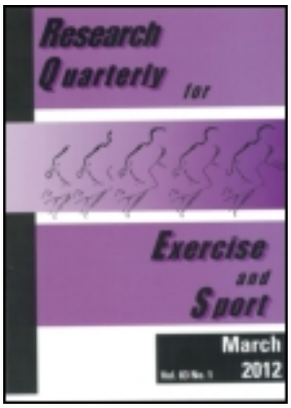

Research Quarterly for Exercise and Sport
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/urqe20

Reliability and Validity of the International Physical Activity Questionnaire for Assessing Walking

Hidde P. van der Ploeg [a], Catrine Tudor-Locke [b], Alison L. Marshall [d], Cora Craig [a c], Maria Hagströmer [e], Michael Sjöström [e] & Adrian Bauman [a c]

[a] Sydney School of Public Health, University of Sydney

[b] Pennington Biomedical Research Center, Baton Rouge, LA

[c] Canadian Fitness and Lifestyle Research Institute, Ottawa

[d] School of Public Health, Queensland University of Technology

[e] Department of Biosciences and Nutrition, Karolinska Institute

Published online: 23 Jan 2013.

To cite this article: Hidde P. van der Ploeg, Catrine Tudor-Locke, Alison L. Marshall, Cora Craig, Maria Hagströmer, Michael Sjöström & Adrian Bauman (2010) Reliability and Validity of the International Physical Activity Questionnaire for Assessing Walking, Research Quarterly for Exercise and Sport, 81:1, 97-101, DOI: 10.1080/02701367.2010.10599632

To link to this article: http://dx.doi.org/10.1080/02701367.2010.10599632

Research Quarterly for Exercise and Sport

Vol. 81, No. 1, pp. 97–101

Reliability and Validity of the International Physical Activity Questionnaire for Assessing Walking

Hidde P. van der Ploeg, Catrine Tudor-Locke, Alison L. Marshall, Cora Craig, Maria Hagströmer, Michael Sjöström, and Adrian Bauman

Key words: adult, measurement properties, physical activity surveillance

Physical inactivity and its accompanying adverse sequelae (e.g., obesity and diabetes) are global health concerns. The single most commonly reported physical activity in public health surveys is walking (Centers for Disease Control and Prevention, 2000; Rafferty, Reeves, McGee, & Pivarnik, 2002). As evidence accumulates that walking is important for preventing weight gain (Levine et al., 2008) and reducing the risk of diabetes (Jeon, Lokken, Hu, & van Dam, 2007), there is increased need to capture this behavior in a valid and reliable manner.

Although the disadvantages of a self-report methodology are well known (Sallis, & Saelens, 2000), it still represents the most feasible approach for conducting population-level surveillance across developed and developing countries. The International Physical Activity Questionnaire (IPAQ) was created and evaluated as a standardized instrument for this purpose. Although two versions of the IPAQ were designed and evaluated (short: nine items; and long: 31 items), the short form was recommended for population monitoring (Craig et al., 2003). However, it has not been recommended for intervention or research studies that require precise physical activity quantification to examine changes in physical activity at the individual level. IPAQ was also not intended to replace instruments that are more responsive to individual changes in activity level, such as objective measures. In addition to walking behaviors, IPAQ also assesses time spent in moderate- and vigorous-intensity activity as well as sitting behaviors, although the latter is not the focus of this analysis. Aggregated IPAQ data have been previously validated compared to accelerometers, and overall reliability was confirmed across 12 countries (Craig et al., 2003). Previous research showed criterion validity Spearman correlations with a median of 0.30 and test-retest reliability Spearman correlations clustered around 0.8 (Craig et al., 2003). The purpose of this study, however, was to reanalyze these data with respect to validity (again compared to an accelerometer) and test-retest reliability specifically for population monitoring of walking.

