

Convergent validity of six methods to assess physical activity in daily life

Duncan J. Macfarlane, Cherry C. Y. Lee, Edmond Y. K. Ho, K. L. Chan, and Dionise Chan

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

Submitted 21 March 2006; accepted in final form 15 June 2006

Macfarlane, Duncan J., Cherry C. Y. Lee, Edmond Y. K. Ho, K. L. Chan, and Dionise Chan. Convergent validity of six methods to assess physical activity in daily life. *J Appl Physiol* 101: 1328–1334, 2006. First published July 6, 2006; doi:10.1152/jappphysiol.00336.2006.—The purpose was to examine the agreement (convergent validity) between six common measures of habitual physical activity to estimate durations of light, moderate, vigorous, and total activity in a range of free-living individuals. Over 7 consecutive days, 49 ethnic Chinese (30 men, 19 women), aged 15–55 yr, wore a Polar heart rate monitor, a uniaxial MTI, and triaxial Tritrac accelerometer, plus a Yamax pedometer for ≥ 600 min/day. They also completed a daily physical activity log and on day 8 a Chinese version of the 7-day International Physical Activity Questionnaire. At each level of activity, there was good agreement between the two questionnaire-derived instruments and the two accelerometry-derived instruments, but wide variation across different instruments, with two- to fourfold differences in mean durations often seen. The heart rate monitor overestimated light activity and underestimated moderate activity compared with all other measures. Spearman correlation coefficients were low to moderate (0.2–0.5) across most measures of activity, with the pedometer showing correlations with total activity that were often superior to the other movement sensors. We conclude that, with the use of commonly accepted cut points for defining light, moderate, vigorous, and total activity, little convergent validity across the instruments was evident, suggesting these measures are sampling different levels of habitual physical activity and care is needed when comparing their results. To provide a more stable comparison of activity among different people, across studies, or against accepted physical activity promotion guidelines, further work is needed to fine tune the different cut points across a range of common activity monitors to provide more consistent results during free-living conditions.

accelerometer; International Physical Activity Questionnaire; log-book; heart rate; pedometer

A SEDENTARY LIFESTYLE IS A recognized risk factor for cardiovascular disease, with sedentary individuals having twice the risk of becoming overweight or obese (37). The amount of habitual physical activity accrued by an individual is also closely associated with all-cause mortality risk (6), yet the majority of people in many countries do not accumulate sufficient exercise to derive health-related benefits (8). Current interpretations of the Centers for Disease Control and Prevention-American College of Sports Medicine guidelines suggest the accumulation of 150 min/wk of moderate exercise to maintain our health (8), but the addition of vigorous-intensity exercise further enhances aerobic power (maximum O_2 uptake), which can decrease both cardiovascular risk (19) and all-cause mortality risk (6). Therefore, being able to accurately quantify not only total activity, but ideally the amount of light, moderate, and vigorous habitual physical activity accrued during daily life, will allow researchers to determine the relative

importance of these independent variables in promoting health and longevity and aid health professionals in designing appropriate exercise prescriptions.

A variety of methods exist to quantify levels of habitual physical activity during daily life, including objective measures such as heart rate, one- and three-dimensional accelerometry, and pedometry (5, 18), as well as subjective recall questionnaires like the International Physical Activity Questionnaire (IPAQ) (8) and physical activity logbooks (PA-log) (2). Yet all possess some important limitations. Heart rate monitors (HRMs) have been widely used to quantify physiological stress, but their efficacy at low intensities has been questioned due to the potential interference of environmental conditions and emotional stress (4, 14, 18). A wide range of self-report activity questionnaires exist that are well suited to large surveillance studies (28) but are limited due to their reliance on subjective recall (18, 20, 30). PA-logs have the benefit of relying only on a 24-h recall, but they do not always provide detailed information on how the participants accumulated their bouts of activity and can take considerable time to process (2, 18). Pedometers are an inexpensive form of body motion sensor, yet many fail to measure slow walking speeds or upper body movements, and most are unable to log data to determine changes in exercise intensity (14, 18, 33). The most common accelerometers used in human activity research measure accelerations either in a vertical plane (uniaxial), or in three planes (triaxial), with excellent data-logging abilities, but they typically do not measure nonambulatory or upper body movements (5, 18, 30) and can show considerable variation in counts per minute, even at similar intensities (26). Neither pedometers, accelerometers, nor HRMs provide any information on the types of physical activities performed.

Each of these devices is often used in isolation to categorize light, moderate, and vigorous levels of habitual physical activity, but to our best knowledge no published study has simultaneously examined whether a range of six such instruments possess convergent validity. Convergent validity is examined by having several different instruments measure the same construct, with a high degree of agreement, or concordance between instruments, indicating good convergent validity. In this study, we chose to examine the convergent validity from four objective instruments (a one-dimensional and a three-dimensional accelerometer, a HRM, and a pedometer), plus two subjective instruments (a 7-day physical activity recall questionnaire and a daily PA-log). Some studies have examined the associations 1) between several of these instruments in controlled laboratory environments (10, 17, 22); 2) in adults during free-living conditions but using a limited range of instruments (2, 4, 25, 30, 33, 34); or 3) have used many

Address for reprint requests and other correspondence: D. Macfarlane, Institute of Human Performance, The Univ. of Hong Kong, 111–113 Pokfulam Road, Pokfulam, Hong Kong (e-mail: DJMAC@HKU.HK).

