Journal of Physical Activity and Health, 2009, 6, 790-804

Global Physical Activity Questionnaire (GPAQ): Nine Country Reliability and Validity Study

Fiona C. Bull, Tahlia S. Maslin, and Timothy Armstrong

Purpose: Instruments to assess physical activity are needed for (inter)national surveillance systems and comparison. ***Methods:*** Male and female adults were recruited from diverse sociocultural, educational and economic backgrounds in 9 countries (total n = 2657). GPAQ and the International Physical Activity Questionnaire (IPAQ) were administered on at least 2 occasions. Eight countries assessed criterion validity using an objective measure (pedometer or accelerometer) over 7 days. ***Results:*** Reliability coefficients were of moderate to substantial strength (Kappa 0.67 to 0.73; Spearman's rho 0.67 to 0.81). Results on concurrent validity between IPAQ and GPAQ also showed a moderate to strong positive relationship (range 0.45 to 0.65). Results on criterion validity were in the poor-fair (range 0.06 to 0.35). There were some observed differences between sex, education, BMI and urban/rural and between countries. ***Conclusions:*** Overall GPAQ provides reproducible data and showed a moderate-strong positive correlation with IPAQ, a previously validated and accepted measure of physical activity. Validation of GPAQ produced poor results although the magnitude was similar to the range reported in other studies. Overall, these results indicate that GPAQ is a suitable and acceptable instrument for monitoring physical activity in population health surveillance systems, although further replication of this work in other countries is warranted.

Keywords: measurement, surveillance, epidemiology, assessment, public health, instrument psychometrics

Bull is with the School of Sport and Exercise Sciences, University of Loughborough, Perth, Australia. Maslin is with the BHF Centre for Physical Activity and Health, Loughborough University, United Kingdom. Armstrong is with the CHP, WHO, Geneva, Switzerland.

Regular physical activity is well recognized as an important lifestyle behavior for the development and maintenance of individual and population health and well-being.[1,2] Globally, high levels of inactivity contribute substantially to the global burden of disease, with heavy resulting economic costs.[3,4] Monitoring population levels of physical activity using a standardized protocol is a core part of a public health response to current concerns regarding levels of physical inactivity and obesity.[5,6]

Measurement of physical activity in large population groups is usually undertaken using self-reported recall, often in the form of a questionnaire conducted either by telephone or household interview. When undertaken using a consistent format over a period of time (eg, years), these data provide an understanding of trends in a population's level of physical activity. Surveys are a particularly useful instrument due to their low cost, ease of administration and adaptability.

Given the importance of physical activity as a major risk factor for non communicable disease, the poor state of measurement and lack of comparable data internationally indicates that there is an urgent need for a standardized measure of physical activity.[3,6] Such a measure needs to be reliable, valid and suitable for use in health surveillance systems. For wide spread international use, such an instrument would have to cater for cultural differences as well as specific subpopulations, including women, individuals with lower socioeconomic status and/or literacy levels ands those from a non-English speaking background.

This paper summarizes the development of the Global Physical Activity Questionnaire (GPAQ) and the methods protocols and results of an international research collaboration undertaken between 2003 to 2005 to test the measurement properties of GPAQ. Specifically, the results from studies designed to assess the test–retest reliability, concurrent validity and the criterion validity of the GPAQ instrument are reported. For concurrent validity GPAQ was compared with the internationally acceptable measure of IPAQ-short version and the selected criterion measure was an objective measure of movement from a motion monitor (either a pedometer or accelerometer).

Methods

GPAQ Instrument

The Global Physical Activity Questionnaire (GPAQ) was developed under the auspices of the World Health Organization (WHO) in 2002 as part of the WHO STEPwise Approach to Chronic Disease Risk Factor Surveillance (STEPS). The STEPS approach has been widely introduced as a feasible approach to monitoring 8 key risk factors of noncommunicable diseases, particularly in developing countries.[7] Both versions of IPAQ (long and short) were considered for use, however the lack of domain-specific estimates from IPAQ-short (9 item) and the total length of IPAQ-long (27 item) prohibited their use within STEPS. Instead a new instrument drawing on the strengths of both IPAQ instruments was developed and required testing.

GPAQ comprises 19 questions grouped to capture physical activity undertaken in different behavioral domains, these are work, transport and discretionary (also known as leisure or recreation). Within the work and discretionary domains, questions assess the frequency and duration of 2 different categories of activity defined by the energy requirement or intensity (vigorous- or moderate-intensity). In the transport domain, the frequency and duration of all walking and cycling for transport is captured but no attempt is made to differentiate between these activities. One additional item collected time spent in sedentary activities.

Reliability and Validity

Reliability is the consistency or reproducibility of a response.[8] It is most frequently assessed for measures of physical activity by the completion of an instrument on 2 separate occasions. Validity is the degree to which a test or instrument measures what it purports to measure.[8] Content (face), construct, concurrent, convergent and criterion are different elements of testing validity. For this research, concurrent validity—the relationship between 2 (similar) measures of interest, and criterion validity—the relationship between the measure (GPAQ) and an objective measure of the behavior were assessed. The IPAQ-short instrument was chosen for the purpose of ascertaining concurrent validity. IPAQ is a well developed and pretested instrument and by evaluating the strength of the correlations between comparable variables from both instruments, the relative ability of GPAQ to measure the same properties can be determined.

The use of an objective measure of physical activity, such as a pedometer or accelerometer, can help to overcome the known limitations of surveys. It is possible for people to over-report or over-estimate intensity and/or duration of activity in questionnaires due to the social desirability of responses.[9,10] The use of motion sensors can overcome these problems by providing an unbiased assessment of a pattern of physical activity level.[10] For this research, reputable devices, already established in the field were used (Yamax Digi-Walker and MTI accelerometers).[11,12]

Study Design

Recruitment of Participating Centers/Collaborators. Using contacts through WHO headquarters and regional offices and other known networks of researchers on physical activity, interest was sought to conduct GPAQ reliability and validity studies. Ten projects were initiated in 2002–03 in the following countries: Bangladesh, Brazil, China (2 sites—Shanghai and Province of Taiwan), Ethiopia, India, Indonesia, Japan, Portugal and South Africa.

