HOSTED BY

ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect

Journal of Sport and Health Science xx (2016) 1–7

www.jshs.org.cn

Original article

Reliability and validity of the French version of the global physical activity questionnaire

Fabien Rivière [a,*,†], Fatima Zahra Widad [a,†], Elodie Speyer [a,b,c], Marie-Line Erpelding [b,c], Hélène Escalon [d], Anne Vuillemin [a]

[a] EA 4360 APEMAC, University of Lorraine, Paris Descartes University, Nancy, 54505, France

[b] Inserm, CIC-1433 Clinical Epidemiology, Nancy, F-54000, France

[c] University Hospital Center of Nancy, Pole S2R, Epidemiology and Clinical Evaluation, Nancy, F-54000, France

[d] The French Public Health Agency, Saint-Maurice, 94410, France

Received 13 January 2016; revised 7 April 2016; accepted 19 June 2016

Available online

Abstract

Background: The Global Physical Activity Questionnaire (GPAQ) has been used to measure physical activity (PA) and sedentary time in France, but no study has assessed its psychometric properties. This study aimed to compare the reliability as well as criterion and concurrent validity of the French version of the GPAQ with the French International PA Questionnaire long form (IPAQ-LF) and use of an accelerometer in a general adult population.

Methods: We included 92 participants (students or staff) from the Medicine Campus at the University of Lorraine, Nancy (north-eastern France). The French GPAQ was completed twice, 7 days apart, to study test-retest reliability. The IPAQ-LF was used to assess concurrent validity of the GPAQ, and participants wore an accelerometer (Actigraph GT3X+) for 7 days to study criterion validity. Reliability as well as concurrent and criterion validity of the GPAQ were tested by the intraclass correlation coefficient (ICC), Spearman correlation coefficient for quantitative variables, and Kappa and Phi coefficients for qualitative variables. Both concurrent and criterion validity of GPAQ were assessed by Bland-Altman plots.

Results: The GPAQ showed poor to good reliability (ICC = 0.37–0.94; Kappa = 0.50–0.62) and concurrent validity (Spearman r = 0.41–0.86), but only poor criterion validity (Spearman r = 0.22–0.42). Limits of agreement for the GPAQ and accelerometer were wide, with differences between 286.5 min/day and 601.3 min/day.

Conclusion: The French version of the GPAQ provides limited but acceptable reliability and validity for the measurement of PA and sedentary time. It may be used for assessing PA and sedentary time in a French adult population.

Keywords: Measurement; Physical activity; Psychometric analysis; Questionnaire; Reliability; Self-report; Sitting time; Validity

1. Introduction

Physical activity (PA) surveillance is a public health preoccupation and is considered by the World Health Organisation (WHO) as a protective factor for non-communicable diseases.[1] A high PA level is associated with reduced mortality and the occurrence of diseases or their consequences and improved quality of life.[2,3] Because of its therapeutic role, PA is also used as adjuvant treatment in chronic diseases.[4,5]

In this context, the measurement of PA is essential to assess strategies promoting PA and to survey and compare PA levels between countries. Questionnaires are the most commonly used instrument in epidemiologic studies to assess PA because they are relatively inexpensive and easy to use both for a large population and in a short time. They can be self-administered, completed during an interview or administered by phone. Many different questionnaires have been developed and used to measure PA, so international comparison is difficult, and overall, their development lacked methodological quality.[6]

In the late 1990s, the International Physical Activity Questionnaire (IPAQ) was developed in 2 forms (short form (IPAQ-SF) and long form (IPAQ-LF)) to create national and international comparable and standardized measures of PA. The

Peer review under responsibility of Shanghai University of Sport.

* Corresponding author.

E-mail address: fbn.riviere@gmail.com (F. Rivière).

† These two authors equally contributed to this work.

http://dx.doi.org/10.1016/j.jshs.2016.08.004

long form of the IPAQ (31 items) was developed to capture information about domains of PA but has been considered too long and too complex to be used in surveillance studies, while the short form (9 items) does not take into account the domains of PA.[7,8] For PA surveillance, the measurement of PA domains is needed to understand the patterns of PA and to develop interventions. Thus, in order to provide an instrument that would address the limits of these questionnaires, the Global Physical Activity Questionnaire (GPAQ) has been developed by the WHO, as part of the WHO STEPwise approach to survey chronic disease risk factors. It is now recommended by the WHO for national surveillance of PA.[1] Since its development, the GPAQ has been translated into and tested in many languages and is used in many countries.[9-16] In France, the GPAQ has been used to describe and analyse PA and sedentary time of the general population.[17] However, it has not been validated in the French language. Evidence for the validity and reliability of the French version of the GPAQ is needed because the results may be affected by the sociocultural specificities of the country.[18]

Rigorous methodology is needed to examine the degree in which an instrument is affected by measurement error (reliability) and measures the construct it intends to measure (validity).[19] Concurrent validity refers to the degree to which the GPAQ measures what it purports to measure, and criterion validity is the degree to which the results of the questionnaire are an adequate reflection of a "gold standard". Because of no satisfying available gold standard measurement for PA behavior, objective measures such as accelerometers and pedometers are commonly used. To appraise the concurrent validity of the GPAQ, a questionnaire measuring the same construct and with similar structure is considered relevant. Even if the IPAQ-LF is more detailed than the GPAQ, it is the most similar in its construct and its structure. For this reason, the IPAQ-LF has been considered relevant to examine the concurrent validity of the GPAQ.

