Examining the validity and reliability of the Chinese version of the International Physical Activity Questionnaire, long form (IPAQ-LC)

Duncan Macfarlane*, Anson Chan and Ester Cerin

Institute of Human Performance, The University of Hong Kong, Pokfulam, Hong Kong

Abstract

Objective: To investigate the reliability and the validity of the long format, Chinese version of the International Physical Activity Questionnaire (IPAQ-LC).
Design: Cross-sectional study, examining the reliability and validity of the IPAQ-LC compared with a physical activity log (PA-log) and objective accelerometry.
Setting: Self-reported physical activity (PA) in Hong Kong adults.
Subjects: A total of eighty-three Chinese adults (forty-seven males, thirty-six females) were asked to wear an ActiTrainer accelerometer (MTI-ActiGraph, Fort Walton Beach, FL, USA) for >10 h over 7 d, to complete a PA-log at the end of each day and to complete the IPAQ-LC on day 8. On a sub-sample of twenty-eight adults the IPAQ-LC was also administered on day 11 to assess its reliability.
Results: The IPAQ-LC had good test–retest reliability for grouped activities, with intra-class correlation coefficients ranging from 0·74 to 0·97 for vigorous, moderate, walking and total PA, with between-test effect sizes that were small (<0·49). The Spearman correlation coefficients were statistically significant for vigorous PA ($r = 0·28$), moderate + walking PA ($r = 0·27$), as well as overall PA ($r = 0·35$), when compared with the accelerometry-based criterion measures, but none of the IPAQ activity categories correlated significantly with the PA-log. In absolute units, only the IPAQ light and overall PA did not differ significantly from the accelerometry measures, yet overall PA was able to faithfully discriminate between quartiles of PA ($P = 0·019$) when compared to accelerometry.
Conclusions: The IPAQ-LC demonstrated adequate reliability and showed sufficient evidence of validity in assessing overall levels of habitual PA to be used on Hong Kong adults.

Keywords
Physical activity
Validity
Reliability
IPAQ
Chinese

Acquiring an adequate level of habitual physical activity (PA) can provide numerous health-enhancing benefits, including reducing the risks of CVD, type II diabetes, obesity and some cancers[1]. While the minimum dose of PA needed to enhance health and prevent hypokinetic conditions is not fully known, the American College of Sports Medicine (ACSM) and American Heart Association recommend undertaking 30 min of moderate-intensity PA lasting at least 10 min on 5 d/week, or 20 min of vigorous-intensity PA on 3 or more d/week[2]; others suggest accumulating 150 min moderate PA/week or 75 min vigorous PA/week[3]. In spite of these recommendations the inhabitants of most countries fail to accrue sufficient PA to derive health-related benefits[4]. In Hong Kong, the population-attributable risk from physical inactivity has recently been shown to exceed that of tobacco smoking[5].

Monitoring whether a population is obtaining the recommended levels of habitual PA necessary to promote health requires a valid and reliable research tool capable of assessing the frequency and duration of common moderate and vigorous activities. Numerous objective methods exit to quantify habitual PA[6,7], such as accelerometers, heart-rate monitors or observation techniques, yet few are easily employed on large samples, making self-report recall questionnaires the method of choice. Many PA questionnaires exist[7], but few have been specifically developed to provide an international standard that can be rigorously translated and used for inter-country comparisons. With this in mind and to aid public health surveillance, a set of standardized international physical activity questionnaires (IPAQ) was developed[4]. The questionnaires were designed to be administered to adults (18–65 years) and in the long format to cover the major activity domains of transportation, work, household and leisure-time PA. The development team devised four variants of IPAQ: a short form (nine items) and a long form (thirty-one items), each of which could be administered by interview or self-completed (see www.ipaq.ki.se); they

Corresponding author: Email djmac@hku.hk

© The Authors 2010

also recently reflected on some of the developments and problems of the IPAQ[8].

A twelve-country validity and reliability study showed that the IPAQ was adequately reliable (Spearman ρ of 0·81 and 0·76 for the long and short version, respectively) and, when compared with a criterion accelerometer, the validity (Spearman ρ of 0·33 and 0·30 for the long and short version, respectively) was comparable to other questionnaires that have used similar validation techniques[9]. Yet in the twelve-country study none examined a Chinese version of IPAQ and the criterion standard was delimited to using accelerometry only, which is known to have numerous limitations[6]. It is also essential to ensure that each localized version of the IPAQ is reliable and valid for the country for which it was adapted, since the recall of physical activities is a complex cognitive process that can generate errors from the interpretation of questions, as well as cultural differences in activities and terminologies[10,11]. The aim of the present study was to examine the reliability and validity of the long self-report version, but using multiple concurrent criterion standards (accelerometry and a physical activity log (PA-log)), as several reviews, including the Surgeon General's Report, state that no single suitable 'gold standard' criterion measure exists for PA comparisons[1,12,13]. Moreover, given that objectively measured PA intensity as captured by accelerometry may not correspond to perceived PA intensity[14], it was important to compare IPAQ estimates of habitual PA with those collected using another subjective but more reliable method (PA-log)[15]. We hypothesized that the IPAQ-LC (long, Chinese self-report version) would be highly reliable, but possess low to moderate validity compared with the objective (accelerometry) and subjective criterion standard (PA-log)[16].