Participating centers were given the option of conducting 1 of 2 protocols. The first protocol incorporated test-retest reliability, concurrent validity and criterion validity. The second protocol was designed to assess test-retest reliability and concurrent validity only. The protocols outlined the recommended time frame between visits and subsequent administrations of the questionnaires.

Ideally, data collected at the initial consultation would be used to obtain concurrent validity and the motion sensor would be introduced to participants. In addition, participants were required to read and sign institutional human subject consent forms. To capture a full week of pedometer or accelerometer data, the time period for visit 2 would be at least 7 days. To assess reliability, a third visit was recommended scheduled preferably 3 to 7 days after the second contact. This specified time frame was important because if less than 3 days there is an increased possibility that the participant may remember their answers rather than attempt to recall the actual activity under taken. On the other hand, if the time gap is more than 7 days the participant may be unable to recall accurately their physical activity levels of the correct time frame. There is no perfect solution to this methodological limitation, but a compromise is to set a recommendation of between 3 to 7 days as the interval between visits as the optimal period to assess reliability.

As some centers introduced modifications to the study design, Table 1 provides details of the specific data collection procedures for each center. Dhaka, Bangladesh; Province of Taiwan, China; Haryana, India; Yogyakarta, Indonesia and Tokyo, Japan completed protocol 1 unchanged. The centers in Shanghai, China and Cape Town, South Africa included three visits, but the initial visit only comprised the administration of the demographics questions and the commencement of the accelerometer data collection. On the 2 subsequent visits GPAQ and IPAQ were administered. Butajira, Ethiopia followed protocol 1 in a slightly different manner, where by the initial sample of 940 participants were used for

Table 1 Country Specific Study Protocols and Data Collection Procedures

	Ethiopia N = 940	Indonesia N = 337	India N = 262	South Africa N = 214	Shanghai, China N = 221	Bangladesh N = 147	Brazil N = 204	Japan N = 148	Portugal N = 67	Taiwan, China N = 141
Sampling method	Random selection from BRHP database	Random	Random	Random	Convenient	Household Stratified systematic sampling	Stratified random sampling	Convenient	Convenient	Convenient
Sample collection	Represents both sexes, various education levels and ethnic groups	360 subjects chosen from rural and urban areas	Although random selection indicated if imbalance in age, gender etc an attempt will be made to rectify	From relevant community based org; civic assoc; rate payers assoc; also government employees. Sample mainly recruited from the workplace	Participants recruited by word of mouth and from announcement and advertisements posted in an urban Shanghai community	Community based org; faith-based org; workplaces; civic organizations	Household	Worksites	Community	Household
Data collection method	Household interview as part of NCD STEPS risk factor surveillance study	Household interviews	Household interviews	Worksite	Household interviews	Interviews at community location	Household interview	Self report survey administered in the worksite	Interview	Household interview
No. times[a]	3	3	3	2	2	3	1	3	1	3
Equipment	Pedometer	Pedometer	Pedometer	CSA/MTI accelerometers	CSA and pedometer	Pedometer	—	Pedometer	—	Pedometer

[a] Number of times GPAQ and IPAQ were administered.

concurrent validity and a sub sample of 241 were selected for the test-retest reliability study. Another sub sample of 186 participants took part in the criterion (pedometer) study. São Paulo, Brazil; and Porto, Portugal both undertook the concurrent validity study only.

Data Collection

Studies were undertaken between September 2003 and January 2005. Each center took responsibility for training interviewers in the administration of the GPAQ instrument as well as the methods for use of the objective measure (ie, pedometer or accelerometer). In addition the GPAQ was translated, then independently back-translated to ensure correct interpretation. Culturally specific examples of different types of physical activity were inserted into the questionnaire and show cards were developed and used to show different kinds of physical activity. Advice and assistance was provided by the GPAQ Research Coordinating Centre at The University of Western Australia (Perth, Western Australia).

Data Management

Each of the 10 sites entered collected data into Microsoft Excel, EPI Info or the Statistical Package for Social Sciences (SPSS version 12.0) program. Sites undertook their own analyses using standard guidelines as well as sending a copy of the data to the coordinating center where country specific and pooled data analyses were conducted. For the purpose of data analysis, all data were transferred into SPSS and pooled to provide overall estimates for reliability and validity.

Scoring of Physical Activity and Data Reduction

All data files were labeled consistently and data were cleaned for missing values and to remove nonplausible responses; for example, those where the days exceeded 7 days per week or where hours/day exceed 24 hours. These cases were excluded from the final database. Data truncation was applied only when the activity reported within any domain or intensity exceeded 16 hours. In accordance with previous IPAQ guidelines, 16 hours was deemed the average maximum amount of time one could spend in any given activity allowing an average of 8 hours for sleep.

Physical activity data were used to compute estimates of total activity for each intensity or domain category captured by each instrument (GPAQ and IPAQ separately); these were derived using reported days and time. Pedometer counts were averaged over the days the pedometer was worn (usually 7 days) to be comparable for criterion validity. No truncation rules were used for these analyses.

Data Analysis

Reliability: Categorical variables such as those with a “Yes/No” response (eg, doing some vigorous activity at work) were assessed for reproducibility using the kappa statistic and percent agreement. Continuous variables such as frequency (days) and duration (time in minutes) of physical activity were assessed within each domain (ie, work, discretionary and transport) and the reliability was assessed using Spearman’s rho coefficient due to the skewed distribution of these data.