This study aimed to assess the test-retest reliability as well as criterion and concurrent validity of the French version of the GPAQ by comparison with the IPAQ-LF and use of an accelerometer in a general adult population in France.

2. Methods

2.1. Patients and study design

A convenient sample was recruited from January 20, 2015 to April 20, 2015, from the Medicine Campus, University of Lorraine, Nancy (north-eastern France), by posting an advertisement on campus and by e-mailing students and staff. Participants had to be ≥18 years old, working or studying at the Medicine Campus, able to read and understand French, and willing to participate in the study. The study protocol was approved by the Legal representative of the French data protection authority (Commission Nationale Informatique et Libertés) of the University of Lorraine, France. All participants were asked to read and sign a consent form. A ratio of 5 subjects per item was used to determine the number of participants to include.[20] Because the GPAQ contained 16 items, a minimum number of 80 participants was required.

Each subject was invited to participate in a face-to-face interview on Day 0 (D0) and receive all explanations about the study and its purpose from an interviewer. After giving consent, participants answered sociodemographic and anthropometric questions, then completed the GPAQ and IPAQ-LF. Then, the interviewer gave the participant an accelerometer and explained its use. Participants were asked to wear the accelerometer for 7 consecutive days. Eight days after the first interview (D8), participants returned the accelerometer and completed the GPAQ and IPAQ-LF a second time. They were also asked if they had changed their activity during the week of the study as compared to a typical week.

2.2. Instruments

We used the French translation of the GPAQ (Version 2.0)[21] to gather information on the time spent in moderate and vigorous PA and in sedentary behavior. At the WHO level, the GPAQ has been translated in French by a professional translator, and back-translated by 2 independent technical experts. The versions were then compared, and where discrepancies existed, these were discussed and a consensus was found. The GPAQ contains 16 items designed to assess the frequency and duration of PA in 3 domains: during work, transportation, and leisure time as well as time spent sitting during a typical week. It distinguishes PA duration by min/day and min/week for each PA domain, which allows for calculating the energy expenditure scored in metabolic equivalent tasks (METs). One MET corresponds to resting energy expenditure. According to duration and energy expenditure, PA level was classified as low, moderate, and high.

The French IPAQ-LF was used to test the concurrent validity of the GPAQ. It contains 27 items designed to assess the frequency and duration of PA in 4 domains: during work, transportation, household activities, and leisure time, then time spent sitting.[22] The IPAQ-LF scores PA in terms of energy expenditure (MET), intensity (low, moderate, high, and sedentary), and duration (min/day, min/week).

The ActiGraph accelerometer, model GT3X+ (ActiGraph, Pensacola, FL, USA), was used as the criterion measure. The device is worn at the waist and measures and records the changes in acceleration and deceleration movements in 3 axes (anteroposterior, superio-inferior, and medial side). Data for measuring acceleration and deceleration are stored in non-volatile flash memory and can be read by using ActiLife software. Accelerometer data were scored using ActiLife 6 Data Analysis Software (ActiGraph) to assess time spent at various PA intensity levels (moderate and vigorous in min/day). Freedson's Adult VM3 (2011) cut-off points were used to determine several PA levels: light, 0–2690 counts per minute (CPM); moderate, 2691–6166 CPM; vigorous, 6167–9642 CPM; and very vigorous, 9643–∞ CPM. Minutes spent at each intensity level were averaged across valid days. Non-wear periods were identified as 60 consecutive minutes with no movement data o (0 counts).[23] All calculations were based on 60 s epochs; an epoch is a user-defined time-sampling interval used to filter the acceleration signal. In this study, we used 7-days PA questionnaires, so only

Please cite this article in press as: Fabien Rivière, et al., Reliability and validity of the French version of the global physical activity questionnaire, Journal of Sport and Health Science (2016), doi: 10.1016/j.jshs.2016.08.004

data with ≥10 h of wear time per day for ≥7 days were considered valid and included in the analysis.[24,25]

Sociodemographic data such as age, sex, and education (high school or higher education) and socioprofessional status (student or staff) were collected. Anthropometric data including height (in cm) and weight (in kg) were reported by each participant for calculating body mass index (BMI, kg/cm^2), then participants were classified by BMI level: underweight (BMI <18.5 kg/cm^2), normal weight (18.5–24.9 kg/cm^2), overweight 25.0–29.9 kg/cm^2), and obese (>30 kg/cm^2). All data (except accelerometer data directly transferred into ActiLife software) were entered into an electronic case report form (CRF) created with Epidata 3.1 (The EpiData Association, Odense, Denmark).

