Original article (short paper)

Mode of administration does matter: comparability study using IPAQ

Felipe de Magalhães Bandeira
Matheus Pintanel Freitas
Mitzi László
Marcelo Cozzensa da Silva
Pedro Curi Hallal
Airton José Rombaldi

Universidade Federal de Pelotas, Pelotas, RS, Brazil

Abstract—This study compared all-domains and domain-specific physical activity scores assessed through four variations of the IPAQ long version: (a) typical week, administered by an interviewer; (b) typical week, self-administered; (c) past seven days, interviewer-administered; (d) past seven days, self-administered. The sample included 38 physical education college students. Self-reported scores were in general twice higher than interview-administered scores, regardless the recall period used. In terms of domain-specific scores, occupational physical activity scores generated by self-report were 6-7 times greater than those originated from interviews. The same trend was observed for household physical activity. Transport physical activity scores did not change according to the mode of administration. In terms of leisure-time physical activity, scores were similar except for the interviewer-administered past seven days, whose scores were lower than the other three versions of IPAQ. In conclusion, the mode of administration of IPAQ does matter; higher scores are obtained through self-report as compared to interviews, probably by misinterpretation of the instrument in self-report mode. The recall period had little effect on physical activity estimates.

Keywords: physical activity, physical activity assessment, validity of tests, reproducibility of results, questionnaires

Resumo—“O modo de administração faz diferença: estudo de comparação usando o IPAQ.” Este estudo comparou o escore total e separado por domínio de atividade física obtidos através de quatro variações do IPAQ versão longa: (a) semana típica, administrado por entrevistador; (b) semana típica, autopreenchido; (c) últimos sete dias, administrado por entrevistador; (d) últimos sete dias, autopreenchido. A amostra incluiu 38 estudantes universitários de educação física. Os escores autopreenchidos eram, em geral, duas vezes maiores do que por entrevista, independentemente do período recordatório. No que se refere aos escores por domínios, o domínio de atividade física ocupacional autopreenchido obteve escores 6-7 vezes maiores do que os originados por entrevista. O mesmo padrão foi observado no domínio de atividades domésticas. Os escores das atividades de deslocamento não modificaram de acordo com o modo de administração. Em relação às atividades de lazer, os escores foram similares, com exceção dos últimos sete dias administrado por entrevistador, que foram inferiores aos outros três. Em conclusão, pontuações mais altas são obtidas através do autopreenchimento em comparação às entrevistas, provavelmente por erro de interpretação no modo autopreenchimento. O período recordatório teve pouco efeito sobre os escores de atividade física.

Palavras-chave: atividade física, avaliação da atividade física, validade dos testes, reproduzibilidade dos resultados, questionários

Resumen—“El modo de administración hace diferencia: estudio comparativo utilizando el IPAQ.” Este estudio comparó la puntuación total y las puntuaciones aisladas de los dominios de actividad física evaluados a través de cuatro variaciones del IPAQ versión larga: (a) semana típica, administrado por un entrevistador; (b) semana típica, autoadministrado; (c) últimos siete días, entrevistador; (d) últimos siete días, autoadministrado. La muestra incluyó a 38 estudiantes universitarios de educación física. Puntajes de auto-reporte eran, en general, dos veces más alto que entrevista, sin tener en cuenta el período de recuerdo. En cuanto a las puntuaciones de los dominios aislados, las puntuaciones de actividad física en el trabajo generados por su propio informe eran 6-7 veces mayores que las originadas a partir de entrevistas. Se observó la misma tendencia para la actividad física de los hogares. Las puntuaciones de actividad física de transporte no cambiaron de acuerdo con el modo de administración. En términos de actividad física en el tiempo libre, las puntuaciones fueron similares a excepción de los últimos siete días administrados por entrevistador, que eran inferiores a los otros tres. En conclusión, puntuaciones más altas se obtienen a través del auto-informe, en comparación con las entrevistas, probablemente por la mala interpretación del instrumento en el modo de auto-informe. El periodo de recuerdo tuvo poco efecto sobre las estimaciones de actividad física.