The costs of publication of this article were defrayed in part by the payment of page charges. The article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

instruments, but on small, homogeneous single-gender groups (20, 21). The overall findings are somewhat equivocal, although better agreement was generally seen between objective measures compared with subjective measures. However, to our knowledge, none has included such a large range of both objective and subjective measures over a full week of normal lifestyle activities. The primary purpose of this study was, therefore, to examine the agreement (convergent validity) between six commonly available measures of habitual physical activity to determine the duration of light, moderate, vigorous, and total activity in a wide range of free-living individuals, thereby meeting a recently espoused research need (5, 17, 32). Examining convergent validity across measures was considered the most appropriate technique, as several reviews, including the Surgeon General's Report, state that no single suitable "gold standard" criterion measure exists for physical activity comparisons (18, 36, 39). We hypothesized that there would be moderate-to-high convergent validity within the four objective instruments and within the two subjective instruments, but low-to-moderate validity between the objective and subjective instruments.

METHODS

Subjects

A convenience sample of 57 apparently healthy, native Chinese speakers were recruited from a large city in China (Hong Kong), with a wide age range (15–55 yr) and of mixed genders (36 men, 21 women). After the study gained approval from the Institute's Ethics Committee, the experimental protocol was explained, and written consent was received from all subjects. Strict quality requirements dictated that the data from eight subjects were not included in the final analysis, leaving a sample of 49 subjects (30 men, 19 women; see Table 1). For each of 7 consecutive days, every subject was required to wear all activity monitors for ≥ 600 min/day during waking hours (except when exposed to water) and to complete a daily PA-log, plus a 7-day physical activity recall questionnaire on day 8. All subjects

Table 1. Subjects' anthropometric data, the average daily pedometer scores, and percentage of time spent in light, moderate, and vigorous activity for the HRM, Tritrac, and MTI

	Combined Total	Male	Female
<i>n</i>	49	30	19
Age, yr	28.6 (9.0)	25.6 (8.5)	33.5 (7.7)
Height, m	1.67 (0.01)	1.71 (0.01)	1.60 (0.01)
Mass, kg	59.9 (8.3)	64.2 (6.7)	53.1 (5.7)
BMI, kg/m ²	21.4 (2.5)	21.9 (2.6)	21.9 (2.4)
Pedometer, steps/day	9,839 (3,088)	9,399 (2,642)	10,534 (3,654)
HRM, %time			
Light	35.7 (13.7)	35.2 (13.1)	36.4 (14.9)
Moderate	3.5 (2.8)	3.2 (2.4)	4.1 (3.2)
Vigorous	1.0 (1.3)	1.1 (1.6)	0.8 (0.8)
Tritrac, %time			
Light	7.2 (2.4)	7.3 (2.6)	7.1 (2.0)
Moderate	5.5 (3.0)	6.0 (3.3)	4.7 (2.5)
Vigorous	0.5 (0.5)	0.4 (0.4)	0.5 (0.7)
MTI, %time			
Light	8.0 (2.2)	7.7 (2.3)	8.4 (2.0)
Moderate	5.3 (3.0)	5.9 (3.2)	4.3 (2.6)
Vigorous	0.6 (0.8)	0.7 (0.9)	0.4 (0.5)

Values are means (SD); *n*, no. of subjects. BMI, body mass index; HRM, heart rate monitor; Tritrac, Tritrac accelerometer; MTI, MTI accelerometer.

were asked to engage in their normal daily habits during the measurement period.

Physical Activity Assessment

HRM. The HRM (Team system, Polar Oy, Kempele, Finland), consisted of an elasticized chest belt that detected and stored the subject's mean heart rate every 5 s, up to a maximum of ~ 11.5 h continuous recording. Each subject used a new chest belt every day, and the heart rate data were downloaded using a Polar proprietary interface and software to a computer for storage. The 5-s epoch HRM data were screened for nonphysiological values (≥ 215 or ≤ 45 beats/min), with any of these aberrant values being replaced by the average of the adjacent pre- and postaberrant value (40), and no individual file was used if the total aberrant data exceeded 3% of the file. The 5-s data were then averaged into 1-min epochs and processed using custom-made Excel Visual Basic Macros to identify the time spent in three activity levels based on published heart rate range (HRR) cut points (16) using $220 - \text{age (yr)}$ to estimate maximum heart rate and the mean of the lowest five daily heart rates to estimate resting heart rate: light activity (20–39.9% HRR), moderate activity (40–59.9% HRR), and vigorous activity ($\geq 60\%$ HRR).

Uniaxial accelerometer (MTI). The MTI accelerometer (model 7164, MTI Actigraph, Fort Walton Beach, FL) was initialized with a time stamp, a 1-min epoch chosen for data storage, and carefully secured in the correct orientation in a small pouch that was worn firmly around the waist on the right side in line with the subject's midaxilla. The MTI data were downloaded and stored on a computer using a proprietary interface and software before being processed using custom-made Excel Visual Basic Macros to identify the time spent in three activity levels based on published cut points (13): light activity [2–2.99 metabolic equivalents (METs) = 693–1,951 counts/min], moderate activity (3–5.99 METs = 1,952–5,724 counts/min), and vigorous activity (≥ 6 METs = $\geq 5,725$ counts/min). Although various studies have used a minimum cut point of zero for light activity (24), we, like some (25, 34), used a higher cut point (693 counts/min = 2 METs), to exclude "very light" activity and to be consistent with the Tritrac, PA-log, and HRM analyses.

