

# BMJ Open Systematic review of the measurement properties of self-report physical activity questionnaires in healthy adult populations

Zoë Silsbury,<sup>1,2</sup> Robert Goldsmith,<sup>1,3</sup> Alison Rushton<sup>2</sup>

**To cite:** Silsbury Z, Goldsmith R, Rushton A. Systematic review of the measurement properties of self-report physical activity questionnaires in healthy adult populations. *BMJ Open* 2015;**5**:e008430. doi:10.1136/bmjopen-2015-008430

► Prepublication history and additional material is available. To view please visit the journal (<http://dx.doi.org/10.1136/bmjopen-2015-008430>).

Received 17 April 2015  
Revised 22 July 2015  
Accepted 13 August 2015



CrossMark

<sup>1</sup>Physiotherapy Department, University Hospital of Wales, Cardiff, UK

<sup>2</sup>School of Sport, Exercise and Rehabilitation Sciences, College of Life and Environmental Sciences, University of Birmingham, Birmingham, UK

<sup>3</sup>School of Physiotherapy, Cardiff University, Cardiff, UK

## Correspondence to

Zoë Silsbury;  
Zoe.Grant@wales.nhs.uk

## ABSTRACT

**Objective:** This systematic review evaluated the measurement properties of current self-report physical activity questionnaires (SRPAQs) completed within healthy adult populations.

**Design:** Two reviewers independently searched seven electronic databases and hand searched for articles investigating measurement properties of a SRPAQ evaluating physical activity over the previous 6 months. Articles published from 1 May 2001 to 4 December 2014 were systematically screened and eligible studies were not limited to English language sources. Articles investigating specific race, gender or socioeconomic populations were excluded.

**Results:** 10 studies investigating 10 SRPAQs were included. The methodological quality of the included studies was evaluated using COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) and ranged from 'poor' to 'good'. The Recent Physical Activity Questionnaire, International Physical Activity Questionnaires and Physical Activity Assessment Tool demonstrated good/excellent test–retest reliability (intra-class coefficient (ICC)=0.76,  $p<0.0001$ ;  $r=0.627-0.91$ ;  $r=0.618$ ,  $p<0.001$ , respectively), but variable criterion validity ( $r=0.67$ ,  $p<0.0001$ ;  $r=-0.02-0.43$ ;  $r=0.392$ ,  $p<0.01$ , respectively). The single-item measure showed significant criterion validity against an accelerometer (for moderate to vigorous physical activity (MVPA)  $k=0.23$ , 95% CI 0.05 to 0.41; and physical activity  $\geq 10$  min bouts 0.39 (95% CI 0.14 to 0.64). Construct validity of the six-point scale and Human Activity Profile varied significantly with age, marital status and presence of comorbidities ( $p<0.05$ ,  $<0.01$ ,  $<0.000$  and  $p<0.05$ ,  $<0.05$ ,  $<0.000$ , respectively). The 1 week Godlin-Shephard recall demonstrated 'moderate' validity with the gold standard measure of accelerometry ( $r=0.43$ ).

**Conclusions:** Inconclusive evidence exists. Further investigation of criterion validity of the short-form International Physical Activity Questionnaire is required, as it demonstrated excellent test–retest reliability.

**PROSPERO number:** CRD42012002484.

## INTRODUCTION

Physical activity (PA) prevents chronic diseases, independent of ethnicity, income,

## Strengths and limitations of this study

- The ambiguity of common terminology may have affected the electronic database search for this systematic review, possibly contributing to omissions in the included studies.
- This review searched for articles, reported in any language, investigating the English language version of a self-report physical activity questionnaire (SPARQ). In one study, this led to the exclusion of a subset of the overall sample that investigated the non-English version(s) of the SRPAQ. This may have diminished the overall confidence in this study's findings.
- Studies investigating use of SRPAQs on healthy subjects were used, which reduces the inference of the conclusions on those suffering from comorbidities or populations with disease.
- The focus of the systematic review was all measurement properties of SRPAQs; however, the eligible studies only investigated the properties of test–retest reliability, internal consistency, criterion, construct and structural validity.
- The COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) methodological quality analysis was self-taught by both the reviewers, using the handbook and articles explaining its use. This may have affected the ranking outcome for the included articles; however, as it was completed independently, this reduced the risk of error and bias.

