Baltic Journal of Health and Physical Activity 2016;8(2):29-41
Journal of Gdansk University of Physical Education and Sport in Gdansk
e-ISSN 2080-9999

ORIGINAL

Authors' Contribution:
A Study Design
B Data Collection
C Statistical Analysis
D Data Interpretation
E Manuscript Preparation
F Literature Search
G Funds Collection

Reliability and validity of the International Physical Activity Questionnaire in Lithuania

Alvydas Kalvenas[1 ABDEFG], Ionuţ Burlacu[2 ACDEF], Karim Abu-Omar[2 ACDEF]

[1] Department of Health, Physical and Social Education, Lithuanian Sports University, Kaunas, Lithuania

[2] Institute of Sport Science and Sport, Friedrich Alexander University, Erlangen, Bavaria, Germany

abstract

Background Purpose of the study was to examine: (1) the criterion validity and test-retest reliability of the IPAQ-LT short-form (SF) and long-form (LF) and (2) its potential over-reporting and energy expenditure over-estimation.

Material/Methods 130 participants, aged 18 - 69 years, wore the Actigraph GT3X accelerometer (ACC) on all waking hours over 7 consecutive days. One day before and after they completed both versions of the back-translated IPAQ-LT. 92 participants were included for the reliability and 81 for the validity tests. Spearman's rho correlation coefficients were calculated as the measurement of agreement.

Results Only the walking category significantly (p <.05) correlated with the ACC, SF (.22), LF (.20). Compared with ACC data IPAQ-LT averaged 997% (SF) and 1512% (LF) more weekly minutes of PA and 864% (SF) and 1477% (LF) more MET-min/week. The classification of participants as sufficiently active was 87.6% (SF), 90.1% (LF), and 8.7% (ACC).

Conclusions The validity for total PA scored relatively low compared with other studies. Substantial PA over-reporting and EE over-estimation were observed. As such, the evidence is very weak to support the use of IPAQ- LT as a relative or an absolute measure of PA and further work in this regard is amended.

Key words physical activity, IPAQ, over-reporting, validity and reliability, surveillance, Lithuania

article details

Article statistics **Word count:** 4,755; **Tables:** 5; **Figures:** 0; **References:** 32
Received: May 2015; **Accepted:** March 2016; **Published:** June 2016

Full-text PDF: http://www.balticsportscience.com

Copyright © Gdansk University of Physical Education and Sport, Poland

Indexation: AGRO, Celdes, CNKI Scholar (China National Knowledge Infrastructure), CNPIEC, De Gruyter - IBR (International Bibliography of Reviews of Scholarly Literature in the Humanities and Social Sciences), De Gruyter - IBZ (International Bibliography of Periodical Literature in the Humanities and Social Sciences), DOAJ, EBSCO - Central & Eastern European Academic Source, EBSCO - SPORTDiscus, EBSCO Discovery Service, Google Scholar, Index Copernicus, J-Gate, Naviga (Softweco, Primo Central (ExLibris), ProQuest - Family Health, ProQuest - Health & Medical Complete, ProQuest - Illustrata: Health Sciences, ProQuest - Nursing & Allied Health Source, Summon (Serials Solutions/ProQuest, TDOne (TDNet), Ulrich's Periodicals Directory/ulrichsweb, WorldCat (OCLC)

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflict of interest: Authors have declared that no competing interest exists.

Corresponding author: Ionuţ Burlacu, Institut für Sportwissenschaft und Sport; Gebbertstraße. 123b D-91058 Erlangen, Germany; e-mail: ionut.burlacu@fau.de; phone: 0049 9131 85 25 456

Open Access License: This is an open access article distributed under the terms of the Creative Commons Attribution-Non-commercial 4.0 International (http://creativecommons.org/licenses/by-nc/4.0/), which permits use, distribution, and reproduction in any medium, provided the original work is properly cited, the use is non-commercial and is otherwise in compliance with the license.

INTRODUCTION

The measurement of physical activity (PA) is of crucial importance for the understanding of its determinants, evaluating the efficacy of interventions as well as informing health-policymakers [1]. In this regard, surveillance systems that allow national and international monitoring and comparisons are mandated [2, 3]. Consequently, standardised approaches to measuring, and collecting and analysing PA data are essential [3, 4]. One prominent instrument for PA surveillance is the International PA Questionnaire (IPAQ). After its development, both IPAQ's short (SF) and long form (LF) measurement properties have been investigated in numerous countries and its use has been recommended for surveillance purposes [5].

Since then, to the extent of our knowledge, the metric properties of the LF have been tested in few studies [5-7] and indicated comparable low correlations of the IPAQ-LF against accelerometer (ACC) data. The IPAQ - SF has been validated in a number of studies, and a recent review of 23 studies concluded that while its reliability has been proven, with respect to its validity, in the large majority of the studies (18) only small correlations with objective measures (ACCs) of activity were found for total PA (range of r = 0.09 to 0.39) while vigorous PA (range r = -0.18 – 0.47 – 9 out of 15 over r = 0.2) and walking (range r = 0.27 – 0.57 – 3 out of 4 over r = 0.4) showing some acceptable correlations [8]. Importantly, it has been noted that the issue of under- or over-reporting is rarely approached [8]. Also, when comparing across countries and studies, there is variability in the validity of the IPAQ-SF [5, 8], and as such testing for its intercultural applicability and adequacy has been called for [2, 5, 8].