Materials and methods

Participants

Two separate groups were recruited for the reliability study and for the validity study. A convenience sample of twenty-eight people was used for the reliability study; while, for the validity study, eighty-eight volunteers were recruited by mailed requests sent to specific residences chosen from thirty-two different neighbourhoods that varied in extremes of socio-economic status and walkability[17,18]. All were native Chinese speakers recruited from a large city in China (Hong Kong). After the study had gained approval from The University of Hong Kong's Ethics Committee, the experimental protocol was explained and written consent was received from all participants. Over seven consecutive days every participant was requested to wear the accelerometer for ≥600 min/d during waking hours (except when exposed to water), to complete a daily PA-log and to complete a 7 d physical activity recall questionnaire (IPAQ-LC) on day 8. Participants taking part in the

reliability study were also asked to complete the IPAQ-LC on day 11. All participants were instructed to engage in their normal daily habits during the measurement period.

Physical activity assessment

Uniaxial accelerometer

The accelerometer (ActiTrainer; MTI-ActiGraph, Fort Walton Beach, FL, USA) was initialized with a time stamp, a 1-min data epoch was chosen, and then it was carefully secured in the correct orientation in a small pouch worn firmly around the waist on the right side in line with the mid-axilla. The accelerometer data were downloaded and stored on a computer using its proprietary software before being processed using custom-made Excel Visual Basic Macros to identify the time spent in three activity levels based on published cut-off points[19]: light activity (2–2·99 MET = 694–2020 counts/min); moderate activity (3–5·99 MET = 2021–5999 counts/min); and vigorous activity (≥6 MET = >5999 counts/min). Although various studies have used a minimum cut-off point of zero for light activity[20], we, like some[21,22], used a higher cut-off point (693 counts/min = 2 MET) to exclude 'very light' activity and to be consistent with the PA-log analysis. The amounts of light, moderate and vigorous activity were reported as MET × min/d (or MET × min/week) using multipliers of 2·5, 4 and 8 MET, respectively. Total step counts were also recorded by the accelerometer using its internal software option.

Physical activity log

At the end of each day participants completed a one-page PA-log, recording all activities with durations ≥10 min, grouped into home, occupation, sitting, moderate leisure, vigorous leisure, transportation and 'other' activities, based on a previous format[15]. This required the participants to circle each activity they took part in, to estimate the duration of each activity and record the time they began each activity. The logs required minimal literacy and were completed in less than 5 min. The logs were collected and each activity scored using metabolic equivalent task (MET) values taken from the most recent Compendium of Physical Activities[23]. For each day the total minutes of activity were aggregated by intensity level into sitting, light (2–2·99 MET), moderate (3–5·99 MET) and vigorous (≥6 MET) activity. Finally, the weekly total duration spent in each intensity level was generated from the seven completed daily logs (MET × min/week).

International Physical Activity Questionnaire – long, Chinese version

The IPAQ-LC is a Chinese version of the long, last 7 d, self-report format[4], available in English (and other languages) at www.ipaq.ki.se. It required the participants to complete thirty-one questions on the frequency and duration of time spent in four activity domains (transportation, work,

household and leisure time), and included sections on walking, moderate, vigorous and sedentary behaviours (sitting and lying awake). The IPAQ-LC was independently translated from English by two bilingual experimenters familiar with questionnaires, then mutually checked and modified by the experimenters for consistency. The Chinese version was then back-translated into English by a third independent bilingual experimenter and checked for any discrepancies by a native English speaker. Each participant completed the self-report IPAQ-LC on day 8, so that its 7 d recall period coincided with the same 7 d of objective data collection and the seven daily PA-logs. In the reliability group, the IPAQ-LC was also re-administered on day 11, with days 4–7 being in common to both recalls (reducing biological variation) but a 3 d gap to reduce the chances of remembering the data first reported. The IPAQ-LC data were presented as the total MET × min/d (or MET × min/week) for walking (shown here as light activity, 3·3 MET), moderate (4 MET) and vigorous (8 MET) activities.

Data analysis

All data were examined for outlying values but no editing was performed unless a clear data input error had been made and checked against field/manual records. Unlike the minimum 5 d requirement of Craig et al.[4] our participants were required to obtain data on 4 d (including one weekend day), but the similar registered time of ≥600 min/d was required before accelerometry analysis. The decision to analyse all participants who completed at least four full days was based on recent reviews[24–26] that suggest this period reliably estimates levels of habitual PA.