Submitted: September 30, 2008
Accepted: February 25, 2009

Hidde P. van der Ploeg, Cora Craig, and Adrian Bauman are with the Sydney School of Public Health at The University of Sydney. Catrine Tudor-Locke is with the Pennington Biomedical Research Center, Baton Rouge, LA. Cora Craig and Adrian Bauman are also with the Canadian Fitness and Lifestyle Research Institute, Ottawa. Alison L. Marshall is with the School of Public Health at Queensland University of Technology. Maria Hagströmer and Michael Sjöström are with the Department of Biosciences and Nutrition at Karolinska Institute.

Method

IPAQ Instrument

The focus of this paper was data collected from the walking items in the IPAQ short form, which assesses the frequency and duration of walking at work and at home, walking to travel from place to place, and any other walking done for recreation, sport, exercise, or leisure using two items (days per week and total time per week). Respondents also reported the frequency and duration of participating in other moderate- and vigorous-intensity physical activity. The questionnaire was designed for two different reference periods: "last 7 days," (the recommended format) or "usual week" and could be self-administered or done by telephone interview. Further details of the questionnaire, data collection, management, scoring, and reduction methods are available in the original

publication (Craig et al., 2003) and at the IPAQ website (www.ipaq.ki.se). The following abbreviations used in this paper differentiate between the various instruments and their administration: S7T = IPAQ short-version 7-day telephone recall; S7S = IPAQ short-version 7-day self-administered recall; and, SUS = IPAQ short-version usual week self-administered.

For this study, data were available from nine countries (Australia, Brazil, Finland, Japan, Portugal, Sweden, the Netherlands, the United Kingdom, and the United States). It is important to emphasize that the walking data were considered separately from data about participation in other moderate-intensity activity, and there was no overlap or double-counting of walking as a moderate-intensity activity. The original design of the IPAQ questionnaire also included questions about walking pace. These have since been removed, given the minimal contribution pace made to reliability and validity estimates (Craig et al., 2003).

Accelerometer

The CSA accelerometer (model 7164, Computer Sciences and Application, Inc., Fort Walton Beach, FL) was the criterion for validation purposes. The CSA is currently distributed as the ActiGraph (Matthews, 2005). Technical specifications and performance properties of this accelerometer have been thoroughly described elsewhere (Chen, & Bassett, 2005; Trost, McIver, & Pate, 2005). To include the CSA data in the present analyses, at least 600 min of registered time had to be recorded each day for at least 5 days (one of which had to be a weekend day). The total amount of physical activity recorded was expressed as total counts per registered time (counts/min). CSA data were considered to be of moderate intensity if they were between 1,952 and 5,724 counts/min (Freedson, Melanson, & Sirard, 1998). The final derived output was time per week in moderate-intensity activity.

Procedures

Test-retest reliability was assessed by completing IPAQ on two occasions—3 or 7 days apart. For the validation component, participants wore the accelerometer for 7 consecutive days and completed the IPAQ at the end of that week. Reliability and validity samples overlap and are neither exactly the same nor completely independent (i.e., not everyone participated in both reliability and validity studies). All centers used a standardized protocol for reliability and validity, and these procedures were overseen for quality control (Craig et al., 2003).

Database

All samples were selected based on convenience, but they represented substantial diversity across countries (Craig et al., 2003). Ethical approval for research was obtained locally in each country, and individual respondents provided informed consent. Complete self-reported walking and accelerometer data were required for this analysis; therefore, it only includes data from the S7T, S7S, and SUS.

Data Analysis

Data were not normally distributed, and, therefore, nonparametric inferential statistical procedures were used (Craig et al., 2003). We conducted two primary analyses:

1. Test-retest reliability of IPAQ estimates of total time walked per week administered either 3 or 7 days apart. Absolute median values were calculated and a paired samples *t* test was used to determine significant differences ($p < .05$) between test and retest. Spearman correlation coefficients were calculated as primary measures of test-retest reliability.
2. Criterion validity, comparing IPAQ estimates of total time walked per week to the CSA measure of time per week in moderate-intensity physical activity. Individuals who reported no other moderate-intensity activity besides walking (regardless of any reported vigorous-intensity activity) were extracted from the dataset. These participants were used to evaluate the relationship between IPAQ estimates of total time walked per week and accelerometer-determined time in moderate-intensity activity to assess the validity of the walking questions. This strategy was necessary to isolate self-reported walking as the only possible contributor to time spent in moderate-intensity activity. As a reference value, the IPAQ walking and the other moderate-intensity activity questions were summed for all participants and compared to the accelerometer-determined time in moderate-intensity activity. Spearman correlation coefficients were the primary validity statistic.