Triaxial accelerometer (Tritrac). The Tritrac accelerometer (model RT3, Stayhealthy, Monrovia, CA) was also initialized with a time stamp, a 1-min epoch chosen for data storage, and carefully secured in the same small pouch with the MTI accelerometer. The Tritrac data were downloaded and stored on a computer using a proprietary interface and software. The dependent variable was vector magnitude, which was the square root of the sum of the squared accelerations of all three axes. The vector magnitude data were processed using custom-made Excel Visual Basic Macros to identify the time spent in three activity levels based on published cut points (27): light activity (650–1,210 counts/min), moderate activity (1,211–2,893 counts/min), and vigorous activity ($\geq 2,894$ counts/min).

Pedometer. A pedometer (Yamax SW-700, New-Lifestyles, Lee's Summit, MO) was attached to each subject's belt immediately adjacent (anterior or posterior) to the pouch containing the two accelerometers. Subjects were shown how to open, record, and reset the number of steps aggregated at the end of each day into a logbook.

PA-log. At the end of each day, each subject completed one page of a seven-page PA-log, recording all activities ≥ 10 min that were grouped into home, occupation, sitting, moderate leisure, vigorous leisure, transportation, plus "others", based on a previous format (2). This required the subjects to circle each activity they participated in, then to estimate the total duration of each activity, plus record the time they began each activity. The logs required minimal literacy and could be completed in < 5 min. The completed logs were collected, and each reported activity was later scored using MET values taken from a Compendium of Physical Activities (3). For each day, the total minutes of activity were aggregated by intensity level into sitting, light (2–2.99 METs), moderate (3–5.99 METs), and vigorous (≥ 6

METs) activity. Finally, the weekly total duration spent at each intensity level was generated from the seven completed daily logs.

IPAQ-Chinese version. The IPAQ-C is a Chinese version of the short, last 7-day interview format questionnaire (8), available in English (and other languages) at www.ipaq.ki.se. It required the subjects to complete seven questions on the frequency and duration of time spent in four domains: walking, in moderate- and vigorous-intensity activity, plus in sedentary behaviors (sitting and lying awake). Initially, the IPAQ-C was independently translated from English by two bilingual experimenters familiar with questionnaires, and then it was 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 subject completed the interview-administered IPAQ-C on day 8 so that its 7-day recall period coincided with the same 7 days of objective data collection and the seven daily PA logs. The IPAQ-C data were presented as the total minutes reported for walking (shown here as light activity, 3.3 METs), moderate (4 METs), and vigorous (8 METs) 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-day requirement of Craig et al. (8), our subjects were required to obtain data on all 7 days, but the similar ≥ 600 min/day of registered time were required before accelerometry (and HRM) analysis. Analysis was only performed if a complete 7-day set of data was available for all measures on all subjects, and this resulted in all data from 8 of the original 57 subjects being deleted.

Our data processing was similar to other published studies that have used these same instruments (2, 8, 13, 15), yet this involved some inconsistency in categorizing intensities across instruments. For example, walking (3.3 METs) was considered a separate and distinct activity from moderate activities (≥ 4 METs) in the IPAQ (8), yet it was classified as moderate activity (3–5.99 METs) by the PA-log (2). For this reason, we have reported IPAQ walking both individually as light activity, and like Ainsworth et al. (2) we included it in moderate IPAQ exercise to permit comparability with the moderate PA-log data. Similar variations occurred, with vigorous activity being defined as ≥ 6 METs by the PA-log (2) but ≥ 8 METs by IPAQ (8). As in the study of Hallal et al. (15), the total IPAQ activity scores (min) were calculated as the weighted sum of moderate activity (including walking), plus twice the minutes of vigorous activity (to reflect the weightings of 4 METs and 8 METs for moderate and vigorous activity, respectively). For comparability, this total volume of health-enhancing physical activity (HEPA-total) was also calculated for the MTI, Tritrac, HRM, and PA-log data (min/wk) as moderate duration plus twice the vigorous duration.

Inspection of our physical activity data confirmed it was not normally distributed; thus Friedman nonparametric tests were used to simultaneously determine whether significant differences existed between the measures. When significance was established, follow-up Wilcoxon nonparametric sign-ranked 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. Nonparametric Spearman correlations were used to examine the associations between data from pairs of measures. Statistical analyses were performed using SPSS 11.0, with data shown for means \pm SD, unless stated.

RESULTS

Descriptive Results

Descriptive characteristics of the subjects are shown in Table 1, together with the average number of steps per day accrued via pedometry and the percentage of time spent in

light, moderate, and vigorous activity measured objectively. Subjects wore the monitors on average 692 ± 58 min (HRM), 840 ± 74 min (Tritrac), and 837 ± 70 min (MTI), which represents 72, 88, and 87% of a 16-h waking day. The lower data collection period for the HRM was due to its 11.5-h maximum data logging, yet it represented 83% of the average duration measured by the Tritrac and MTI. Inspection of the synchronized HRM and MTI records showed that most recordings began around 8:00–8:30 AM, and hence the HRM terminated around 8:00 PM. After 8:00 PM, the MTI records typically showed very little activity, especially health-enhancing moderate or vigorous activity. Consequently, the 11.5-h HRM data captured virtually all of the important daily activity, allowing a fair comparison between the durations of time (min/wk) spent in light, moderate, and vigorous with other measures, although a longer HRM capture period would have been ideal. The relative percentage of time spent in each activity level for these three monitors is also shown in Table 1, with the percentage of time spent in moderate (3–6%) and vigorous (0.5–1.0%) activity being comparable across all instruments, yet light activity determined from the HRM was much higher (36%) compared with the MTI and Tritrac (7–8%).