Palabras claves: actividad física, evaluación de la actividad física, validez de las pruebas, reproducibilidad de resultados, cuestionarios
Introduction

Alongside urbanization, economic development, technological advance, globalization and the consequent stress of modern life, societies have undergone many changes in the patterns of morbidity and mortality, leading to a new epidemiological profile of disease, now characterized by a high burden of non-communicable diseases (NCDs) (World Health Organization [WHO], 2010). Physical inactivity is one of the major risk factors for NCDs (WHO, 2011). A 2012 study showed that if a quarter of physical inactivity was eliminated, 1.3 million deaths could be prevented every year worldwide (Lee et al., 2012).

Therefore, monitoring physical activity at the population level is essential for designing appropriate policies for behavior change. However, measuring population physical activity levels is challenging (Gabriel, Morrow, & Woolsey, 2012). More objective techniques, such as accelerometry, double labeled water, calorimetry and heart rate monitoring are usually too expensive for use in large-scale studies (Tremblay, 2010) particularly in low and middle income countries. Therefore, questionnaires (Tremblay, 2010) continue to be the most frequent instruments used for the assessment of physical activity, particularly for surveillance purposes.

Questionnaires have been widely used to measure population physical activity in Brazil (Hallal, Matsudo, & Farias Jr, 2012; Rombaldi, Menezes, Azevedo, & Hallal, 2010; Siqueira et al., 2011) and internationally. Using self-report physical activity data from 122 countries, Hallal and colleagues showed that 1/3 of the adults worldwide are physically inactive (Hallal et al., 2012), strengthening the message that the use of reliable questionnaires is essential for surveillance purposes (Siqueira et al., 2011).

A series of validation studies so far assessed the validity of the International Physical Activity Questionnaire - IPAQ (Craig et al., 2003). Because there are different ways of applying IPAQ (e.g., length of IPAQ - short vs. long; reference period - last 7 days vs typical week; mode of administration - interviewer vs self-reported) (Kim, Park, & Kang, 2012; Craig et al., 2003) and this questionnaire is used worldwide (Hallal et al., 2012, 2007) it is important to check for agreement among these strategies. Previous studies on this topic are available (Farias Jr, Siqueira, Nahas, & Barros, 2011; Kim et al. 2012). Kim et al. (2012) examining the convergent validity of IPAQ showed no differences in physical activity levels according to the recall period used, but detected differences between self-report and interviewer-administered versions. Farias Jr et al. (2011) reported modest differences in the prevalence of inactivity according to the recall period used (54.6% in typical week vs. 60.8% in the last 7 days). We found no studies assessing whether the different methodologies of IPAQ agree among themselves (Cevero et al., 2009; Hallal et al., 2010; Kim et al., 2012).

The aim of this study was to compare physical activity scores assessed through four variations of IPAQ long version: (a) typical week, administered by an interviewer; (b) typical week, self-administered; (c) past seven days, administered by an interviewer; (d) past seven days, self-administered.

Method

Participants and data collection

This study comprised a convenience sample of students (18+ years old) (n = 38) enrolled in the Physical Education School of the Federal University of Pelotas. We invited students that had not yet attended to a course on the measurement of physical activity in epidemiological studies to participate.

Measures

In order to characterize the participants, we inquired them about sex, age, year of enrollment in the physical education school, as well as measured their weight and height. To verify the levels of physical activity, we used IPAQ long version (Craig et al., 2003). This questionnaire covers four physical activity domains: work-related physical activity (paid employment, as well as voluntary work), transportation physical activity (PA), household physical activity, and leisure-time physical activity (LTPA). For LTPA and occupational physical activity, a score was calculated as follows: minutes per week of walking + minutes per week of moderate-intensity physical activity + (minutes per week of vigorous-intensity physical activity × 2), in accordance with previous publications (Bicalho et al., 2010; Hallal, Victora, Wells, & Lima, 2003). The housework physical activity score was calculated without including walking, which is not part of the housework section of the IPAQ. Similarly, the transport-related score took into account only the number of minutes per week of walking and cycling (Hallal et al., 2003).