Results

Table 1 shows the country source of data and sample sizes as well as participants' descriptive characteristics and their physical activity levels as assessed by the IPAQ and accelerometer for the reliability and validity analyses. Participants across countries were generally middle-aged, well educated, and employed (Craig et al., 2003). Total time walked per week was comparable across the samples; median estimates of IPAQ total time walked per week ranged from 150 to 210 min. The total time walked per week was also higher in all samples than time spent in other moderate-intensity physical activities (range: 15–90 min). It is also notable that the 25th percentile for total

time walked per week was 60 min (the singular exception was 80 min for the last 7-day recall self-administered version), compared to 0 across all samples for time spent in other moderate-intensity physical activities. This difference reflects the underlying frequency of at least one third of participants reporting no other moderate-intensity physical activity (range across samples: 33–49%) to only 8–18% of participants who reported no walking. The validity samples, however, based on accelerometer data recorded between 201 and 334 median min per week in moderate-intensity activities. For participants who only walked, the median time spent walking was 105, 205, and 150 min for the S7T, S7S, and SUS, respectively. This was significantly lower than the time spent in moderate physical activity assessed by the accelerometer, which was 200, 302, and 350 min, respectively.

The absolute values of the first reliability measurement of the IPAQ walking question did not significantly differ from the retest (see Table 1). Test-retest reliability correlations of the IPAQ walking question ranged from 0.69 to 0.91 (see Table 2). The 3-day test-retest period showed a higher correlation (reflecting a stronger repeatability) than the 7-day test-retest data for the SUS; 7-day test-retest data were not collected for the S7T or S7S.

The criterion validity correlations for the IPAQ walking question ranged from 0.18 to 0.39 for those who did not participate in any other moderate-intensity physical activities (see Table 2). These correlations were generally higher than for the combined walking and moderate-intensity questions (Spearman = 0.04–0.24) in the unedited relevant samples. The IPAQ versions that asked participants to recall their physical activity over the previous 7 days (S7T and S7S) generally showed higher criterion validity correlations (with the exception of the S7T walking only) than the SUS, which asked participants to report their physical activity behavior in a usual week.

Discussion

Although walking is the single most commonly reported physical activity in public health surveys (Centers for Disease Control and Prevention, 2000; Rafferty et al., 2002), few studies have examined the reliability and validity of questions designed to measure this behavior specifically. The findings herein supported the reliability and validity of the IPAQ short-form questionnaire expressly for national monitoring of walking behaviors.

Table 1. Participant characteristics and physical activity levels

	Reliability samples			Validity samples		
	S7T[a]	S7S[b]	SUS[c]	S7T[d]	S7S[e]	SUS[f]
N	303	278	763	80	446	202
Gender, *N* men (%)	115 (38)	124 (45)	342 (45)	45 (56)	203 (46)	93 (46)
Age (*M*/*SD*)	45/12	36/11	39/13	34/11	40/13	37/12
Education (*M*/*SD*)	14/4	16/4	14/5	14/3	15/4	16/3
Employed, *N* (%)	234/77	250 (90)	430 (76)	79 (99)	376 (84)	161 (80)
Time walking IPAQ, median min/week (P_{25}; P_{75})	205 (60; 420)	180 (60; 420)	200 (60; 420)	210 (60; 630)	205 (80; 420)	150 (60; 360)
Time other moderate PA IPAQ, median min/week (P_{25}; P_{75})	15 (0; 185)	90 (0; 300)	90 (0; 240)	43 (0; 180)	60 (0; 240)	30 (0; 120)
Time walking IPAQ retest, median min/week (P_{25}; P_{75})	210 (60; 420)	180 (60; 360)	180 (60; 420)	N/A	N/A	N/A
Time moderate PA CSA, median min/week (P_{25}; P_{75})	N/A	N/A	N/A	201 (115; 267)	312 (198; 432)	334 (233; 420)