education and body morphology.<sup>1</sup> The WHO estimated that, in 2008, 31% of the global healthy adult population failed to achieve the recommended 150 min of moderate-intensity aerobic PA, or 75 min of vigorous intensity PA, a week.<sup>2</sup> This has been described as the greatest public health problem of the 21st century.<sup>3</sup> PA encompasses 'any bodily movement produced by contraction of skeletal muscle resulting in energy expenditure above basal level'.<sup>4</sup>

A recent systematic review<sup>5</sup> described a positive association between PA and physical as well as psychological health; however, heterogeneity of study designs reduces confidence in these findings. This was supported by a systematic review evaluating the use of PA in managing the pathogenesis, physical strength and fitness, quality of life and symptoms of patients with chronic disease.<sup>6</sup> The review found strong evidence for effectiveness within cardiovascular disease, type 2 diabetes and obesity; and strong evidence for improved symptom management and quality of life for patients with cancer, osteoarthritis, osteoporosis, fibromyalgia and depression. The evidence supporting the role of PA in the prevention and management of chronic disease assists advancement of the public health research agenda,<sup>7</sup> thus improving quality of life and healthcare cost-effectiveness.<sup>8</sup> Research investigating PA in healthy participants is therefore valuable to inform prevention of chronic disease.

The International Classification of Functioning, Disability, and Health (ICF)<sup>9</sup> considers the individual's PA level and participation alongside external factors (eg, environment) for disease management. The integration of the biopsychosocial model and the ICF into physiotherapy practice has enabled a focus on physical ability and graded return to function rather than pain. Healthcare professionals commonly see patients for pain relief prior to the development of chronic disease and are best placed to address issues surrounding risk factors, including lack of PA. Therefore, the demand on physiotherapists to accurately evaluate PA for health status and treatment efficacy is increasing. Physiotherapists use a wide range of outcome measures to inform and evaluate their clinical practice, but selection of the most appropriate measure is challenging.<sup>10</sup> PA outcome measures assess the actual or perceived ability of an individual to carry out a variety of daily tasks and recreational or competitive sport.<sup>11</sup> However, the multidimensional and individual-specific nature of PA has resulted in a diverse range of outcome measures, contributing to a lack of consensus from clinicians/researchers regarding the best measure.

Patient-reported PA outcome measures are relatively inexpensive and easy to administer. It is acknowledged that self-report PA questionnaires (SRPAQs) comprise a detailed assessment to allow the detection of clinically relevant change in diverse populations.<sup>12</sup> However, due to the breadth of the activity dimensions analysed, it is argued that SRPAQs may lead to misclassification.<sup>12</sup> The advancement of real-time data acquisition from performance-based outcome measures including accelerometers, and the doubly labelled water (DLW) technique, are now considered the most accurate methods for determining energy expenditure; as PA is defined as energy expended above the metabolic rate, these methods are considered the gold standard for assessing PA. However, as the DLW technique cannot predict patterns or type of activity performed, and accelerometers are not able to provide information on swimming,

step/inclined activity, or strength training, there is a risk of under-estimating the energy expended in these activities.<sup>13</sup> Furthermore, performance-based outcome measures tend to use expensive equipment and require data analysis from trained professionals, which is generally beyond the scope of departments interested in evaluating PA. Therefore, real-time data acquisition is used in combination with SRPAQs that consider the multidimensional nature of the activity completed.<sup>13</sup> This informs the development of government guidelines.

Limited resources for implementing performance-based outcome measures reinforce the need to identify the best SRPAQ for clinical practice and research, as they are the most practical and economical outcome measures for heterogeneous populations.<sup>14</sup> It is important that the measurement properties of SRPAQs are evaluated to reduce the risk of data misinterpretation and bias.<sup>15</sup> An International Delphi study of experts formulated the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN), defining the essential properties of outcome measures as: reliability, measurement error, validity, responsiveness and interpretability.<sup>16</sup>

One systematic review<sup>11</sup> has evaluated the measurement properties of SRPAQs in patients aged 18–55 years. The authors incorporated all SRPAQs regardless of time frame, administration type and whether the participants were already experiencing chronic disease. Heterogeneity and ambiguity in terminology may have contributed to their diverse data, making synthesis difficult; and hence identification of the best SRPAQ was not possible. Therefore, further study utilising a more homogenous population was advocated; results of which may allow for focused data analysis and synthesis, enabling conclusions to be drawn.