Activity durations. Comparisons of the mean \pm SD duration of time (min/wk) for the five methods capable of discerning light, moderate, vigorous, and HEPA-total activity are shown in Table 2, along with the 25th, 50th (median), and 75th percentiles due to the nonparametric nature of the data, plus the *P* values from Wilcoxon paired comparisons. Knowledge of the 25th and 75th percentiles can be helpful in defining intensity cut points (24).

The HRM significantly overestimated the mean time spent in light activity (1,730 min) compared with all other measures, while the PA-log and IPAQ-C showed similar estimates (978 and 708 min, respectively), and neither differed significantly from the lower estimates by the Tritrac and MTI (which were also similar at 426 and 471 min, respectively). In direct contrast, the HRM produced significantly lower estimates of mean moderate activity (173 min) compared with all other measures, while the PA-log and IPAQ-C again were similar (952 and 854 min, respectively), but significantly higher than the comparable Tritrac and MTI estimates (311 and 302 min, respectively). Few significant differences were seen in estimates of vigorous activity, with again the Tritrac and MTI producing the lowest rank, yet very similar, mean values of 26 and 29 min, respectively. The HRM produced a middle rank of 48 min, with the PA-log and IPAQ-C (84 and 126 min, respectively) producing the highest mean estimates of vigorous activity. When HEPA-total activity was calculated (moderate duration plus twice vigorous duration), again three differing groupings resulted. The subjective recall-derived data from the PA-log (1,120 min) and IPAQ-C (1,107 min) were very similar, but they were both significantly higher than the almost identical estimates of 363 and 360 min produced by the Tritrac and MTI, respectively. The HRM total of 270 min was significantly lower than all other measures and represented $\sim 75\%$ of the objective accelerometry-derived values and 25% of the subjective recall-derived values.

Correlations between instruments. The Spearman rank-order correlation coefficients between each instrument for light, moderate, vigorous, and HEPA-total activity are found in

Table 2. Duration of time spent in each intensity range as determined by the five measurement devices, together with the 25th, 50th, and 75th percentiles, plus the significance test between pairs of measures

		Percentiles			Significance Testing (Wilcoxon <i>P</i> Values)			
	Min/wk [mean (SD)]	25th	50th	75th	IPAQ-C	HRM	Tritrac	MTI
<i>Light activity</i>								
PA-log	975.5 (1474.8)	0.0	30.0	1,782.5	0.773	0.004*	0.395	0.401
IPAQ-C‡	708.0 (557.2)	290.0	480.0	1,050.0		0.000*	0.011	0.044
HRM	1,730.1 (672.0)	1,227.0	1,654.8	2,366.8		0.000*	0.000*	
Tritrac	425.7 (144.1)	320.2	408.0	534.6		0.003*		
MTI	471.2 (133.4)	395.2	497.0	565.5				
<i>Moderate activity</i>								
PA-log	952.7 (638.3)	535.0	770.0	1,397.5	0.036	0.000*	0.000*	0.000*
IPAQ-C§	854.3 (603.2)	360.0	720.0	1,340.0		0.000*	0.000*	0.000*
HRM	172.7 (135.4)	62.5	123.0	270.2		0.000*	0.000*	
Tritrac	311.3 (158.3)	196.5	298.2	399.1		0.778		
MTI	301.8 (152.0)	164.2	343.7	408.6				
<i>Vigorous activity</i>								
PA-log	83.8 (151.4)	0.0	0.0	105.0	0.027	0.907	0.120	0.052
IPAQ-C	126.2 (207.9)	0.0	30.0	180.0		0.156	0.007	0.002*
HRM	48.4 (66.9)	3.6	19.6	58.0		0.070	0.009	
Tritrac	26.1 (30.5)	2.1	16.1	38.5		0.459		
MTI	28.9 (39.9)	1.0	20.0	28.4				
<i>HEPA-total activity†</i>								
PA-log	1,120.2 (619.6)	645.0	945.0	1,500.0	0.601	0.000*	0.000*	0.000*
IPAQ-C	1,106.7 (637.8)	630.0	900.0	1,590.0		0.000*	0.000*	0.000*
HRM	269.5 (237.9)	102.2	175.7	443.0		0.015*	0.005*	
Tritrac	363.6 (196.1)	227.5	332.0	512.0		0.960		
MTI	359.6 (200.9)	186.2	354.0	496.6				

PA-log, physical activity logbook; IPAQ-C, Chinese version of the International Physical Activity Questionnaire; HEPA-total, total health-enhancing physical activity. *Significant using Holm's sequential Bonferroni adjustment. †Moderate + (2 × vigorous); ‡walking only; §(walking + moderate activity).