Concurrent Validity: Summary variables of total physical activity were computed from GPAQ and IPAQ to assess their association using Spearman’s rho coefficients with particular care taken to ensure the computed variables were indeed comparable. The following formulas were used: For GPAQ: total vigorous-intensity activity = sum of vigorous at work plus vigorous in discretionary domain; total moderate-intensity activity = sum of moderate-intensity activity at work plus total transport-related plus total discretionary moderate-intensity activity;

For IPAQ: total moderate-intensity and total walking activity were combined to compute a total moderate-intensity variable that would be comparable to GPAQ. In addition, a variable ‘inactive’ (yes/no) was computed using the following definitions: for GPAQ ‘no physical activity in the work, transport and discretionary domains’ and for IPAQ no physical activity reported in vigorous, moderate and walking domains.

Criterion Validity: The association between total minutes of physical activity (from GPAQ) and total pedometer counts per day was assessed using Spearman’s rho. For those studies using accelerometers, data were aggregated to provide measures of average time spent in moderate-intensity activity and vigorous-intensity activity using available formula. The correlation between these measures and total self-reported physical activity, total vigorous-intensity and total moderate-intensity activity derived from GPAQ was assessed. In addition, the relationship between the computed variable ‘time spent sitting’ (a proxy measure of sedentary behavior) was compared with measures from both pedometers and MTI’s.

The following standards were applied to interpret the agreement coefficients: 0 to 0.2 = poor; 0.21 to 0.40 = fair; 0.41 to 0.60 = moderate/acceptable; 0.61 to 0.80 = substantial; 0.81 to 1.0 = near perfect.

Results

Sample sizes across the study ranged from 67 (Portugal) to 940 participants (Ethiopia). Overall the sample populations were similar across the studies in terms of gender and age with approximately equal numbers of males and females, and a slightly higher prevalence of younger participants (18–44 year olds) compared with the older category (45–75 year olds) (Table 2). Just over two-thirds of the sample was from urban regions, and only Ethiopia, India, Japan and Indonesia included participants from both rural and urban regions. For the countries that recorded BMI, the majority of participants

Table 2 Demographic Profile of Participants: By Country

Total sample N = 2122	Ethiopia N = 940	Indonesia N = 337	India N = 262	South Africa N = 214	Shanghai, China N = 221	Bangladesh N = 147	Brazil N = 204	Japan N = 148	Portugal N = 67	Taiwan, China N = 141
Sex										
Male	473	164	116	92	112	78	100	86	17	63
Female	467	173	118	122	109	69	104	52	50	78
Age groups										
18–44 yrs	746	170	131	118		103	117	103	56	71
45–75 yrs	194	167	103	96		44	87	45	9	70
Education										
<13 Years	912	—	—	124	138	63	—	34	—	68
≥13 Years (Tertiary or equivalent)	28	—	—	70	83	83	—	114	—	73
BMI status										
Underweight <18.5	324	—	41	1	11	5	—	30	—	6
Acceptable weight	576	—	122	58	172	68	—	90	—	76
18.5–24.9										
Overweight 25-29.9	36	—	52	78	36	58	—	24	—	47
Obese >30	1	—	19	77	2	16	—	4	—	5
Population category										
Urban	404	196	125	200	221	147	204	104	—	—
Rural	536	140	137	—	—	—	—	6	—	—

were within the healthy weight range except for South Africa where a higher proportion of overweight and obese participants were recruited. Education status was recorded by 6 centers and in three the majority of participants had less than 13 years of education.

Reliability

Data collected on visit 2 and visit 3 were pooled and used to assess test-retest reliability. Overall, for categorical variables (ie, those items with a response of 'Yes' or 'No') kappa statistics ranged from 0.67 (for doing some discretionary-related vigorous-intensity activity) to 0.76 (for being sedentary at work) indicating a substantial association (Table 3). Percent agreement ranged from 85.6% (for sedentary in discretionary time) to 92.1% (for doing some vigorous-intensity activity in discretionary time). Across all three domains, items in the work domain produced the strongest kappa statistics while percent agreement was higher for items on discretionary physical activity compared with the work and transport activity.

When looking at the reproducibility of items within different sub populations, some patterns clearly emerged (data not shown). Across all variables, kappa statistics for males were consistently stronger compared with females although the difference was mostly very small in magnitude. Similarly, kappa statistics were consistently slightly stronger for the urban (0.72 to 0.78) population compared with the rural populations (0.58 to 0.71). The pattern of results by level of education was more mixed. The correlations for vigorous-intensity work activity (0.66) and for transport activity (0.69) were stronger for those with <13 years of education compared with those with ≥13 years. However, the reverse pattern was seen for items assessing sedentary at work and all the discretionary-related items (sedentary, moderate- and vigorous-intensity activity). Analysis by BMI category also produced varied results. For most variables the kappa statistics were similar or higher for those classified as overweight compared with healthy weight and were notably lower for those in the underweight category (data not shown).

Subanalysis by country showed several clear patterns. Results from Shanghai, China were consistently higher than other countries while the results from Taiwan, China were consistently lower than other countries. The kappa statistic could not be calculated for Bangladesh vigorous-intensity discretionary activity as no participants reported doing this activity. Data from India were excluded from all reliability analyses due to an error in data collection methodology.