2.3. Statistical analysis

Data analysis involved use of SAS 9.4 (SAS Inst., Cary, NC, USA). Qualitative variables were reported as relative frequency and quantitative variables as mean ± SD or median. The Kolmogorov-Smirnov test was used to assess the normality of data distribution. For participants who declared changing their PA, paired Student's *t* test was used to evaluate the difference in total PA between the 2 visits. Because the activity measured by the GPAQ includes work and household activities, it was compared to the sum of work and household PA measured by the IPAQ-LF.

Test–retest reliability was tested by the kappa coefficient for categorical data and the intraclass correlation coefficient (ICC) for quantitative data. Spearman correlation was also calculated for quantitative data to compare with previous studies.[11-14] Non parametric correlation coefficient was used because of non-Gaussian distribution for most of PA-score. For one of the GPAQ's question, one answer modality was overrepresented and the correlation was not concordant with the observed agreement (when visualizing the data, the agreement seems good but it was not observed when assessed with ICC and Spearman correlation). Thus the variable was converted into a discrete variable, and the prevalence-adjusted and bias-adjusted kappa (PABAK) was used to assess the agreement.[26] ***Concurrent validity*** was examined by comparing data for the GPAQ and IPAQ-LF at D0 and D8[19] with the Spearman correlation coefficient and its 95%CI for quantitative data and the Phi coefficient for qualitative data. ***Criterion validity*** was examined by comparing minutes of PA obtained with the GPAQ to accelerometer-obtained data at D8 by the Spearman correlation coefficient and its 95%CI.

Both the concurrent and criterion validity of the GPAQ were assessed by Bland-Altman plots to measure the agreement and bias for total PA and sedentary time between questionnaire's answers and results from accelerometer.[27] Correlation assesses the degree to which 2 variables are related. However, a high correlation does not necessary imply that there is good agreement between the 2 methods. Thus, Bland-Altman was used to quantify the agreement between 2 measurements by plotting the difference between the 2 measurements against the average obtained with each of the 2 methods.

Kappa and Phi coefficients were classified by the ratings suggested by Landis and Koch:[28] poor, <0.00; slight, 0.00–0.20; fair, 0.21–0.40; moderate, 0.41–0.60; substantial, 0.61–0.80; and almost perfect, 0.81–1.00. ICC and Spearman correlation <0.50 were considered as poor, between 0.50 and 0.75 as moderate, and >0.75 were as good.[29]

3. Results

3.1. Participant characteristics

In total, 92 subjects participated in the study (mean age 30.1 ± 10.7 years, range 19–58 years; 67 (72.8%) females); 56.5% were students, 95.6% had higher education, 9.8% had chronic disease, and 76.9% had normal BMI (Table 1). Overall, 25% of participants declared having changed their activity between the 2 visits, but the difference between the total PA means measured by the GPAQ was not statistically significant ($p = 0.49$).

3.2. Descriptive statistics for the GPAQ, IPAQ, and accelerometer

All descriptive statistics for GPAQ, IPAQ, and accelerometer are presented in Table 2.

3.3. Test-retest reliability

The ICCs ranged from 0.37 to 0.94, with the highest ICC for vigorous leisure PA. Only total vigorous and vigorous leisure PA showed good reliability, whereas all other PA scores were poor to moderate, with the lowest value for moderate leisure PA(ICC = 0.37, 95%CI: 0.15–0.56). A good reliability for total sitting time was also observed (ICC = 0.80, 95%CI: 0.69–0.87) whereas it was moderate for total PA (ICC = 0.58, 95%CI: 0.40–0.72). For PA level, the kappa coefficient showed moderate to substantial correlation, varying from 0.50 to 0.62 for moderate and low PA levels, respectively. For vigorous activity at work, the GPAQ showed an almost perfect reliability (PABAK = 0.91). Except for total PA, with ICC = 0.58, 95%CI: 0.40–0.72 and Spearman's r = 0.82, 95%CI: 0.72–0.88, most Spearman values were similar to the ICC (Table 3).

Table 1
Sociodemographic and anthropometric characteristics of participants.

	Total sample ($n = 92$, %)
Sex	
Male	25 (27.7)
Female	67 (72.8)
Socio-professional status	
Student	52 (56.5)
Staff	40 (43.5)
Education level	
High school	4 (4.4)
Higher education	88 (95.6)
Age (year)*	30.1 ± 10.7
BMI (kg/cm^2)*	22.6 ± 3.5
BMI classes (kg/cm^2)*	
Underweight <18.5	3 (3.3)
Acceptable weight 18.5-24.9	71 (76.9)
Overweight 25.0–29.9	14 (15.4)
Obese >30	4 (4.4)

* Data are presented as mean ± SD.
Abbreviation: BMI = body mass index.

Table 2
Data for PA measured by the Global Physical Activity Questionnaire GPAQ, IPAQ and an accelerometer at day 0 (D0) and day 8 (D8) in 92 participants.

