Public Health Nutrition: 16(3), 440–452 doi:10.1017/S1368980012002996

Convergent validity of the International Physical Activity Questionnaire (IPAQ): meta-analysis

Youngdeok Kim[1,*], Ilhyeok Park[2] and Minsoo Kang[1]
[1]Department of Health and Human Performance, Middle Tennessee State University, 1500 Greenland Drive, PO Box 96, Murfreesboro, TN 37132, USA: [2]Department of Physical Education, Seoul National University, Seoul, South Korea

Submitted 21 October 2011: Final revision received 6 March 2012: Accepted 3 May 2012: First published online 2 July 2012

Abstract

Objective: The purpose of the present study was to use a meta-analytic approach to examine the convergent validity of the International Physical Activity Questionnaire (IPAQ).
Design: Systematic review by meta-analysis.
Setting: The relevant studies were surveyed from five electronic databases. Primary outcomes of interest were the product-moment correlation coefficients between IPAQ and other instruments. Five separate meta-analyses were performed for each physical activity (PA) category of IPAQ: walking, moderate PA (MPA), total moderate PA (TMPA), vigorous PA (VPA) and total PA (TPA). The corrected mean effect size (ESρ) unaffected by statistical artefacts (i.e. sampling error and reliability) was calculated for each PA category. Selected moderator variables were length of IPAQ (i.e. short and long form), reference period (i.e. last 7 d and usual week), mode of administration (i.e. interviewer and self-reported), language (i.e. English and translated) and instruments (i.e. accelerometer, pedometer and subjective measure).
Subjects: A total of 152 ESρ across five PA categories were retrieved from twenty-one studies.
Results: The results showed small- to medium-sized ESρ (0·27–0·49). The highest value was observed in VPA while the lowest value was found in MPA. The ESρ were differentiated by some of the moderator variables across PA categories.
Conclusions: The study shows the overall convergent validity of IPAQ within each PA category. Some differences in degree of convergent validity across PA categories and moderator variables imply that different research conditions should be taken into account prior to deciding on use of the appropriate type of IPAQ.

Keywords
IPAQ
Convergent validity
Meta-analysis
Physical activity

Physical activity (PA) has been regarded as one of the most important habitual behaviours which leads to a healthy life by preventing diseases and increasing health benefits[1–5]. As the importance of PA has been emphasized, attempts have been made to develop appropriate measurement tools, including objective and subjective measurement tools, to quantify the amount of PA in daily life. Of these, questionnaires remain the most widely used measurement tool in large-scale studies due to their efficiency of measuring PA levels in large populations[6].

The International Physical Activity Questionnaire (IPAQ) is an instrument which was developed by the International Consensus Group in 1998–1999 to establish a standardized and culturally adaptable measurement tool across various populations in the world[7]. IPAQ is designed to assess the levels of habitual PA for individuals ranging from young to middle-aged adults (i.e. 15–69 years old). In addition, there are different forms of IPAQ depending on several variations which include length of questionnaire (i.e. short or long form), reference period (i.e. last 7 d or usual week) and mode of administration (i.e. self-report or interviewer-based).

Soon after IPAQ was developed it was translated into several different languages and numerous studies have been conducted to examine the reliability and validity of these versions across countries. In these studies one of the most commonly applied approaches to establish the validity evidence of IPAQ is the convergent validity, which indicates the extent to which different measurement tools measure the same construct. However, the extent to which the estimates from IPAQ linearly relate to other counterpart instruments has varied depending on the different characteristics of IPAQ examined (i.e. translation, length, reference period and mode of administration) and the instrument used for the comparison[8], yet quantification of the exact extent of variations is still undefined.

*Corresponding author: Email yk2k@mtmail.mtsu.edu

To the best of our knowledge, no studies to date have examined the sources and magnitudes of factors that may explain such discrepancies in convergent validity of IPAQ across studies. With high prevalence of usage of IPAQ in measuring levels of PA at the population level and limited information for convergent validity of IPAQ in various formats, synthesizing all empirical evidence on convergent validity of IPAQ would provide more comprehensive information. The purpose of the present study was therefore to apply a meta-analytic method to quantifying the overall convergent validity of IPAQ across different studies and to investigate the sources and magnitudes of moderator factors that may affect the overall convergent validity of IPAQ.

Methods

Search strategy and selection criteria

The relevant studies for examining convergent validity of IPAQ were obtained from five electronic databases (i.e. SPORTDiscus, Medline, Google Scholar, PubMed and EBSCOhost). The main keywords used to identify the appropriate studies were 'International Physical Activity Questionnaire', 'IPAQ', 'validity', 'convergent validity', 'comparison' and 'validation'. All of these keywords were entered with several combinations.

The primary outcome of interest was the correlation coefficient between IPAQ and another instrument. The following criteria were used to select potential studies for inclusion: (i) a study that used IPAQ as either a main instrument to be validated or an instrument to be compared with; (ii) a study in which the participants were not physically or emotionally challenged or disabled; (iii) a study in which the mean age of participants fell between 15 and 69 years old; (iv) in circumstances where IPAQ was translated into other languages, no changes in the structure occurred; (v) a study had a precise definition of PA intensity derived from the instrument; (vi) a study that reported statistical results in sufficient detail to estimate effect size (ES*r*); and (vii) a peer-reviewed article published in English. Using these criteria, potentially relevant studies were screened by two independent reviewers and full texts of all studies meeting the inclusion criteria were further assessed for methodological quality and for data extraction. Consensus was achieved through discussion when disagreements occurred between the two reviewers.

Methodological quality

Two reviewers independently assessed the methodological quality of studies using the modified version of the Downs and Black checklist[9], which was used in recent systematic reviews[10,11]. The modified checklist consisted of fifteen items within three domains (i.e. reporting, external validity and internal validity), and possible scores ranged between 0 and 15 (e.g. higher scores indicated better methodological quality). Any study that scored relatively low on methodological quality (i.e. Z-score $< -1{\cdot}96$) was not considered for inclusion in the meta-analyses.

Data extraction and coding

The long form of IPAQ examines the habitual PA in daily life using twenty-seven items across four PA domains (i.e. leisure time, domestic and gardening, occupational and transport-related activities), while the short form of IPAQ consists of seven summarized items that measure the comprehensive level of PA regardless of the domains to be measured. In both forms, the participants are asked to report the durations and frequencies of three specific PA categories, i.e. walking, moderate PA (MPA) and vigorous PA (VPA). Total amount of time spent engaging in or energy expenditure for each PA category can be estimated as main outcomes using metabolic equivalent of task (MET) values of 3·3, 4 and 8 for walking, MPA and VPA, respectively. Because the MET value of walking is within a range for moderate-intensity PA (i.e. 3–6 MET)[12], it has also been recommended to combine the estimates of walking and MPA to obtain the total MPA (denoted as TMPA)[13]. Total PA (TPA) can be simply estimated by summation of all estimates from each category (i.e. walking + MPA + VPA). Therefore, there are a total of five PA categories that can be derived from IPAQ (i.e. walking, MPA, TMPA, VPA and TPA).