Our data processing was similar to other published studies that have used these same instruments[4,15,19,27], yet this involved some slight inconsistencies in categorizing intensities across instruments. For example, walking (3·3 MET) was considered a separate and distinct activity from moderate activities (≥4 MET) in the IPAQ[4], yet it has been traditionally classified as moderate activity (3–5·99 MET) by the PA-log[15]. For this reason we have reported IPAQ–walking both (i) individually, as light activity, and (ii) like Ainsworth et al.[15] we included it in IPAQ–moderate PA to permit comparability with the moderate PA-log data. Similar variations occurred with vigorous activity being defined as ≥6 MET by the PA-log[15] but ≥8 MET by IPAQ[4].

Inspection of our PA data confirmed they were not normally distributed; thus for validity analysis Friedman's non-parametric test for dependent samples was used to simultaneously determine if significant differences existed between the measures. When significance was established, follow-up Wilcoxon signed-rank tests were used to determine where differences between individual pairs of data existed, with Holm's sequential Bonferroni adjustment used to control for type 1 errors. Non-parametric Spearman correlations were used to examine the associations between

data from pairs of measures. Statistical analyses were performed using JMP v8 software (SAS Institute, Cary, NC, USA), with data shown as mean and standard deviation unless stated otherwise. The reliability measures recommended by Hopkins[28] included the unbiased typical error (TE) determined from the SD of the test–retest change score divided by $\sqrt{2}$, with the CV% being the TE expressed as a percentage of the overall mean score; the intra-class correlation coefficient (ICC) and the effect size indicate the magnitude of the difference between the test–retest estimates of habitual PA, and were interpreted similar to Saunders et al.[18].

Results

The reliability sample contained twelve males and sixteen females with an average age of 26·2 (SD 9·9) years, height of 1·65 (SD 0·08) m, weight of 58·3 (SD 10·7) kg and BMI of 21·3 (SD 3·0) kg/m². The validity study began with eighty-eight volunteers, but only eighty-three produced data that were acceptable (five volunteers reported outlying data deemed to be unacceptable, defined when daily averages for walking >6 h, or moderate PA >4·5 h, or vigorous PA >2 h). This resulted in analysing data from forty-seven males and thirty-six females with an average age of 40·9 (SD 11·1) years, height of 1·65 (SD 0·08) m, weight of 62·8 (SD 12·6) kg and BMI of 22·9 (SD 3·5) kg/m².

Reliability of the IPAQ-LC

Table 1 shows that the test–retest reliability of the domains (working, active transport, domestic, leisure and sitting) of PA were in generally acceptable, although domestic activity showed an unacceptably low ICC (0·22) and high CV% even though the effect size remained quite small (0·31). When categorized according to the intensity of the activity (walking, moderate, walking + moderate, vigorous, total activity), all group activities showed moderately high ICC values (0·74–0·95) with reasonable CV% and either trivial or small effect sizes (<0·50).

Validity of the IPAQ-LC

Table 2 presents the commonly used Spearman correlation coefficients to assess the correspondence of data acquired using the IPAQ-LC with the accelerometry, PA-log and total step counts (for overall PA only). Significant correlations of $r = 0·35$ and $0·36$ were found between the IPAQ-LC and the accelerometer and average step counts per day, respectively. However, the IPAQ-LC was only weakly correlated with the PA-log ($r = 0·13$). When total PA was examined in its sub-components (light, moderate, vigorous), vigorous PA and moderate (including moderate + walking) PA were the only components that correlated significantly with the accelerometry data. No correlations between the IPAQ-LC and the PA-log data reached statistical significance, although vigorous PA approached this ($P = 0·056$).

Table 1 Reliability of the IPAQ-LC measures, showing total values over 7 d in MET × min/week, in a sample of Hong Kong adults

PA measure (MET × min/week)	Test 1		Test 2		ICC	TE	CV (%)	ES
	Mean	SD	Mean	SD				
Working	311·4	771·4	291·1	774·6	0·97	131·9	43·8	0·11 (trivial)
Active transport	478·5	560·5	455·3	579·5	0·88	202·7	43·4	0·08 (trivial)
Domestic	52·8	131·0	15·6	39·4	0·22	87·0	254·5	0·31 (small)
Leisure	928·2	943·6	750·3	813·1	0·88	315·4	37·6	0·40 (small)
Sitting	3754·4	1138·5	4115·6	963·6	0·71	588·9	15·0	0·44 (small)
All walking activities	994·9	889·6	961·9	1108·8	0·95	238·8	24·4	0·10 (trivial)
All moderate activities	297·5	338·9	197·8	319·4	0·74	174·2	70·3	0·41 (small)
All (walking + moderate)	1292·4	1002·9	1159·7	1229·1	0·93	313·3	25·6	0·30 (small)
All vigorous activities	478·5	819·4	352·6	613·5	0·80	334·9	80·6	0·27 (small)
All activities (walking + moderate + vigorous)	1770·9	1344·3	1512·3	1328·4	0·93	373·8	22·8	0·49 (small)

IPAQ-LC, long format, Chinese version of the International Physical Activity Questionnaire; PA, physical activity; MET, metabolic equivalent task; ICC, intra-class correlation coefficient; TE, typical error of measurement; ES, effect size.
Values are means and standard deviation, *n* 28. Tests 1 and 2 were conducted within 3 d for all subjects. TE is the error associated with biological and technical variation, in order to show when a true change occurs for an individual. CV% is TE expressed as a percentage of mean score. ES indicates magnitude of differences between tests: <0·2 = trivial; 0·2–0·6 = small; 0·6–1·2 = moderate; >1·2 = large.