Variable	GPAQ				IPAQ				Accelerometer	
	D0		D8		D0		D8			
	Mean ± SD	Median	Mean ± SD	Median	Mean ± SD	Median	Mean ± SD	Median	Mean ± SD	Median
Total PA (MET min/week)	2011.1 ± 1940.5	1580.0	1818.0 ± 1478.2	40.7	2648.3 ± 2099.8	2251.5	2484.1 ± 2268.0	1777.5		
PA by domain										
Work										
Vigorous	31.3 ± 300.3	0	33.0 ± 230.6	0	34.8 ± 300.8	0	15.6 ± 85.7	0		
Moderate	467.4 ± 1575.3	0	321.1 ± 965.4	0	203.5 ± 758.1	0	212.4 ± 871.2	0		
Transport	375.9 ± 410.8	240.0	378.5 ± 426.2	250.0	306.8 ± 295.5	242.5	351.3 ± 414.0	260.7		
Household										
Vigorous	n/a	n/a	n/a	n/a	4.5 ± 35.4	0	22.1 ± 126.1	0		
Moderate	n/a	n/a	n/a	n/a	475.9 ± 785.2	150.0	356.7 ± 594.3	160.0		
Work + household										
Vigorous	n/a	n/a	n/a	n/a	39.3 ± 302.4	0	37.8 ± 171.8	0		
Moderate	n/a	n/a	n/a	n/a	695.0 ± 1080.1	240.0	596.1 ± 1189.1	190.0		
Leisure										
Vigorous	852.2 ± 1073.3	680.0	772.6 ± 955.9	480.0	868.7 ± 1085.9	600.0	691.3 ± 1011.5	0		
Moderate	284.3 ± 366.0	240.0	312.8 ± 382.9	240.0	193.9 ± 265.4	0	218.9 ± 415.4	340.0		
Sitting time (min/day)	570.0 ± 152.8	600.0	588.6 ± 146.4	600.0	554.5 ± 138.5	584.3	583.6 ± 143.2	597.1	843.6 ± 134.5	814.0
PA duration by intensity (min/week)										
Vigorous	883.5 ± 1090.1	720.0	805.6 ± 977.7	480.0	903.5 ± 1102.4	720.0	707.0 ± 1015.6	360.0	72.0 ± 67.2	46.7
Moderate	751.7 ± 1659.8	360.0	633.9 ± 990.3	360.0	903.8 ± 1131.4	480.0	860.6 ± 1266.9	370.0	426.2 ± 139.5	429.4
PA level (%)										
Low	29.4		22.8		8.7		15.2			
Moderate	44.6		45.6		60.9		54.3			
High	26.1		22.8		30.4		30.4			

Abbreviation: GPAQ = global physical activity questionnaire; IPAQ = international physical activity questionnaire; MET = metabolic equivalent task; n/a = not assessed by the questionnaire; PA = physical activity.

3.4. Concurrent validity

For both measurement times, we observed good correlations between the GPAQ and IPAQ for vigorous activity during leisure, total vigorous activity, and sitting time ($r = 0.76$–0.89) (Table 4). The values at D0 and D8 seemed almost identical, but important discrepancies were observed between vigorous work at D0 ($r = 0.58$, 95%CI: 0.43–0.70) and at D8 ($r = 0.81$, 95%CI: 0.73–0.87). Overall, total PA showed moderate

Table 3
Test-retest reliability of the GPAQ ($n = 68$).

Variables	ICC (95%CI)	Spearman's Rho (95%CI)	Kappa coefficient
Total PA	0.58 (0.40–0.72)	0.82 (0.72–0.88)	
PA by domain			
Work			
Vigorous			0.91(+)
Moderate	0.48 (0.28–0.64)	0.52 (0.33–0.68)	
Transport	0.67 (0.52–0.79)	0.69 (0.53–0.79)	
Leisure			
Vigorous	0.94 (0.91–0.96)	0.89 (0.84–0.94)	
Moderate	0.37 (0.15–0.56)	0.53 (0.33–0.68)	
Sitting time	0.80 (0.69–0.87)	0.78 (0.67–0.86)	
PA by intensity			
Total vigorous	0.84 (0.76–0.90)	0.80 (0.70–0.88)	
Total moderate	0.48 (0.28–0.65)	0.56 (0.38–0.71)	
PA level			
Low			0.62
Moderate			0.50
High			0.57

(+): Adjusted kappa (PABAK).
Abbreviations: GPAQ = global physical activity questionnaire; PA = physical activity; ICC = intraclass correlation coefficient; 95%CI = 95% confidence interval.

Table 4
Concurrent validity between the GPAQ and IPAQ-LF data at day 0 (D0) and day 8 (D8) ($n = 92$).