Given the need of harmonization of health instruments in the European context [2, 9], the use of IPAQ for PA surveillance purposes across this region [10], the lack of its validation in the Lithuanian population and the restrictive use of leisure-time PA as an indicator in Lithuanian FinBalt based surveys [11], examining the reliability and validity of IPAQ for this nation is of crucial importance. Moreover, the importance of this study is accentuated by its comprehensive nature defined by the simultaneous validation of the SF and the LF, its focus on the over-reporting issue [8,12] and the classification of participants according to physical activity levels defined by the IPAQ scoring protocol [13]. Against this background, this study aims (1) to examine the criterion validity and test-retest reliability of the IPAQ-LTSF and LF and (2) to assess its potential over-reporting and EE over-estimation in a Lithuanian sample of healthy adults.

MATERIAL AND METHODS

PARTICIPANTS

The study was performed in 2010 and involved a convenience sample of 130 participants aged 18–69 years, living in an urban area and considering themselves healthy. Participants were recruited among the employees of a university and a private company. The ethical approval was obtained from the Ethics Supervision Committee in Social Sciences. Verbal consent was obtained from all subjects participating in the study.

PROCEDURE

Researchers and participants met twice. During the first visit participants completed both versions of IPAQ-LT along with indicators regarding their health and demographic information. The order of the SF and LF administration was randomized. Upon completion they each received an ACC and were instructed to wear it on their right hip (iliac crest) for seven consecutive days during all waking hours except when showering, bathing or swimming. The ACC was programmed to collect data 1-2 days after the first questionnaire administration starting at 00.00 A.M. The second visit occurred the day following the seventh ACC wearing day when the device was retrieved and both versions of IPAQ-LT were randomly administered.

INSTRUMENTS AND DATA PROCESSING

IPAQ-LT. Both versions of the IPAQ- LT were translated and back-translated following the recommendations of the IPAQ committee [14] and data scoring and processing following the IPAQ – Guidelines for Data Processing and Analysis [13].

GT3X ACTIGRAPH ACCELEROMETER

In line with most IPAQ validation studies [5–8] accelerometry was employed as the criterion measurement method. Data was collected using the uniaxial measurement of the tri-axial GT3X Actigraph Model (Actigraph, Pensacola, Florida, USA). The technical specifications of this ACC are presented elsewhere [15]. The ACC was programmed to record counts in 60 seconds intervals. Invalid data was represented as counts larger than 16,000/ min and treated as missing values [16].Non-wear time was established as 60 min of consecutive values of zero [16] with allowance of two activity counts between the zero values [17]. The ACC compliance requirements were: a minimum of 600 min. of wear time for establishing a valid day [16,17] with a minimum of 4 days for considering data to be valid [16]. Following the current PA recommendations [18,19], the IPAQ explicitly asks information regarding PA lasting at least 10 minutes in each activity intensity level. Thus, ACC data was operationalised as 10 or more consecutive epochs with allowance of a maximum two min dip under the cut-point [16]. The accelerometer data was downloaded using the ActiGraphActiLife v4.4.1 software firmware v3.1.1 and processed using the ActigraphActilife v6.7.1 software.

The cut-off points suggested by Freedson et al. [20] were employed. As the IPAQ - Guidelines for Data Processing and Analysis [13] classifies physical activities into walking, moderate, and vigorous intensity levels defined as EE ranging from 3.3 to 3.99 METs, 4.00 to 7.99 METs and ≥ 8 METs, the Freedson [20] cut-off points were modified by taking 1921 counts/min. as the 3.00 METs reference value and 1258 counts/min. as 1 MET change between 3 and 9 METs [21]. The following cut-off points emerged: walking 2328 to 3208 counts/min., moderate PA 3209 to 8240 counts/min., vigorous PA ≥ 8241 counts/min., and sedentary behaviour as <100 counts/min [17].

For ACC data, the total PA category was computed as all activity over 3.3 MET and not by summing the values of the walking, moderate and vigorous categories. Similarly, the walking + moderate category contained all PA between 3.3 MET and 7.99 MET (2328 to 8294 counts/min) stemming from a single measurement and not by adding the separate measurements of the walking

(3.3-3.99 MET) and the moderate (4.0–7.99 MET) categories. We opted to do so as, during the data processing phase, a difference in reported mean min/week (17.1) and estimated Met-min/week (195.6) between the two approaches was observed, the latter scoring lower.

RELIABILITY AND VALIDITY

The full completion of both questionnaires administration represented the selection criteria for the reliability analysis while for the validity test the participants also had to meet the ACC compliance requirements. As such, 92 (70.7%) of the total 130 participants were considered for reliability analysis and, as 11 participants did not meet the ACC compliance requirements, only 81 (62.3%) for the criterion validity one. Variables included in both tests were min/week spent in PA according to walking, moderate, vigorous, walking + moderate and total intensity categories. The sitting variable was expressed in weekly minutes during weekdays (SF) and complete week (LF).

