes; for factors. $R = .77$ family; $R = .79$ friends (n = 52)
Demographics section	
#1–22 and 27–30 and 44 (adapted from National Household Travel Survey 2001) ¹⁴	Adapted ^a
#23–26 (Bauman, 2001, 2004) ¹⁵	No
# 32–42 (based on National Survey of Pedestrian and Bicyclist Attitudes and Behaviors 2002) ²³	Adapted ^a
#43 (Oakes and Rossi, 2003) ²⁴	No

^a *Adapted* indicates that the Twin Cities Walking Study version of the survey was adapted from the source and so reliability measures of the source survey might not apply.

environmental conditions, which again, poses critically important threats to validity. Accordingly, this article reports on the test–retest reliability of that survey.

In previous evaluations, Craig et al¹¹ report a 12-country study that examined the test–retest reliability of several draft versions of the IPAQ. Only 6 sites used the long form that examined the last 7 days, the

version used in this study. Some were self-report and others over the phone. Given the hybrid approach in our study, as is explained in the following, participants followed along with a paper copy while being interviewed over the phone; we consider both in Table 1. Test–retest correlation coefficients in these studies ranged from .70 to .91 for the version used in this study. In the Craig et

al study, the highest associations were achieved between the second and third visits/administrations. A number of studies have examined the short form as well, and there are a number of published reliability analyses of the short form.^{36,37}

Brownson et al³⁸ reported on percent agreement and intraclass correlations (ICCs) for 93 variables from the NEWS survey. These included many questions used in the Twin Cities Walking Survey (sections e through m). Although many of their measures of agreement were modest, they achieved higher levels of agreement when combining individual questions into scales, as predicted by psychometric theory. This built on earlier reliability testing by Saelens et al,¹³ who examined 106 people and reported significant intraclass correlations greater than .58 for all scales with most .75 or greater (reliability was not reported for individual items, but rather scales based on sections of the NEWS inventory). Reliability testing had also been conducted with a draft version of NEWS.³⁹ In contrast, Cerin et al examined validity but not reliability of NEWS.⁴⁰

Some questions in the TCWS were adapted from the NHTS. However, to our knowledge reliability statistics have not been reported on that survey. Some of the other parts of the survey had reliability testing but often on factors rather than questions (see Table 1).

Methods

The TCWS recruitment and methods are documented elsewhere.¹⁻³ Participants came from 36 study areas, each 805 m square and selected to vary in terms of 2 dimensions thought to be important for physical activity—street pattern and population density. The larger study region was divided into a grid, and the cells were divided into low, medium, and high density and small, medium, and large block size (a block is an area bounded by streets). Participants were selected from residents of the extreme areas—low density, large block, low density, small block, and so on.¹⁻³

In each area, approximately 20 participants over 25 years of age were recruited; participants were in town during the week of data collection and were able to walk unaided for 20 minutes. The study was conducted in 4 waves from April to November; to minimize seasonal bias, approximately 5 respondents were examined from each study area each 2 months. Of participants, 74% were randomly sampled and recruited using letters and phone calls after a media campaign; the remainder were recruited to reach the area quota of 20. Recruiting into the study was a challenge, and a limitation of the study is that dropouts were not quantified—they were under 20%. Table 2 presents descriptive statistics about the sample population for the entire study.

Four trained study team staff conducted all measurements. The height and weight of participants were measured in person, and people were given a sealed copy of the survey, a copy of the travel diary, and their

accelerometer. The survey was conducted over the phone after a week of wearing the accelerometer and completing the travel diary. Participants were asked to open the paper copy at the time of the interview and to have it in front of them to follow along. Study staff completed the survey on a computer using telephone headsets. It took 45 minutes to an hour to complete. The research was approved by the Institutional Review Board of the University of Minnesota. In almost all cases, the same staff member conducted the first and second surveys.

All participants were asked if they would take the survey and do all other measures a second time. The first 20% of people who agreed were selected to do the retest resulting in a retest sample of 158 subjects. The initial sampling aimed to get equal numbers of people from each of 36 study areas (called focus areas). The retest sample was not required to have this distribution. All 158 subjects were retested within 1 to 4 weeks of their original survey. In survey test–retest analyses, a key issue is to retest after subjects have forgotten what they said but while the answers to questions are likely to be the same. The time period of 1 to 4 weeks seemed appropriate.

Table 3 presents mean scores and standard deviations for the retest sample at times 1 and 2. The table is broken into 4 main parts corresponding to parts of the survey process. First are measured variables such as height and weight included for comparison—similar to the accelerometer readings and the travel diaries, these were not part of the actual survey. Second is the first part of the survey, which focuses on perceptions of individual issues and the social and physical environment. Headings in capitals indicate the main subjects of the survey, other lines represent more detailed topics, and as mentioned earlier, the overall survey is available online.⁹ Some questions were open-ended, such as a description of one's neighborhood (question "b"), or had very few responses and are not reported here. Others were collapsed into scales as is indicated. A third set of questions came from the long form of the IPAQ. Finally, a fourth set of questions asked about demographic issues.