Throughout the systematic review of selected studies, ES*r* values were extracted separately for each of the five PA categories to avoid dependency issues in the meta-analysis. In addition, each ES*r* was extracted only if the compared PA categories from both IPAQ and the other instrument were consistent or reasonably consistent (see outcome domains in Table 1). For example, estimates in walking, MPA and TMPA from IPAQ should be compared with estimates for MPA obtained from the other instrument. Likewise, ES*r* values were extracted for VPA and TPA only if they were compared with the same PA categories from the other instrument. However, because a pedometer does not provide the information of step counts within specific PA categories, ES*r* that were estimated between total step counts of pedometers and each PA category of IPAQ were also extracted. If a single study reported more than one ES*r* within the same PA category, but from different subpopulations, we assumed each ES*r* from different subpopulations to be independent from each other and included them in a single meta-analysis[14]. The units or scales of estimated value within each study were not considered because the primary outcome of interest in the present study was the correlation coefficient, which is a scale invariant coefficient in itself[15].

Moderator variables which may affect overall convergent validity of IPAQ were obtained from different characteristics of IPAQ used in each study: (i) length of IPAQ (i.e. short and long forms); (ii) reference period (i.e. last 7 d and usual week); (iii) mode of administration

https://doi.org/10.1017/S1368980012002996 Published online by Cambridge University Press

Table 1 Features of the studies included in the meta-analyses of convergent validity of IPAQ and outcome domains examined

		Type of IPAQ					Outcome domains§
Study	Population (*n*)†	Language	Length	Period	Administration	Instrument‡ (cut-off standard)	IPAQ – instrument
Boon *et al.*[26]	New Zealand (64)	English	Long	–	Self-reported	Accelerometer (ActiGraph GT1 M) ● Moderate: 1952–5724 counts (3–5·9 MET) ● Vigorous: >5724 counts (≥6 MET)	Moderate – moderate Vigorous – vigorous
Bull *et al.*[28]	Bangladesh (147), Brazil (204), China (221), Ethiopia (940), Indonesia (337), India (234), Japan (148), Portugal (67), South Africa (214), Taiwan (141)	Translated	Short	–	Interviewer	Subjective (GPAQ) ● Moderate: summation of moderate-intensity activity at work, transport-related and discretionary activity ● Vigorous: summation of vigorous-intensity activity at work and discretionary activity	Total moderate – moderate Vigorous – vigorous Total PA – total PA
Craig *et al.*[7]	UK (151), Finland (84), USA (26), Netherland (30), Sweden (49)	English (UK and USA) Translated (Fin, Net, and Swe)	Long	Last 7 d	Self-reported	Accelerometer (MTI model 7164) ● Total activity counts	Total PA – total counts
	USA (29), Guatemala (61), South Africa (107)	English (USA) Translated (Gua and SA)	Long	Usual week	Interviewer		
Craig *et al.*[7]	Japan (127), USA (26), Brazil (28)	English (USA) Translated (Jap and Bra)	Long	Usual week	Self-reported	Accelerometer (MTI model 7164) ● Total activity counts	Total PA – total counts
	Australia (62)	English	Short	Usual week	Self-reported		
	Finland (84), USA (26), Netherland (28), Sweden (49), UK (151)	English (UK and USA) Translated (Fin, Net, and Swe)	Short	Last 7 d	Self-reported		
	USA (29), Guatemala (61), South Africa (107)	English (USA) Translated (Gua and SA)	Short	Usual week	Interviewer		
	Japan (127), USA (26), Brazil (28)	English (USA) Translated (Jap and Bra)	Short	Usual week	Self-reported		
De Cocker *et al.*[16]	Belgium (1239)	Translated	Long	–	–	Pedometer (Yamax SW-200) ● Step counts	Walking – step counts Moderate – step counts Vigorous – step counts

Table 1 *Continued*

		Type of IPAQ					Outcome domains§
Study	Population (*n*)†	Language	Length	Period	Administration	Instrument‡ (cut-off standard)	IPAQ – instrument
De Cocker *et al.*[17]	Belgium (310)	Translated	Long Short	Usual week	Self-reported	Pedometer (Yamax SW-200) ● Step counts Subjective (MLTPAQ) ● Walking: structured walking ● Moderate: 3–5·9 MET ● Vigorous: ≥6 MET ● Total PA Subjective (Baecke-Q) ● Total PA	Pedometer Walking – step counts Moderate – step counts Vigorous – step counts Total PA – step counts Subjective (MLTPAQ) Walking – walking Moderate – moderate Vigorous – vigorous Total PA – total PA Subjective (Baecke-Q) Total PA – total PA
Deng *et al.*[18]	China (224)	Translated	Short	Last 7 d	Interviewer	Pedometer (Yamax SW-200) ● Step counts	Walking – step counts Moderate – step counts Vigorous – step counts Total PA – step counts
Dinger *et al.*[19]	USA (123)	English	Long	–	Self-reported	Accelerometer (MTI model 7164) ● Moderate: 1952–5724 counts (3–5·9 MET) ● Vigorous: >5724 counts (≥6 MET) ● Total activity counts Pedometer (Yamax SW-200) ● Step counts	Accelerometer Walking – moderate Moderate – moderate Vigorous – vigorous Total PA – total counts Pedometer Walking – step counts Moderate – step counts Vigorous – step counts Total PA – step counts
Ekelund *et al.*[34]	Sweden (185)	Translated	Short	Last 7 d	Self-reported	Accelerometer (MTI model 7164) ● Total activity counts	Total PA – total counts
Gauthier *et al.*[20]	Canada (31)	Translated	Long	Last 7 d	Self-reported	Pedometer (Yamax SW-200) ● Step counts	Walking – step counts Moderate – step counts Vigorous – step counts Total PA – step counts
Hagstromer *et al.*[29]	Sweden (46)	Translated	Long	Last 7 d	Self-reported	Accelerometer (MTI) ● Moderate: 1952–5724 counts (3–5·9 MET) ● Vigorous: >5724 counts (≥6 MET) ● Total activity counts	Total moderate – moderate Vigorous – vigorous Total PA – total counts