Table 2 Non-parametric correlations of the IPAQ-LC PA estimates with accelerometry-based estimates, self-reported PA-log and total step counts (overall PA only) in a sample of Hong Kong adults

MET × min/d	Pairs of PA estimates compared	Spearman ρ	P
Overall PA	IPAQ *v.* accelerometer	0·35	0·001*
	IPAQ *v.* PA-log	0·13	0·248
	IPAQ *v.* step counts (total)†	0·36	0·001*
Light PA	IPAQ–walk *v.* accelerometer	0·21	0·061
	IPAQ–walk *v.* PA-log	−0·18	0·111
Moderate PA	IPAQ–moderate *v.* accelerometer	0·10	0·370
	IPAQ–moderate *v.* PA-log	−0·04	0·734
	IPAQ–moderate + walk *v.* accelerometer	0·27	0·013*
	IPAQ–moderate + walk *v.* PA-log	0·13	0·255
Vigorous PA	IPAQ *v.* accelerometer	0·28	0·010*
	IPAQ *v.* PA-log	0·21	0·056

IPAQ-LC, long format, Chinese version of the International Physical Activity Questionnaire; PA, physical activity; MET, metabolic equivalent task.
*Correlation was significant ($P < 0.05$).
†Step counts = average steps/d.

Table 3 Non-parametric test of differences between IPAQ-LC and accelerometry-based and self-report PA-log estimates in a sample of Hong Kong adults

| MET × min/d	Pairs of PA estimates compared	P
Overall PA	IPAQ (350·9, SD 307·2) *v.* accelerometer (329·3, SD 126·3)	0·643
	IPAQ (350·9, SD 307·2) *v.* PA-log (572·4, SD 305·2)	0·001*
Light PA	IPAQ–walk (189·0, SD 196·8) *v.* accelerometer (174·5, SD 64·6)	0·679
	IPAQ–walk (189·0, SD 196·8) *v.* PA-log (70·6, SD 117·3)	0·001*
Moderate PA	IPAQ–moderate (95·0, SD 140·0) *v.* accelerometer (148·5, SD 87·4)	0·001*
	IPAQ–moderate (95·0, SD 140·0) *v.* PA-log (464·8, SD 290·2)	0·001*
	IPAQ–moderate + walk (284·0, SD 249·0) *v.* accelerometer (148·5, SD 87·4)	0·001*
	IPAQ–moderate + walk (284·0, SD 249·0) *v.* PA-log (464·8, SD 290·2)	0·001*
Vigorous PA	IPAQ (66·9, SD 6·3) *v.* accelerometer (6·3, SD 22·9)	0·001*
	IPAQ (66·9, SD 6·3) *v.* PA-log (37·0, SD 100·5)	0·045*

IPAQ-LC, long format, Chinese version of the International Physical Activity Questionnaire; PA, physical activity; MET, metabolic equivalent task.
All original data in units of MET × min/d, with mean and standard deviation values, and associated *P* values from Wilcoxon signed-rank tests.
*Significant ($P < 0.05$).

Comparison of the mean MET × min/d data in Table 3 showed no significant difference between IPAQ-LC and the accelerometry data for overall PA (difference = 21·6 MET × min/d), as well as for light PA (difference = 14·4 MET × min/d). However, all other comparisons with accelerometry, including all comparisons with the PA-log,
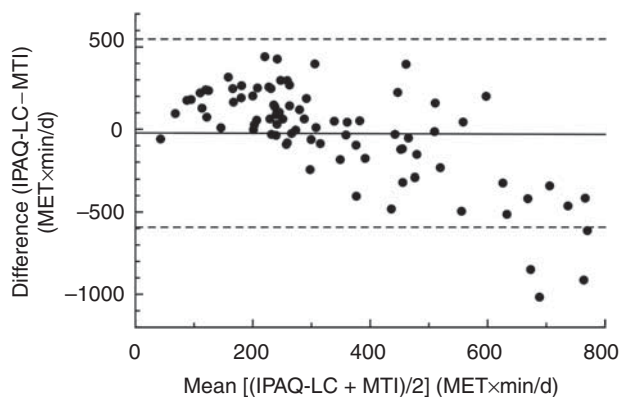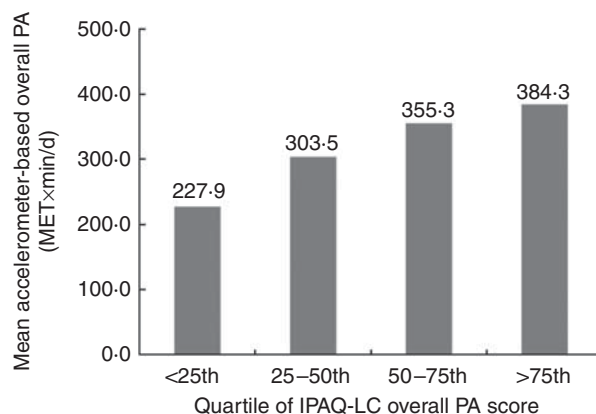
Fig. 1 Modified Bland–Altman plot for overall physical activity in a sample of Hong Kong adults (*n* 83), showing the mean value estimated by the long format, Chinese version of the International Physical Activity Questionnaire and the accelerometer (ActiTrainer; MTI-ActiGraph, Fort Walton Beach, FL, USA), Mean [(IPAQ-LC + MTI)/2] (MET × min/d), plotted against the difference between the two methods, Difference (IPAQ-LC − MTI) (MET × min/d). Mean bias (−21·6) is indicated by ——; − − − indicates 95 % limits of agreement (553·9, −597·1)