The global recommendation for health documents the dose–response relationship of PA to prevent non-communicable diseases in healthy people in specific age categories.<sup>1</sup> This recognises the differing PA levels to evidentially achieve optimum results,<sup>17</sup> and addresses the health benefits of PA in healthy populations to prevent chronic disease. Research investigating PA in an elderly population found that motivation was the most significant barrier.<sup>18</sup> As motivation is not a component in most SRPAQs, specific PA questionnaires were developed for the elderly, and their role and measurement properties have already been evaluated.<sup>19</sup> Furthermore, it has been reported that assessing PA for greater than a 6-month period and across seasons is unreliable due to poor subject recall.<sup>19</sup> Consequently, a systematic review investigating the best existing SRPAQ for the healthy adult population of 18–60 years will improve accuracy of PA reporting. As WHO guidelines stipulate the health benefits gained from PA over 1 week and the subjectivity of self-reported PA in greater than a 6-month period, the systematic review targeted daily, weekly and monthly PA to inform health assessments and recommendations for disease prevention.

## METHOD

### Objective

To evaluate the measurement properties of existing SRPAQs to ascertain the optimum PA outcome measure for use in a healthy adult population.

### Study design

The systematic review followed a predefined and published protocol adapted from the Cochrane Handbook,<sup>20</sup> the Centre for Reviews and Dissemination Group,<sup>21</sup> and Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA).<sup>22</sup>

### Search strategy

The following electronic databases were searched to maximise the identification of appropriate articles<sup>20</sup>:

- ▶ OVID Medline without Revisions
- ▶ CINAHL Plus with Full Text Database
- ▶ EMBASE
- ▶ CAB Abstracts
- ▶ Health Management Information Consortium
- ▶ Journals @ OVID Full Text
- ▶ PubMed

Databases were searched from 1 May 2001, when the WHO first highlighted the importance of PA evaluation,<sup>9</sup> to 4 December 2014. Citation tracking of reference lists ensured all relevant articles were obtained<sup>21</sup> (boxes 1 and 2).

#### Box 1 Inclusion criteria for this systematic review: Inclusion criteria were articles

- ▶ Completed on the adult human population (>18 and <60 years old); The National Institute for Health and Care Excellence guidelines recognise different activity thresholds for varying age groups and so specific questionnaires were developed for the adult population.<sup>23</sup> Further research suggests that motivation and the ability to perform activities is more of a limitation for the elderly, rather than the amount.<sup>18</sup> Therefore the age group of 60 years old and above have specific physical activity (PA) questionnaires which were beyond the scope of this review.<sup>19</sup>
- ▶ Reporting assessment of  $\geq 1$  measurement property; Including content, construct or criterion validity, internal consistency, interpretability, responsiveness, reliability or absolute measurement error; justified for self-report PA questionnaires (SRPAQs) by an International Delphi study.<sup>16</sup>
- ▶ Reporting assessment of  $\geq 1$  written self-report standardised original English version outcome measure(s) focusing on PA; SRPAQs were used as they are practical for both cost and participant convenience, and feasible for use and analysis by clinicians.<sup>24</sup>
- ▶ Investigating English SRPAQs as this was the aim of this systematic review.
- ▶ Utilising a participant group representative of the general population, non-exclusive of race, gender or socioeconomic group. This allows transfer of the findings to the general population.<sup>25</sup>