Table 1 *Continued*

Study	Population (*n*)†	Type of IPAQ				Instrument‡ (cut-off standard)	Outcome domains§
		Language	Length	Period	Administration		IPAQ – instrument
Hagstromer *et al.*[21]	Sweden (980)	Translated	Long	–	Self-reported	Accelerometer (MTI model 7164) ● Moderate: 760–5724 counts ● Vigorous: >5724 counts ● Total activity minutes	Walking – Moderate Moderate – Moderate Total moderate – moderate Vigorous – vigorous Total PA – total PA
Kolbe-Alexander *et al.*[22]	South Africa (male: 42, female: 61)	Translated	Short	Usual week	Self-reported	Accelerometer (MTI model 7162) ● Moderate: 1952–5724 counts ● Vigorous: >5724 counts	Walking – moderate Moderate – moderate Vigorous – vigorous
Kurtze *et al.*[23]	Norway (108)	Translated	Short	Last 7 d	Self-reported	Accelerometer (ActiReg) ● Moderate (3–5·9 MET) ● Vigorous (≥6 MET)	Walking – moderate Moderate – moderate Vigorous – vigorous
Lachat *et al.*[30]	Vietnam (188)	Translated	Short	Usual week	Self-reported	Accelerometer (MTI GT256) ● Moderate (3–5·9 MET) ● Vigorous (≥6 MET) ●Total activity counts	Total moderate – moderate Vigorous – vigorous Total PA – total PA
Macfarlane *et al.*[31]	China (49)	Translated	Short	Last 7 d	Interviewer	Accelerometer (MTI model 7164) ● Moderate: 1952–5724 counts ● Vigorous: >5724 counts Accelerometer (Tritrac model RT3) ● Moderate: 1211–2893 counts ● Vigorous: >2893 counts Subjective (PA-log) ● Moderate (3–5·9 MET) ● Vigorous (≥6 MET)	Accelerometer (MTI 7164) Total moderate – moderate Vigorous – vigorous Accelerometer (Tritrac RT3) Total moderate – moderate Vigorous – vigorous Subjective (PA-log) Total moderate – moderate Vigorous – vigorous
Mader *et al.*[24]	Switzerland (35)	Translated	Short	Usual week	Interviewer	Accelerometer (MTI model 7164) ● Moderate 574–4944 counts ● Vigorous: >4944 counts Subjective (QIMO) ● Total activities (MET-min/week)	Accelerometer Walking – moderate Moderate – moderate Total moderate – moderate Vigorous – vigorous Total PA – total counts Subjective (QIMO) Total PA – total PA
Roman-Vinas *et al.*[27]	Spain (54)	Translated	Long	Last 7 d	Self-reported	Accelerometer (MTI ActiGraph) ● Moderate: 1952–5724 counts ● Vigorous: >5724 counts ● Total activity counts	Total moderate – moderate Moderate – moderate Vigorous – vigorous Total PA – total counts

Table 1 *Continued*

		Type of IPAQ					Outcome domains§
Study	Population (*n*)†	Language	Length	Period	Administration	Instrument‡ (cut-off standard)	IPAQ – instrument
Thuy *et al.*(35)	Vietnam (122)	Translated	Long	Last 7 d	Interviewer	Pedometer (Yamax SW-200) ● Step counts Questionnaire (GPAQ) ● Total PA	Total PA – total PA Total PA – step counts
Timperio *et al.*(32)	Austrailia (97)	English	Short Long	Last 7 d	Interviewer	Accelerometer (MTI model 7164) ● Moderate: 1952–5724 counts ● Vigorous: >5724 counts ● Total activity minutes	Total moderate – moderate Vigorous – vigorous Total PA – total PA
van der Ploeg *et al.*(25)	Mixed (884)	–	Short	Last 7 d Usual week	Interviewer Self-reported	Accelerometer (MTI model 7164) ● Moderate: 1952–5724 counts	Walking – moderate Total moderate – moderate
Vandelanotte *et al.*(33)	Belgium (53)	Translated	Long	Usual week	Self-reported	Accelerometer (MTI model 7164) ● Moderate: 1952–5724 counts ● Vigorous: >5725 counts ● Total activity minutes Subjective (PA-log) ● Moderate (3–5·9 MET) ● Vigorous (≥6 MET) ● Total activity minutes	Accelerometer Total moderate – moderate Vigorous – vigorous Total PA – total PA Subjective (PA-log) Total moderate – moderate Vigorous – vigorous Total PA – total PA

IPAQ, International Physical Activity Questionnaire; MET, metabolic equivalent of task; PA, physical activity.
†Regions where the participants were recruited (sample size); '-' indicates no moderator variables were extracted.
‡Types of instrument and cut-off standards compared with IPAQ: GPAQ, Global Physical Activity Questionnaire; MLTPAQ, Minnesota Leisure Time Physical Activity Questionnaire; Baecke-Q, Baecke questionnaire; OIMQ, Office In Motion Questionnaire.
§Outcome domains for meta-analyses (PA categories).

Table 2 Stem-and-leaf plots of correlation coefficients (ES*r*) of IPAQ

Walking (*n* 17)		MPA (*n* 17)		TMPA (*n* 23)		VPA (*n* 35)		TPA (*n* 60)	
Stem	Leaf	Stem	Leaf	Stem	Leaf	Stem	Leaf	Stem	Leaf
0·9		0·9		0·9		0·9		0·9	2
0·8		0·8		0·8		0·8		0·8	
0·7		0·7		0·7	1 5	0·7	2 9 9	0·7	
0·6		0·6	8	0·6	0 8 8	0·6	0 3 7 8	0·6	0 6
0·5	1 6	0·5		0·5	0 5	0·5	0 1 1 2	0·5	2 2 2 3 3 3 4 5 6 7 9
0·4	9	0·4	1	0·4	0	0·4	0 1 2 2 2 3 4 5 6 6 7	0·4	0 3 5 6 7 7
0·3	2 8 9	0·3	1 1 3	0·3	0 2 9	0·3	0 1 8	0·3	0 0 1 1 2 2 2 3 4 4 4 6 6 7 8 8 9 9
0·2	0 0 4 5 6	0·2	3 7 7	0·2	4 8 9 9 9	0·2	0 2 5 8 9	0·2	0 1 1 1 3 4 5 6 7 8 8 9 9 9
0·1	0 2 5 7 8 9	0·1	2 5 5 7 9	0·1	0 2 3 7 9	0·1	8	0·1	2 3 6
0·0		0·0	5 6	0·0	4	0·0	5 5	0·0	2 5
−0·0		−0·0	6 9	−0·0	1	−0·0	3 9	−0·0	2
−0·1		−0·1		−0·1		−0·1		−0·1	2
−0·2		−0·2		−0·2		−0·2		−0·2	7
−0·3		−0·3		−0·3		−0·3		−0·3	

IPAQ, International Physical Activity Questionnaire; *n*, number of ES*r*; MPA, moderate physical activity; TMPA, total moderate physical activity; VPA, vigorous physical activity; TPA, total physical activity.

(i.e. interviewer and self-reported); and (iv) language (i.e. English and translated). In addition, the instruments which were used for comparison with IPAQ within each study were also extracted as a moderator variable: (v) instruments (i.e. accelerometer, pedometer and subjective measure).