STATISTICAL ANALYSIS

For the series of questions related to participants Body Mass Index (BMI), age, and subjective health status, descriptive statistics were used. BMI was computed as weight in kilograms divided by height in meters squared.

Data from IPAQ-LT and the ACC was non-normally distributed; therefore, Spearman's rank correlation coefficients were computed as the measurement of agreement between both administrations of IPAQ-LT and association between the ACC and the IPAQ-LT. Following Razali's et al. 2011 [22] study on the comparison of different normality tests, the Shapiro-Wilk test was considered appropriate for our sample size and thus employed. For all sets of variables involved in the reliability and validity assessments, the Shapiro-Wilks test found a significant (mostly at the $p < .001$ level) departure from normality. To quantify the magnitude of the Spearman correlation, the following thresholds were used (negligible (±) .00 to.30, low (±) .30 to .50, moderate (±) .50 to 70, high (±) .70 to .90 and very high (±) .90 to .100) [23]. The Statistical Product and Service Solutions version 16.0 (SPSS Inc. Chicago, IL, USA) was employed for all statistical analyses. The accepted level of significance was $p<0.05$.

MEASUREMENT BIAS

To assess the potential over-reporting and EE over-estimation, means and standard deviations of total and intensity specific time spent (min/week) in PA and PA EE estimation (Met-min/week) were computed. To determine whether the potential difference between the ACC's and IPAQ-LT's PA measurement and PA EE estimation was significant, the Wilcoxon signed-rank test was employed.

The sensitivity and specificity of IPAQ-LT was approached by assessing the distribution of participants according to the "low", "moderate" and "high" category levels compared with the ACC one. The category levels for both measurement instruments were computed following the IPAQ scoring protocol [13].

RESULTS

Table 1 presents the demographic and self-reported health status characteristics of the samples involved for the reliability (n = 92) and validity (n = 81) analysis.

Table 1. Description of the reliability and validity samples according to their size

	Reliability (n = 92)		Validity (n = 81)	
	n	%	n	%
Age (years)				
18-25	27	30.4	21	25.9
26-44	31	33.7	30	37.0
45-69	33	35.8	30	37.0
Gender				
male	29	31.5	23	28.4
female	63	68.4	58	71.6
BMI[1]				
underweight (<18.5)	4	4.3	32	3.7
normal (18.5 – 24.9)	54	58.7	49	60.5
overweight (25.0-29.9)	24	26.1	22	27.2
obese (> 30.0)	10	10.9	7	8.6
Subjective health status				
very good	10	10.9	8	9.9
good	60	65.2	54	66.7
moderate	22	23.9	19	23.5
bad	0	0	0	0
very bad	0	0	0	0

[1] BMI: body mass index computed weight (kg) / height (m)^2.

Test-retest reliability was examined by comparing the reported time spent in each PA intensity category of the two SF and LF administrations (Table 2). The Spearman correlation coefficients ranged from r =.46 to r = .70 for the SF and the LF revealing overall moderate reliability of the questionnaire items. The walking category correlated the lowest for both SF and LF, while the vigorous (SF) and the sitting categories (LF) the highest. All reported correlation coefficients were significant at the p >.001 level.

Table 2. Test–retest reliability of IPAQ-LT (SF/ LF) based on Spearman-rank correlation coefficients (n = 92)

	IPAQ-LT(SF) r[1] (p[1])	IPAQ-LT(LF) r[1](p[1])
Vigorous PA (min/week)	.67 (.00)***	.67 (.00)***
Moderate PA (min/week)	.53 (.00)***	.61 (.00)***
Walking (min/week)	.46 (.00)***	.49 (.00)***
Sitting (min/weekday/week[2])	.60 (.00)***	.70 (.00)***
Total PA (min/week)	.51 (.00)***	.56 (.00)***

*** Significant at p < .001 level.
[1] Values are r : correlation coefficients; (p): significance value.
[2] For (LF) sum sitting during weekdays and week-end days.

Table 3 presents the association between the IPAQ–LT SF and LF and the ACC measured PA expressed in min/week. For both versions of the IPAQ-LT only the walking category significantly positively correlated (p <.05) with the ACC the nature of the correlation being low (r. = .22 (SF); r. = .20 (LF). Additionally, the sitting category for the SF correlated low (p <.001) with the ACC (r = .28).

Table 3. Validity of IPAQ-LT (SF/LF) based on Spearman-rank correlation coefficients against accelerometry as the criterion measure (min/week) (n = 81)

	IPAQ-LT(SF)[4] r[1] (p[1])	IPAQ-LT(LF)[4] r[1](p[1])
Vigorous PA	.04 (.35)	.14 (.10)
Moderate + Walking PA[3]	.02 (.41)	-.14 (.09)
Moderate PA	-.03 (.37)	-.20 (.05)
Walking PA	.22 (.02)*	.20 (.03)*
Sitting [2]	.28 (.00)***	.17 (.06)
Total PA	-.11 (.14)	.02 (.40)

*** Significant at p < .001 level. * Significant at p < .05 level;
[1] Values are r: correlation coefficients; (p): significance value;
[2] For (LF) sum mean minutes sitting during weekdays and week-end days;
[3] For the ACC moderate + walking category comprised of a single measurement between 3.3 and 7.99 MET ; the total PA category is composed by a single measurement above 3.3;
[4] Modified Freedson et al. (1998) cut-points according to IPAQ scoring protocol (walking 3.3-3.99 MET, moderate 4-7.99 MET, vigorous 8 and above MET); PA recorded in bouts of 10 min with an allowance of 2 min. dip under the cut-point.