#### Box 2 Exclusion criteria for this systematic review: Exclusion criteria were articles

- ▶ That were systematic reviews as original articles were required for methodological quality analysis.<sup>26</sup>
- ▶ Completed on patients where reduced physical activity (PA) was due to neurological conditions or psychological factors. This reduces the variables that may have affected PA or completion of the self-report PA questionnaires (SRPAQs).<sup>27</sup>
- ▶ Completed on a specific population group as this reduces transferability of the conclusions drawn to the general population.<sup>27</sup>
- ▶ Utilising occupational PA or sedentary questionnaires or those conducted via the telephone. PA should be explicitly measured rather than being defined by involuntary occupational activity or a lack of sedentary behaviour.<sup>28</sup> Telephone conducted questionnaires introduce the potential for inconsistent administration from the interviewer and potentially limit the number of participants from low socioeconomic groups.<sup>29</sup> As the International Classification of Functioning, Disability, and Health emphasises the importance of PA for health promotion this is the concept of interest.<sup>30</sup>
- ▶ Examining outcome measures focusing on pain or specific limb injury as the causative factor for activity modification. Assessment of PA due to pain or potentially short-term inactivity reduces transfer of the results to the general public.<sup>31</sup>
- ▶ Examining PA logbooks rather than questionnaires. PA logbooks are less likely to reflect usual behaviour as PA levels may vary between seasons or as a result of illness or time constraints.<sup>24</sup>
- ▶ Investigating long-term recall of PA (classified as >6 months). As the recall accuracy for vigorous and less intense activity over this timescale has been deemed unreliable.<sup>32</sup> Furthermore, seasonal variations of PA are difficult to account for in the same survey.<sup>19</sup>
- ▶ Investigating SRPAQs written in a language other than English. As investigating the consistency and accuracy of SRPAQ translation was beyond the scope of the review.<sup>33</sup>

#### Keywords:

“Self Report\*” AND “Motor Activity” OR “Physical Activity” AND “Outcome Assessment” AND “Healthcare” OR “Outcome Measure\*” OR Questionnaire\* AND [“Physical Activity” OR “GPAQ” OR “International Physical Activity Questionnaire”] AND Valid\* OR Reliab\* OR “Measurement Propert\*” NOT Obesity NOT Girl NOT Boy NOT Psych\* NOT Environment\* NOT Elderly.

#### Eligibility criteria

Included articles were original studies investigating  $\geq 1$  measurement property of an English language SRPAQ focusing on PA within the past 6 months,<sup>34</sup> in a non-exclusive group. Articles published in any language (to ensure comprehensiveness<sup>20</sup>), investigating English language SRPAQs in participants aged 18–60 years,<sup>19</sup> were included. Articles investigating SRPAQs focusing on pain, specific injury or pathological conditions were excluded.



## Study selection

Two reviewers (ZS and RG) completed an independent search of electronic databases using the keywords. Titles and abstracts were evaluated independently by both reviewers using the eligibility criteria, removing duplicate articles. The two reviewers then independently evaluated the full-text article for eligibility.<sup>20</sup> Any disagreements were resolved through discussion, and inconsistencies discussed with a third reviewer (AR).<sup>35</sup> Level of agreement was evaluated using Cohen's  $\kappa$ . Details of study inclusion were recorded using the PRISMA flow diagram.<sup>22</sup>

## Assessment of methodological quality and data extraction

The COSMIN checklist of measurement properties<sup>26</sup> was used to evaluate the methodological quality of included studies. When studies are deemed to have good methodological quality, it indicates that their conclusions are more trustworthy.<sup>26</sup> The COSMIN checklist was developed in an International Delphi study<sup>36</sup> and comprises 12 components. Nine components analyse the standard of the included measurement property and one component assesses the interpretability of the study. Finally, two components question the generic requirements needed for the studies where Item Response Theory methods are utilised. Furthermore, there is a section to record the general requirements of results found to summarise the study's findings.

Training in use of the COSMIN analysis involved reading the background to its development and the manual on its use and scoring system.<sup>36 26</sup> The COSMIN scoring system and manual were referred to when interpreting the checklist for each measurement property.<sup>16 37</sup> The two independent reviewers completed the COSMIN analysis, with disagreement resolved by the third reviewer who was experienced in using COSMIN. Agreement was reached before data extraction. Level of agreement between reviewers was calculated at each stage of the process using Cohen's  $\kappa$ .<sup>38</sup> The two independent reviewers extracted the data from the included studies,<sup>39</sup> and the third reviewer mediated any discrepancies in data.