Variable	D0		D8	
	Spearman's Rho (95%CI)	Phi coefficient	Spearman's Rho (95%CI)	Phi coefficient
Total PA	0.66 (0.53–0.76)		0.67 ((0.54–0.77)	
PA by domain				
Work				
Vigorous	0.58 (0.43–0.70)		0.81 (0.73–0.87)	
Moderate	0.56 (0.40–0.68)		0.61 (0.46–0.72)	
Transport	0.52 (0.35–0.65)		0.69 (0.57–0.79)	
Leisure				
Vigorous	0.86 (0.79–0.90)		0.79 (0.70–0.85)	
Moderate	0.46 (0.28–0.61)		0.53 (0.36–0.66)	
Sitting time	0.85 (0.78–0.90)		0.89 (0.84–0.93)	
PA by intensity				
Total vigorous	0.86 (0.79–0.90)		0.76 (0.66–0.84)	
Total moderate	0.41 (0.22–0.56)		0.58 (0.42–0.70)	
PA level				
Low		0.22		0.49
Moderate		0.27		0.27
High		0.57		0.54

Abbreviations: GPAQ = global physical activity questionnaire; IPAQ-LF = international physical activity questionnaire-long form; PA = physical activity.

Please cite this article in press as: Fabien Rivière, et al., Reliability and validity of the French version of the global physical activity questionnaire, Journal of Sport and Health Science (2016), doi: 10.1016/j.jshs.2016.08.004

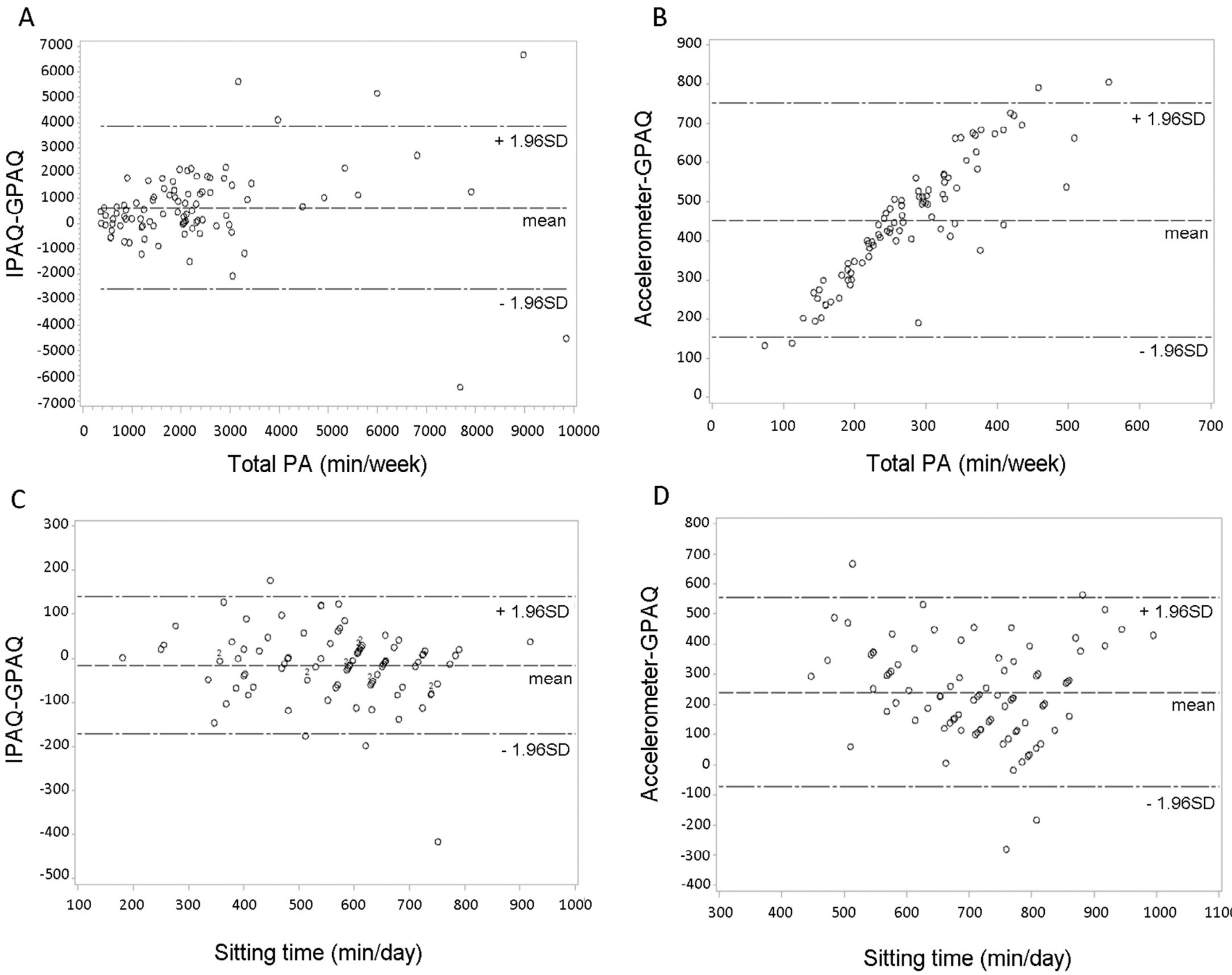

Fig. 1. Bland-Altman plots of the validity of the Global Physical Activity Questionnaire (GPAQ). A&C: Agreement of GPAQ with IPAQ for total PA(A), sitting time (B) at D0; B&D: Agreement of GPAQ with accelerometer for total PA (C), sitting time (D) at D8. IPAQ = international physical activity questionnaire; PA = physical activity.