Study characteristics

A total of sixty-seven potentially relevant studies were considered for further review. By systematic review based on inclusion criteria, a total of twenty-eight studies were excluded due to their inability to meet criteria and duplication. Full texts of the remaining thirty-nine studies were reviewed for a detailed assessment. Of these, twenty-one studies met all inclusion criteria and secured relatively higher methodological quality (mean 13·2; SD 1·3). A total of 152 ES*r* values across five PA categories in IPAQ were retrieved (i.e. seventeen ES*r* from ten studies for walking[16–25], seventeen ES*r* from twelve studies for MPA[16–24,26,27], twenty-three ES*r* from ten studies for TMPA[21,24,25,27–33], thirty-five ES*r* from seventeen studies for VPA[16–24,26–33] and sixty ES*r* from sixteen studies for TPA[7,17–21,24,25,27–30,32–35]). See Table 2 for stem–leaf plots of ES*r* extracted across PA categories. Total sample sizes for each PA category ranged from a low of 4453 in TMPA to a high of 8867 in TPA.

Computation of effect sizes

The measure of ES*r* in the present study was the product-moment correlation coefficients (e.g. Pearson r and Spearman ρ), which represent the strength of associations between the estimates of IPAQ and other counterpart instruments as an indication of convergent validity of IPAQ. The psychometric meta-analytic method proposed by Hunter and Schmidt[36,37] was conducted to obtain the population-level estimates unaffected by statistical artefacts, such as sampling error and measurement error. The 'bare-bone' mean ES*r* (i.e. ES$\bar{r}$), corrected for only sampling error, was calculated by weighting each ES*r* with the respective sample size when aggregating them into ES$\bar{r}$. In order to correct for the measurement errors of IPAQ in addition to sampling error, the reliability coefficients of IPAQ with respect to each PA category (e.g. intra-class correlation coefficients) were further extracted. There were eleven reliability coefficients available for walking (mean 0·74; SD 0·15), nine for MPA (mean 0·63; SD 0·22), eight for TMPA (mean 0·62; SD 0·21), twelve for VPA (mean 0·67; SD 0·23) and thirty-two for TPA (mean 0·77; SD 0·13). Because the reliability coefficients were not available for all of the included studies, the artefact distributions were calculated for each PA category to obtain the corrected mean ES*r* at the population level (i.e. ESρ) that was unaffected by sampling error and measurement error. 95 % confidence intervals (CI) were produced on the basis of the standard error of ESρ and 95 % credibility intervals (CV) were also yielded using the residual standard deviation of ESρ. According to Cohen's guidelines, ESρ was interpreted as small (<0·30), medium (0·31–0·49) and large (≥0·50)[38].

Moderator analysis

For determining the presence of moderator effects on ESρ, three different criteria (i.e. the percentage of variance components attributed to statistical artefacts, the Q homogeneity statistic and 95 % CV) were simultaneously examined as recommended by Hunter and Schmidt[37]. To be specific, we concluded that moderators exist if: (i) the percentage of variance accounted for by statistical artefacts is less than 75 % of the observed variance in ES*r*; (ii) the Q homogeneity statistic is significant; and (iii) the 95 % CV is either relatively large or includes zero. However, due to the imprecise meaning of 'large' CV, we focused mainly on first two criteria to examine the moderator effects unless disagreement occurred.

Results

Overall effect sizes

The ESρ corrected for artefacts of sampling error and measurement error across each PA category is presented in Table 3. There were positive relationships between IPAQ and other instruments across all PA categories (ESρ range = 0·27–0·49) in which all 95 % CI did not include zero. According to Cohen's guideline, medium-sized ESρ were retrieved for walking (ESρ = 0·32), TMPA (ESρ = 0·45), VPA (ESρ = 0·49) and TPA (ESρ = 0·39), while MPA had a small-sized effect size with an ESρ of 0·27. The proportions of variance accounted by artefacts among the total variance of observed ESr for each PA category were all less than 75 % and statistical significances were found in Q homogeneity tests for all PA categories (all $P < 0·05$). Therefore, follow-up moderator analyses were conducted using predefined moderators as hypothesized in the present study.

Moderator analyses

Moderator analyses were conducted to examine the effects of language (i.e. English and translated), length of IPAQ (i.e. short and long form), reference period (i.e. last 7 d and usual week), mode of administration (i.e. interviewer and self-reported) and instruments (i.e. accelerometer, pedometer and subjective measure) on overall ESρ for each PA category (see Table 4). Collectively, substantial differences in ESρ were detected by different levels of included moderators across all PA categories.

In terms of language of IPAQ, there were consistent trends in the rank of ESρ for all PA categories in studies which used translated versions. These studies had significantly greater ESρ compared with those in which the English version was applied. Using the 75 % rule and Q homogeneity statistic, the observed ESr values obtained from the English-version IPAQ studies for walking, MPA, TMPA and VPA were shown to be homogeneous, while there was still a large amount of unexplained variance in ESρ, mostly for which the translated versions were used.

The variations in ESρ for all PA categories were also not significantly explained by different length of IPAQ with the exception of walking and MPA. The percentage of variance accounted for by artefacts increased dramatically for the studies where the long form was used (83·9 % and 56·6 % for walking and MPA, respectively). Non-significant Q statistics were detected for the long forms of walking and MPA (Q(df = 6) = 8·34; $P > 0·05$ and Q(df = 7) = 14·06; $P > 0·05$, respectively). Although the length of IPAQ accounted for a relatively small percentage of variance in ESρ for most PA categories, the ESρ values by different length of IPAQ differed significantly in walking and TMPA, where the 95 % CI for ESρ were not overlapped between the long and short form. Moreover, systematic trends for the rank of ESρ were detected in the studies which used short-form versions of IPAQ. These studies had greater ESρ for all PA categories.

Moderator analyses by reference period did not significantly increase the percentage of variance accounted for by artefacts or the non-significant Q homogeneity statistic for all PA categories. Moreover, there were no observable trends for rank of ESρ values across PA categories.

With respect to the mode of administration, the results showed that interviewer-administered studies had greater ESρ values for all PA categories with the exception of MPA, in which 80·9 % of the variation in ESρ for self-reported studies was attributed to artefacts with a non-significant Q statistic (Q(df = 10) = 13·60; $P > 0·05$). The ESρ values differed significantly by different mode of administration in TMPA, VPA and TPA. Interviewer-administered studies had a greater ESρ than those which utilized the self-reported measure of IPAQ.

The type of instrument moderately increased the percentage of variance accounted for by artefacts in walking, VPA and TPA, in which non-significant Q homogeneity statistics for respective types of instruments were also detected. The studies which utilized subjective measures had greater ESρ values than studies utilizing objective measures in all PA categories with the exception of walking, in which opposite results were yielded.