Table 3, with the statistically significant correlations remaining weak to moderate ($r = 0.3$ – 0.7). Not surprisingly, similar patterns of agreement were seen to the activity durations, with the recall-based IPAQ-C and PA-log showing significant and moderate correlations ($r = 0.6$ – 0.7), except during light activity ($r = 0.25$). Similarly, the accelerometry-based Tritrac and MTI showed significant and moderate correlations ($r = 0.5$ – 0.8) across all ranges of activities. The HRM was generally poorly correlated with the other measures ($r < 0.3$), except during vigorous activity ($r = 0.4$ – 0.6) and with the MTI during HEPA-total activity ($r = 0.7$). When comparing estimates of HEPA-total activity, the only paired comparisons that showed reasonable agreement ($r > 0.5$) were the recall-derived pairs of PA-log and IPAQ-C and the accelerometry-derived pairs of Tritrac and MTI.

All Spearman correlations between the average daily pedometer scores (steps/day) and the HEPA-total activity (min) from the other measures were statistically significant (Table 3). Moderately weak correlations were found between the pedometer scores and the IPAQ-C, the PA-log, and HRM, while quite strong correlations were found between the pedometer and both the Tritrac and the MTI.

DISCUSSION

The uniqueness of this study is that it not only simultaneously compared six different measures (pedometer, PA-log, IPAQ, HRM, Tritrac, and MTI) that are often used to estimate levels of habitual physical activity but that it also acquired a high-quality data set from a reasonable number ($n = 49$) of

diverse and free-living individuals, each over 7 consecutive days with >8 h/day of successful monitoring. However, the participants comprised a convenience sample and were not necessarily representative of the normal Hong Kong Chinese population. Many were health professionals (nurses, physical or occupational therapists, and trainers), although some were high school students and others were sedentary office workers. Owing to the high proportion of subjects being active health professionals, it was neither surprising that the mean number of steps/day recorded by the pedometer in our study approached nearly 10,000, nor that only 24% were classified overweight and none obese (41).

Although converting raw heart rate, pedometer, and accelerometer values into units of energy expenditure has some benefits and is frequently performed (17, 20, 21), this approach has also been questioned (12, 20), especially as adequate conversion equations are not available for all ages, and most have not been validated using free-living activities (26). Since the conversion into energy expenditure adds a variable error that does not exist in the raw data (12, 14), we have followed recommendations (26) by analyzing our data using raw units and reporting the times accrued above specific intensity cut points (2, 20). Examining the time accrued above intensity cut points also allows direct comparisons with common recommendations on physical activity. The disadvantage of this approach is that it requires intensity cut points that are comparable across different instruments, and, until adequate guidelines are developed, comparisons across instruments and between studies will be difficult (23). Existing data suggest

Table 3. Spearman correlations between the six measures of physical activity

	Spearman Correlations (<i>P</i> Values)				
	IPAQ-C	HRM	Tritrac	MTI	Pedometer
<i>Light activity</i>					
PA-log	0.251 (0.082)	0.062 (0.670)	0.217 (0.134)	0.213 (0.143)	
IPAQ-C ^d		0.238 (0.099)	0.345 (0.015 ^a)	0.371 (0.009 ^b)	
HRM			0.099 (0.496)	0.142 (0.330)	
Tritrac				0.586 (0.000 ^b)	
<i>Moderate activity</i>					
PA-log	0.679 (0.000 ^b)	−0.260 (0.071)	0.322 (0.024 ^a)	0.095 (0.515)	
IPAQ-C ^e		−0.108 (0.460)	0.149 (0.307)	−0.059 (0.687)	
HRM			0.143 (0.325)	0.236 (0.102)	
Tritrac				0.772 (0.000 ^b)	
<i>Vigorous activity</i>					
PA-log	0.722 (0.000 ^b)	0.605 (0.000 ^b)	0.215 (0.138)	0.454 (0.001 ^b)	
IPAQ-C		0.555 (0.000 ^b)	0.176 (0.226)	0.441 (0.002 ^b)	
HRM			0.358 (0.012 ^a)	0.634 (0.000 ^b)	
Tritrac				0.503 (0.000 ^b)	
<i>HEPA-total activity^c</i>					
PA-log	0.596 (0.000 ^b)	−0.017 (0.910)	0.317 (0.026 ^a)	0.176 (0.226)	0.371 (0.009 ^b)
IPAQ-C		0.241 (0.096)	0.166 (0.254)	0.091 (0.532)	0.300 (0.036 ^a)
HRM			0.274 (0.057)	0.462 (0.001 ^b)	0.371 (0.009 ^b)
Tritrac				0.728 (0.000 ^b)	0.668 (0.000 ^b)
MTI					0.699 (0.000 ^b)

P values are in parentheses: ^a*P* < 0.05; ^b*P* < 0.01. ^cModerate + (2 × vigorous); ^dwalking only; ^e(walking + moderate activity).

concordance between these intensity cut points can be wide (24, 39), leading to requests for further studies examining the agreement between different monitors during free-living conditions (17, 32, 39), using raw units rather than those of energy expenditure (12, 20), and to which we feel this paper contributes significantly.