Statistical Methods

Pearson/Spearman correlations and kappa coefficients were estimated as measures of temporal stability. Correlation coefficients were estimated for all of the continuous or ordinal categorical measures.^{41,42} For this report, we considered a correlation coefficient of .9 and greater *excellent*, .8 to .9 *very good*, .7 to .8 *adequate/acceptable*, and below .7 *low*. For this study we were interested in seeing which of the measures had reliability (point estimates) of .7 or greater. Kappa coefficients were estimated for all of the nominal categorical and dichotomous measures. The kappa coefficient is a measure of reliability calculated by adjusting the percent agreement between the 2 time periods using the agreement expected due to chance. Kappa coefficients above

Table 2 Sample Characteristics for Entire Study

Measure	N for which data available	% in category	Min	Max	Mean	Median	SD
Covariates							
males	702	35.19					
white persons	713	81.21					
college degree	706	28.90					
married	705	58.87					
own home	704	75.28					
age in years	703		24.00	86.00	47.04	45.00	13.73
body mass index	693		16.23	66.20	28.36	27.18	6.60
Exposures							
density (persons/hectare)	36		3.55	48.91	21.72	22.30	12.39
block size (hectares)	36		1.01	10.21	3.07	2.64	1.96

.7 were also considered acceptable. All analysis was conducted using Stata, version 9.2.

Results

Table 4 shows the retest sample was largely similar to the main sample with a few exceptions. The retest sample had more participants from low-density, large-block areas, with college degrees, with better self-reported health, who walked longer distances each day, and who were overall more active each day. Other differences were not statistically significant. All subsequent tables/analyses are limited to the $N = 158$ subjects who participated in the test–retest reliability evaluation rather than the entire study sample.

Table 5 then presents Pearson r , Spearman's rho, and kappa statistics, where appropriate, along with corresponding confidence intervals and standard errors. Table 5 is also broken into 4 main parts corresponding to parts of the TCWS survey itself. Bold-font point estimates indicate correlation point estimates above or including .70 or kappa statistics above or including .70 (ie, high reliability) as described earlier.

Measured variables, height and weight, were determined by trained researchers and are highly reliable as would be expected. Acceptable reliability was observed for quality of life/health (section a), social cohesion, and social life in the neighborhood (sections c and d). Most of the NEWS survey (e–m) had adequate to excellent reliability, reported as scales, except for part e (types of residences in the neighborhood).

Questions on the exercise environment (questions n and o) had a complex pattern. There was high reliability for the exercise-oriented items in the home environment (question n) but low reliability for convenience to various facilities and access to free or low-cost facilities (question o).

Questions related to reasons for moving to the neighborhood (p) had adequate or low reliability except for closeness to public transport and quality of schools, which had good reliability.

The activity checklist (questions in section q); questions about self-confidence and readiness for exercise (questions r–v); and physical activity enjoyment, barriers, and social support (w–z) typically had low to adequate reliability using Spearman correlations as most appropriate to these survey questions. There were a few exceptions: notably, good reliability on enjoyment of vigorous activity (question u), the benefits of regular physical activity (question x), and adequate to good reliability on social supports for exercise (question z). There was high reliability on time spent watching videogames, perhaps because there was little of such activity (question q).

The IPAQ answers had low reliability. We compared IPAQ results in MET minutes per week with MET minutes using the algorithms from the IPAQ developers.¹¹

In general, demographic questions had excellent correlation coefficients or kappa statistics. There were a few exceptions including some of the dog-walking questions and a number of questions about cycling and needed changes for cycling and walking.

Discussion

The reliability of the IPAQ was low. Part of the reason for the very low IPAQ-LF reliability might be that this study had a slightly longer period between test and retest and patterns of physical activity might have changed. In this study of environmental effects on physical activity, however, people were still living in the same places 4 weeks after the first administration and weather patterns were similar. This potentially causes problems for those interested in investigating environmental effects on habitual physical activity given that activity might vary over time between people living in the same places.

It is also possible that respondents prepared themselves for the second administration of the survey by paying more careful attention to their movements. This might be the reason that the Craig et al¹¹ study found the highest associations between the second and third visits—that is people were more consistent once they knew

Table 3 Summary Statistics at Time 1 and Time 2 for Reliability Evaluation Subjects