Table 3 Reliability[a] of GPAQ Categorical Variables: Pooled Data and by Country

		Work domain			Transport domain	Discretionary domain		
	Data collection center	Sedentary	Vigorous intensity	Moderate intensity	Walking & cycling	Sedentary	Vigorous intensity	Moderate intensity
Kappa	Pooled	0.73	0.73	0.70	0.72	0.68	0.67	0.72
(% agreement)	N = 1524	(86.9)	(91.3)	(86.3)	(88.3)	(85.6)	(92.1)	(89.9)
By country	Bangladesh	0.55	0.72	0.57	0.49	0.48	—	0.74
	N = 147	(80.0%)	(96.6%)	(84.1%)	(74.5%)	(94.5%)		(98.6%)
	Shanghai, China	0.98	0.91	0.98	1.00	1.00	1.00	1.00
	N = 221	(99.5%)	(99.6%)	(99.5%)	(100%)	(100%)	(100%)	(100%)
	Ethiopia	0.54	0.63	0.49	0.43	0.52	0.44	0.52
	N = 339	(80.2%)	(43.1%)	(75.2%)	(93.2%)	(76.1%)	(82.9%)	(81.1%)
	Indonesia	0.78	0.66	0.73	0.70	0.44	0.61	0.44
	N = 337	(89.3%)	(89.9%)	(86.4%)	(85.8%)	(85.5%)	(98.52%)	(92.3%)
	South Africa	0.70	0.66	0.72	0.75	0.70	0.71	0.78
	N = 214	(91.6%)	(93.9%)	(92.9%)	(87.9%)	(85.2%)	(88.3%)	(89.3%)
	Japan	0.74	0.88	0.87	0.75	0.82	0.89	0.88
	N = 148	(87.8)	(97.9)	(94.6)	(91.2)	(91.6)	(95.9)	(94.6)
	Taiwan, China	0.48	0.56	0.34	0.44	0.37	0.47	0.46
	N = 120	(74.2)	(83.3)	(73.3)	(72.5)	(69.8)	(77.5)	(74.2)

Note. India was excluded from analysis due to methodological error within their reliability study.
[a] Data from visit 2 compared with visit three.

Coefficients for test-retest reliability of GPAQ continuous variables are presented in Table 4. Spearman's rho correlation ranged from 0.67 for total time spent in vigorous-intensity discretionary-related activity to 0.81 for total transport-related activity. Similarly to the categorical variables, the work domain produced the strongest correlation coefficients.

Analyses by sub population showed some similar patterns to those observed with the categorical variables. Slightly stronger correlations were seen across all domains for males compared with females, with the exception of transport where there was little difference. Similarly, slightly associations were observed for the urban sub population (0.71 to 0.85) in comparison with rural sub population (0.60 to 0.71). Stronger correlations on total work physical activity and transport activity were found for those with lower education (<13 years) compared with those with ≥13 years. However, this pattern was reversed for all discretionary-related variables. Analyses by BMI, showed consistently weaker correlations for those in the 'underweight' category compared with the other BMI categories for all items except discretionary activity time.

Analyses by country showed some variability. Shanghai, China showed the strongest correlations across all variables and Taiwan, China recorded the lowest correlations with the exception of total transport time for which Ethiopia reported the lowest correlations, and total discretionary moderate time for which Bangladesh reported the lowest correlation. The coefficients ranged from 0.31 to 0.99 with most over 0.55 (Table 4).

Overall, the results indicate a high level of repeatability between the 2 administrations of GPAQ with moderate to substantial correlations for the pooled data across the variables (range *r* = .67 to 0.81).

Concurrent Validity

Concurrent validity between GPAQ and IPAQ was assessed using the following summary variables: total physical activity time; total vigorous-intensity activity; total moderate-intensity activity; total sedentary time and inactive.

The pooled results showed an acceptable level of association with Spearman's rho coefficients ranging from 0.45 for total moderate-intensity physical activity to 0.57 for total vigorous-intensity physical activity (Table 5). The measure of sedentary behavior had the strongest correlation with a coefficient of 0.65 which was in part expected because the question was identical on both the IPAQ and GPAQ questionnaires. Kappa coefficient for inactive category was 0.22 suggesting only poor agreement on this derived measure from IPAQ and GPAQ.

Analysis by sub populations showed a broadly similar pattern to the pooled results (data not shown). A stronger correlation was seen for sedentary-time for

Table 4 Reliability[a] of GPAQ Continuous Variables: Pooled Data and by Country

	Work domain			Transport domain	Discretionary domain		
Spearman's rho	**Vigorous total**	**Moderate total**	**Work total**	**Transport total**	**Vigorous total**	**Moderate total**	**Discretionary total**
Pooled N = 2221	0.73**	0.74**	0.77**	0.81**	0.67**	0.71**	0.78**
Bangladesh N = 147	0.72**	0.57**	0.58**	0.57	—	0.31**	—
Shanghai, China N = 221	0.92**	0.99**	0.99**	0.98	1.00**	1.00**	1.00**
Ethiopia N = 339	0.64**	0.50**	0.56**	0.53	0.46**	0.52**	0.73**
Indonesia N = 147	0.68**	0.78**	0.80**	0.70	0.61**	0.45**	0.52**
South Africa N = 214	0.69**	0.75**	0.76**	0.75	0.71**	0.77**	0.71**
Japan N = 148	0.88**	0.85**	0.83**	0.90**	0.89**	0.83**	0.88**
Taiwan, China N = 141	0.48**	0.40**	0.53**	0.54**	0.49**	0.50**	0.52**

[a] Data from visit 2 compared with visit three.
**Correlations are significant at the 0.01 level (2-tailed).

women (0.78) compared with men (0.64) but the reverse pattern was seen for time reported for vigorous-intensity activity (0.47 and 0.62, respectively). In addition, a notably higher correlation was observed for sedentary time for the urban population (0.73) compared with rural (0.55).

The country-specific analysis showed some wide variations, including higher correlations for India (total activity and sedentary time) and China (Province of Taiwan) (vigorous- and moderate- intensity activity), and a lower correlation for Brazil on most variables compared with the other countries. Overall a systematic difference in results between countries or sub population groups is not evident although there was a tendency for a slightly stronger association for the measure of vigorous-intensity activity compared with moderate-intensity activity.

Kappa statistics for the correlation between the variable 'inactive' showed a poor correlation between the 2 instruments (0.22) although the percent agreement was quite high for the pooled data (86%) and ranged between 62% to 95% for analyses by country. The low kappa is possibly due to skewed data and low cell counts.