Fig. 2 Mean accelerometer-based estimate for overall physical activity (PA) in MET × min/d in each quartile of overall PA score (MET × min/d) estimated from the long format, Chinese version of the International Physical Activity Questionnaire (IPAQ-LC) in a sample of Hong Kong adults (*n* 83)

showed significant differences from the IPAQ-LC data. The Bland–Altman plot (Fig. 1) also showed a small bias between the overall PA mean and the differences data in MET × min/d when comparing accelerometry and the IPAQ-LC (bias = −21·6), but large 95 % limits of agreement of −597·1 and 553·9. Also, the difference between the two estimates of PA appeared to depend on the level of PA. Specifically, as compared with the accelerometer, the IPAQ-LC overestimated overall PA in individuals with low levels of PA and underestimated overall PA in individuals with high levels of PA. Yet when the mean overall accelerometry scores were compared against quartiles of overall PA from the IPAQ-LC, there was a relatively clear and linear increase in the mean values (227·9, 303·5, 355·3 and 384·3 MET × min/d) as one progressed from the <25th to the >75th percentile (Fig. 2). The ability of IPAQ-LC to appropriately screen respondents who did (true positives = sensitivity) or did not (true negatives = specificity) meet current ACSM PA guidelines[29] was also undertaken[30]. The 'moderate' category of the standardized IPAQ scoring protocol (www.ipaq.ki.se) reflects current guidelines[2] and all those who met or exceeded this category were compared with those who accumulated activity above the moderate accelerometry threshold (2021 counts/min) of at least 30 min/d. The analysis showed IPAQ-LC had a sensitivity of 90 % and a specificity of 29 %.

Discussion

To our best knowledge, the present study is the only one to examine the reliability and validity of the long version of IPAQ that has been modified specifically for

the Cantonese-speaking group of Chinese who live in the most southern regions of China. Although several other studies have examined aspects of the validity/reliability of IPAQ on Chinese subjects, these were either performed using the short version on Cantonese speakers[31,32] or have been delimited to Mandarin speakers from Beijing[33], Chengdu[34] or Taiwan[35], whose dialect and written characters differ from those commonly used in Hong Kong and whose geographical locations have cooler climates.

Unlike the short format, the long version of IPAQ allows respondents to report the frequency, duration and intensity of all activities (>10 min) across a variety of contexts, which has been a limitation of previous self-report questionnaires[9]. Being able to monitor the domain in which the activity is performed is important not only in studies using ecological models to examine the associations between activity and the physical environment[36], but also in prospective studies to examine which domains of activity may have responded to an intervention or whether direct compensation from one domain to another occurs (e.g. increased active transport leading to decreased leisure activity) without a net change in total activity.

In the process of being considered valid, a questionnaire should first be reliable. The results in Table 1 show that IPAQ-LC produced ICC values for each domain that were consistently above 0·7, a level of reproducibility considered acceptably good for questionnaire data[37], with the exception of domestic activity (which also showed an unacceptably high CV % in part due to the low mean score). The ICC for each activity domain compare favourably with other detailed reliability data on the IPAQ long format[37], although Levy and Readdy showed a much higher ICC for total domestic activity (0·69). The poor reliability for domestic activity in our Hong Kong study is suspected to be related to the infrequent and varied

household activities undertaken by most Hong Kong residents (Table 2 shows means of 52·8 and 15·5 MET × min/week for the test and retest). The vast majority of Hong Kong residents live in multi-storey apartments[38] that require no garden or outdoor maintenance and many families have full-time domestic helpers to take care of indoor domestic activities, which may have contributed to the low reliability of self-reported domestic activities. Yet all of the effect sizes, indicating the magnitude of the PA differences between assessments, were small or trivial for each specific domain of activity or when similar intensities of activity were combined (walking, moderate, walking + moderate, vigorous, total activity). These results suggest that the IPAQ-LC is adequately reliable for use on Cantonese-speaking respondents.