	Time 1		Time 2	
	Mean	SD	Mean	SD
Measured Variables				
height (m)	1.70	0.09	1.70	0.09
weight (kg)	83.35	22.19	83.82	22.28
body mass index	29.00	7.41	29.16	7.47
TCWS Survey (Questions used varying 5-point scales unless otherwise noted)				
A. QUALITY OF LIFE				
a. Single question on overall health (5-point scale + NA)	3.80	0.85	3.73	0.83
B. NEIGHBORHOOD DEFINITION (Open-ended question not tested for reliability here)				
C. SOCIAL COHESION OF NEIGHBORHOOD				
c. Social Cohesion of Neighborhood (Composed of 5 questions)	19.37	3.34	19.33	3.50
D. SOCIAL LIFE IN NEIGHBORHOOD				
d.1 days of interaction (Composed of 9 questions, each asking the number of days in the past month when respondent has interacted with neighbors)	55.32	38.12	43.91	35.52
d.2 sense of community (Composed of 3 questions)	11.71	1.81	11.69	1.72
d.3 activism (Composed of 5 questions)	19.50	3.43	20.01	3.35
E. TYPES OF RESIDENCES IN NEIGHBORHOOD				
e.1 detached single family homes	4.00	0.76	3.99	0.66
e.2 townhouses	2.05	0.87	2.20	0.84
e.3 apartments, 1–3 stories	2.14	0.89	2.10	0.83
e.4 apartments, 4–6 stories	1.34	0.58	1.32	0.61
e.5 apartments, 7–12 stories	1.12	0.34	1.10	0.36
e.6 apartments, >13 stories	1.03	0.22	1.05	0.24
F. STORES AND FACILITIES				
f. Stores, facilities, and other things (Composed of 19 questions about distance in minutes to named stores and facilities)	75.33	18.25	71.73	21.33
G–M NEIGHBORHOOD COMBINED DIMENSIONS (Each original question uses a 4-point scale unless noted otherwise)				
g. Access to services (Composed of 6 questions)	18.95	3.07	18.79	3.06
h. Streets in my neighborhood (Composed of 5 questions with coding reversed)	13.85	2.31	14.12	1.98
i. Places for walking and cycling (Composed of 5 questions)	15.78	2.93	15.51	2.69
j. Neighborhood surroundings (Composed of 6 questions)	19.78	3.10	19.33	3.16
k. Safety from traffic (Composed of 8 questions)	22.66	4.48	22.89	4.39
l. Safety from crime (Composed of 6 questions)	19.39	3.13	19.25	3.04
m. Neighborhood satisfaction (Composed of 17 questions each on a 5-point scale)	61.89	9.56	62.65	8.94
N–O EXERCISE ENVIRONMENT				
n. Home environment for exercise (Composed of questions about 15 items, each yes, no/don't know)	6.36	2.62	6.56	2.75
o. Convenient exercise facilities (Composed of questions about 18 facilities, each yes, no/don't know)	11.38	3.77	11.41	3.64
o.19 free/low-cost rec facilities (4-point scale)	3.57	0.61	3.37	0.70
P. REASONS FOR MOVING TO NEIGHBORHOOD				
p.1 affordability/value	4.45	0.97	4.51	0.77
p.2 closeness to open space	3.49	1.20	3.50	1.09
p.3 closeness to job or school	3.60	1.40	3.84	1.25
p.4 closeness to public transport	2.36	1.48	2.55	1.44

(continued)

Table 3 (continued)

	Time 1		Time 2	
	Mean	SD	Mean	SD
p.5 desire for nearby shops/service	2.62	1.14	2.81	1.10
p.6 ease of walking	3.05	1.22	3.16	1.19
p.7 sense of community	3.11	1.20	3.19	1.18
p.8 safety from crime	4.10	1.06	4.08	1.05
p.9 quality of schools	3.24	1.66	3.39	1.52
p.10 closeness to rec facilities	2.73	1.15	2.72	1.11
p.11 access to freeways	3.24	1.31	3.30	1.23
Q. ACTIVITY CHECKLIST				
q.1 computer min per week	219.85	374.31	193.71	283.61
q.2 videogames min per week	148.33	563.78	62.96	163.84
q.3 reading min per week	304.34	446.53	273.73	411.76
q.4 sitting min per week	248.06	414.31	189.32	333.07
q.5 phone min per week	163.07	230.58	157.55	361.53
q.6 TV min per week	621.02	569.13	595.96	535.42
q.7 car min per week	156.95	189.66	135.00	199.19
R–V SELF-CONFIDENCE AND READINESS				
r. Stage of change for commuting (Composed of 8 levels [8-point scale] of active commuting)	3.56	2.04	3.40	1.89
s. Stages of readiness for change in vigorous exercise (Composed of a 5-point scale—lower is better)	3.17	1.58	3.18	1.56
t. Self-confidence for vigorous exercise (Composed of 5 questions)	11.04	2.95	10.88	2.76
u. Enjoyment of vigorous exercise (Composed of 3 questions)	11.78	2.75	11.94	2.63
v. Self-confidence for moderate physical activity (Composed of 3 questions)	12.55	2.51	11.91	2.57
W–Z PHYSICAL ACTIVITY—ENJOYMENT, BARRIERS, AND SOCIAL SUPPORT				
w. Enjoyment of moderate physical activity (Not compiled)				
x. Benefits of regular physical activity (Composed of 10 questions)	43.22	5.09	42.84	5.00
y. Barriers to regular physical activity				
y.1 self-conscious about looks	1.74	0.89	1.73	0.90
y.2 lack of interest—exercise	2.38	1.02	2.61	1.07
y.3 lack of self-discipline	3.20	1.17	3.10	1.18
y.4 lack of time	3.21	1.24	3.29	1.22
y.5 lack of energy	2.87	1.04	2.99	0.92
y.6 lack of company	2.04	1.18	2.14	1.19
y.7 lack of enjoyment	2.08	1.03	2.07	0.95
y.8 discouragement	2.02	1.07	2.08	0.94
y.9 lack of equipment	2.01	1.13	2.02	1.02
y.10 lack of good weather	2.39	1.14	2.44	1.07
y.11 lack of skills	1.79	0.86	1.87	0.84
y.12 lack of facilities	1.94	1.07	2.08	1.05
y.13 lack of knowledge	1.78	0.91	1.74	0.86
y.14 lack of good health	1.88	0.98	1.95	1.02
y.15 fear of injury	1.63	0.86	1.69	0.93
z. Exercise social support				
z.1 family support (Composed of 3 questions)	8.58	3.84	7.92	3.45
z.2 friend support (Composed of 3 questions)	7.03	3.56	6.88	3.18