Criterion Validity

For criterion validity, pooled analyses were undertaken using data from 6 studies providing objective estimates of physical activity using pedometers (this excluded South Africa and Shanghai, China because they used accelerometers). For the variable total physical activity, there was fair agreement between GPAQ and pedometers with pooled data giving a correlation of 0.31 (Table 6). The measure of sedentary behavior (time spent sitting) showed only poor agreement in the expected negative direction (-0.20).

Analyses by sub population groups revealed similar correlations for those with more/less education and for men compared with women. The analysis by urban/rural showed a stronger correlation for rural (0.43) compared with urban sub populations (0.23). There were also marked differences between those classified as overweight/obese (0.08) and those classified as a healthy BMI (0.34) and underweight (0.52; Table 6).

South Africa and Shanghai, China both used MTI accelerometers and assessed criterion validity by comparing measures of minutes of total moderate- and vigorous-intensity physical activity, and total sedentary time from GPAQ with measures of time derived from accelerometer counts. The results shown in Table 7 reveal the agreement between reported sedentary time and the accelerometer (minutes of counts of <100) was poor in South Africa (-0.02) and fair-moderate in Shanghai, China (0.40). Correlation coefficients for moderate-intensity activity was fair in China (0.23) and poor in South Africa (-0.03) while the results on vigorous-intensity activity showed a fair correlation in both countries (0.23 for Shanghai, China; 0.26 for South Africa).

Table 5 Concurrent Validity Between GPAQ and IPAQ: Pooled Data and by Country

	Sample size	Total vigorous Spearman's rho	Total moderate[a] Spearman's rho	Total physical activity Spearman's rho	Sedentary time Spearman's rho	Inactive Kappa (% agreement)
Pooled	2657	0.57**	0.45**	0.54**	0.65**	0.22 (85.2%)
Bangladesh	147	0.51**	0.55**	0.53**	0.48**	0.15 (61.9%)
Brazil	204	0.42**	0.29**	0.46**	0.47**	0.32 (81.4%)
China (Shanghai)	221	0.60**	0.60**	0.60**	0.81**	0.23 (81.45%)
China (Province of Taiwan)	141	0.79**	0.75**	0.53**	0.56**	0.25 (82.3%
Ethiopia	940	0.45**	0.40**	0.47**	0.45**	0.10 (95.0%)
India	234	0.51**	0.71**	0.92**	0.98**	—[a]
Indonesia	337	0.67**	0.29**	0.34**	0.83**	0.10 (72.7%)
Japan	148	0.68**	0.68**	0.29**	0.68**	0.52 (89.2%)
Portugal	67	0.52**	0.50**	0.23	0.56**	0.42 (83.1%)
South Africa	214	0.46**	0.30**	0.34**	0.60**	0.11 (82.2%)

[a] IPAQ had all active participants hence kappa statistics could not be calculated.
** Correlation is significant at the 0.01 level (2-tailed).

Table 6 Criterion Validity for Centers Using Pedometers as the Objective Measure

	Country	Sample size	Total physical activity time Spearman's rho	Total sedentary time Spearman's rho
	Pooled	1507	0.31**	–0.20**
By country[a]	Bangladesh	146	0.06	–0.12
	Province of Taiwan, China	220	0.35**	–0.37**
	Ethiopia	186	0.31**	–0.20**
	India	234	0.35**	–0.29**
	Indonesia	337	0.30**	–0.12*
	Japan	148	0.23**	0.00

[a] Criterion validity results excludes Brazil and Portugal as they did not collect objective measures.
** Correlation is significant at the 0.01 level (2-tailed).

Table 7 Criterion Validity for Centers Using Accelerometers as the Objective Measure

GPAQ 2nd visit compared with accelerometer counts	N	Average sedentary counts/day	Average moderate counts/day	Average vigorous counts/day
Shanghai, China	215			
Total physical activity across all domain (minutes)		–0.24	0.24**	0.04
Total vigorous intensity activity across all domains		0.01	0.06	0.23**
Total moderate intensity activity across all domains		–.20**	0.23**	–0.01
Time spent sitting per day (minutes)		0.40**	–0.19**	0.01
South Africa	83			
Total physical activity across all domain (minutes)		–0.16	–0.01	0.03
Total vigorous intensity activity across all domains		–0.10	0.06	0.26*
Total moderate intensity activity across all domains		0.12	–0.03	–0.03
Time spent sitting per day (minutes)		–0.02	–0.02	0.15

** Correlation is significant at the .05 level (2-tailed).

Qualitative Feedback

On completion of the GPAQ research program, participating centers were asked to provide feedback on the ease of use and feasibility of GPAQ from their field work experience and for suggestions on possible improvements. The feedback received included adding more descriptions of moderate-intensity and vigorous-intensity physical activities to increase applicability for diverse populations groups and cultures.

There was also concern about the clarity of the definition of the work domain and it should be improved to clarify that volunteer work and other unpaid work (eg, housework), were included. Other recommendations included the removal of 'screening questions' (P1 and P9 in GPAQ original version) to avoid apparent repetition. The show cards were deemed very important and helped both in the training of interviewers and also for responders. To improve data recording only 1 response format should be provided not 2 optional formats. All suggestions received were considered at an expert meeting on physical activity surveillance convened by WHO in February 2005. In light of these suggestions, a modified version of GPAQ has been developed (See Appendix A for revised GPAQ version recommended for use).

Discussion

There is an urgent need to commence or improve the collection of data on physical activity in population health monitoring systems. For physical activity, it is the lack of a suitable measure that has in part been responsible for the limited collection of population level data in many countries. The WHO Global Strategy for Diet, Physical Activity and Health calls for all countries to undertake monitoring of diet and physical activity.[6] The development of the GPAQ instrument for assessing physical activity is an important and timely contribution to this agenda.