## Data synthesis within and across studies

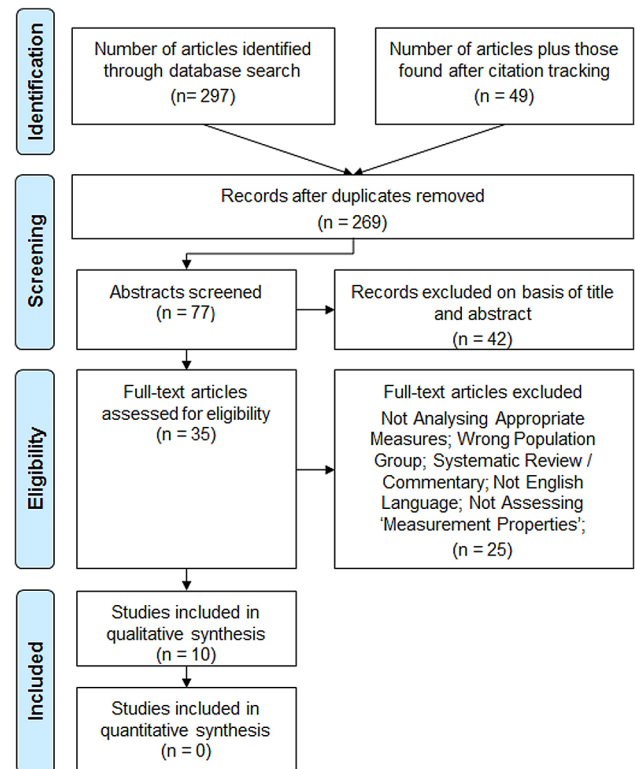
The SRPAQ investigated by each study was documented along with the measurement property assessed and the study's design and aim. The demographics of the population examined within each study were outlined and the COSMIN ranking for each measurement property methodology was recorded.<sup>37</sup>

For SRPAQs investigated in more than one study, the results and observed trends across the studies were collated. Consequently, a narrative synthesis of the data was used to discuss the relationship between the studies and their findings.

## RESULTS

### Included articles

Ten studies investigating the measurement properties of SRPAQs, comprising of over 2500 participants of



**Figure 1** Flow diagram depicting study identification (adapted from Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)).

multiple ethnicities across 11 countries, were included. **Figure 1** depicts the PRISMA process of study eligibility (adapted from Moher *et al*<sup>22</sup>). All 346 articles electronically identified were written in English, with 269 removed as they were duplicates or did not satisfy the inclusion criteria from title and abstract assessment. Forty-nine articles were retrieved by hand searching reference lists but all were excluded. The full text of the remaining articles (n=77) was screened. Following full text screening, a further 67 articles were excluded. The most common reason for exclusion was that the article did not analyse an appropriate outcome measure (n=26), or population group (n=16). Complete agreement (with discussion between only n=3 articles that were subsequently excluded) did not necessitate use of Cohen's  $\kappa$ .

## PA questionnaires and measurement properties

The data extraction for each article is presented in online supplementary table S1. Consensus between reviewers was reached at each stage following discussion on each article and reference to the COSMIN handbook, providing perfect agreement.<sup>38</sup> Online supplementary table S2 summarises data synthesis across multiple studies investigating an SRPAQ by combining data analyses from individual studies.

## Self-report physical activity questionnaires

Ten SRPAQs were investigated over the 10 studies, with four versions of the International Physical Activity

Questionnaire (IPAQ) identified. The studies analysed test–retest reliability and criterion validity of SRPAQs, except for the human activity profile (HAP), six-point scale and Godlin-Shephard 1 week (G-S) recall. Comparison of the six-point scale, HAP and IPAQ long form ‘Past 7 days’ (IPAQ-L7S), was evaluated, and specific demographic groups were assessed for differing PA.<sup>40</sup> The COSMIN methodological rank ranged between poor and good.

### International Physical Activity Questionnaire

Both the written IPAQ short form ‘Usual Week’ (IPAQ-SUS) and IPAQ short form ‘Past 7 days’ (IPAQ-S7S) demonstrated excellent test–retest reliability over 7 days ( $r=0.79$  and  $0.75$ , respectively).<sup>41</sup> Pooled correlation between the short and long forms, and within the short versions was moderate ( $p=0.67$ ; 95% CI 0.64 to 0.70; and  $p=0.58$ ,  $0.51$ ,  $0.64$ , respectively).<sup>41</sup> Assessment of the IPAQ short form (IPAQ-SF) walking component demonstrated excellent reliability over 3 days ( $r=0.77$  for IPAQ-S7S,  $0.72$  for IPAQ-SUS) and was very good over 7 days ( $r=0.91$  for IPAQ-SUS).<sup>42</sup> The criterion validity of the IPAQ-SUS and its walking component against an accelerometer was poor ( $r=0.13$ <sup>31</sup> and  $0.26$ ,<sup>35</sup> respectively). Whereas the IPAQ-S7S and its separate walking component demonstrated small to moderate ( $r=0.26$ – $0.4$ )<sup>41</sup> and moderate criterion validity ( $r=0.39$ ),<sup>42</sup> respectively.