Table 3 Results of meta-analyses for overall weighted mean correlation coefficients (ESr) across PA categories of IPAQ

PA category	K	n	N	ES$\bar{r}$†	ESρ‡	% of variance accounted for§	95 % CI	95 % CV	Q statistic
Walking	10	17	4453	0·28	0·32	36·5	0·27, 0·37	0·14, 0·51	46·52*
MPA	12	17	3854	0·21	0·27	48·9	0·23, 0·32	0·08, 0·47	34·72*
TMPA	10	23	4983	0·35	0·45	26·6	0·37, 0·54	0·05, 0·85	86·39*
VPA	17	35	7684	0·40	0·49	27·1	0·43, 0·56	0·13, 0·87	129·26*
TPA	16	60	8867	0·34	0·39	25·9	0·35, 0·43	0·09, 0·69	231·33*

PA, physical activity, IPAQ, International Physical Activity Questionnaire; K, number of studies; n, number of ESr; N, total sample size; CV, credibility interval; MPA, moderate physical activity; TMPA, total moderate physical activity; VPA, vigorous physical activity; TPA, total physical activity.
*$P < 0·05$.
†Averaged ESr corrected for sampling error only.
‡Averaged ESr corrected for sampling error and measurement errors of IPAQ.
§Percentage of variance accounted for by statistical artefacts including sampling error and measurement error of IPAQ.

https://doi.org/10.1017/S1368980012002996 Published online by Cambridge University Press

Table 4 Results of moderator analyses across all PA categories of IPAQ

Moderator	Effect	*K*	*n*	*N*	ES$\bar{r}$†	ESρ‡	% of variance accounted for§	95 % CI	95 % CV	*Q* statistic
Language										
Walking	English	1	2	226	0·11	0·12	100·0	0·12, 0·12	0·12, 0·12	0·03
	Translated	8	12	3960	0·31	0·37	28·0	0·30, 0·43	0·15, 0·59	42·86*
MPA	English	2	3	310	0·20	0·24	100·0	0·24, 0·24	0·24, 0·24	0·24
	Translated	10	14	3544	0·22	0·28	46·6	0·23, 0·34	0·07, 0·49	30·14*
TMPA	English	1	2	192	0·22	0·31	100·0	0·31, 0·31	0·31, 0·31	0·44
	Translated	7	18	4139	0·39	0·55	44·6	0·46, 0·63	0·18, 0·91	40·50*
VPA	English	3	5	502	0·40	0·43	100·0	0·43, 0·43	0·43, 0·43	3·37
	Translated	14	30	7182	0·40	0·52	26·1	0·44, 0·59	0·12, 0·91	115·31*
TPA	English	3	15	1016	0·27	0·29	77·7	0·26, 0·33	0·16, 0·42	19·32
	Translated	8	45	7851	0·37	0·43	24·2	0·38, 0·48	0·12, 0·74	185·76*
Length										
Walking	Long	5	7	3116	0·23	0·24	83·9	0·23, 0·26	0·20, 0·28	8·34
	Short	6	10	1337	0·31	0·38	32·1	0·29, 0·47	0·10, 0·66	31·14*
MPA	Long	7	8	2927	0·24	0·26	56·6	0·23, 0·30	0·16, 0·36	14·06
	Short	6	9	927	0·20	0·28	31·3	0·14, 0·42	−0·15, 0·71	28·73*
TMPA	Long	5	6	1283	0·19	0·23	49·5	0·16, 0·29	0·07, 0·39	12·12*
	Short	6	17	3655	0·41	0·55	38·2	0·45, 0·65	0·13, 0·97	44·47*
VPA	Long	11	13	3483	0·43	0·46	10·7	0·38, 0·55	0·15, 0·78	121·87*
	Short	9	22	4201	0·39	0·56	21·5	0·45, 0·67	0·04, 0·99	102·66*
TPA	Long	10	28	3690	0·32	0·35	51·6	0·32, 0·38	0·18, 0·52	54·33*
	Short	8	32	5177	0·36	0·43	21·1	0·37, 0·50	0·07, 0·79	151·69*
Reference period										
Walking	Last 7 d	4	5	526	0·35	0·41	37·8	0·30, 0·53	0·15, 0·68	13·22*
	Usual week	4	8	1462	0·29	0·35	48·6	0·28, 0·41	0·17, 0·52	16·46*
MPA	Last 7 d	5	7	567	0·21	0·26	34·2	0·11, 0·40	−0·13, 0·64	20·49*
	Usual week	3	5	758	0·23	0·33	42·5	0·21, 0·46	0·06, 0·61	11·77*
TMPA	Last 7 d	5	7	767	0·25	0·29	52·2	0·22, 0·37	0·10, 0·49	13·43*
	Usual week	4	5	510	0·17	0·24	46·4	0·10, 0·39	−0·07, 0·56	10·78*
VPA	Last 7 d	7	10	802	0·35	0·39	12·6	0·20, 0·58	−0·21, 0·99	79·51*
	Usual week	5	10	1672	0·32	0·45	49·7	0·37, 0·54	0·18, 0·72	20·13*
TPA	Last 7 d	7	22	1864	0·32	0·34	64·3	0·31, 0·38	0·19, 0·50	34·25*
	Usual week	5	24	3042	0·32	0·37	51·2	0·33, 0·41	0·18, 0·57	46·86*
Administration										
Walking	Self-reported	7	12	2845	0·27	0·31	48·7	0·26, 0·35	0·16, 0·45	24·62*
	Interviewer	3	4	369	0·33	0·40	41·8	0·27,0 ·53	0·13, 0·67	9·57*
MPA	Self-reported	8	11	2209	0·22	0·28	80·9	0·25, 0·31	0·18, 0·39	13·60
	Interviewer	3	5	406	0·22	0·26	23·4	0·05, 0·47	−0·21, 0·73	21·41*
TMPA	Self-reported	6	7	1738	0·17	0·23	44·5	0·14, 0·31	0·00, 0·46	15·74*
	Interviewer	5	16	3200	0·43	0·53	12·1	0·42, 0·63	0·10, 0·95	133·33
VPA	Self-reported	11	17	3166	0·37	0·45	53·3	0·40, 0·50	0·24, 0·66	31·93*
	Interviewer	6	18	4518	0·43	0·56	37·3	0·47, 0·66	0·15, 0·98	48·27*
TPA	Self-reported	10	35	4850	0·27	0·31	59·8	0·28, 0·33	0·16, 0·46	58·56*
	Interviewer	6	25	4017	0·44	0·52	21·4	0·45, 0·59	0·18, 0·86	116·62*
Instrument										
Walking	Accelerometer	6	9	1596	0·29	0·35	71·6	0·31, 0·39	0·24, 0·46	12·56
	Pedometer	5	6	2237	0·28	0·30	16·4	0·20, 0·39	0·07, 0·53	36·60*
	Subjective	1	2	620	0·22	0·27	87·6	0·23, 0·30	0·22, 0·32	2·28
MPA	Accelerometer	9	11	2804	0·15	0·21	60·7	0·17, 0·26	0·06, 0·36	18·11
	Pedometer	4	5	1001	0·25	0·28	34·2	0·19, 0·37	0·07, 0·49	14·64*
	Subjective	–	–	–	–	–	–	–	–	–
TMPA	Accelerometer	8	11	2155	0·18	0·23	51·0	0·17, 0·29	0·03, 0·43	21·57*
	Pedometer	–	–	–	–	–	–	–	–	–
	Subjective	3	12	2783	0·51	0·64	56·9	0·57, 0·72	0·38, 0·91	21·09*
VPA	Accelerometer	12	15	2044	0·31	0·42	69·5	0·37, 0·46	0·24, 0·59	21·57
	Pedometer	4	5	998	0·25	0·26	14·7	0·11, 0·42	−0·07, 0·60	34·08*
	Subjective	5	15	4642	0·54	0·60	4·6	0·49, 0·70	0·19, 0·99	340·37*
TPA	Accelerometer	10	35	3404	0·30	0·34	49·5	0·30, 0·38	0·12, 0·56	70·81*
	Pedometer	5	7	1213	0·34	0·36	65·0	0·32, 0·40	0·25, 0·46	10·77
	Subjective	5	18	4250	0·43	0·53	22·4	0·45, 0·62	0·18, 0·89	144·72*