There has been a notable shift in focus from measures of total physical activity to an interest in tracking trends within specific domains, in particular the transportation domain. Other domains are also of interest such as sport and recreation, particularly among those sectors responsible for provision and delivery of programs, services and facilities; and the work domain, which is of particular interest in countries experiencing rapid economic transition.

The GPAQ instrument is a set of items aimed at the measurement of physical activity at a population level. GPAQ was developed after a review of available instruments and in consultation with experts on physical activity as part of the WHO STEPWISE approach to chronic disease risk factor surveillance.[13] Using the experiences of the IPAQ instrument and research, the GPAQ research program was undertaken across 10 countries to allow for variation in culture and patterns of work, transport and leisure.

The results show that overall GPAQ items have good reproducibility with pooled correlations of mostly greater than 0.72. These results are encouraging as evidence on the performance of physical activity instruments in the populations included in this study is limited. Comparison of results across the three domains revealed some evidence of slightly higher correlation coefficients for items in the work and transport domains compared with those assessing the discretionary domain. In contrast, the reverse pattern was seen for percent agreement. Given that these variations were small it is unlikely that they reflect any important trends. Some differences were seen in the sub analyses conducted by age, sex, education, urban/rural and notably by country. Reliability was observed to be consistently lower, but remained in the "moderate" range, in the samples from Bangladesh, Ethiopia and Taiwan, China. This suggests some potential differential performance of GPAQ in these countries that would warrant further research. The results on reliability from the sample from Shanghai, China were also puzzling, specifically because they suggest for several items almost perfect agreement, which is rarely expected nor reported. Despite a thorough review of implementation of the study protocols, and data collection, recording and entry, no source of error could be identified that might explain the apparent 'perfect' agreement. It is recommended that replication work is undertaken to substantiate these specific findings.

The reproducibility of continuous measures from GPAQ was found to be of substantial strength. But again there were some differences in results from different sub populations. Slightly stronger reliability coefficients were observed for men compared with women, and urban comparison with rural populations.

Concurrent validity of GPAQ was assessed using IPAQ, the only other physical activity instrument designed with the same purpose of population monitoring. Overall, the results showed an acceptable level of association (0.45 to 0.57). Analysis by sub populations showed some stronger correlations for variables in men compared with women, and, for sedentary time, in the urban population compared with rural. Country-specific analysis showed consistently higher correlations in India, consistently lower correlations in Indonesia and sometimes also for Ethiopia on selected variables. No systematic differences between countries or sub population groups was evident although the degree of variation would suggest that further testing is warranted.

Criterion validity was assessed by comparing self reported levels of activity with an objective assessment captured by motion monitors, either a pedometer or an accelerometer. For those studies using pedometers, pooled results showed there was fair agreement between GPAQ and pedometer counts on total physical activity. Secondary analyses showed some differences; slightly stronger correlation for men compared with women; for those with higher education compared with those with <13 years; and a stronger correlation for rural compared with urban. There were more marked differences between those classified as overweight/obese and those classified as a healthy BMI and underweight. It is possible that those classified as overweight/obese are less likely to be physically active compared with their counterparts and potentially more likely to over report physical activity due to social desirability. This could explain the direction of these results. It is also possible that the pedometer does not assess physical activity as well in overweight/obese adults resulting in a poor association between these measures of activity.

For studies using accelerometers, the overall results revealed lower levels of agreement between self reported time spent in activity and accelerometer counts, particularly in the sample from South Africa. Low correlations have also been reported from other studies, including those testing IPAQ[14] and may reflect inherent limitations of these measurement devices to act as a 'gold standard' for self report questionnaires.[2]

The results found for criterion validity are consistent with expectations and results from similar studies testing different physical activity instruments.[14–17] Taken overall the findings from this set of diverse international studies, undertaken as part of the development of the GPAQ instrument, indicate that the GPAQ instrument performs well. The reproducibility and validity of items are comparable, and in some cases better than other similar sets of questions on physical activity.

Recommendations

There are several areas which could benefit from further work. Any changes made to the version of GPAQ used in this research would require further reliability and validity testing. It is also desirable to conduct further GPAQ testing in other countries to build a wider evidence base for its use internationally, in particular, GPAQ has not yet been tested in the Eastern Mediterranean or European regions (with the exception of Portugal). In addition, such work should include robust samples of urban, peri urban and rural populations. For several countries involved in this study, these are the first data of this kind. Feedback included the importance of using good quality, culturally appropriate show cards and that further research was needed to explore respondents understanding and reporting of activities of different intensity and time in minutes. A strong understanding and confirmation of the use of GPAQ is vital to help build support for and advance efforts in the implementation of public health approaches to the promotion of activity as part of national chronic disease prevention programs. Finally, the need to share population-level data on physical activity is particularly important within the current phase of implementation of the WHO Global Strategy on Diet, Physical Activity and Health.

Acknowledgments

The authors would like to thank the following research study groups for their assistance and contribution to the development and testing of GPAQ research program: Bangladesh: Mahmudur Rahman; Brazil: Sandra Matsudo; China: Fu Hua and Li Yang (Shanghai, China), Yiing Mei Liou and Jin Chong Chen Taiwan, China); Ethiopia: Fikru Tesfaye and Zerihun Tadesse; India: Puneet Misra, Anand Krishnan; Indonesia: Nawi Ng; Japan: Norio Murase; Portugal: Jorge Mota; South Africa: Vicki Lambert and Ezera Tshabangu. In addition the author wish to thank Dr Ruth Bonita, then based at WHO, for her support for this project and the development of the GPAQ instrument. This article contains the views of its authors and does not necessarily represent the decisions or stated policy of WHO.