The total and intensity specific reported PA mean min/week and estimated EE mean MET-min/week obtained by ACC and IPAQ-LT are presented in Table 4. In comparison with the measurement by ACC, the IPAQ-LT reported an average of 552.4 (~10 fold) (SF) and 838.1 (~15 fold) (LF) more min/week of total PA. Regarding PA intensity, categories the IPAQ reported in average with 121.3 (~101 fold) (SF) and 63.6 (53 fold) (LF) more min/week of vigorous PA, 126.1 (~4 fold) (SF) and 371.1 (12 fold) (LF) of moderate, 322.1 (~53 fold) (SF) and 419.5 (~70 fold) (LF) of walking. When the walking and moderate categories were added, the IPAQ-LT averaged with 435.8 (~10 fold) (SF) and 773.1 (~14 fold) (LF) more min/week. Similarly, the IPAQ-LT substantially overestimated total PA EE averaging 2171 (~9 fold) (SF) and 3708.4 (~15 fold) (LF) more MET-min/week when compared with ACC. Regarding PA intensity, categories IPAQ-LT estimated in average 814.4 (~70 fold) (SF) and 661.3 (~56 fold) (LF) more MET-min/week for the vigorous category, 411.8 (~3 fold) (SF) and 1836.6 (~12 fold) (LF) for the moderate, and 1019.3 (~46 fold) (SF) and 1302 (~60 fold) (LF) for walking. When the walking and moderate categories were added, the IPAQ-LT estimated in average 1355.6 (~6 fold) (SF) and 3055.6 (~12 fold) (LF) more MET-min/week.

Table 4. Total and intensity specific reported mean minutes and estimated MET-minutes per week of physical activity by IPAQ-LT(SF and LF)and measured by ACC (n = 81)

		ACC [2]	IPAQ-LT (SF)	P[3]	IPAQ-LT (LF)	P[4]
Vigorous	(min/week)[1]	1.2 ± 8.4	122.5 ± 204.8	.00***	64.8 ± 189.7	.00***
	(MET-min/week)[1]	11.7 ± 79.0	826.1 ± 1308.3	.00***	673.0 ± 1618.5	.00***
Moderate +	(min/week)[1]	55.5 ± 83.1[5]	491.3 ± 472.6	.00***	828.6 ±851.7	.00***
Walking	(MET-min/week)[1]	240.3 ± 364.0[5]	1595.9 ± 1489.7	.00***	3295.9 ± 2756.6	.00***
Moderate	(min/week)[1]	31.0 ±56.3	157.1 ± 225.6	.00***	403.1 ±537.5	.00***
	(MET-min/week)[1]	142.7 ± 262.4	554.5 ± 741.4	.00***	1978.6 ± 1982.3	.00***
Walking	(min/week)[1]	6.0 ± 16.8	328.1 ± 380.7	.00***	425.5 ± 443.8	.00***
	(MET-min/week)[1]	22.0 ± 61.9	1041.3 ± 1188.7	.00***	1324.0 ± 1321.4	.00***
Total	(min/week)[1]	55.4 ± 75.9[5]	607.8 ±579.2	.00***	893.5 ± 941.4	.00***
(VMW)	(MET-min/week)[1]	251.0 ± 356.0[5]	2422.0 ± 2340.2	.00***	3959.4 ± 3938.2	.00***

*** Significant at $p < .001$ level.
[1] Values are mean ± standard deviation;
[2] Modified Freedson et al. (1998) cut-points according to IPAQ scoring protocol (walking 3.3-3.99 MET, moderate 4-7.99 MET, vigorous 8 and above MET); PA recorded in bouts of 10 min.with an allowance of 2 min. dip under the cut-point;
[3,4] Between group p-value determined by the Wilcoxon rank-test: ACC and SF[3]; ACC and LF[4];
[5] For the ACC the moderate. + walking category is composed by a single measurement between 3.3 and 7.99 MET; the total PA category is composed by a single measurement above 3.3 MET.

Table 5 presents the classification of participants according to their PA level by SF, LF and ACC and categorised in concordance with the IPAQ scoring protocol [13]. Both IPAQ-LT SF (87.6%) and LF (90.1%) classified sufficiently active participants differently when compared with the classification by ACC (8.7%). With respect to activity levels, both versions of IPAQ classified participants differently when compared to ACC with 46.9% (SF), 49.4% (LF) and 0% (ACC) of the sample as highly active, 40% (SF and LF) and 8.7% (ACC) as moderately active and 12.3% (SF), 9.9% (LF) and 91.3% (ACC) as insufficiently active.