On day 1, participants visited the Laboratory of Biochemistry and Physiology of Exercise to collect the variables of interest, and answered to the self-reported typical week version. On day 8, participants answered to the (a) administered by interviewer typical week; (b) self-administered past seven days; (c) administered by interviewer past seven days versions. On day 8, all participants completed the self-report version first. We chose this strategy in order to avoid learning of the questionnaire after the interviewer’s administration. Only two interviewers administered the questionnaires, in order to avoid error. Those subjects had experience in applying IPAQ in surveys before, and underwent a 20 hours training protocol. Due to the fact that the sample was composed of physical education students, we standardized that college practical classes should be reported in the occupational domain.

Body weight was collected through a digital electronic Filizola scale, with 0.1 kg of resolution. Height was measured with a wall stadiometer, with resolution of 0.1 cm. We calculated body mass index (BMI) and used the World Health Organization cut-off points (WHO, 1995).

Data analysis

We entered the scores in an Excel spreadsheet and, after checking for errors, we transferred the data to the
statistical software STATA 12.0. Initially, we used the Shapiro-Wilk test to check whether the scores were normally distributed. For the calculation of means and standard deviations (SD), we used descriptive statistics. We used the Kruskal-Wallis test to determine differences among the scores, with Dunn post-hoc test. The level of significance was set at \( p < .05 \).

The study was approved by the ethics committee of the physical education school (protocol 029/2012). All participants signed a consent form after the explanation of the methodology and aims of the study.

**Results**

The study included 38 participants, 26 men, with a mean age of 21.1 years (SD 2.2), height 170.6 cm (SD 8.8), body mass of 67.1 kg (SD 12.5), BMI of 22.9 kg/m\(^2\) (SD 2.6) (Table 1).

Table 1. Mean and standard deviation of the variables describing the sample (n=38).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male</td>
</tr>
<tr>
<td>Age (years)</td>
<td>20.7 ± 1.7</td>
</tr>
<tr>
<td>Height (cm)</td>
<td>174.6 ± 7.0</td>
</tr>
<tr>
<td>Body Mass (kg)</td>
<td>73.1 ± 10.0</td>
</tr>
<tr>
<td>Body Mass Index</td>
<td>23.9 ± 2.2</td>
</tr>
</tbody>
</table>

Table 2 shows the means and SD of the total physical activity scores in the four versions of IPAQ. Regardless the recall period use, the self-report mode led to scores that were significantly higher than those obtained through interview (\( p < .001 \)). The same pattern was observed in both genders (\( p < .001 \) for men and women, respectively) (Table 3).

Table 2. Mean and standard deviation of the total PA scores (minutes/week) measured using the long version of IPAQ, in a habitual week and in the last seven days obtained via self-report and via trained interviewer (n=38).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male</td>
</tr>
<tr>
<td>Total PA score in a habitual week via self-report</td>
<td>1679.7*</td>
</tr>
<tr>
<td>Total PA score in a habitual week via interview</td>
<td>743.4*</td>
</tr>
<tr>
<td>Total PA score in the last seven days via self-report</td>
<td>1364.1*</td>
</tr>
<tr>
<td>Total PA score in the last seven days via interview</td>
<td>609.3*</td>
</tr>
</tbody>
</table>

* Kruskal Wallis test; different letters (a or b) determine significant differences (\( p < .001 \)).

Figure 1 presents mean physical activity scores in the occupational, household, leisure-time and commuting (minutes/week) domains. In terms of the occupational physical activity score (\( p < .001 \)), values were higher according to self-report as compared to interviews (6 times higher when considered habitual week; 7 times higher when considered last seven days). The same pattern was observed for household (\( p < .02 \)) and LTPA (\( p < .002 \)) although the magnitude of the difference was smaller. Commuting physical activity did not considerably change according to the mode of administration or recall period (\( p > .05 \)).

**Discussion**

To determine population physical activity levels, simple but accurate and inexpensive instruments are required. It is necessary
to use valid standardized instruments such as the IPAQ, which allows comparison among surveys conducted in different locations (Bauman et al., 2011; Zanichetta et al., 2010). IPAQ in Latin America has shown to present high reliability and moderate validity when compared to accelerometers (Hallal et al., 2010). However, our study shows that the self-report long version of IPAQ, particularly in the household and occupational domains, leads to surprisingly high scores, a finding that has been reported before (Hallal et al., 2010). Interestingly, we found this level of disagreement, even though most of our participants had high levels of education and were attending to a physical education college course.