Note. PA = physical activity; IPAQ = International Physical Activity Questionnaire; S7T = IPAQ short version 7-day telephone recall; S7S = IPAQ short version 7-day self-administered recall; SUS = IPAQ short version usual week self-administered; education is missing for The Netherlands (*n* = 74), Portugal (*n* = 196), and United States (San Diego, *n* = 28). Employment is missing for Portugal.

[a]Data from Sweden and United Kingdom (Cambridge).

[b]Data from The Netherlands, United Kingdom (Bristol), and United States (San Diego).

[c]Data from Brazil, Finland, Japan, Portugal, Sweden, and United States (San Diego).

[d]Data from Australia.

[e]Data from Finland, The Netherlands, Sweden, United Kingdom (Bristol), and United States (San Diego and South Carolina).

[f]Data from Brazil, Japan, and United States (San Diego).

Yore and colleagues (2007) reported reliability (three administrations over 22 days) and validity (against the CSA) of the Behavior Risk Factor Surveillance System (BRFSS) questions, which include items to assess walking frequency and duration (including occupational reasons, recreation, exercise, or transportation) in a usual week. Median time reported walking in that study was similar to our own; across the BRFSS administrations, participants reported walking between 178 and 240 min per week, compared with 150–210 min per week in our study. The BRFSS study participants (Yore et al., 2007) reported more moderate-intensity activity, with medians ranging from 150 to 180 min per week, compared to 15–90 min per week in our data. The difference in nonwalking moderate-intensity physical activity may be that the university sample in the BRFSS study included more active participants.

The BRFSS study (Yore et al., 2007) reported agreement (Kappa) between standards for minimal physical activity levels (e.g., walking ≥ 30 min/day on ≥ 5 days/week) as indicated by repeated questionnaire administrations (reliability) and concurrent questionnaire administrations and accelerometer (validity). Yore et al. (2007) reported the walking data reliability was fair to moderate (Kappa range: 0.34–0.56). Although they calculated Spearman's correlations on the continuous data, they did not report these, as the overall findings were the same as those based on Kappa interpretations. The correlations between IPAQ administrations in the current study also supported 3-day (Spearman = 0.69–0.91) and 7-day (Spearman = 0.71) reliability of walking questions. Three-day reliability was higher than 7-day reliability, which was likely due to the time overlap in the recall periods with 3 days in between.

Pearson correlations between BRFSS and accelerometer estimates of time spent in moderate-intensity activity per week ranged from 0.16 to 0.27 across three administrations (Yore et al., 2007) and were similar to correlations in the combined IPAQ walking and moderate-intensity data versus accelerometer-recorded time (Spearman = 0.04–0.24). With regard to validity of the walking items, the latest BRFSS study (Yore et al., 2007) reported Kappa values of 0.19–0.23 (interpreted as fair by the authors) versus the accelerometer, which are consistent with the Spearman correlations (0.18–0.39) observed between the IPAQ walking and accelerometer data in this study.

Interestingly, an earlier developmental version of the BRFSS questions showed similar correlation coefficients between the 7-day recall walking data and all moderate-intensity activity assessed with the CSA (Spearman = 0.41) and between combined walking and moderate-intensity activity data against CSA assessed moderate-intensity physical activity (Spearman = 0.26; Ainsworth et al., 2000). Although these correlations were similar to the findings in this study, the BRFSS study did not conduct separate analyses with data from participants who only reported walking and no other moderate physical activities. That is a strength of the current study. Finally, another study (Carter-Nolan et al., 2006) reported a Spearman correlation of 0.26 between the CSA total counts per week and self-reported metabolic equivalent hours of walking per week; however, it did not report the correlation with accelerometer-derived time in moderate-intensity only.