PA, physical activity, IPAQ, International Physical Activity Questionnaire; *K*, number of studies; *n*, number of ES*r*; *N*, total sample size; CV, credibility interval; MPA, moderate physical activity; TMPA, total moderate physical activity; VPA, vigorous physical activity; TPA, total physical activity.
*$P < 0·05$.
†Averaged ES*r* corrected for sampling error only.
‡Averaged ES*r* corrected for sampling error and measurement errors of IPAQ.
§Percentage of variance accounted for by statistical artefacts including sampling error and measurement error of IPAQ.

Discussion

To our knowledge, the present study is the first comprehensive attempt to synthesize the scientific evidence on convergent validity of IPAQ using meta-analysis. The first purpose of the study was to examine the overall convergent validity of IPAQ. The results showed that the overall ESρ for each PA category were all positive, which supports the convergent validity evidence of IPAQ, but they varied from small-to-medium effect size according to Cohen's definitions[38]. Walking, TMPA, VPA and TPA of IPAQ secured medium-sized ESρ, while MPA had a small-sized ESρ. Such variations in ESρ by different categories of IPAQ may be due to the inherent property of IPAQ as a subjective measure. Measuring PA in IPAQ relies on the recall of diverse activities for a 7 d period, which requires participants to utilize their cognitive ability for the recall process. The greatest ESρ observed in VPA can be explained by the evidence which shows that vigorous-intensity PA tends to be more structured, which may positively affect participant recall. On the other hand, walking and moderate-intensity activity are not typically structured but rather accumulated gradually during daily life[29]. This may result in participants not recalling the exact amount of walking and activities involved in MPA[33,39,40]. Another possible explanation for varying results in ESρ across PA categories is that variations in individual perceptions with respect to the intensity of each PA category may occur due to insufficient information for each specific category[41]. For example, IPAQ defines VPA as an activity causing harder than usual breathing and MPA as an activity causing somewhat harder breathing[42]. In order to clarify this gap between MPA and VPA, IPAQ offers some examples of activity according to MET values for each type of intensity; however, different perceived exertions may exist with respect to the specific examples given by IPAQ considering that IPAQ covers a broad range of ages from 15 to 69 years. Hallal *et al.*[8] noted that specific examples linked to physiological signs or culturally adapted examples should be provided to aid participants in distinguishing MPA from VPA; we suggest that stratifying age-relevant examples would be beneficial to obtain more valid measures for MPA and VPA.

In IPAQ, participants are instructed to report time spent in MPA that lasted for at least 10 min except while walking, which is asked in separate questions. Walking and MPA that are defined as MET values of 3·3 and 4 in IPAQ fall within the same boundary of moderate-intensity PA (i.e. 3–6 MET)[12]. Our finding suggests that TMPA, which is the sum of walking and MPA, has a greater ESρ than walking and MPA, indicating that TMPA has secured more strong convergent validity than sole measures of walking and MPA. This may imply that IPAQ has secured its initial intention of discriminating walking from MPA, in that summation of the estimates from walking and MPA would yield more valid estimates for TMPA. Some researchers argue that separation of walking and MPA in the same questionnaire may confuse participants about time spent in walking under MPA[19]; however, the results of the present study indicated, collectively, that participants may well conceive time spent in walking separate from MPA.

The second purpose of the present study was to investigate the effects of moderator variables on overall validity of IPAQ across all PA categories. IPAQ was developed with the aim of international monitoring and national comparison[7]; however, variation incurred by language translation still remained questionable due to the different cultural atmospheres[42]. In our study, we attempted to synthesize a total of 152 ESr from different cultures. There were 120 ESr retrieved from translated versions of IPAQ, which yielded greater ESρ values compared with English versions of IPAQ across all PA categories. These findings supported that IPAQ secured comparable convergent validity across different cultures without any structural changes in IPAQ. Although we agree that some examples or words should be adapted in accordance with the cultural atmosphere where IPAQ would be used, following well-established translation protocols suggested by the IPAQ consensus group would be promising for positive convergent validity of IPAQ in different cultures.

IPAQ has two different versions (i.e. long or short form). The long form measures the habitual PA in three intensity-specific categories across four domains, while the short form examines only generic PA within three intensity-specific categories without any separation of specific domains. The short form has been recommended for population-based study due to its feasibility and preferences over the long form[7]; however, the estimates from the short form tend to overestimate actual PA due to the lack of sufficient information for specific domains[43]. Bauman *et al.*[42] noted that the large variances in PA measures estimated from the short form could be caused by using the short form as a means of estimating continuous levels of PA, while the primary purpose of the short form is categorical reporting. In the current meta-analyses, levels of PA with the forms of continuous measures obtained from the short form have ESρ comparable to or even larger than that of the long form. From this, we can conclude that using a short form to estimate the amount of PA as a form of continuous measures seems to be acceptable if the primary interest of the study is not domain-specific measures. However, 95 % CV for ESρ obtained from the studies where the short form was used were shown to be relatively large *v.* the estimates from the long form. One should bear in mind that PA estimates from the short form can be varied dramatically by unexplained moderators or factors, while the long form may provide more stable measures.

Measuring generic PA using questionnaires relies heavily on recall processes that may require the appropriate retrieval cues for stimulating the search of the participant's memory[6,42]. There are two cues with respect to reference period (i.e. last 7 d or usual week) that one can utilize to aid the participant's recall process. In the original development study of IPAQ[7], the International Consensus Group found the comparability of both 'last 7 d' and 'usual week' reference periods in terms of reliability and validity and suggested to use the last 7 d reference period based on the preferences in participating countries of their study. In the current analyses, no particular patterns for the rank of ESρ by different reference periods were observed across all PA categories. It could be expected to have stronger convergent validity when using the last 7 d reference period, since most studies have implemented the IPAQ right after they finished collecting objective data for a 7 d period. The comparable results between the last 7 d and usual week may reflect the fact that people tend to conceive the reference period of usual week as the last 7 d and subsequently respond in a common way as they regarded.