(continued)

Table 3 (continued)

	Time 1		Time 2	
	Mean	SD	Mean	SD
IPAQ				
IPAQ Part 1: Job Related (MET min per week)				
IPAQ1.1 Vigorous	15.54	43.28	23.51	54.90
IPAQ1.2 Moderate	22.66	49.79	27.22	57.32
IPAQ1.3 Walking	27.10	57.47	32.69	64.76
IPAQ Part 2: Transport Physical Activity				
IPAQ2.1 Bike	5.84	19.99	5.44	21.49
IPAQ2.2 Walk	20.57	35.02	25.65	46.23
IPAQ Part 3: Housework				
IPAQ3.1 Vigorous	30.19	55.00	40.57	62.67
IPAQ3.2 Outside Moderate	44.97	53.16	51.80	60.45
IPAQ3.3 Inside Moderate	54.94	51.99	58.45	55.21
IPAQ Part 4: Recreation Physical Activity				
IPAQ4.1 Walk	29.23	31.44	37.83	47.91
IPAQ4.2 Vigorous	15.82	33.46	26.23	49.40
IPAQ4.3 Moderate	13.07	27.41	23.77	49.83
IPAQ Part 5: Sitting				
IPAQ5.1 Weekday	347.24	213.78	319.18	209.68
IPAQ5.2 Weekend	254.19	180.94	248.69	164.10
IPAQ TOTALS (in MET min per week)				
work total	1260.64	3128.65	898.88	2523.01
transport total	349.89	639.23	661.27	5178.25
domestic total	1560.74	1818.50	1455.36	1923.66
leisure total	844.68	1179.31	758.32	923.99
total walk	935.22	1089.84	815.06	1071.70
total physical activity	4015.96	3902.37	3773.84	6089.90
Demographics				
Dem.2 Time at address	162.38	132.79	164.98	134.59
Dem.4 People in household	2.64	1.29	2.64	1.31
Dem.6 Age	46.30	12.82	46.47	12.92
Dem.12a Number of children	1.88	0.89	1.90	0.93
Dem.19 Work address	8.64	6.12	8.22	5.98
Dem.20 Minutes home to work	19.80	12.93	19.93	12.86
Dem.23 Amount work from home	2.71	1.20	2.64	1.15
Dem.27 Time spent walking dog	3.25	0.87	3.42	0.94
Dem.28 Vehicles in household	1.97	1.03	1.98	1.02
Dem.29 Type of Vehicle	1.73	0.99	1.78	1.03
Dem.31 Number of bikes	1.92	1.69	1.89	1.43
Dem.33 Most recent time riding bike	2.87	1.18	2.99	1.07
Dem.39 Satisfaction with community design for bikes	2.31	1.17	2.21	1.14
Dem.43 Friends who are professionals	2.48	1.06	2.45	1.10
Dem.44 Household income	6.63	3.06	6.52	3.06

what to remember for the survey. However, this still presents a problem for those wishing to use a *single* administration of the IPAQ.

Some of this low reliability might be the result of the character of the study. Although survey researchers were highly trained, there might have been some differ-

ences in the way they conducted the parts of the survey that showed differences in reliability. In addition, the study had a large age distribution. However, we performed Pearson and Spearman correlations between physical activity totals at time 1 and time 2 for several domains of physical activity (work, transport, domestic,

Table 4 Comparison of Characteristics of Subjects In and Out of Reliability Evaluation

	Subjects not in reliability evaluation		Subjects in reliability evaluation		Difference	Test statistic	P
	N	% or Mean ^a	N	% or Mean ^a			
Total	558		158				
Women	365	65.41	101	64.33	-1.08		
Men	193	34.59	56	35.67	1.08	$\chi^2 = 0.06$.80
HDLB (strata)	145	25.99	36	22.78	-3.20		
LDLB	128	22.94	52	32.91	9.97		
HDSB	141	25.27	37	23.42	-1.85		
LDSB	144	25.81	33	20.89	-4.92	$\chi^2 = 6.75$.08
Randomly selected	407	72.94	123	77.85	4.91		
Self-selected	151	27.06	35	22.15	-4.91	$\chi^2 = 1.54$.21
White	447	80.40	132	84.08	3.68	$\chi^2 = 1.09$.3
College degree	146	26.21	58	38.93	12.71	$\chi^2 = 9.25$.002
Married	321	57.73	94	63.09	5.35	$\chi^2 = 1.39$.24
Own home	416	74.82	114	77.03	2.21	$\chi^2 = 0.31$.58
Body mass index	536	28.21	157	28.88	0.67	$t = -1.05$.29
Age	555	47.45	148	45.51	-1.94	$t = 1.62$.11
Self-rated overall health (5 = Excellent)	556	3.62	149	3.77	0.15	$t = -1.81$.07
Household income (median)	506	56.40	136	59.30	2.90	$t = -0.90$.37
IPAQ Travel walking (met minutes per week)	558	257.52	158	259.51	1.99	$t = -0.04$.97
IPAQ Leisure walking (met minutes per week)	558	316.00	158	315.44	-0.56	$t = 0.01$.99
Diary miles walked per day (miles)	555	0.75	158	1.20	0.45	$t = -3.29$.00
Accelerometer, total count per day (1000s)	555	219.89	158	237.85	17.96	$t = -1.83$.07

^a Unless otherwise stated (eg, median).

leisure, total walk, and total physical activity) for both the total population and separately for those age 25 to 39, 40 to 59, and 60 to 80. We found no systematic differences in IPAQ measures.