It was not unexpected that quantifying light physical activity would create the greatest variation between the five measures, with over a fourfold difference between both the mean and median times determined by Tritrac and HRM. That the HRM produced values that were statistically much higher than all other measures suggests the current HRM cut points for light-intensity activities might be too generous, although it is also possible that periods of emotional excitement may have supplemented this category, especially with younger subjects (4). The higher duration of light activity measured by the HRM may also have included significantly more upper body movements that are often missed by waist-mounted accelerometers (17). However, during moderate-intensity exercise, the 40–59% HRR cut points from the HRM produced a mean duration that was significantly much smaller and represented only 20–60% of the mean durations from all other instruments. Although it may appear these results suggest that consideration could be given to raising the lower HRR threshold for light activity from the current 20% HRR (16), to the previous 25% HRR (36), and lowering the 40% HRR threshold for moderate activity, a considerable amount of supportive evidence would first be needed. It is possible that the subjects' level of aerobic fitness, which was not controlled for in this study, may have led to greater variation in the HRM scores, since a more fit person may have a lower heart rate at a given work rate, even though the accelerometry scores may not differ. The discrepancies between these different instruments are also shown by the percentage of the cohort meeting the guidelines of accumulat-

ing 150 min/wk of moderate activity: HRM 39%, MTI 76%, Tritrac 82%, PA-log 92%, and IPAQ-C 94%. Although comparable inconsistencies in estimating the duration of moderate activity using three similar methods have been reported (2), others have shown much better agreement (20, 31), reflecting a large variability between studies in estimating these types of activity in free-living subjects (30).

The typically large variances seen in all measures of vigorous activity (our SD values always exceeding the mean values, Table 2) may have contributed to lack of significant differences in mean scores, even though they often differed by two- to threefold. Although the recall-derived measures of vigorous activity had the highest mean values, they often failed to reach a significant level of overreporting that has been documented (31). The HEPA-total activity again showed good consistency within each type of instrument, but not across different instruments, with the recall-derived scores consistently higher than the accelerometry-derived scores, and the HRM producing the lowest scores. It appears our subjects may have consistently overreported all levels of activity, which differs from the overreporting of vigorous activity and underreporting of light activity by similar recall instruments compared with accelerometry data (31). While several studies have data suggesting overreporting of activity can occur for the IPAQ (7, 11, 29), others have documented few differences between a common 7-day recall questionnaire and the time accrued in light, moderate, and vigorous activity via MTI and Tritrac (20). Yet when Leenders et al. (21) converted their data to units of energy expenditure, both the MTI and Tritrac underestimated the 7-day recall values, further compounding the problems associated with converting raw units to those of energy expenditure. Since no gold standard criterion method exists to record the time spent at various intensities (18, 30, 36), we can, therefore, only conclude that the questionnaires, accelerome-

ters, and HRMs (and their associated intensity cut points) used in this study do not precisely measure the same activity patterns and lack convergent validity.

The correlation coefficients between each measure showed considerable variation, but were typically modest in size at best, and not too dissimilar to other studies (2, 28, 30, 31), but not as strong as some (20). Of note were the relatively strong correlations between the PA-log and IPAQ-C, indicating that the IPAQ-C is an acceptably valid subjective measure of activity in ethnic Chinese, although its objective validity was less impressive. The correlation coefficients between the average daily pedometer scores and the HEPA-total activity estimated by all other instruments were often as good, and frequently larger, than the other between-instrument coefficients. This finding agrees with previous studies (5, 20) that a simple and inexpensive pedometer can provide an equally respectable estimate of total physical activity as other more expensive movement sensors and supports the contention that these pedometers are well suited to large cross-sectional or interventional studies.

Our results show that the IPAQ-C estimates of time accrued in light, moderate, vigorous, or HEPA-total activity were not significantly different from those from the PA-log. The consistency between two different questionnaires involving daily (PA-log) and weekly (IPAQ-C) recalls might suggest that overreporting was less likely to have occurred, but it does not remove this possibility, as activity self-reports are frequently higher than objective measures (2, 9). Our subjects may have overreported all activities for both the IPAQ-C and PA-log by failing to report only activities with durations that exceeded 10 min (29). The high levels of inactivity in Hong Kong population (1) may also have contributed to recall overreporting, since their inactivity could lower fitness levels and lead to raising their perceived current activity level (2, 31). Thus our subjects' interpretation of exercise intensity may not have been consistent with our objective definitions of activity thresholds and, therefore, reflects the difficulties of applying the same rigid intensity cut points to a sample varying in age, gender, and habitual physical activity level (9, 30).

The relative consistency between the Tritrac and MIT supports recent reviews (12, 32) that these accelerometers provide comparable information on physical activity patterns, yet the HRM often failed to produce values that were consistent with the other instruments. Such inconsistencies in the objective measurement of physical activity are quite possibly due to inequalities in the intensity cut points between instruments. Until researchers delineate core activities and intensities that can be used to produce comparable results across different monitoring devices (24), researchers risk making conclusions about the activity status of sample populations that could be biased and quite method specific. The lack of interinstrument agreement in our study also questions the efficacy of applying intensity cut points derived from Occidental populations to an Oriental sample. It seems prudent that population-specific cut points may be justified in order to avoid the risk of drawing conclusions based on inappropriate thresholds (18, 30, 38). The large number of significant differences seen between measures in Table 2, together with the relatively weak correlations seen in Table 3, indicates an overall lack of convergent validity across all measures (with a possible exception for the pedometer compared with other HEPA-total activity scores). This

result not only highlights the risks associated with drawing conclusions based on only one measure of physical activity, but also it makes it difficult to ascertain which measure or measures are in error, especially as no internationally accepted criterion or gold standard exists for the measurement of habitual physical activity. It has been suggested that the Intelligent Device for Energy Expenditure and Activity monitor may have the potential to serve as a criterion measure (39). The Intelligent Device for Energy Expenditure and Activity monitor is a portable device that uses five sensors and complex algorithms to determine dynamically the type, duration, intensity, and energy expenditure of the current activity with a high degree of accuracy (42). However, further work with this type of technology is still required before it is widely accepted.