correlation at both D0 ($r = 0.66$, 95%CI: 0.53–0.76) and D8 ($r = 0.67$, 95%CI: 0.54–0.77). Results of Bland-Altman analysis (Fig. 1A, C) for the GPAQ and IPAQ demonstrated a mean difference of 637.2 ± 1641.5 MET min/week. The limits of agreement for the 2 instruments were wide, with the difference between 1004.3 and 2580.1. For sedentary time, the mean difference of sedentary time was −15.5 ± 79.2 min/day. Overall, the classification by level of PA with the 2 questionnaires, at both times, was only poorly to moderately correlated, with a Phi coefficient ranged from 0.22 to 0.57.

3.5. Criterion validity

Accelerometer data were considered valid for 87 of the 92 participants (5 participants did not wear an accelerometer for at least 10 h per day over 7 days). Criterion validity was assessed by comparing total PA time spent in vigorous-intensity activity, or in moderate-intensity activity, or sitting per day reported with the GPAQ and derived from accelerometer counts. Poor but significant correlations for sedentary time ($r = 0.42$, $p < 0.01$) and total vigorous PA ($r = 0.38$, $p < 0.01$) were observed (Table 5).

Bland-Altman findings revealed that the GPAQ underreported total PA, with a mean difference between the GPAQ and accelerometer data of 443.95 ± 157.46 min/week (Fig. 1B, D). Limits of agreement for the 2 instruments were wide, with the difference between 286.5 and 601.3 min/week GPAQ underestimated sedentary time as compared with the accelerometer, with a mean difference between the 2 instruments of 251.2 ± 161.1 min/day. Limits of agreement for the 2 instruments ranged from 90.1 to 412.3 min/day.

4. Discussion

This study provides results, for the first time in a French population, for the reliability and validity of the GPAQ.

Table 5
Criterion validity of the GPAQ: Spearman's correlation between the GPAQ and accelerometer data at day 8 (D8) ($n = 87$).

GPAQ	Accelerometer		
	Average sedentary counts/day	Average moderate counts/day	Average vigorous counts/day
Total vigorous PA (min)	0.02	0.19	0.38**
Total moderate PA (min)	−0.20	0.10	−0.10
Total PA across all domains (min)	−0.20	0.40**	0.24*
Time spent sitting (min)	0.42**	−0.22*	0.30**

* $p < 0.05$,** $p < 0.01$, compared GPAQ with accelerometer's values.
Abbreviations: GPAQ = global physical activity questionnaire; PA = physical activity.

For reliability, we found poor to good correlation, with highest value obtained for vigorous leisure PA, which indicates the stability of this type of PA. This result is consistent with the findings by Matthews et al.[30] who observed no significant variation in vigorous leisure time activity over 1 year in 580 healthy adults.[30] Overall, our results are comparable to other studies testing the psychometric properties of the GPAQ. Herrmann et al.[13] demonstrated short- and long-term reliability with ICC values from 0.54 to 0.92. Bull et al.[11] reported test–retest correlation coefficients from 0.67 to 0.81 and kappa coefficients from 0.67 to 0.73 for pooled data.

Whereas Bull et al.[11] and Herrmann et al.[13] showed a poor to moderate correlation between the GPAQ and IPAQ (with coefficients from 0.45 to 0.57 and 0.26 to 0.63, respectively), our results indicate a poor to good concurrent validity. A reason of this difference could be the use, by the former studies, of the IPAQ short-form (IPAQ-SF) as compared with our use of the long form. Unlike the GPAQ and IPAQ-LF, which measure PA in different domains, the IPAQ-SF measures overall PA duration and frequency, which may explain the differences. In measuring the concurrent validity of the GPAQ, the IPAQ-LF may be more relevant than the IPAQ-SF. However, despite an acceptable concurrent validity, the agreement between the GPAQ and the IPAQ-LF to classify participants by PA levels was only poor to moderate (Phi coefficients 0.22 to 0.57), with the highest agreement attributable to high PA level. In addition, the Bland-Altman analysis revealed wide discrepancies in total PA measured by the 2 questionnaires, with a mean difference of 637.2 ± 1641.5 MET min/week. A possible explanation could be that the IPAQ-LF contains detailed items dedicated to household activities, whereas in the GPAQ, household activities are included in work activities. Also, the IPAQ-LF measures time spent walking, which is not considered by the GPAQ if it is not brisk walking (considered moderate activity). These differences may explain the gap in total PA measured by the 2 questionnaires. These results indicate the difficulty in comparing different questionnaires and thus the need to use the same questionnaire in a population surveillance study to be able to interpret the pattern of PA over the years.