Given these issues, the low test-retest reliability points to potential problems and limitations of a *single* administration of the IPAQ to characterize *habitual* physical activity. Further work assessing the reliability of the IPAQ compared with the travel diary and accelerometer data is ongoing.

Apart from the IPAQ-LF, the Twin Cities Walking Survey includes highly reliable demographic sections and many moderately to highly reliable questions about perceptions of the built environment as well as social and personal perceptions. Of those that were less reliable, some questions might well be able to be dropped because of the increased capacity of GIS—for example,

questions about perceptions of housing type. Although it is interesting to compare the effects of perceived and measured environments and the interactions between the two, in predicting physical activity, other measures might be more useful such as perceived accessibility to facilities or perceived difficulty of parking. Overall, although many measures have clearly acceptable reliability, there is more work to do to develop measures of the perceived environment related to physical activity.

Acknowledgments

Thanks to the Active Living Research program of the Robert Wood Johnson Foundation, which funded this study. Thanks also for excellent work on statistics and surveys by Becky Lien, Dan Pepper, Gail Babes, Kimberly Carolan, Nathan Mitchell, and Patricia Battin.

Table 5 Test-Retest Correlations for Reliability Sample

Question	Pearson <i>r</i>	95% CI	Spearman rho	95% CI	Kappa	SE (Kappa)	Derived scale or individual item
Measured Variables							
mean height	.99	.99, 1.00	.99	.99, 1.00			individual
mean weight	1.00	.99, 1.00	1.00	.99, 1.00			individual
body mass index	1.00	.99, 1.00	1.00	.99, 1.00			individual
TCWS Survey							
a. Single question on overall health	.76	.68, .82	.76	.67, .82			individual
c. Social Cohesion of Neighborhood	.77	.69, .84	.76	.67, .82			scale
d. Social life in neighborhood							
d.1 days	.81	.75, .87	.78	.71, .84			scale
d.2 community	.79	.72, .85	.79	.72, .85			scale
d.3 activism	.77	.69, .84	.73	.63, .81			scale
NEWS							
e. Types of residences in neighborhood							
e.1 detached single family homes	.61	.50, .71	.57	.45, .67			individual
e.2 townhouses	.56	.44, .66	.51	.38, .62			individual
e.3 apartments 1–3 stories	.59	.48, .69	.56	.44, .66			individual
e.4 apartments 4–6 stories	.52	.40, .63	.56	.44, .66			individual
e.5 apartments 7–12 stories	.47	.33, .59	.56	.44, .67			individual
e.6 apartments >13 stories	.75	.67, .81	.77	.69, .83			individual
f. Stores, facilities, and other things	.91	.82, .96	.89	.78, .95			individual
g. Access to services	.79	.72, .85	.77	.69, .83			scale
h. Home environment for exercise	.67	.53, .77	.69	.55, .79			scale
i. Places for walking and cycling	.68	.56, .77	.69	.58, .78			scale
j. Neighborhood surroundings	.75	.66, .82	.76	.67, .83			scale
k. Safety from traffic	.71	.60, .79	.67	.56, .76			scale
l. Safety from crime	.81	.74, .86	.78	.71, .84			scale
m. Neighborhood satisfaction	.82	.67, .90	.81	.65, .90			scale
n. Home environment for exercise	.90	.86, .93	.89	.85, .93			scale
o. Convenient exercise facilities	.68	.44, .83	.65	.39, .81			scale
o.19 free/low-cost rec facilities	.51	.37, .63	.49	.35, .61			individual
p. Reasons for moving to neighborhood							
p.1 affordability/value	.44	.30, .57	.42	.27, .55			individual
p.2 closeness to open space	.55	.42, .66	.55	.42, .66			individual
p.3 closeness to job or school	.60	.47, .70	.56	.43, .67			individual
p.4 closeness to public transport	.80	.73, .86	.80	.72, .85			individual
p.5 desire for nearby shops/service	.63	.51, .72	.62	.51, .72			individual
p.6 ease of walking	.71	.61, .78	.69	.58, .77			individual
p.7 sense of community	.70	.61, .78	.73	.64, .80			individual
p.8 safety from crime	.57	.44, .68	.58	.45, .68			individual
p.9 quality of schools	.78	.71, .84	.80	.73, .86			individual
p.10 closeness to rec facilities	.41	.26, .55	.41	.26, .55			individual
p.11 access to freeways	.71	.61, .79	.72	.62, .79			individual
Q. Activity checklist							
q.1 computer min per week	.63	.49, .73	.72	.62, .81			individual
q.2 videogames min per week	.85	.70, .93	.94	.87, .97			individual
q.3 reading min per week	.62	.50, .72	.66	.54, .75			individual
q.4 sitting min per week	.69	.60, .77	.40	.25, .53			individual