The self-reported results were higher than those obtained through interviews. Probably this occurred due to a lack of understanding of the IPAQ questions, because in the interview methodology, the interviewers are able to help explain the questionnaire during the application. This kind of misinterpretation was also reported in other publication (Lawlor, Taylor, Bedford, & Ebrahim, 2002), which showed that population physical activity scores were overestimated in the occupational and household contexts. Many of the tasks performed in these two domains are sparse through the day, vary considerably from day to day, and are shorter than 10 consecutive minutes of duration, the minimum required for IPAQ. As shown previously (Hallal et al., 2010; Pardini et al., 2001) commuting and leisure-time physical activity were more stable and did not vary considerably according to the mode of administration or recall period.

Regarding the recall period, we found no major differences across versions of IPAQ. Although information about the past seven days being less likely to be influenced by recall bias or social desirability (i.e., individuals answering about what they would like to do instead of about what they actually do), scores were similar to those obtained using a typical week. Further, it has been shown that a recall period of one week may not reflect the habitual weekly physical activity levels (Farias Jr et al., 2011) (i.e. an active person might have been inactive in the past week only due to travelling commitments, for example). However, a meta-analysis suggested that people tend to answer the questions about a typical week, thinking about the last seven days (Kim et al., 2012). In fact, the meta-analysis showed that the reference period was unrelated to the percentage of variance explained, whereas the mode of administration did influence results for total, vigorous-intensity and transport physical activity. The interview-administered version resulted in higher validity scores as compared to the self-administered version.

In order to overcome the challenge of the questionnaire administration mode, researchers proposed averaging values obtained through the past seven days (Rombaldí et al., 2010; Teychenne, Ball, & Salmon, 2008) and a typical week (Cardoso, Rombaldí, & Silva, 2013). This strategy might be useful in some cases, but would create logistic challenges for large-scale surveys, in which physical activity is only one more indicator measured among several others. Difficulties in determining the ideal recall period suggest more research on this topic is needed, including those that bring expertise from other fields that deal with similar problems, or those that examine the validity and reliability of IPAQ for subgroups of the population (Altschuler et al., 2009; Timperio, Salmon, & Crawford, 2003).

The recall period did not influence the levels of physical activity by IPAQ, even when stratified by gender. Although many studies have found that males have higher levels of physical activity than females (Farias Jr et al., 2011; Hallal et al., 2011) our study showed that the recall period is not related to such differences.

Conclusions

The mode of administration of IPAQ does matter; higher scores were obtained through self-report as compared to interviews. Leisure-time and commuting physical activity were more stable and varied less according to administration mode or recall period. Gender-stratified analysis confirmed the findings of the overall sample. Given the importance of measuring physical activity, inexpensive measures such as IPAQ are desirable. However, we should exercise care when interpreting findings of studies using different modes of administration and reference periods.

Limitations and future research

The major limitation of this study was the sample composed by physical education students only. Participants were more active than the general population. A positive aspect of our paper is that, to our knowledge, this is the first study that showed clear inconsistencies depending on the mode of administration of IPAQ. Given the importance of this questionnaire for population studies, especially in low and middle-income countries, further work is needed on the validity and reliability of the different versions of IPAQ.

References


Motriz, Rio Claro, v.21 n.4, p.370-374, Oct./Dec. 2015


**Authors’ note**

Felipe de Magalhães Bandeira is affiliated with the Physical Education School, Federal, University of Pelotas, Brazil.

Matheus Pintanel Freitas, Marcelo Cozzensa da Silva, Pedro Curi Hallal and Airton José Rombaldi are affiliated with the graduate program in physical education at the Federal University of Pelotas, Brazil.

Mitzi László is affiliated with the Study Group in Epidemiology of Physical Activity, Federal University of Pelotas, Brazil.

**Corresponding author**

Matheus Pintanel Freitas
Avenida Brasil, 582, Pelotas 96025-000, RS, Brazil
Phone(fax): + 55 53 84260556
E-mail: matheus.pintanel@hotmail.com

**Manuscript received on March 18, 2015**

**Manuscript accepted on October 14, 2015**

Motriz. The Journal of Physical Education. UNESP. Rio Claro, SP, Brazil - eISSN: 1980-6574 – under a license Creative Commons - Version 3.0