This study demonstrated that validity coefficients for walking alone were better than for combined walking and other moderate-intensity activities, when compared with the accelerometer criterion standard; it seems that the IPAQ short-form walking items are more valid than those focused on other moderate-intensity activities. This finding may also be an artifact of the methodological decision

Table 2. Test-retest reliability and criterion validity of walking assessed with the International Physical Activity Questionnaire

	IPAQ S7T		IPAQ S7S		IPAQ SUS	
	N	Spearman correlation	*N*	Spearman correlation	*N*	Spearman correlation
Three-day test-retest reliability	301	0.69[a]	240	0.77[b]	78	0.91[c]
Seven-day test-retest reliability	N/A	N/A	N/A	N/A	684	0.72[d]
Criterion validity walking only (reported no other moderate PA)	26	0.18[e]	137	0.39[f]	84	0.26[g]
Criterion validity walking + moderate PA	62	0.19[e]	364	0.24[f]	181	0.04[g]

Note. PA = physical activity; IPAQ = International Physical Activity Questionnaire; S7T = IPAQ short version 7-day telephone recall; S7S = IPAQ short version 7-day self administered recall; SUS = IPAQ short version usual week self administered
[a]Data from Sweden and United Kingdom (Cambridge).
[b]Data from The Netherlands, United Kingdom (Bristol), and United States (San Diego).
[c]Data from Sweden, and United States (San Diego).
[d]Data from Brazil, Finland, Japan, and Portugal.
[e]Data from Australia.
[f]Data from Finland, The Netherlands, Sweden, United Kingdom (Bristol), United States (San Diego and South Carolina).
[g]Data from Brazil, Japan, Portugal, and United States (San Diego).

to use the Freedson et al. (1998) accelerometer cut points (i.e., minimally 1,952 counts/min), which were specifically calculated using lab-based ambulatory activities. The comparable BRFSS study chose a slightly more conservative (i.e., higher) cut point (i.e., 2,020 counts/min) for classifying minimally moderate-intensity activity, in line with the U.S. National Health and Nutrition Examination Survey accelerometer data (Troiano et al., 2008).

The findings of this study were similar across the IPAQ short-form formats that were examined. However, the validity estimates for the SUS were not as strong as those using data from the past 7-day versions (S7T and S7S). This might be partly due to the fact that the accelerometer data were collected over the same 7-day period as recalled in the S7T and S7S IPAQ questionnaires. However, these findings also supported the recommendation that the 7-day recall versions should be preferred over the usual week versions (Craig et al., 2003).

As noted (Craig et al., 2003), the IPAQ has been tested in developed and developing countries, and, therefore, it can be used with confidence in most national surveillance efforts. However, some questions remain about cultural differences in implementing the IPAQ protocol, which could result in some differences in physical activity population estimates between countries. Admittedly, the great reliance on convenience sampling in the original study led to motivated volunteers, albeit across different geographies and cultures. Further, at least one sample in this study was limited in size (validity analysis of S7T, $n = 26$ reporting any walking). Regardless, this study supported the ability of the IPAQ short form to provide reliable and valid estimates of time spent in walking behaviors. Consistent use of the IPAQ short form across studies and populations will allow for more systematic comparison of physical activity behaviors, particularly in walking, which is an emerging focus of physical activity surveillance.

References

Ainsworth, B. E., Bassett, D. R., Jr., Strath, S. J., Swartz, A. M., O'Brien, W. L., Thompson, R. W., et al. (2000). Comparison of three methods for measuring the time spent in physical activity. *Medicine & Science in Sports & Exercise, 32*(Suppl. 9), S457–S464.

Carter-Nolan, P. L., Adams-Campbell, L. L., Makambi, K., Lewis, S., Palmer, J. R., & Rosenberg, L. (2006). Validation of physical activity instruments: Black Women's Health Study. *Ethnicity & Disease, 16,* 943–947.