(continued)

Table 5 (continued)

Question	Pearson <i>r</i>	95% CI	Spearman		Kappa	SE (Kappa)	Derived scale or individual item
			rho	95% CI			
q.5 phone min per week	.80	.73, .86	.62	.49, .72			individual
q.6 TV min per week	.74	.65, .81	.69	.58, .77			individual
q.7 car min per week	.40	.10, .63	.53	.27, .72			individual
r. Stage of change for commuting	.68	.58, .77	.61	.49, .71			individual
s. Stages of readiness for change in vigorous exercise	.74	.66, .81	.73	.64, .80			individual
t. Self-confidence for vigorous exercise	.65	.54, .75	.66	.54, .75			scale
u. Enjoyment of vigorous exercise	.69	.58, .77	.80	.73, .86			scale
v. Self-confidence for moderate physical activity	.52	.39, .64	.49	.35, .61			scale
x. Benefits of regular physical activity	.83	.76, .88	.82	.75, .88			scale
Y. Barriers to regular physical activity							
y.1 self-conscious about looks	.71	.61, .78	.69	.59, .77			individual
y.2 lack of interest—exercise	.60	.47, .70	.57	.44, .67			individual
y.3 lack of self-discipline	.78	.70, .84	.78	.70, .84			individual
y.4 lack of time	.75	.66, .82	.75	.66, .82			individual
y.5 lack of energy	.59	.46, .69	.56	.43, .67			individual
y.6 lack of company	.70	.60, .78	.68	.58, .77			individual
y.7 lack of enjoyment	.59	.46, .69	.62	.50, .72			individual
y.8 discouragement	.46	.32, .59	.47	.32, .59			individual
y.9 lack of equipment	.59	.46, .69	.55	.42, .66			individual
y.10 lack of good weather	.54	.41, .65	.53	.39, .64			individual
y.11 lack of skills	.55	.42, .66	.57	.44, .68			individual
y.12 lack of facilities	.61	.49, .71	.57	.44, .68			individual
y.13 lack of knowledge	.72	.63, .79	.69	.59, .77			individual
y.14 lack of good health	.63	.51, .72	.52	.39, .64			individual
y.15 fear of injury	.42	.27, .55	.38	.22, .52			individual
z. Exercise social support							
z.1 family support	.79	.72, .85	.80	.73, .86			scale
z.2 friend support	.78	.71, .84	.77	.69, .83			scale
IPAQ							
IPAQ Part 1: Job Related							
IPAQ1.1 Vigorous	.22	.07, .36	.32	.17, .45			individual
IPAQ1.2 Moderate	.38	.24, .51	.48	.35, .59			individual
IPAQ1.3 Walking	.42	.28, .54	.36	.22, .49			individual
IPAQ-Part 2: Transport Physical Activity							
IPAQ2.1 Bike	.30	.15, .44	.36	.21, .49			individual
IPAQ2.2 Walk	.22	.06, .36	.29	.14, .43			individual
IPAQ-Part 3: Housework							
IPAQ3.1 Vigorous	.25	.10, .39	.35	.20, .48			individual
IPAQ3.2 Outside Moderate	.29	.14, .43	.43	.29, .55			individual
IPAQ3.3 Inside Moderate	.11	-.05, .26	.09	-.07, .24			individual
IPAQ-Part4: Recreation Physical Activity							
IPAQ4.1 Walk	.14	-.02, .29	.20	.04, .34			individual
IPAQ4.2 Vigorous	.26	.11, .40	.51	.39, .62			individual
IPAQ4.3 Moderate	.22	.07, .37	.36	.21, .49			individual
IPAQ-Part 5: Sitting							
IPAQ5.1 Weekday	.63	.52, .72	.62	.51, .71			individual

(continued)

Table 5 (continued)