Table 5. Distribution of study participants based on the IPAQ physical activity category levels by self-report and accelerometry (n = 81)

	IPAQ-LT(SF)[1]	IPAQ-LT(LF)[1]	ACC[1]
High[2]	38 (46.9)	40 (49.4)	0 (0.0)
Moderate[2]	33 (40.7)	33 (40.7)	8 (8.7)
Low[2]	10 (12.3)	8 (9.9)	73 (91.3)

[1] Values are n (%) of participants;
[2] Levels of physical activity according to the IPAQ scoring protocol.

DISCUSSION

To the extent of our knowledge, this is the first study to examine the reliability and validity of the IPAQ in a Lithuanian sample as well as going beyond that to address the issue of over-reporting and EE over-estimation and further physical activity level classification comparison according to the IPAQ scoring protocol [13].

RELIABILITY

The Spearman correlation coefficients stemming from the test-retest reliability analysis of the IPAQ-LT were generally smaller when compared with other IPAQ validation studies [5, 7, 8] evidentiating only a moderate reliability of the questionnaires. Concerning total PA, the SF scored lower (r = .51) when compared with the studies of Craig et al. [5] (r = .66) and Macfarlane et al. [21] (r = .89). Similarly, the LF scored moderate correlation coefficients for total PA (r = .56) still lower than the studies of Macfarlane et al. [7] (ICC = .79), and Craig et al. (r = 0.70–0.91) [5]. One possible explanation lies in the fact that, whilst our reliability analysis used min/week data, most of the above mentioned studies used MET-min/week to express the intensity of PA which inherently requires further data processing and thus potential alteration in the direction of diminishing accuracy. Secondly, the comparison between studies might also have been influenced, as Hallal et al. [24] notes, by the statistical test employed, as some studies used ICC [21]while others, like ours, Spearman correlation coefficients [5, 23, 24].

Similarly to the study by Craig et al. [5], the LF seems to be more reliable than the SF in measuring total PA. The reliability of the intensity specific categories for both SF and LF ranged from highest r = .67 (vigorous) to lowest r = .46 (moderate) with the LF correlating higher for moderate and walking when compared to SF. This is comparable with other SF studies where the vigorous category seems to be more reliable than the walking and moderate categories [2], while different from the LF study by Macfarlane et al. [7], where the walking category was the most reliable and the moderate one the least. Furthermore, in contrast with other studies [21] IPAQ-LT seems to be more reliable in reporting intensity specific sub-categories than total PA. Moreover, considerable variations in the reliability of IPAQ-LT sub-categories have been observed. This could be due to measurement error, true variation in activity or both [21] and, potentially, by the fact that with rising intensity of activity participants might recall the time and days when this occurred more precisely. For example, the frequency of engaging into vigorous PA might be lower than moderate and/or walking, as it is more often planned and easier to recall.

VALIDITY

Overall, the Spearman correlation coefficients between IPAQ-LT and the ACC were quite poor. The total PA correlation coefficients for both SF (r = -.05) and LF(r = -.10) were negligible, while only the walking PA category significantly ($p < .005$) correlated with the ACC, the nature of it being low for both SF (r = .22) and LF (r = .20). Additionally, the SF sitting category significantly ($p < .001$) correlated with the ACC the value being also low (r = .28).

Our results scored comparably lower correlation coefficients than the ones stemming from a recent SF review [8] and several LF studies [5–7] where almost all found a small degree of correlation against some form of objectively measured total PA (mostly ACCs). When discussing intensity categories, regarding the SF, only one study included in the review by Lee et al. [8] found walking as being associated with ACC, the correlation being higher r = .42 (male)/.57 (female) [25]. As for the LF, the above mentioned outcome was also reported by Macfarlane et al., [7] the correlation coefficient being (r = .21) similar. In contrast with the above mentioned studies [5–8], the vigorous and moderate PA categories of both IPAQ-LT versions did not significantly correlate with the ACC. One possible explanation for this lies in a different methodological approach employed by our study, where the ACC cut-off points proposed by Freedson et al. [20] were modified to encompass vigorous (≥ 8MET) and moderate (4–7.99 MET) PA in line with the IPAQ scoring protocol [13] which, to the extent of our knowledge, provides reasoned methodological novelty with respect to intensity sub-category validation. A second explanation could be attributed to the fact that we took in consideration only the registered ACC data occurring in at least 10 minute bouts (with an allowance of 2 min. dip under the cut-point), which created two discrepancies: 1) between the number of participants reporting intensity specific PA with IPAQ-LT (81) and those registering activity when measured with the ACC (i.e. 2 for vigorous, 45 for moderate, 20 for walking and 60 for total PA) and 2) between the values of the intensity specific PA reported by each participant with IPAQ-LT and those measured with the ACC (e.g. for ACC assessed vigorous PA only 2 participants registered min/week of activity the values for the rest being assumed to be 0).