It has been widely recognized that interviewer administration would minimize the possible errors in implementing subjective measurement tools that are due to participant's misinterpretation and/or misunderstanding of the questions being asked[44,45]. The findings of the current meta-analyses were mostly in agreement with previous understandings that the greater ESρ values were found from the studies in which interviewer administration was applied across all PA categories with the exception of MPA. Interviewer administration may have several advantages in that it prevent respondents from skipping questions and also could provide more opportunities to obtain more detailed information on each question *v.* the self-administrated questionnaire[7]. Moreover, it allows the researchers to obtain more reliable estimates of PA levels among less educated populations who cannot fully understand the context being asked[8]. Despite the benefits of interviewer administration, the self-reported approach may be more preferred in a large epidemiological study due to time or budget limitations; however, there would be a strong possibility to obtain more accurate measures of PA when an interviewer administered the IPAQ.

Objective measurement tools to quantify levels of PA have been highly recognized for their capability to provide more precise and accurate estimates of PA levels over subjective measurement tools[46]. There has been an increase in using objective measurement tools as a means of criterion for validating PA questionnaires. In the current meta-analyses, three types of instrument (i.e. accelerometer, pedometer and subjective measure) have been used for comparison with IPAQ. The studies featuring subjective measurement tools used as a counterpart instrument to IPAQ resulted in the greatest ESρ values for most of the PA categories. These findings are broadly in agreement with the notion that subjective measurement tools tend to share similar psychometric properties based on common subjective recall processes[10]. In other words, similar systematic errors such as cognitive biases or social desirability might occur for subjective measurement tools, by which stronger linear relationships of the estimates from IPAQ with other subjective measurement tools could be estimated. While the systematic errors within the estimates from objective measurement tools are more likely to occur by different measurement conditions, such as seasons and months[47] or number of monitoring days[48], that may result in lower convergent validity of IPAQ when comparing with objective measurement tools. In addition, such inconsistency between the estimates from IPAQ and objective measurement tools may also be attributed to the fact that IPAQ is intended to measure activities longer than 10 min in duration, whereas the accelerometer and pedometer tend to measure every form of physical movement. The concept of 10 min in IPAQ may result in unreliably large variations within individual PA levels, which may worsen the linear relationship of estimates of IPAQ with other objective measures[8,34].

There were several limitations that should be considered when examining the results of the present study. First, variations by different cut-off standards set to determine PA categories of accelerometer data across studies were not considered, which may influence varying results in ESρ, especially in MPA and VPA that are based on those standards. However, considering that there is no single 'gold standard' measure as a criterion for PA comparison, we believe that the results from our study may be generalized as overall convergent validity of IPAQ. Another area of concern is that the measure of effect size aggregated for the current meta-analysis was the correlation coefficients, which are not capable of detecting the agreements on the estimates between IPAQ and other criterion measures. Correlation coefficients would provide sufficient information for convergent validity of IPAQ as a form of linear relationship; however, examining the agreements would give an insight into the extent to which the IPAQ over- or underestimates the actual level of PA. Thus, we suggest future studies to conduct the meta-analytic review on the agreements between IPAQ and other criterion instruments. In addition, 95 % CV around ESρ values in moderator analyses showed that there was still a large amount of unexplained variance after controlling for artefacts and predefined moderators. Hierarchical moderator analyses may be a more appropriate approach to resolve this problem[37]; however, more effect sizes would be needed for each level of moderators. Lastly, some of the moderator analyses were conducted based on the small number of ESr; which may affect the generalizability of the current findings. Small-sized meta-analysis (i.e. $<$200 ESr) may only be capable of summarizing the evidence or generating hypotheses for future research[49]. The process of confirming validity evidence for a certain measurement tool is regarded as a 'never ending process'[50]; therefore, more evidence not only for convergent validity but also diverse aspects of validity of IPAQ should be continuously accumulated across different populations or measurement conditions.

Conclusion

The present study attempted to synthesize all scientific evidence to examine the overall convergent validity of IPAQ. The findings indicated that IPAQ is a reasonably valid measurement tool for measuring habitual PA. However, the variations in convergent validity across different PA categories and moderator variables imply that different research conditions should be taken into account prior to deciding on use of the appropriate type of IPAQ.

Acknowledgements

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors. There are no conflicts of interest. Study concept and design: Y.K., I.P. and M.K. Acquisition of data: Y.K. and M.K. Statistical analysis and interpretation of data: Y.K., I.P. and M.K. Drafting of manuscript: Y.K. Critical revision of manuscript: Y.K., I.P. and M.K. Study supervision: M.K.