Question	Pearson <i>r</i>		Spearman		Kappa	SE (Kappa)	Derived scale or individual item
	<i>r</i>	95% CI	rho	95% CI			
IPAQ5.2 Weekend	.30	.15, .44	.30	.15, .44			individual
IPAQ TOTALS							
work total	.32	.17, .45	.51	.38, .61			individual
transport total	.26	.11, .40	.33	.18, .46			individual
domestic total	.33	.19, .46	.27	.12, .41			individual
leisure total	.58	.46, .67	.50	.37, .61			individual
total walk	.39	.25, .52	.27	.12, .41			individual
total physical activity	.15	.00, .30	.31	.16, .45			individual
Demographics							
Dem.1 type of house					.92	.06	individual
Dem.2 time at address	.99	.99, .99	.99	.98, .99			scale
Dem.3 own or other					.95	.09	individual
Dem.4 people in household	.95	.94, .97	.96	.94, .97	n/a	n/a	individual
Dem.5 relatives					.83	.10	individual
Dem.6 age	1.00	.99, 1.00	1.00	.99, 1.00	n/a	n/a	individual
Dem.7 sex					.98	.09	individual
Dem.8 Hispanic/Latino					.80	.09	individual
Dem.9 race					1.00	.08	individual
Dem.10 education					.86	.05	individual
Dem.11 marital status					.96	.06	individual
Dem.12 parent of children in house					.85	.09	individual
Dem.12a number of children	.94	.89, .96	.97	.94, .98			individual
Dem.14 work for pay	n/a	n/a	n/a	n/a	.76	.09	individual
Dem.18 drive at work	n/a	n/a	n/a	n/a	.78	.11	individual
Dem.19 address of workplace	.89	.84, .93	.88	.82, .92			individual
Dem.20 minutes home to work	.89	.83, .93	.85	.78, .90			individual
Dem.22 work from home in past 2 months					.73	.11	individual
Dem.23 how often work from home	.75	.37, .92	.73	.33, .91			individual
Dem.24 own dog					1.00	.09	individual
Dem.25 walk dog					.48	.14	individual
Dem.26 how often walk dog					.60	.12	individual
Dem.27 time to walk dog	.74	.55, .86	.72	.51, .85			individual
Dem.28 number of vehicles	.91	.87, .93	.94	.91, .96			individual
Dem.29 vehicle type	.76	.68, .82	.78	.70, .83			individual
Biking questions							
Dem.31 number bikes	.90	.86, .93	.91	.88, .94			individual
Dem.32 ridden bike in past 2 years					.85	.09	individual
Dem.33 most recent bike in neighborhood	.67	.53, .78	.67	.53, .77			individual
Dem.34 personal safety biking					.38	.11	individual
Dem.37 injured biking?					.90	.11	individual
Dem.38 injured by vehicle biking?					1.00	.45	individual
Dem.39 satisfied design for biking	.73	.63, .80	.68	.58, .77			individual
Dem.40 change needed for biking					.66	.09	individual
Dem.41a need bike paths/trails					.46	.17	individual
Dem.41b need bike lanes					.10	.17	individual
Dem.41c allow bikes on sidewalks					1.00	.29	individual
Dem.41d don't allow bikes on sidewalks					.53	.27	individual

(continued)

Table 5 (continued)

Question	Pearson <i>r</i>		Spearman		Kappa	SE (Kappa)	Derived scale or individual item
	95% CI		rho	95% CI			
Dem.42a need more crosswalks					.55	.13	individual
Dem.42b need sidewalks					.74	.13	individual
Dem.42c need street lights					.58	.13	individual
Dem.42d need lights on paths/trails					.44	.15	individual
Dem.43 professionals close friends	.81	.74, .86	.80	.73, .85	n/a	n/a	individual
Dem.44 household income	.88	.83, .91	.88	.83, .91	n/a	n/a	individual

References

- Forsyth A, Hearst M, Oakes JM, Schmitz MK. Design and destinations: factors influencing walking and total physical activity. *Urban Stud.* 2008;45(9):1973–1996.
- Oakes JM, Forsyth A, Schmitz MK. The effect of neighborhood density and street connectivity on walking behavior: The Twin Cities Walking Study. *Epidemiol Perspect Innov.* 2007;4:16. <http://www.epi-perspectives.com/content/4/1/16/abstract>.
- Forsyth A, Oakes JM, Schmitz MK, Hearst M. Does residential density increase walking and other physical activity? *Urban Stud.* 2007;44(4):679–697.
- Bound J, Brown C, Mathiowetz N. Measurement error in survey data. In: Heckman JJ, Leamer E, eds. *Handbook of Econometrics*. Vol 5. New York, NY: Elsevier; 2001:3706–3843.
- Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu CM. *Measurement Error in Nonlinear Models: A Modern Perspective*. New York, NY: Chapman & Hall/CRC; 2006.
- Fuller WA. *Measurement Error Models*. New York, NY: Wiley; 1987.
- Gustafson P. *Measurement Error and Misclassification in Statistics and Epidemiology*. Boca Raton, LA: Chapman & Hall/CRC; 2004.
- Nunnally JC, Bernstein I. *Psychometric Theory*. New York, NY: McGraw-Hill; 1994.
- Twin Cities Walking Study. Twin Cities Walking Survey. 2005a. <http://www.activelivingresearch.org/node/10619>. Accessed November 2008.
- Twin Cities Walking Study. Twin Cities Walking Survey References. 2005b. <http://www.activelivingresearch.org/node/10619>. Accessed November 2008.
- Craig CL, Marshall AL, Sjöström M, et al, and the IPAQ Consensus Group and the IPAQ Reliability and Validity Study Group. International Physical Activity Questionnaire (IPAQ): 12-country reliability and validity. *Med Sci Sports Exerc.* 2003;35:1381–1395.
- International Physical Activity Questionnaire. 2007. IPAQ long form for last seven days. <http://www.ipaq.ki.se/>. Accessed November 2008.
- Saelens BE, Sallis JF, Black JB, Chen D. Neighborhood-based differences in physical activity: an environment scale evaluation. *Am J Public Health.* 2003;93:1552–1558.
- National Household Travel Survey. 2001. Users guide. http://nhts.ornl.gov/2001/usersguide/appendix_m.pdf. Accessed November 2008.
- Bauman A, Russell SJ, Furber S, Dobson AJ. The epidemiology of dog walking: an unmet need for human and canine health. *Med J Aust.* 2001;175:632–634.
- Bauman A. 2004. Personal communication to James Sallis and shared with Active Living Research grantees.
- Calfas KJ, Sallis JF, Lovato CY, Campbell J. Physical activity and its determinants before and after college graduation. *Med Exerc Nutr Health.* 1994;3:323–334.
- Frank L, Leerssen C, Chapman J, Contrino H. *Strategies for Metropolitan Atlanta's Regional Transportation and Air Quality (SMARTRAQ)*. Atlanta, GA: Georgia Institute of Technology; 2001.
- Hovell MF, Sallis JF, Hofstetter CR, Spry VM, Faucher PF, Caspersen CJ. Identifying correlates of walking for exercise: an epidemiologic prerequisite for physical activity promotion. *Prev Med.* 1989;18:856–866.
- Marcus B, Rossi J, Selby V, Niaura R, Abrams D. The stages and processes of exercise adoption and maintenance in a worksite sample. *Health Psychol.* 1992a;11:386–395.
- Marcus B, Selby V, Niaura R, Rossi J. Self-efficacy and the stages of exercise behavior change. *Res Q Exerc Sport.* 1992b;63:60–66.
- Miilunpalo S, Nupponen R, Laitakari J, Marttila J, Paronen O. Stages of change in two modes of health-enhancing physical activity: methodological aspects and promotional implications. *Health Educ Res.* 2000;15:435–448.
- National Survey of Pedestrian and Bicyclist Attitudes and Behaviors. 2002. Survey questionnaire. http://www.bts.gov/programs/omnibus_surveys/targeted_survey/2002_national_survey_of_pedestrian_and_bicyclist_attitudes_and_behaviors/survey_questionnaire/. Accessed November 2008.
- Oakes JM, Rossi PH. The measurement of SES in health research: current practice and steps toward a new approach. *Soc Sci Med.* 2003;56:769–784.
- Parker EA, Lichtenstein RL, Schulz AJ, et al. Disentangling measures of individual perceptions of community social dynamics: results of a community survey. *Health Educ Behav.* 2001;28:462–486.
- Sallis JF, Grossman RM, Pinski RB, Patterson TL, Nader PR. The development of scales to measure social support