OVER-REPORTING AND EE OVER-ESTIMATION

It had been argued [8] that very few studies go beyond a correlation-based criterion validity analysis to also comprise an accuracy evaluation, i.e. comparisons of absolute values stemming from IPAQ and objectively measured PA. In doing so, this study assessed the potential mean bias for overall and intensity specific time spent in PA (min/week) and estimated EE (MET-min/week) of the IPAQ-LT (SF and LF) when compared with accelerometry. Both the SF and LF substantially over-reported the total mean time spent in PA and over--estimated the mean EE when compared with the ACC. In comparison with over-reporting values stemming from five SF studies averaging 106% (36%–173%) [8], more mean min/week of total PA, the IPAQ –LT (SF) substantially over-reported total PA (~15 fold) and, with respect to intensity categories, the largest measurement discrepancy being observed for the vigorous PA category. Comparably, two other studies [21, 26] found the SF vigorous category to report the largest discrepancy in mean min/week of PA when compared with ACC. One possible explanation for the large discrepancies in reporting intensity specific min/week of PA and their respective EE estimations lies in the above mentioned methodological novelty regarding the adjustment of cut-off points according to the IPAQ scoring protocol. Secondly, taking into consideration only the registered ACC data occurring in at least 10 minute bouts (with an allowance of 2 min. dip under the cut-point) influenced the total and intensity specific PA prevalence in the direction of diminishing the values when compared to minute by minute ACC total and intensity specific PA prevalence. For example, the vigorous PA registered minute by minute with the ACC was 22 mean min/week, now only ~5 fold (456.8%) more than those by IPAQ-LT (SF),~2 fold (194.5%) more than those by IPAQ-LT (LF) and ~18 fold less (1733%) less than the mean min/week by ACC registered as 10 min bouts. Similarly, the moderate PA registered minute by minute with the ACC was 97 mean min/week, now only ~1 fold (61.9%) more than those by IPAQ-LT (SF), ~3 fold (315.5 %) more than those by IPAQ-LT (LF) and ~3 fold less (312.9%) less than the mean min/week by ACC registered as 10 min bouts. The walking PA registered minute by minute with the ACC was 88.7 mean min/week, now only ~3 fold (269.8%) more than those by IPAQ-LT (SF), ~4 fold (379.7 %) more than those by IPAQ-LT (LF) and ~15 fold (1478.3%) less than the mean min/week by ACC registered as 10 min bouts. Finally, the total PA registered minute by minute with the ACC was 88.7 mean min/week, ~4 fold (379.7 %) more than those by IPAQ-LT (LF), ~15 fold (318.5%) less than the mean min/week by ACC registered as 10 min bouts and ~2 fold (244.3%) more than those by IPAQ-LT (SF) value, which, surprisingly, is similar to the range of over-reporting mentioned in the review by Lee et al. [8] Thirdly, it is difficult to pinpoint the real magnitude of over-reporting and EE over-estimation with IPAQ when compared to ACC due the usual under-estimation of the latter [27]. Despite this shortcoming, it is hard to imagine that such large discrepancies might mainly arise from this characteristic of accelerometry as, when compared to doubly labelled water, the employed Freedson [20] equation underestimates 24-h PA by only 59% [28].

In this vein, we conclude that IPAQ-LT substantially over-reports time spend in PA as well over-estimates EE. Consequently, when compared with ACC, it differs in classifying respondents according to activity levels with both forms classifying around 90% of the sample as being sufficiently active whereas the ACC under 10% the largest discrepancy being for the highest category, followed by the low and moderate ones. In terms of its sensitivity IPAQ-LT behave perfectly in identifying 100% of those participants being sufficiently active but very poor, SF (13.7%) and LF (10.9%), in doing so for those insufficiently active (specificity). In this regard, our results come in agreement with other studies reporting on the sensitivity and specificity with IPAQ, Macfarlane et al. [21] reporting a sensitivity of 90% and a specificity of 29% with the LF while Ekelund et al. [29] reporting a 77% sensitivity in depicting sufficiently active individuals and a 45% specificity in screening insufficiently active ones with the SF.