References

1. World Health Organization (2010) *Global Recommendations on Physical Activity for Health*. Geneva: WHO; available at http://whqlibdoc.who.int/publications/2010/9789241599979_eng.pdf
2. National Institutes of Health Consensus Development Panel on Physical Activity and Cardiovascular Health (1996) Physical activity and cardiovascular health. *JAMA* **276**, 241–246.
3. Shiroma EJ & Lee I (2010) Physical activity and cardiovascular health: lessons learned from epidemiological studies across age, gender, and race/ethnicity. *J Am Heart Assoc* **122**, 743–752.
4. Warburton DER, Nicol CW & Bredin SSD (2006) Health benefits of physical activity: the evidence. *CMAJ* **174**, 801–809.
5. Blair SN, Cheng Y & Holder S (2001) Is physical activity or physical fitness more important in defining health benefits? *Med Sci Sports Exerc* **33**, 6 Suppl., S379–S399.
6. Sallis JF & Saelens BE (2000) Assessment of physical activity by self-report: status, limitations, and future directions. *Res Q Exerc Sport* **71**, 2 Suppl., S1–S14.
7. Craig CL, Marshall AL, Sjostrom M *et al.* (2003) International Physical Activity Questionnaire: 12-country reliability and validity. *Med Sci Sports Exerc* **35**, 1381–1395.
8. Hallal PC, Gomez LF, Parra DC *et al.* (2010) Lessons learned after 10 years of IPAQ use in Brazil and Colombia. *J Phys Act Health* **7**, Suppl. 2, S259–S264.
9. Downs SH & Black N (1998) The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *J Epidemiol Community Health* **52**, 377–384.
10. Prince SA, Adamo KB, Hamel ME *et al.* (2008) A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. *Int J Behav Nutr Phys Act* **5**, 56.
11. Warburton DR, Charlesworth S, Ivey A *et al.* (2010) A systematic review of the evidence for Canada's Physical Activity Guidelines for Adults. *Int J Behav Nutr Phys Act* **7**, 39.
12. Ainsworth BE, Haskell WL, Whitt MC *et al.* (2000) Compendium of physical activities: an update of activity codes and MET intensities. *Med Sci Exerc Sports* **32**, Suppl. 9, S498–S516.
13. International Physical Activity Questionnaire (2005) Guidelines for data processing and analysis. http://www.ipaq.ki.se/scoring.pdf (accessed May 2010).
14. Lipsey MW & Wilson DB (2001) *Practical Meta-analysis*. Newbury Park, CA: Sage.
15. Filliben JJ (1975) The probability plot correlation coefficient test for normality. *Technometrics* **17**, 111–117.
16. De Cocker KA, Cardon G & De Bourdeaudhuij IM (2007) Pedometer-determined physical activity and its comparison with the International Physical Activity Questionnaire in a sample of Belgian adults. *Res Q Exerc Sport* **78**, 429–437.
17. De Cocker KA, De Bourdeaudhuij IM & Cardon GM (2009) What do pedometer counts represent? A comparison between pedometer data and data from four different questionnaires. *Public Health Nutr* **12**, 74–81.
18. Deng HB, Macfarlane DJ, Thomas GN *et al.* (2008) Reliability and validity of the IPAQ-Chinese: the Guangzhou Biobank Cohort study. *Med Sci Sports Exerc* **40**, 303–307.
19. Dinger MK, Behrens TK & Han JL (2006) Validity and reliability of the International Physical Activity Questionnaire in college students. *Am J Health Promot* **37**, 337–343.
20. Gauthier AP, Lariviere M & Young N (2009) Psychometric properties of the IPAQ: a validation study in a sample of northern Franco-Ontarians. *J Phys Act Health* **6**, Suppl. 1, S54–S60.
21. Hagstromer M, Ainsworth BE, Oja P *et al.* (2010) Comparison of a subjective and an objective measure of physical activity in a population sample. *J Phys Act Health* **7**, 541–550.
22. Kolbe-Alexander TL, Lambert EV, Harkins JB *et al.* (2006) Comparison of two methods of measuring physical activity in South African older adults. *J Aging Phys Act* **14**, 98–114.
23. Kurtze N, Rangul V & Hustvedt BE (2008) Reliability and validity of the international physical activity questionnaire in the Nord-Trøndelag health study (HUNT) population of men. *BMC Med Res Methodol* **8**, 63.
24. Mader U, Martin BW, Schutz Y *et al.* (2006) Validity of four short physical activity questionnaires in middle-aged persons. *Med Sci Sports Exerc* **38**, 1255–1266.
25. van der Ploeg HP, Tudor-Locke C, Marshall AL *et al.* (2010) Reliability and validity of the international physical activity questionnaire for assessing walking. *Res Q Exerc Sport* **81**, 97–101.
26. Boon RM, Hamlin MJ, Steel GD *et al.* (2010) Validation of the New Zealand Physical Activity Questionnaire (NZPAQ-LF) and the International Physical Activity Questionnaire (IPAQ-LF) with accelerometry. *Br J Sports Med* **44**, 741–746.
27. Roman-Vinas B, Serra-Majem L, Hagstromer M *et al.* (2010) International Physical Activity Questionnaire: reliability and validity in a Spanish population. *Eur J Sport Sci* **10**, 297–304.
28. Bull FC, Maslin T & Armstrong T (2009) Global physical activity questionnaire (GPAQ): nine country reliability and validity. *J Phys Act Health* **6**, 790–804.
29. Hagstromer M, Oja P & Sjostrom M (2006) The International Physical Activity Questionnaire (IPAQ): a study of concurrent and construct validity. *Public Health Nutr* **9**, 755–762.
30. Lachat CK, Verstraeten R, Khanh le NB *et al.* (2008) Validity of two physical activity questionnaires (IPAQ and PAQA) for Vietnamese adolescents in rural and urban areas. *Int J Behav Nutr Phys Act* **5**, 37.

31. Macfarlane DJ, Lee CC, Ho EY *et al*. (2006) Convergent validity of six methods to assess physical activity in daily life. *J Appl Physiol* **101**, 1328–1334.
32. Timperio A, Salmon J, Rosenberg M *et al*. (2004) Do logbooks influence recall of physical activity in validation studies? *Med Sci Sports Exerc* **36**, 1181–1186.
33. Vandelanotte C, De Bourdeaudhuij I, Sallis JF *et al*. (2005) Reliability and validity of a computerized International Physical Activity Questionnaire (IPAQ). *J Phys Act Health* **2**, 63–75.
34. Ekelund U, Sepp H, Brage S *et al*. (2006) Criterion-related validity of the last 7-day, short form of the International Physical Activity Questionnaire in Swedish adults. *Public Health Nutr* **9**, 258–265.
35. Thuy AB, Blizzard L, Schmidt M *et al*. (2010) Reliability and validity of the global physical activity questionnaire in Vietnam. *J Phys Act Health* **7**, 410–418.
36. Hunter JE, Schmidt FL & Jackson GB (1982) *Meta-analysis: Cumulating Research Findings Across Studies*. Beverly Hills, CA: Sage.
37. Hunter JE & Schmidt FL (2004) *Methods of Meta-analysis: Correcting Error and Bias in Research Findings*, 2nd ed. Newbury Park, CA: Sage.
38. Cohen JA (1992) Power primer. *Psychol Bull* **112**, 155–159.
39. Montoye HJ, Kemper HCG, Saris WHM *et al*. (1996) *Measuring Physical Activity and Energy Expenditure*. Champaign, IL: Human Kinetics.
40. Washburn RA, Heath GW & Jackson AW (2000) Reliability and validity issues concerning large-scale surveillance of physical activity. *Res Q Exerc Sport* **71**, 2 Suppl., S104–S113.
41. Shephard JR (2003) Limits to the measurement of habitual physical activity by questionnaires. *Br J Sport Med* **37**, 197–206.
42. Bauman A, Ainsworth BE, Bull F *et al*. (2009) Progress and pitfalls in the use of the International Physical Activity Questionnaire (IPAQ) for adult physical activity surveillance. *J Phys Act Health* **6**, Suppl. 1, S5–S8.
43. Hallal CP, Victora GC, Wells CKJ *et al*. (2004) Comparison of short and full-length International Physical Activity Questionnaires. *J Phys Act Health* **1**, 227–234.
44. Heesch CK, van Uffelen GZ, Hill LR *et al*. (2010) What do IPAQ questions mean to older adults? Lessons from cognitive interviews. *Int J Behav Nutr Phys Act* **7**, 35.
45. Vuillemin A, Oppert J, Guillemin F *et al*. (2000) Self-administered questionnaire compared with interview to assess past-year physical activity. *Med Sci Sports Exerc* **32**, 1119–1124.
46. Bassett DR (2000) Validity and reliability issues in objective monitoring of physical activity. *Res Q Exerc Sport* **71**, 2 Suppl., S30–S36.
47. Kang M, Bassett DR, Tudor-Locke C *et al*. (2012) Measurement effects of seasonal and monthly variability on pedometer-determined data. *J Phys Act Health* **9**, 336–343.
48. Kang M, Bassett DR, Tudor-Locke C *et al*. (2009) How many days are enough? A study of 365 days of pedometer monitoring. *Res Q Exerc Sport* **80**, 445–453.
49. Flather MD, Farkouh ME, Pogue JM *et al*. (1997) Strengths and limitations of meta-analysis: larger studies may be more reliable. *Control Clinical Trials* **18**, 568–579.
50. Shepard LA (1993) Evaluating test validity. *Rev Res Educ* **19**, 405–450.