Similarly, the IPAQ-L7S demonstrated very good/excellent repeatability over 1 week ( $r=0.627$ ,  $p<0.001$ <sup>43</sup>;  $p=0.7$ – $0.88$ <sup>41</sup>;  $r=0.74$ – $0.79$ ,  $p<0.0001$ <sup>44</sup>) and 2 weeks ( $r=0.74$ – $0.79$ ;  $p<0.0001$ ), even when tested on a small sample ( $n=36$ ).<sup>44</sup> Furthermore, no significant difference was found between repeated IPAQ-L7S administrations for separate intensities of PA ( $r=0.74$ – $0.84$ ) in a study of good methodological quality.<sup>45</sup> The IPAQ-L7S demonstrated moderate validity with both the accelerometer ( $p=0.05$ – $0.43$ )<sup>42</sup> and DLW technique (for activity-related energy expenditure (AEE)  $r=0.31$ ,  $p=0.06$  and metabolic energy equivalent per week  $r=0.33$ ,  $p<0.05$ ).<sup>44</sup> Moreover, classification of ‘active’ and ‘non-active’ participants demonstrated good agreement between the IPAQ-L7S and accelerometer ( $p<0.001$ ).<sup>43</sup>

The test–retest reliability of the IPAQ long form ‘Usual Week’ (IPAQ-LUS) was excellent ( $r=0.91$ ), but poor correlation was found with an accelerometer ( $p=0.91$ ), although the sample size was small ( $n=28$ ).<sup>41</sup> The IPAQ-LUS demonstrated better agreement with the Actigraph accelerometer for combined moderate and vigorous PA ( $r=0.3$ ), and solely vigorous PA ( $r=0.42$ ), than for moderate PA ( $r=0.19$ ).<sup>46</sup> Above this intensity, the IPAQ-LUS tends to over-estimate moderate/vigorous PA, which correlates to a 165% increase in PA,<sup>46</sup> whereas the IPAQ-L7S has been shown to under-estimate PA by 27%.<sup>44</sup> This component of the study demonstrated fair methodological quality.

### Recent Physical Activity Questionnaire

The Recent Physical Activity Questionnaire (RPAQ) possesses high test–retest reliability for physical AEE (PAEE)

(ICC=0.76), with the separate PA domains demonstrating poor (home ICC=0.62) to high reliability (work ICC=0.85).<sup>47</sup> Validity of the estimated PAEE was associated significantly with the DLW technique ( $r=0.39$ ;  $p=0.0004$ ), and strongly with vigorous PA ( $r=0.67$ ;  $p<0.0001$ ).<sup>47</sup> The study was deemed to be of good COSMIN rank.

### PA Assessment Tool

A single study of good COSMIN methodological quality evaluating the PA Assessment Tool (PAAT) demonstrated significant test–retest reliability ( $r=0.618$ ,  $p<0.001$ ).<sup>43</sup> When assessing the PAAT against the accelerometer, significant correlation ( $r=0.38$ – $0.392$ ;  $p<0.01$ ) and fair agreement regarding participant classification as ‘active’ ( $k=0.338$ ) was found.<sup>43</sup>

### Six-point scale

Fair agreement was found between the six-point scale and IPAQ-L7S ( $k=0.46$ ), and good agreement with the HAP ( $k=0.57$ ), in a study of poor COSMIN rank.<sup>40</sup> They found that increasing age and presence of comorbidities had a significant negative effect on PA ( $p<0.05$  and  $0.000$ , respectively). Gender and smoking had no effect ( $p>0.05$ ) in an experiment of fair COSMIN methodological quality.<sup>40</sup>

### Human activity profile

One study found that the HAP poorly correlated with the IPAQ-L7S ( $k=0.38$ ), with this portion being of poor COSMIN ranking. Age ( $p<0.05$ ) and the presence of comorbidities ( $p<0.000$ ) significantly affected PA, while gender and smoking did not ( $p>0.05$ ), which was similar to the six-point scale.<sup>40</sup> The findings demonstrated that occupation significantly affected PA categorisation using the HAP ( $p<0.000$ ), and this portion of the study was deemed to have fair COSMIN ranking.<sup>37</sup>