The IPAQ-LC showed reasonable evidence of validity for overall (total) PA as it was significantly correlated with the criterion accelerometer, with a Spearman correlation ($r = 0.35$, $P < 0.001$) that is very similar to the one obtained in the multi-national validation study by Craig *et al.*[4] and in other studies on the long version of IPAQ[33,39–42]. Although validity correlations around 0·35 for total activity from objective criteria are not ideal, they are frequently reported for many other widely used self-report PA questionnaires used for PA surveillance[4,7,9]. In comparison, none of the activity categories from the long version of IPAQ used in our study was significantly correlated with those from the self-reported PA-log. However, the light and moderate sub-categories of IPAQ-LC were relatively poorly correlated with the criterion accelerometer, with only vigorous activity showing a clear significant result (along with moderate PA when compared with 'moderate + walking' IPAQ activity).

It was not unexpected that the IPAQ-LC results correlated poorly with light-intensity accelerometry scores, as the lowest intensity of activity measured by IPAQ-LC is walking, which is arguably a moderate form of activity with MET = 3·3 and thus strictly not a form of light activity (which normally encompasses the 2–2·99 MET range[2]). In comparison, it was interesting to see the IPAQ-LC scores for 'walking and moderate PA combined' (arguably a more comparable measure of moderate activity) being significantly correlated with the accelerometry-based estimates of moderate PA, as also occurred for vigorous activity. However, the fact that the activity categories from IPAQ-LC consistently failed to correlate with the PA-log suggests these two self-reported instruments may not be measuring the same constructs and may reflect differential ability to recall activities (the IPAQ recalled the last 7 d, while the PA-log recalled events at the end of each day). However this cannot fully explain the results as others have shown good correlations between long versions of IPAQ and a PA-log[40]. It is possible that the respondents did not fully comply with the PA-log protocol and did not regularly record their PA at the end of each study day. Data collection using personal digital assistants or electronic mail systems might have yielded more consistent results as they motivate protocol compliance by automatically recording the time of data entry.

In terms of absolute comparisons, the IPAQ-LC showed reasonable evidence of validity for overall (total) PA, with the mean MET × min/d value being a non-significant 6·5% higher than the mean accelerometer value (but significantly 39% lower than the mean value recorded from the PA-log). The modified Bland–Altman plot in Fig. 1 supports the finding of a relatively small mean bias for overall PA between the IPAQ-LC and accelerometry data (21·6 MET × min/d). However, the large 95% limits of agreement suggest that there can be considerable individual errors, although these wide limits appear to have been partly influenced by three outliers that were in the range of 800–1000 MET × min/d. Some care is clearly needed when interpreting the IPAQ data, particularly as the bias was more pronounced at high to very high activity levels (Fig. 1), although such errors are likely to affect only the most active respondents.

Significant differences were seen between every intensity sub-category in IPAQ-LC and the PA-log, with no consistent pattern; respondents reported more light and vigorous IPAQ activity, but less moderate activity. This inconsistency may again be due to IPAQ only having walking as the lowest form of activity, but also partly due to assigning a single MET value to each IPAQ intensity, while the PA-log allowed individualized MET values for each reported activity. Previous research has also shown that the completion of a daily PA-log does not appear to influence the estimates of validity for instruments such as the IPAQ[42]. Despite the inability of IPAQ-LC to accurately measure light, moderate and vigorous activity when compared with criterion accelerometry, it remains a useful epidemiological tool since it can accurately assess total PA, which is often the most common requirement in many activity studies. This epidemiological value of IPAQ-LC is further shown by its ability to accurately rank each quartile of the respondents using the overall MET × min/d value. Figure 2 shows that there was a statistically significant linear trend ($P = 0.019$) in the criterion accelerometry readings (mean overall MET × min/d) as the quartiles progressed from the <25th percentile up to the >75th percentile IPAQ score. A similar ability to appropriately rank respondents into quartiles of self-reported activity has also been reported for the IPAQ short form in a group of Swedish adults[43].

The IPAQ-LC was very commendable in correctly screening 90% of those participants who achieved moderate exercise of at least 30 min/d (sensitivity), but was very poor in classifying only 29% of participants who were unable to meet this target (specificity). One other study reporting the sensitivity and specificity of the IPAQ long form has produced respective percentages of 71% and 59%[41]. In general, it appears that the long version of IPAQ is relatively good at identifying active members of the community, possibly due to the typical over-reporting of IPAQ data[41,44], but is relatively poor at identifying those

who need to accrue greater levels of PA. When compared with IPAQ, the 7 d Physical Activity Recall (PAR)[45] has been shown to provide markedly higher levels of specificity and sensitivity, which was attributed to the PAR focusing more on leisure activity compared with the four domains of activity in the IPAQ[41]. As an important aim of public health is to promote adequate activity levels at a community level, the fact that IPAQ-LC was poor at identifying those truly in need of greater activity remains a limitation of IPAQ-LC as a surveillance tool.