References

1. U.S. Department of Health and Human Services (USDHHS). *Physical Activity and Health: A Report of the Surgeon General*. Atlanta, GA: US Department of Health and Human Services; 1996:81–172.
2. World Health Organization. *Preventing Chronic Diseases: A Vital Investment*. Geneva, Switzerland: World Health Organization; 2005:106.
3. Bull FC, Armstrong T, Dixon T, Ham S, Neiman A, Pratt M. Physical Inactivity. In: Ezzati M, Lopez A, Rodgers A, Murray C, eds. *Comparative Quantification of Health Risks: Global and Regional Burden of Disease due to Selected Major Risk Factors*. Geneva, Switzerland: World Health Organization; 2005:729–881.
4. Pratt M, Macera CA, Wang G. Higher direct medical costs associated with physical inactivity. *Phys Sportsmed*. 2000;28:63–70.
5. Levi F, Luuchini F, Negri E, Lavecchia C. Trends in mortality from cardiovascular and cerebrovascular diseases in Europe and other areas of the world. *Heart*. 2002;88:119–124.
6. World Health Organization. Global Strategy on Diet, Physical Activity and Health. WHA57.17. *In Proceedings of the 57th World Health Assembly*. Geneva, Switzerland: World Health Organization. 2004:2-18.
7. Armstrong T, Bonita R. Building capacity for an integrated noncommunicable disease risk factor surveillance system in developing countries. *Ethn Dis*. 2003;13:2–13.
8. Thomas JR, Nelson JK. *Research Methods in Physical Activity*. 4th ed. Champaign, Ill: Human Kinetics; 2001:181–185.
9. Rzewnicki R, Vanden Auweele Y, De Bourdeaudhuij I. Addressing over reporting on the International Physical Activity Questionnaire (IPAQ) telephone survey with a population sample. *Public Health Nutr*. 2003;6:299–305.
10. Shephard RJ. Limits to the measurement of habitual physical activity by questionnaires. *Br J Sports Med*. 2003;37:197–206.
11. Le Masurier GC. Motion sensor accuracy under controlled and free-living conditions. *Med Sci Sports Exerc*. 2004;36:905–910.
12. Melanson EL, Freedson PS. Validity of the Computer Science and Applications (CSA) activity monitor. *Med Sci Sports Exerc*. 1995;27:934–940.
13. World Health Organization. *WHO STEPS Surveillance Manual: The WHO STEPwise Approach to Chronic Disease Risk Factor Surveillance*. Geneva, Switzerland: World Health Organization; 2005:7.
14. Craig CL, Marshall AL, Sjostrom M, et al. International physical activity questionnaire: 12-country reliability and validity. *Med Sci Sports Exerc*. 2003;35:1381–1395.
15. Brown W, Bauman A, Chey T, Trost S, Mummery K. Comparison of surveys used to measure physical activity. *Aust N Z J Public Health*. 2003;28:128–134.
16. Craig CL, Russell SJ, Cameron C. Reliability and validity of Canada's physical activity monitor for assessing trends. *Med Sci Sports Exerc*. 2002;34:1462–1467.
17. Timperio A, Salmon J, Rosenberg M, Bull FC. Do logbooks influence recall of physical activity in validation studies? *Med Sci Sports Exerc*. 2004;36:1181–1186.

Appendix

Global Physical Activity Questionnaire (GPAQ) Version 1

CORE Physical Activity (Section P)				
Next I am going to ask you about the time you spend doing different types of physical activity. Please answer these questions even if you do not consider yourself to be an active person. Think first about the time you spend doing work. Think of work as the things that you have to do such as paid or unpaid work, household chores, harvesting food, fishing or hunting for food, seeking employment. *[Insert other examples if needed]*				
P 1	Does your work involve mostly sitting or standing, with walking for no more than 10 minutes at a time?	Yes 1 No 2	☐	*If Yes, go to P6*
P 2	Does your work involve vigorous activity, like [*heavy lifting, digging or construction work*] for at least 10 minutes at a time? *INSERT EXAMPLES & USE SHOWCARD*	Yes 1 No 2	☐	*If No, go to P4*
P 3a	In a typical week, on how many days do you do vigorous activities as part of your work?	Days a week	☐☐	
P 3b	On a typical day on which you do vigorous activity, how much time do you spend doing such work?	In hours and minutes OR in Minutes only	hrs ☐☐ : mins ☐☐ or minutes ☐☐☐	
P 4	Does your work involve moderate-intensity activity, like brisk walking *[or carrying light loads]* for at least 10 minutes at a time? *INSERT EXAMPLES & USE SHOWCARD*	Yes 1 No 2	☐	*If No, go to P6*
P 5a	In a typical week, on how many days do you do moderate-intensity activities as part of your work?	Days a week	☐☐	
P 5b	On a typical day on which you did moderate-intensity activities, how much time do you spend doing such work?	In hours and minutes OR in Minutes only	hrs ☐☐ : mins ☐☐ or minutes ☐☐☐	
P 6	How long is your typical work day?	Number of hours	hrs ☐☐	
Other than activities that you've already mentioned, I would like to ask you about the way you travel to and from places. For example to work, for shopping, to market, to church. [*insert other examples if needed*]				
P 7	Do you walk or use a bicycle (*pedal cycle*) for at least 10 minutes continuously to get to and from places?	Yes 1 No 2	☐	*If No, go to P9*
P 8a	In a typical week, on how many days do you walk or bicycle for at least 10 minutes to get to and from places?	Days a week	☐☐	
P 8b	How much time would you spend walking or bicycling for travel on a typical day?	In hours and minutes OR in Minutes only	hrs ☐☐ : mins ☐☐ or minutes ☐☐☐	
The next questions ask about activities you do in your leisure time. Think about activities you do for recreation, fitness or sports *[insert relevant terms]*. Do not include the physical activities you do at work or for travel mentioned already.				
P 9	Does your [*recreation, sport or leisure time*] involve mostly sitting, reclining, or standing, with no physical activity lasting more than 10 minutes at a time?	Yes 1 No 2	☐	*If Yes, go to P 14*
P 10	In your [*leisure time*], do you do any vigorous activities like [*running or strenuous sports, weight lifting*] for at least 10 minutes at a time? *INSERT EXAMPLES & USE SHOWCARD*	Yes 1 No 2	☐	*If No, go to P 12*
P 11a	If Yes, In a typical week, on how many days do you do vigorous activities as part of your [*leisure time*]?	Days a week	☐☐	

continued

P 11b	How much time do you spend doing this on a typical day?	In hours and minutes OR in Minutes only	hrs ☐☐ : mins ☐☐ or minutes ☐☐☐