Centers for Disease Control and Prevention. (2000). Prevalence of leisure-time and occupational physical activity among employed adults—United States, 1990. *Morbidity and Mortality Weekly Report, 49,* 420–424.

Chen, K. Y., & Bassett, D. R., Jr. (2005). The technology of accelerometry-based activity monitors: Current and future. *Medicine & Science in Sports & Exercise, 37*(Suppl. 11), S490–S500.

Craig, C. L., Marshall, A. L., Sjöström M., Bauman, A. E., Booth, M. L., Ainsworth, B. E., et al. (2003). International Physical Activity Questionnaire: 12-country reliability and validity. *Medicine & Science in Sports & Exercise, 35,* 1381–1395.

Freedson, P. S., Melanson, E., & Sirard, J. (1998). Calibration of the Computer Science and Applications, Inc. accelerometer. *Medicine & Science in Sports & Exercise, 30,* 777–781.

Jeon, C. Y., Lokken, R. P., Hu, F. B., & van Dam, R. M. (2007). Physical activity of moderate intensity and risk of type 2 diabetes: A systematic review. *Diabetes Care, 30,* 744–752.

Levine, J. A., McCrady, S. K., Lanningham-Foster, L. M., Kane, P. H., Foster, R. C., & Manohar, C. U. (2008). The role of free-living daily walking in human weight-gain and obesity. *Diabetes, 57,* 548–554.

Matthews, C. E. (2005). Calibration of accelerometer output for adults. *Medicine & Science in Sports & Exercise, 37*(Suppl. 11), S512–S522.

Rafferty, A. P., Reeves M. J., McGee H. B., & Pivarnik J. M. (2002). Physical activity patterns among walkers and compliance with public health recommendations. *Medicine & Science in Sports & Exercise, 34,* 1255–1261.

Sallis, J. F., & Saelens, B. E. (2000). Assessment of physical activity by self-report: Status, limitations, and future directions. *Research Quarterly for Exercise and Sport, 71*(Suppl. 2), S1–S14.

Troiano, R. P., Berrigan D., Dodd, K. W., Mâsse L. C., Tillert, T., & McDowell, M. (2008). Physical activity in the United States measured by accelerometer. *Medicine & Science in Sports & Exercise, 40,* 181–188.

Trost, S. G., McIver, K. L., & Pate, R. R. (2005). Conducting accelerometer-based activity assessments in field-based research. *Medicine & Science in Sports & Exercise, 37*(Suppl. 11), S531–S543.

Yore, M. M., Ham, S. A., Ainsworth, B. E., Kruger, J., Reis, J. P., Kohl, H. W., III, et al. (2007). Reliability and validity of the instrument used in BRFSS to assess physical activity. *Medicine & Science in Sports & Exercise, 39,* 1267–1274.

Authors' Notes

We would like to thank the Public Health Agency of Canada for partial funding of this project and the following persons: members of the IPAQ Reliability and Validity Study Group—Barbara Ainsworth, Adrian E. Bauman, Michael Booth, Cora L. Craig, Ulf Ekelund, Pekka Oja, Michael Pratt, Michael Sjöström, James F. Sallis, and Alison L. Marshall; IPAQ Reliability and Validity Study Group coordinators—Fiona Bull, Victor Matsudo, Sandra Matsudo, Pekka Oja, Toshihito Katsumura, Willem van Mechelen, Jorge Mota, Luis Sardinha, Michael Sjöström, Ulf Ekelund, Barbara Ainsworth, James F. Sallis, Mark Davis, Angie Page, Ashley Cooper, and Nicholas J. Wareham. Please address all correspondence concerning this article to Hidde P. van der Ploeg, Center for Physical Activity and Health, Level 2, Medical Foundation Building (K25), University of Sydney, New South Wales, Australia 2006.

E-mail: hiddep@health.usyd.edu.au