A number of methodological limitations were contained within the present study. Due to the small size of the validity study (n 83) and especially the reliability study (n 28), an examination of how demographic factors such as age, gender or education affected the validity and reliability of the IPAQ-LC was not considered. Participants in the validity study were part of a larger study on the built environment (convenience sample of 334 citizens), and those volunteering to have their activity objectively assessed may have introduced a self-selection bias (e.g. being more active or more aware of their activity habits). In comparison, the reliability study was performed on a slightly younger group that included university students and postgraduates; this may have contributed to the lower reliability in the domestic activity domain, as some of these duties may have been done by domestic helpers or on a rotation basis when in shared student accommodation. Thus the generalizability of these results to the wider community may be limited.

As occurs frequently in validations of PA questionnaires, an accelerometer was used as the criterion measure even though it is known to have its own limitations. Accelerometers are well known to underestimate not only several forms of PA[13], but also the energy cost of free-living activities, especially when using regression equations derived from moderate and vigorous intensity cut-off points that vary within the literature[20,46]. Nevertheless accelerometers are capable of precisely measuring the frequency, duration and intensity of an activity[47] and will remain a common criterion until more acceptable criterion measures can be routinely used on large number of free-living members of the community.

Overall, the present study suggests that the IPAQ-LC is a sufficiently reliable and valid measure of total PA, as well as in ranking overall PA in a Cantonese-speaking Chinese population. However, since the domains and sub-categories of activity of the IPAQ-LC generally had an unacceptably low level of validity (particularly moderate activity), the reliable and valid shorter version of the Chinese IPAQ[32] might be more appropriate and time-efficient for many studies, especially in those where total PA is the primary outcome variable.

Acknowledgements

References

1. US Department of Health and Human Services (1996) *Physical Activity and Health: A Report of the Surgeon General.* Atlanta, GA: US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion.
2. Haskell WL, Lee IM, Pate RR *et al.* (2007) Physical activity and public health: updated recommendation for adults from the American College of Sports Medicine and the American Heart Association. *Med Sci Sports Exerc* **39**, 1423–1434.
3. US Department of Health and Human Services (2008) *2008 Physical Activity Guidelines for Americans.* Rockville, MD: US Department of Health and Human Services.
4. 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.
5. Lam TH, Ho SY, Hedley AJ *et al.* (2004) Leisure time physical activity and mortality in Hong Kong: case–control study of all adult deaths in 1998. *Ann Epidemiol* **14**, 391–398.
6. Welk GJ (2002) *Physical Activity Assessments for Health-Related Research.* Champaign, IL: Human Kinetics.
7. Pereira MA, FitzerGerald SJ, Gregg EW *et al.* (1997) A collection of Physical Activity Questionnaires for health-related research. *Med Sci Sports Exerc* **29**, 6 Suppl., S1–S205.
8. 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.
9. 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.
10. Meriwether RA, McMahon PM, Islam N *et al.* (2006) Physical activity assessment: validation of a clinical assessment tool. *Am J Prev Med* **31**, 484–491.
11. Masse LC (2000) Reliability, validity, and methodological issues in assessing physical activity in a cross-cultural setting. *Res Q Exerc Sport* **71**, 2 Suppl., S54–S58.
12. LaMonte MJ, Ainsworth BE & Tudor-Locke C (2003) Assessment of physical activity and energy expenditure. In *Obesity: Etiology, Assessment, Treatment and Prevention*, pp. 111–137 [RE Andersen, editor]. Champaign, IL: Human Kinetics.
13. Welk GJ (2005) Principles of design and analyses for the calibration of accelerometry-based activity monitors. *Med Sci Sports Exerc* **37**, 11 Suppl., S501–S511.
14. Dale D, Welk GJ & Mathews CE (2002) Methods for assessing physical activity and challenges for research. In *Physical Activity Assessments for Health-Related Research*, pp. 19–34 [GJ Welk, editor]. Champaign, IL: Human Kinetics.
15. Ainsworth BE, Bassett Jr DR, Strath SJ *et al.* (2000) Comparison of three methods for measuring the time

spent in physical activity. *Med Sci Sports Exerc* **32**, 9 Suppl., S457–S464.

16. Matthews CE (2002) Use of self-report instruments to assess physical activity. In *Physical Activity Assessments for Health-Related Research*, pp. 107–123 [GJ Welk, editor]. Champaign, IL: Human Kinetics.

17. Cerin E, Macfarlane DJ, Ko H-H *et al.* (2007) Measuring perceived neighbourhood walkability in densely-populated urban areas in Asia. *Cities* **24**, 204–217.

18. Saunders PU, Pyne DB, Telford RD *et al.* (2004) Reliability and variability of running economy in elite distance runners. *Med Sci Sports Exerc* **36**, 1972–1976.

19. Freedson PS, Melanson E & Sirard J (1998) Calibration of the Computer Science and Applications, Inc. accelerometer. *Med Sci Sports Exerc* **30**, 777–781.