A poor criterion-related validity for the GPAQ as compared with accelerometer data was shown. These results are comparable to Cleland et al.[12] and Bull et al.,[11] who demonstrated correlations with accelerometer data ranging from 0.19 to 0.48 and −0.20 to 0.40, respectively, whereas results from Hoos et al.[14] showed correlations from 0.32 to 0.52. According to Bland-Altman analysis, the GPAQ seems to underestimate total PA as compared with the accelerometer. This finding can be explained by the GPAQ including only PA that lasts at least 10 min, whereas the accelerometer measures all activities regardless of duration. This result was already found in studies comparing questionnaires to objective measures of PA.[31] In this study and according to Bland-Altman analysis, the GPAQ seemed to underestimate sedentary time as measured by the accelerometer. This finding can be justified most likely by difficulty to accurately recall sitting time as well as by a response bias due to social desirability, which may affect the degree of reporting the time spent sitting by subjects.[31] Future research is needed to identify whether a bias does exist and if so, whether it differs by gender or socioprofessional status, and to what extent.

This study had several strengths, beginning with the adherence to standardized WHO protocols in administering questionnaires (GPAQ was always administered before the IPAQ) and the concordant measurement period (the same 7 days) for both questionnaires and the accelerometer. Also, we used Bland-Altman analysis, a useful and recommended approach to assess the level of agreement, as compared with usual correlation coefficients assessing only the strength of the relationship between the measures.[27] Finally, the use of the IPAQ-LF seems relevant because it induced better concurrent validity with the GPAQ than in previous studies.

The major limitation of this study was the use of accelerometer as an alternative to the gold standard. However, in the absence of a gold standard, accelerometer may be used to measure PA in daily life.[32,33]

5. Conclusion

This study adds important and new information in testing the psychometric properties of the GPAQ in France. The results suggest that the GPAQ is a reliable questionnaire for use in the French population. The overall validity was poor to good but remained acceptable and was similar to previous studies.[11,12] Another important highlight is the need to use the same questionnaire in surveillance studies to allow for comparison and follow-up of the PA level of the study population and for PA surveillance in general.

Acknowledgments

This study was undertaken in the University of Lorraine. Interviews were conducted in a local area serviced by the faculty for this purpose. The authors thank the following for their assistance and contribution to the development and achievement of this research:

Marc Braun: Dean of the Faculty of Medicine of Nancy;

Nathalie Richard: HR Manager of the Faculty of Medicine of Nancy, who participated in the dissemination of the information message to staff;

Angelo Tonelli: Responsible for service real estate, furniture, maintenance and security of the Faculty of Medicine of Nancy, who ensured the place for interviews;

Lorent Phialy: Responsible for the publication on dynamic screens of the Faculty of Medicine of Nancy, who participated in the development of the video message;

Elisabeth Schmitt: Responsible for the education office of the Faculty of Medicine of Nancy, who participated in the dissemination of the information about the study to students;

All volunteers who participated in this study.

Authors' contributions

FZW, AV, ES, and FR participated in the design of the study. FZW and FR contributed to data collection. FZW and MLE contributed to data reduction/analysis. HE contributed to data analysis and interpretation of results. All authors contributed to the manuscript writing. All authors have read and approved the final version of the manuscript, and agree with the order of presentation of the authors.

Competing interests

None of the authors declare competing financial interests.