A number of factors may contribute to the lack of convergent validity in the measurement of habitual physical activity, including several limitations in the present study. Our study involved a relatively small convenience sample of Chinese, with a limited body mass index range, who were generally quite active. This lack of sample heterogeneity not only limits the applicability of the results to wider populations but it also may have reduced the agreement seen between measures. Inconsistency in the definition of what constitutes light, moderate, and vigorous activity (even when using METs) among different measurement devices will also decrease the concordance between measures, as will the fact that all of these devices monitor slightly different aspects of activity (18, 35). For example, pedometers do not quantify the magnitude of the vertical movement captured by the one-dimensional MTI, yet the Tritrac captures three-dimensional movements, but only the HRM can measure the net cardiovascular (and emotional) stress of the activity. While the indirect subjective IPAQ-C and PA-log measures rely heavily on recall and would be expected to show lower concordance with the objective measures (35), the lack of an internationally recognized gold standard measure of habitual physical activity also limits such studies to examining convergent, rather than criterion, validity.

In summary, these findings show that there is generally an acceptable consistency in the times accrued at low, moderate, vigorous, and HEPA-total activity thresholds within similar types of activity monitor (questionnaire derived; accelerometry derived), but there is poor agreement (convergent validity) across the different types of monitor (PA-log, IPAQ-C, HRM, MTI, and Tritrac). As an estimate of HEPA-total activity for larger studies, the standard pedometer appears no less valid than more expensive motion sensors. To provide a stable template to compare activity profiles among different people, across studies, or against common activity promotion guidelines, further work is needed to fine-tune the cut points across a range of common activity monitors to ensure that they provide more consistent estimates of physical activity during free-living conditions.

ACKNOWLEDGMENTS

Technical assistance was provided by the generous loan of several MTI accelerometers from Dr. Michael Sjostrom, Karolinska Institute, Sweden.

GRANTS

No external financial assistance was provided for this study.