The implications of these findings are as follows. For one, they stem as being crucial for PA surveillance and further policy-making as the IPAQ has been recommended and used in PA surveillance systems (e.g. Eurobarometer, World Health Survey) [10] and other studies [30]. In this regard, at least to some concern, the IPAQ-LT seems to behave similarly across-nations in terms of its reliability though, in line with the conclusion by Lee et al. [8], its validity scored negligible or low and significant only for the walking category. Furthermore, the IPAQ-LT seems to over-report and over-estimate EE fact that might have paramount implications in terms of depicting accurate PA levels. One such example might arise by comparing the results of our study to the one of Bauman et al. [30], where IPAQ-SF similarly categorised the Lithuanian representative population based sample (n = 2227) the difference being in that 15% (12.3% IPAQ-LT) of the sample was categorised as insufficiently active and 85% (87.7% IPAQ-LT) as sufficiently. Although this comparison might strengthen the idea regarding IPAQ producing consistent results [31], the implications might be more severe in terms of depicting real PA levels and thus potentially diminish the magnitude of the issue. In this regard, we conclude that IPAQ-LT might indeed over-estimate PA levels and further research to adjust for this is needed. Moreover, despite its high sensitivity, the IPAQ--LT seems not to behave as accurately in terms of its specificity. This could have greater implications for public health practice and research in terms of identifying individuals most in need of PA as well as assessing trends and potential changes in PA behaviour. Secondly, with respect to its usefulness for national PA surveillance in Lithuania the IPAQ-LT, despite its moving beyond a leisure-time PA indicator as used by the FinBalt Monitor [11], the evidence is very weak to support the use of IPAQ-LT as either a relative or an absolute measure of PA and further work in this regard is amended. Finally, given the multitude of methodological approaches across IPAQ validation studies and their inherent implications with respect to inter-study comparability [8], future work towards harmonization of validation protocols should be amended. In this regard we recommend that future ACC based equations employed in classifying PA according to intensity categories should be modified in accordance to the IPAQ protocol classification. Furthermore, when validating the total and moderate and walking categories as an aggregate indicator, we recommend the inclusion of the whole spectrum of activity when creating the moderate plus walking (3.3 MET to 7.99 MET) and the total PA (over 3.3 MET) categories and not generate them as a product of the separate sub-category-values (see Table 4). This recommendation is supported by the difference between the two approaches mentioned in the methods section and observed while processing ACC data (Table 4). Finally, for increased accuracy when examining the metric proprieties of IPAQ we recommend the less processed min/week when measuring the association of IPAQ, at least against ACC.

LIMITATIONS

The first series of limitations of our study lie in the healthy, high socio-economic background, solely urban based and convenience nature of our sample and as such it may not be representative for the Lithuanian population and might reflect a higher-level of self-selection. Despite this, our sample size was similar to most healthy adult population samples of other SF [8] IPAQ validation studies and generally larger than most LF ones [5–7]. Furthermore, the sample was quite diverse in terms of age groups, although, over two-thirds were females and all participants were residing in one of the largest urban areas in Lithuania. Secondly, although ACCs are the most commonly used tools to examine the measurement properties of IPAQ [8], they come with their inherent limitations of capturing only ambulatory activity [26] out of which some activities are not assessed (e.g. cycling, upper-body movements, swimming) [32]. Furthermore, the regression equation defining activity intensity cut-off points used in this study is based on laboratory studies on U.S. population [20] and nourishing a series of limitations with respect to their use for assessing free-living PA, its tendency to over-estimate energy costs of walking while under-estimate all other activities [27] as well as its appropriateness for use in the Lithuanian population. Fourthly, walking was assessed by ACC in terms of its EE value (3.3.MET) provided by the IPAQ scoring protocol [13] and not by step count assessment, which might have influenced the real time spent in this intensity category due to the potential inclusion of other activities with the same EE values.

CONCLUSION

This study explored the measurement properties of the IPAQ with respect to reliability, validity, over-reporting and EE estimation, compared PA prevalence with IPAQ against accelerometry and touched upon issues of sensitivity and specificity. Regarding the reliability analysis IPAQ scored moderate correlation coefficients for total PA and across intensity levels for both short and long instruments. With respect to its validity, for both versions of IPAQ-LT only the walking category significantly correlated with the ACC, the nature of the correlation being negligible. Substantial over-reporting and EE over estimation were also observed for overall and intensity specific time spent in PA (min/week) and estimated EE (MET-min/week) of the IPAQ-LT (SF and LF). Furthermore, the IPAQ-LT seems to behave well in identifying sufficiently active participants though quite poorly with respect to its specificity. Against this background, we conclude that the evidence is quite weak to support the use of IPAQ-LT as either a relative or an absolute measure of PA and further work in this regard is amended both with respect to adjusting the measurement properties of the IPAQ as well as in the direction of accelerometer based validation.

REFERENCES

[1] Bauman A, Phongsavan P, Schoeppe S, et al. Physical activity measurement – a primer for health promotion. Promot Educ. 2006;13:92-103.

[2] Rütten A, Vuillemin A, Ooijendijk WTM, et al. Physical activity monitoring in Europe. The European Physical Activity Surveillance System (EUPASS) approach and indicator testing. Public Health Nutr. 2003;6:377-384.

[3] World Health Organization. Global Strategy on diet, physical activity, and health. Geneva: WHO, 2004. Available at http://www.who.int/dietphysicalactivity/strategy/eb11344/strategy_english_web.pdf

[4] Rütten A, Rzewnicki R, Ziemainz H, et al. Towards a European Health Monitoring System. Results of a pilot study on physical activity. In: McQueen D and Puska P, eds. Global Behavioral Risk Factor Surveillance. New York: Kuwer Academic/ Plenum Publishers; 2003, 73-92.

[5] Craig CL, Marshall AL, Sjöström M, et al. International Physical Activity Questionnaire: 12-country reliability and validity. Med Sci Sports Exerc. 2003;35:1381-1395.

[6] Boon RM, Hamlin MJ, Steel GD, et al. Validation of the New Zealand Physical Activity Questionnaire (NZPAQ-LF) and the International Physical Activity Questionnaire (IPAQ-LF) with accelerometry. Br J Sports Med. 2010;44:741-746.