References

1. World Health Organization. Global strategy on diet, physical activity and Realth. In: *Proceedings of the 57th World Health Assembly*. Geneva: World Health Organization; 2004.
2. O'Donovan G, Blazevich AJ, Boreham C, Cooper AR, Crank H, Ekelund U, et al. The ABC of Physical Activity for Health: a consensus statement from the British Association of Sport and Exercise Sciences. *J Sports Sci* 2010;**28**:573–91.
3. Reiner M, Niermann C, Jekauc D, Woll A. Long-term health benefits of physical activity–a systematic review of longitudinal studies. *BMC Public Health* 2013;**13**:813. doi: 10.1186/1471-2458-13-813
4. Kujala UM. Evidence on the effects of exercise therapy in the treatment of chronic disease. *Br J Sports Med* 2009;**43**:550–5.
5. Pedersen BK, Saltin B. Evidence for prescribing exercise as therapy in chronic disease. *Scand J Med Sci Sports* 2006;**16**:3–63.
6. van Poppel MN, Chinapaw MJ, Mokkink LB, Van Mechelen W, Terwee CB. Physical activity questionnaires for adults. *Sports Med* 2010;**40**: 565–600.
7. Armstrong T, Bull F. Development of the world health organization global physical activity questionnaire (GPAQ). *J Public Health (Bangkok)* 2006; **14**:66–70.
8. Craig CL, Marshall AL, Sjöström M, Bauman AE, Booth ML, Ainsworth BE, et al. International physical activity questionnaire: 12-country reliability and validity. *Med Sci Sports Exerc* 2003;**35**:1381–95.
9. World Health Organization. *WHO STEPS surveillance manual: the WHO STEPwise approach to chronic disease risk factor surveillance*. Geneva: World Health Organization; 2005.
10. Thuy AB, Blizzard L, Schmidt M, Hung Luc P, Magnussen C, Dwyer T. Reliability and validity of the global physical activity questionnaire in Vietnam. *J Phys Act Health* 2010;**7**:410.
11. Bull FC, Maslin TS, Armstrong T. Global physical activity questionnaire (GPAQ): nine country reliability and validity study. *J Phys Act Health* 2009;**6**:790.
12. Cleland CL, Hunter RF, Kee F, Cupples ME, Sallis JF, Tully MA. Validity of the Global Physical Activity Questionnaire (GPAQ) in assessing levels and change in moderate-vigorous physical activity and sedentary behaviour. *BMC Public Health* 2014;**14**:1255. doi: 10.1186/1471-2458-14-1255
13. Herrmann SD, Heumann KJ, Der Ananian CA, Ainsworth BE. Validity and reliability of the global physical activity questionnaire (GPAQ). *Meas Phys Educ Exerc Sci* 2013;**17**:221–35.
14. Hoos T, Espinoza N, Marshall S, Arredondo EM. Validity of the global physical activity questionnaire (GPAQ) in adult Latinas. *J Phys Act Health* 2012;**9**:698–705.
15. Soo K, Manan WWA, Suriati WW. The bahasa melayu version of the global physical activity questionnaire reliability and validity study in Malaysia. *Asia Pac J Public Health* 2015;**27**:NP184–93.
16. Trinh OT, Do Nguyen N, van der Ploeg HP, Dibley MJ, Bauman A. Test-retest repeatability and relative validity of the Global Physical Activity Questionnaire in a developing country context. *J Phys Act Health* 2009; **6**(Suppl 1):S46–53.
17. Escalon H, Bossard C, Beck F, Bachelot-Narquin R. *Baromètre santé nutrition 2008*. Saint-Denis: INPES; 2009.
18. Arredondo EM, Mendelson T, Holub C, Espinoza N, Marshall S. Cultural adaptation of physical activity self-report instruments. *J Phys Act Health* 2012;**9**(Suppl 1):S37–43.
19. Terwee CB, Mokkink LB, van Poppel MN, Chinapaw MJ, van Mechelen W, de Vet HC. Qualitative attributes and measurement properties of physical activity questionnaires. *Sports Med* 2010;**40**:525–37.
20. Anthoine E, Moret L, Regnault A, Sébille V, Hardouin J-B. Sample size used to validate a scale: a review of publications on newly-developed patient reported outcomes measures. *Health Qual Life Outcomes* 2014; **12**:176. doi: 10.1186/s12955-014-0176-2
21. World Health Organization. Global. *physical activity questionnaire (GPAQ) analysis guide*. Geneva: World Health Organization; 2012.
22. Crinière L, Lhommet C, Caille A, Giraudeau B, Lecomte P, Couet C, et al. Reproducibility and validity of the French version of the long International Physical Activity Questionnaire in patients with type 2 diabetes. *J Phys Act Health* 2011;**8**:858–65.
23. Wyatt J. *ActiLife 6 user's manual*. ActiGraph Software Department; 2012.
24. Pedišić Ž, Bauman A. Accelerometer-based measures in physical activity surveillance: current practices and issues. *Br J Sports Med* 2015;**49**:219–23.
25. Matthews CE, Ainsworth BE, Thompson RW, Bassett DR Jr. Sources of variance in daily physical activity levels as measured by an accelerometer. *Med Sci Sports Exerc* 2002;**34**:1376–81.
26. Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Phys Ther* 2005;**85**:257–68.
27. Bland JM, Altman D. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;**327**:307–10.
28. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;159–74.
29. Portney LG, Watkins MP. *Foundations of clinical research: applications to practice (Vol. 2)*. Upper Saddle River, NJ: Prentice Hall; 2000.
30. Matthews CE, Freedson PS, Hebert JR, Stanek EJ, Merriam PA, Rosal MC, et al. Seasonal variation in household, occupational, and leisure time physical activity: longitudinal analyses from the seasonal variation of blood cholesterol study. *Am J Epidemiol* 2001;**153**:172–83.
31. Prince SA, Adamo KB, Hamel ME, Hardt J, Gorber SC, Tremblay M. A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. *Int J Behav Nutr Phys Act* 2008; **5**:56. doi: 10.1186/1479-5868-5-56
32. LaMonte MJ, Ainsworth BE. Quantifying energy expenditure and physical activity in the context of dose response. *Med Sci Sports Exerc* 2001; **33**(6 Suppl):S370–8.
33. Westerterp KR. Assessment of physical activity: a critical appraisal. *Eur J Appl Physiol* 2009;**105**:823–8.