P 12	In your [*leisure time*], do you do any moderate-intensity activities like brisk walking,[*cycling or swimming*] for at least 10 minutes at a time? *INSERT EXAMPLES & USE SHOWCARD*	Yes 1 No 2	☐	*If No, go to P 14*
P 13a	If Yes In a typical week, on how many days do you do moderate-intensity activities as part of [*leisure time*]?	Days a week	☐☐	
P 13b	How much time do you spend doing this on a typical day?	In hours and minutes OR in Minutes only	hrs ☐☐ : mins ☐☐ or minutes ☐☐☐	
The following question is about sitting or reclining. Think back over the past 7 days, to time spent at work, at home, in *[leisure]*, including time spent sitting at a desk, visiting friends, reading, or watching television, but do not include time spent sleeping.				
P 14	Over the past 7 days, how much time did you spend sitting or reclining on a typical day?	In hours and minutes OR in Minutes only	hrs ☐☐ : mins ☐☐ or minutes ☐☐☐	

Global Physical Activity Questionnaire (GPAQ) Revised, Version 2

GPAQ V2				
Next I am going to ask you about the time you spend doing different types of physical activity in a typical week. Please answer these questions even if you do not consider yourself to be a physically active person. Think first about the time you spend doing work. Think of work as the things that you have to do such as paid or unpaid work, study/training, household chores, harvesting food/crops, fishing or hunting for food, seeking employment. *[Insert other examples if needed]*. In answering the following questions 'vigorous-intensity activities' are activities that require hard physical effort and cause large increases in breathing or heart rate, 'moderate-intensity activities' are activities that require moderate physical effort and cause small increases in breathing or heart rate.				
		Response	**Coding Column**	
P 1	Does your work involve vigorous-intensity activity that causes large increases in breathing or heart rate like *[carrying or lifting heavy loads, digging or construction work]* for at least 10 minutes continuously? *INSERT EXAMPLES & USE SHOWCARD*	Yes 1 No 2	☐	*If No, go to P3*
P 2a	In a typical week, on how many days do you do vigorous-intensity activities as part of your work?	Days a week	☐☐	
P 2b	How much time do you spend doing vigorous-intensity activities at work on a typical day?	In hours and minutes	hrs ☐☐ : mins ☐☐	
P 3	Does your work involve moderate-intensity activity, that causes small increases in breathing or heart rate such as brisk walking *[or carrying light loads]* for at least 10 minutes continuously? *INSERT EXAMPLES & USE SHOWCARD*	Yes 1 No 2	☐	*If No, go to P5*
P 4a	In a typical week, on how many days do you do moderate-intensity activities as part of your work?	Days a week	☐☐	
P 4b	How much time do you spend doing moderate-intensity activities at work on a typical day?	In hours and minutes	hrs ☐☐ : mins ☐☐	
The next questions exclude the physical activities at work that you have already mentioned. Now I would like to ask you about the usual way you travel to and from places. For example to work, for shopping, to market, to place of worship. [*insert other examples if needed*]				
P 5	Do you walk or use a bicycle (*pedal cycle*) for at least 10 minutes continuously to get to and from places?	Yes 1 No 2	☐	*If No, go to P7*
P 6a	In a typical week, on how many days do you walk or bicycle for at least 10 minutes continuously to get to and from places?	Days a week	☐☐	
P 6b	How much time do you spend walking or bicycling for travel on a typical day?	In hours and minutes	hrs ☐☐ : mins ☐☐	

continued

Global Physical Activity Questionnaire (GPAQ) Revised, Version 2 *(continued)*

		Response	Coding Column	
The next questions exclude the work and transport activities that you have already mentioned. Now I would like to ask you about sports, fitness and recreational activities (*leisure*),[*insert relevant terms*].				
P 7	Do you do any vigorous-intensity sports, fitness or recreational (*leisure*) activities that cause large increases in breathing or heart rate like [*running or football,*] for at least 10 minutes continuously? *INSERT EXAMPLES & USE SHOWCARD*	Yes 1 No 2	☐	*If No, go to P9*
P 8a	In a typical week, on how many days do you do vigorous-intensity sports, fitness or recreational (*leisure*) activities?	Days a week	☐☐	
P 8b	How much time do you spend doing vigorous-intensity sports, fitness or recreational activities on a typical day?	In hours and minutes	hrs ☐☐ : mins ☐☐	
P 9	Do you do any moderate-intensity sports, fitness or recreational *(leisure*) activities that causes a small increase in breathing or heart rate such as brisk walking,[*cycling, swimming, volleyball*] for at least 10 minutes continuously? *INSERT EXAMPLES & USE SHOWCARD*	Yes 1 No 2	☐	*If No, go to P 11*
P 10a	In a typical week, on how many days do you do moderate-intensity sports, fitness or recreational (*leisure*) activities?	Days a week	☐☐	
P 10b	How much time do you spend doing moderate-intensity sports, fitness or recreational (*leisure*) activities on a typical day?	In hours and minutes	hrs ☐☐ : mins ☐☐	
The following question is about sitting or reclining at work, at home, getting to and from places, or with friends including time spent [sitting at a desk, sitting with friends, travelling in car, bus, train, reading, playing cards or watching television], but do not include time spent sleeping. INSERT EXAMPLES & USE SHOWCARD				
P 11	How much time do you usually spend sitting or reclining on a typical day?	In hours and minutes	hrs ☐☐ : mins ☐☐	