REFERENCES

- Adab P and Macfarlane DJ. Exercise and health—new imperatives for public health policy in Hong Kong. *HK Med J* 4: 389–393, 1998.
- Ainsworth BE, Bassett DR Jr, Strath SJ, Swartz AM, O'Brien WL, Thompson RW, Jones DA, Macera CA, and Kimsey CD. Comparison of three methods for measuring the time spent in physical activity. *Med Sci Sports Exerc* 32: S457–S464, 2000.
- Ainsworth BE, Haskell WL, Whitt MC, Irwin ML, Swartz AM, Strath SJ, O'Brien WL, Bassett DR Jr, Schmitz KH, Emplaincourt PO, Jacobs DR Jr, and Leon AS. Compendium of physical activities: an update of activity codes and MET intensities. *Med Sci Sports Exerc* 32: S498–S504, 2000.
- Allor KM and Pivarnik JM. Stability and convergent validity of three physical activity assessments. *Med Sci Sports Exerc* 33: 671–676, 2001.
- Bassett DR Jr. Validity and reliability issues in objective monitoring of physical activity. *Res Q Exerc Sport* 71: S30–S36, 2000.
- Blair SN, Cheng Y, and Holder S. Is physical activity or physical fitness more important in defining health benefits? *Med Sci Sports Exerc* 33: S379–S399, 2001.
- Brown W, Bauman A, Chey T, Trost S, and Mummery K. Comparison of surveys used to measure physical activity. *Aust NZ J Public Health* 28: 128–134, 2004.
- Craig CL, Marshall AL, Sjostrom M, Bauman AE, Booth ML, Ainsworth BE, Pratt M, Ekelund U, Yngve A, Sallis JF, and Oja P. International physical activity questionnaire: 12-country reliability and validity. *Med Sci Sports Exerc* 35: 1381–1395, 2003.
- Duncan TE and McAuley EM. Social support and efficacy cognitions in exercise adherence: a latent growth curve analysis. *J Behav Med* 16: 199–218, 1993.
- Eston RG, Rowlands AV, and Ingledew DK. Validity of heart rate, pedometer, and accelerometry for predicting the energy cost of children's activities. *J Appl Physiol* 84: 362–371, 1998.
- Fogelholm M, Malmberg J, Suni J, Santtila M, Kyrolainen H, Manttysaari M, and Oja P. International Physical Activity Questionnaire: validity against fitness. *Med Sci Sports Exerc* 38: 753–760, 2006.
- Freedson P, Pober D, and Janz KF. Calibration of accelerometer output for children. *Med Sci Sports Exerc* 37: S523–S530, 2005.
- Freedson PS, Melanson E, and Sirard J. Calibration of the Computer Science and Applications, Inc. accelerometer. *Med Sci Sports Exerc* 30: 777–781, 1998.
- Freedson PS and Miller K. Objective monitoring of physical activity using motion sensors and heart rate. *Res Q Exerc Sport* 71: S21–S29, 2000.
- Hallal PC, Victora CG, Wells JC, and Lima RC. Physical inactivity: prevalence and associated variables in Brazilian adults. *Med Sci Sports Exerc* 35: 1894–1900, 2003.
- Howley ET. Type of activity: resistance, aerobic and leisure versus occupational physical activity. *Med Sci Sports Exerc* 33: S364–S369, 2001.
- King GA, Torres N, Potter C, Brooks TJ, and Coleman KJ. Comparison of activity monitors to estimate energy cost of treadmill exercise. *Med Sci Sports Exerc* 36: 1244–1251, 2004.
- LaMonte MJ, Ainsworth BE, and Tudor-Locke C. Assessment of physical activity and energy expenditure. In: *Obesity: Etiology, Assessment, Treatment and Prevention*, edited by Andersen RE. Champaign, IL: Human Kinetics, 2003, p. 111–137.
- Lee IM, Hsieh CC, and Paffenbarger RS Jr. Exercise intensity and longevity in men. The Harvard Alumni Health Study. *JAMA* 273: 1179–1184, 1995.
- Leenders N, Sherman WM, and Nagaraja HN. Comparisons of four methods of estimating physical activity in adult women. *Med Sci Sports Exerc* 32: 1320–1326, 2000.
- Leenders NY, Sherman WM, Nagaraja HN, and Kien CL. Evaluation of methods to assess physical activity in free-living conditions. *Med Sci Sports Exerc* 33: 1233–1240, 2001.
- Louie L, Eston RG, Rowlands AV, Tong KK, Ingledew DK, and Fu FH. Validity of heart rate, pedometer, and accelerometry for estimating the energy cost of activity in Hong Kong Chinese boys. *Pediatr Exerc Sci* 11: 229–239, 1999.
- Masse LC, Fuemmeler BF, Anderson CB, Matthews CE, Trost SG, Catellier DJ, and Treuth M. Accelerometer data reduction: a comparison of four reduction algorithms on select outcome variables. *Med Sci Sports Exerc* 37: S544–S554, 2005.
- Matthews CE. Calibration of accelerometer output for adults. *Med Sci Sports Exerc* 37: S512–S522, 2005.
- Matthews CE, Ainsworth BE, Hanby C, Pate RR, Addy C, Freedson PS, Jones DA, and Macera CA. Development and testing of a short physical activity recall questionnaire. *Med Sci Sports Exerc* 37: 986–994, 2005.
- McMurray RG, Baggett CD, Harrell JS, Pennell ML, and Bangdiwala SI. Feasibility of the Tritrac R3D accelerometer to estimate energy expenditure in youth. *Pediatr Exerc Sci* 16: 219–230, 2004.
- Nichols JF, Morgan CG, Sarkin JA, Sallis JF, and Calfas KJ. Validity, reliability, and calibration of the Tritrac accelerometer as a measure of physical activity. *Med Sci Sports Exerc* 31: 908–912, 1999.
- Pereira MA, FitzGerald SJ, Gregg EW, Joswiak ML, Ryan WJ, Suminski RR, Utter AC, and Zmuda JM. A collection of Physical Activity Questionnaires for health-related research. *Med Sci Sports Exerc* 29: S1–S205, 1997.
- Rzewnicki R, Vanden Auweele Y, and De Bourdeaudhuij I. Addressing overreporting on the International Physical Activity Questionnaire (IPAQ) telephone survey with a population sample. *Public Health Nutr* 6: 299–305, 2003.
- Schmidt MD, Freedson PS, and Chasan-Taber L. Estimating physical activity using the CSA accelerometer and a physical activity log. *Med Sci Sports Exerc* 35: 1605–1611, 2003.
- Sirard JR, Melanson EL, Li L, and Freedson PS. Field evaluation of the Computer Science and Applications, Inc. physical activity monitor. *Med Sci Sports Exerc* 32: 695–700, 2000.
- Trost SG, McIver KL, and Pate RR. Conducting accelerometer-based activity assessments in field-based research. *Med Sci Sports Exerc* 37: S531–S543, 2005.
- Tudor-Locke C, Williams JE, Reis JP, and Pluto D. Utility of pedometers for assessing physical activity: construct validity. *Sports Med* 34: 281–291, 2004.
- Tudor-Locke C, Ainsworth BE, Thompson RW, and Matthews CE. Comparison of pedometer and accelerometer measures of free-living physical activity. *Med Sci Sports Exerc* 34: 2045–2051, 2002.
- Tudor-Locke C, Williams JE, Reis JP, and Pluto D. Utility of pedometers for assessing physical activity: convergent validity. *Sports Med* 32: 795–808, 2002.
- US Department of Health and Human Services. *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, 1996.
- US Department of Health and Human Services. *The Surgeon General's Call to Action to Prevent and Decrease Overweight and Obesity*. Rockville, MD: US Department of Health and Human Services, Public Health Services, Office of the Surgeon General, 2001, p. 8–15.
- Ward DS, Evenson KR, Vaughn A, Rodgers AB, and Troiano RP. Accelerometer use in physical activity: best practices and research recommendations. *Med Sci Sports Exerc* 37: S582–S588, 2005.
- Welk GJ. Principles of design and analyses for the calibration of accelerometer-based activity monitors. *Med Sci Sports Exerc* 37: S501–S511, 2005.
- Welk GJ and Corbin CB. The validity of the Tritrac-R3D Activity Monitor for the assessment of physical activity in children. *Res Q Exerc Sport* 66: 202–209, 1995.
- WHO Expert Consultation. Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies. *Lancet* 363: 157–163, 2004.
- Zhang K, Pi-Sunyer FX, and Boozer CN. Improving energy expenditure estimation for physical activity. *Med Sci Sports Exerc* 36: 883–889, 2004.