20. Matthews CE (2005) Calibration of accelerometer output for adults. *Med Sci Sports Exerc* **37**, 11 Suppl., S512–S522.

21. Matthews CE, Ainsworth BE, Hanby C *et al.* (2005) Development and testing of a short physical activity recall questionnaire. *Med Sci Sports Exerc* **37**, 986–994.

22. Tudor-Locke C, Ainsworth BE, Thompson RW *et al.* (2002) Comparison of pedometer and accelerometer measures of free-living physical activity. *Med Sci Sports Exerc* **34**, 2045–2051.

23. Ainsworth BE, Haskell WL, Whitt MC *et al.* (2000) Compendium of physical activities: an update of activity codes and MET intensities. *Med Sci Sports Exerc* **32**, 9 Suppl., S498–S504.

24. Masse LC, Fuemmeler BF, Anderson CB *et al.* (2005) Accelerometer data reduction: a comparison of four reduction algorithms on select outcome variables. *Med Sci Sports Exerc* **37**, 11 Suppl., S544–S554.

25. Ward DS, Evenson KR, Vaughn A *et al.* (2005) Accelerometer use in physical activity: best practices and research recommendations. *Med Sci Sports Exerc* **37**, 11 Suppl., S582–S588.

26. Trost SG, McIver KL & Pate RR (2005) Conducting accelerometer-based activity assessments in field-based research. *Med Sci Sports Exerc* **37**, 11 Suppl., S531–S543.

27. Hallal PC, Victora CG, Wells JC *et al.* (2003) Physical inactivity: prevalence and associated variables in Brazilian adults. *Med Sci Sports Exerc* **35**, 1894–1900.

28. Hopkins WG (2000) Measures of reliability in sports medicine and science. *Sports Med* **30**, 1–15.

29. American College of Sports Medicine (2005) *ACSM's Guidelines for Exercise Testing and Prescription*, 7th ed., pp. 133–173. Baltimore, MD: Lippincott, Williams and Wilkins.

30. Bland M (1987) *An Introduction to Medical Statistics*. Oxford: Oxford Medical Publications.

31. 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.

32. Macfarlane DJ, Lee CC, Ho EY *et al.* (2007) Reliability and validity of the Chinese version of IPAQ (short, last 7 days). *J Sci Med Sport* **10**, 45–51.

33. Qu NN & Li KJ (2004) Study on the reliability and validity of international physical activity questionnaire (Chinese Version, IPAQ). *Zhonghua Liu Xing Bing Xue Za Zhi* **25**, 265–268.

34. Jia YJ, Xu LZ, Kang DY *et al.* (2008) Reliability and validity regarding the Chinese version of the International Physical Activity Questionnaires (long self-administrated format) on women in Chengdu, China. *Zhonghua Liu Xing Bing Xue Za Zhi* **29**, 1078–1082.

35. Liou YM, Jwo CJ, Yao KG *et al.* (2008) Selection of appropriate Chinese terms to represent intensity and types of physical activity terms for use in the Taiwan version of IPAQ. *J Nurs Res* **16**, 252–263.

36. Giles-Corti B, Timperio A, Bull F *et al.* (2005) Understanding physical activity environmental correlates: increased specificity for ecological models. *Exerc Sport Sci Rev* **33**, 175–181.

37. Levy SS & Readdy RT (2009) Reliability of the International Physical Activity Questionnaire in research settings: last 7-day self-administered long form. *Meas Phys Educ Exerc Sci* **13**, 191–205.

38. Sallis JF, Bowles HR, Bauman A *et al.* (2009) Neighborhood environments and physical activity among adults in 11 countries. *Am J Prev Med* **36**, 484–490.

39. Hagstromer M, Bergman P, De Bourdeaudhuij I *et al.* (2008) Concurrent validity of a modified version of the International Physical Activity Questionnaire (IPAQ-A) in European adolescents: the HELENA Study. *Int J Obes (Lond)* **32**, Suppl. 5, S42–S48.

40. 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.

41. Johnson-Kozlow M, Sallis JF, Gilpin EA *et al.* (2006) Comparative validation of the IPAQ and the 7-Day PAR among women diagnosed with breast cancer. *Int J Behav Nutr Phys Act* **3**, 7.

42. 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.

43. 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.

44. Rzewnicki R, Vanden Auweele Y & De Bourdeaudhuij I (2003) Addressing overreporting on the International Physical Activity Questionnaire (IPAQ) telephone survey with a population sample. *Public Health Nutr* **6**, 299–305.

45. Sallis JF, Haskell WL, Wood PD *et al.* (1985) Physical activity assessment methodology in the Five-City Project. *Am J Epidemiol* **121**, 91–106.

46. Metzger JS, Catellier DJ, Evenson KR *et al.* (2008) Patterns of objectively measured physical activity in the United States. *Med Sci Sports Exerc* **40**, 630–638.

47. Bassett DR Jr (2000) Validity and reliability issues in objective monitoring of physical activity. *Res Q Exerc Sport* **71**, 2 Suppl., S30–S36.