- for diet and exercise behaviors. *Prev Med.* 1987;16:825–836.
27. Sallis JF, Johnson MF, Calfas KJ, Caparosa S, Nichols JF. Assessing perceived environment variables that may influence physical activity. *Res Q Exerc Sport.* 1997;68:345–351.
 28. Sallis JF, McKenzie TL, Elder JP, Conway T. Middle School Physical Activity and Nutrition (MSPAN) Student Survey, 1999. <http://www.drjamesallis.sdsu.edu/measures.html>. Accessed November 2008.
 29. Sallis JF, McKenzie TL, Elder JP, Conway T. MSPAN Student/Parent Survey Reliabilities, 1997. <http://www.drjamesallis.sdsu.edu/measures.html>. Accessed November 2008.
 30. Sallis JF, Pinski RB, Grossman RM, Patterson TL, Nader PR. The development of self-efficacy scales for health-related diet and exercise behaviors. *Health Educ Res.* 1988;3:283–292.
 31. Sampson RJ, Raudenbush SW, Earls F. Neighborhoods and violent crime: a multilevel study of collective efficacy. *Science.* 1997;277:918–924.
 32. US Department of Health and Human Services, National Center for Health Statistics. National Health Interview Survey 1989 [Computer file]. Hyattsville, MD: National Center for Health Statistics [producer]; 1989. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor]; 1991.
 33. Neighborhood Quality of Life Survey. <http://www.drjamesallis.sdsu.edu/measures.html>. Accessed November 2008.
 34. Frank L, Sallis J, Conway T, Chapman J, Saelens B, Bachman W. Many pathways from land use to health: associations between neighborhood walkability and active transportation, body mass index, and air quality. *J Am Plann Assoc.* 2006;72:75–87.
 35. Owen N, Cerin E, Leslie E, et al. Neighborhood walkability and the walking behavior of Australian adults. *Am J Prev Med.* 2007;33:387–395.
 36. Brown WJ, Trost SG, Bauman A, Mummery K, Owen N. Test-retest reliability of four physical activity measures used in population surveys. *J Sci Med Sport.* 2004;7(2):205–215.
 37. Mader U, Martin B, Schutz Y, Marti B. Validity of four short physical activity questionnaires in middle-aged persons. *Med Sci Sports Exerc.* 2006;38:1255–1266.
 38. Brownson RC, Chang JJ, Eyler AA, et al. Measuring the environment for friendliness toward physical activity: a comparison of the reliability of 3 questionnaires. *Am J Public Health.* 2004;94(3):473–483.
 39. Bourdeaudhuij ID, Sallis JF, Saelens BE. Environmental correlates of physical activity in a sample of Belgian adults. *Am J Health Promot.* 2003;18:83–92.
 40. Cerin E, Saelens B, Sallis J, Frank L. Neighborhood Environment Walkability Scale: validity and development of a short form. *Med Sci Sports Exerc.* 2006;1682–1691.
 41. Kline RB. *Principles and Practice of Structural Equation Modeling.* 2nd ed. New York, NY: Guilford Press; 2005.
 42. Netemeyer RG, Bearden WO, Sharma S. *Scaling Procedures: Issues and Applications.* Thousand Oaks, CA: Sage Publications; 2003.