[7] Macfarlane DJ, Chan A, Cerin E. Examining the validity and reliability of the Chinese version of the International Physical Activity Questionnaire, long form (IPAQ-LC). Public Health Nutr. 2011;4:443-450.

[8] Lee PH, Macfarlane DJ, Lam TH, et al. Validity of the International Physical Activity Questionnaire short form (IPAQ-SF): A systematic review. Int J BehavNutr Phys Act. 2011;8:115-126.

[9] Bullinger M. International comparability of health interview surveys: An overview of methods and approaches. In: Nosikov A, Gudex C, eds. EUROHIS: Developing common instruments for health surveys. Amsterdam: IOS Press; 2003, 1-12.

[10] World Health Organization. Review of physical activity surveillance. Copenhagen: WHO Regional Office for Europe, 2010.
Available at http://www.euro.who.int/__data/assets/pdf_file/0005/148784/e95584.pdf

[11] Puska P, Helasoja V, Prättälä R. Health behaviour in Estonia, Finland and Lithuania 1994–1998 standardized comparison. Eur J Public Health. 2003;13:11-17.

[12] Rzewnicki R, Auweele YV, De Bourdeaudhuij I. Addressing overreporting on the International Physical Activity Questionnaire (IPAQ) telephone survey with a population sample. Public Health Nutr. 2003;6:299-305.

[13] International Physical Activity Questionnaire. Guidelines for Data Processing and Analysis of the International Physical Activity Questionnaire (IPAQ) – Short and Long Forms, 2005.
http://www.ipaq.ki.se/scoring.pdf [accessed 15 February 2014].

[14] International Physical Activity Questionnaire. Cultural Adaptation, 2005.
http://www.ipaq.ki.se/cultural.htm [accessed 15 February 2014].

[15] Santos-Lozano A, Santín-Medeiros F, Cardon G, et al. Actigraph GT3X: Validation and determination of physical activity intensity cut points. Int J Sport Nutr. 2013;34:975-982.

[16] Masse LC, Fuemmeler BF, Anderson CB, et al. Accelerometer data reduction: a comparison of four reduction algorithms on select outcome variables. Med Sci Sports Exerc. 2005;37:S544-S554.

[17] Matthews CE, Chen KY, Freedson PS, et al. Amount of time spent in sedentary behaviors in the United States, 2003–2004. Am J Epidemiol. 2008;167:875-881.

[18] Haskell WL, Lee IM, Pate RR, et al. Physical activity and public health: updated recommendation for adults from the American College of Sports Medicine and the American Heart Association. Circulation. 2007;116:1081-1093.

[19] World Health Organization. Global Recommendations on Physical Activity for Health. Geneva: WHO, 2010. Available at http://whqlibdoc.who.int/publications/2010/ 9789241599979_eng.pdf.

[20] Freedson PS, Melanson E, Sirard J. Calibration of the Computer Science and Applications, Inc. accelerometer. Med Sci Sports Exerc. 1998;30:777-781.

[21] Macfarlane DJ, Lee CC, Ho EY. Reliability and validity of the Chinese version of IPAQ (short, last 7 days). J Sci Med Sport. 2007;10:45-51.

[22] Razali NM, Wah YB. Power Comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling Tests. J Stat Model Anal. 2011;2:21-33.

[23] Mukaka MM. A guide to appropriate use of correlation coefficient in medical research. Malawi Med J. 2012;24(3):69-71.

[24] Hallal PC, Victora CG. Reliability and validity of the International Physical Activity Questionnaire (IPAQ). Med Sci Sports Exerc. 2004;36:556-556.

[25] Kolbe-Alexander TL, Lambert EV, Harkins JB, et al. Comparison of two methods of measuring physical activity in South African older adults. J Aging Phys Act. 2004;14:99-114.

[26] Dinger MK, Behrens TK, Han JL. Validity and reliability of the International Physical Activity Questionnaire in college students. Am J Health Edu. 2006;37:337-343.

[27] Crouter SE, Churilla JR, Bassett Jr, DR. Estimating energy expenditure using accelerometers. Eur J Appl Physiol. 2006;98:601-612.

[28] Leenders NY, Sherman WM, Nagaraja HN. Energy expenditure estimated by accelerometry and doubly labeled water: do they agree? Med Sci Sports Exerc. 2006;38:2165-2172.

[29] EkelundU, Sepp H, Brage S. Criterion-related validity of the last 7-day, short form of the International Physical Activity Questionnaire in Swedish adults. Public Health Nutr. 2006;9:258-265.

[30] Bauman A, Bull F, Chey T, et al. The international prevalence study on physical activity: results from 20 countries. Int J Behav Nutr Phys Act. 2009;6:21.

[31] Guthold R, Ono T, Strong KL, et al. Worldwide variability in physical inactivity: a 51-country survey. Am J Prev Med. 2008;34:486-494.

[32] Freedson P, Pober D, Janz KF. Calibration of accelerometer output for children. Med Sci Sports Exerc. 2005;37:S523-530.

Cite this article as:
Kalvenas A, Burlacu I, Abu-Omar K. Reliability and validity of the International Physical Activity Questionnaire in Lithuania. Balt J Health Phys Act. 2016;8(2):29-41.