### Single-item measure

One study of good COSMIN methodological quality reported significant correlation between the accelerometer and the single-item measure for all moderate-vigorous PA over 30 min ( $r=0.46$ ;  $p<0.001$ ). Stronger correlation was shown when this intensity of PA was taken at 10 min intervals ( $r=0.57$ ;  $p<0.001$ ).<sup>48</sup> Participants were found to under-report activity using the single-item measure ( $-1.59$  days) compared with all objectively measured moderate-vigorous PA, although when compared with recorded PA in 10 min bouts, there was stronger correlation of data ( $0.38$  days).

### G-S 1 week recall

Moderate correlation of the G-S recall with an accelerometer when the PA was of moderate to moderate-vigorous intensity ( $r=0.3$  and  $0.4$ , respectively) was shown but a stronger association for vigorous PA ( $r=0.5$ ) was exhibited. It was demonstrated that the G-S recall data were not significantly altered by their use in participants

of different genders, weight classifications or ethnicity ( $p>0.05$ ).<sup>49</sup> Furthermore, the subgroup analysis demonstrated that males performed more moderate-vigorous activity than females ( $p<0.001$ ), and non-overweight participants completed more moderate and moderate-vigorous PA than overweight/obese individuals ( $p<0.05$ ). The COSMIN rank for this study was poor.

## DISCUSSION

### Measurement properties

#### Self-report physical activity questionnaires

This systematic review identified available English language SRPAQs, updating and focusing on the previously completed research.<sup>11</sup> Single studies of good methodological quality each evaluated the RPAQ,<sup>47</sup> PAAT,<sup>43</sup> single-item measure<sup>48</sup> and G-S recall,<sup>49</sup> giving confidence to the demonstrated significant test–retest reliability and criterion validity. The six-point scale demonstrated a moderate level of agreement with the HAP and IPAQ-L7S, whereas the HAP only demonstrated a fair level of agreement with the IPAQ-L7S.<sup>40</sup> The HAP was able to distinguish PA differences between participant subgroups, however, fair COSMIN methodological quality was calculated for this study,<sup>40</sup> due to limited statistical analysis, and the sample size was small, which reduces the confidence in this study's results. The G-S recall identified females as less active than males and the overweight/obese subgroup as less active than those who are non-overweight.<sup>49</sup> This study also demonstrated that gender, ethnicity and weight did not significantly alter the data derived from the accelerometer or G-S recall questionnaire. This testing of different participant subgroups supports the construct validity of the SRPAQs, and consequently their use in heterogeneous sample groups, which is representative of healthcare patient populations.

#### International Physical Activity Questionnaire

The most investigated SRPAQ was the IPAQ, specifically the IPAQ-L7S, which is the most extensively used SRPAQ worldwide, due to its varying formats and translation into many different languages. All forms of the IPAQ demonstrated very good/excellent test–retest reliability,<sup>41–45</sup> with the results for the IPAQ-L7S and short form IPAQ 'walking only' component being classified as excellent. The English IPAQ-SUS was evaluated with a small sample, which limits its generalisability. However, cross-cultural comparisons involving non-English speaking countries corroborate the excellent reliability of the IPAQ-SF<sup>50</sup> and, consequently, their use in clinical practice. Repeated administrations for separate PA intensities demonstrated no significant difference on IPAQ-L7S value, demonstrating excellent test–retest reliability. Unfortunately, this was completed on a group of university students of limited age range,<sup>45</sup> which reduces transferability to the general population, and the authors stated poor recording consistency of the data collected.

However, total PA was correlated, albeit weakly, to motivation and competency, demonstrating an attempt to explain PA scores within the excellent sample size.

The criterion validity of the IPAQ against an accelerometer or DLW technique was variable for the differing intensities of PA across the studies, with poorer correlation being found for total PA.<sup>41–42</sup> The use of limited accelerometer data on a small convenience sample may have caused this and, with increased numbers, the validity of the IPAQ-S7S did significantly improve.<sup>41</sup> Furthermore, a systematic review analysing the IPAQ-SF reported an over-estimation of PA by approximately 84%,<sup>51</sup> so despite being reliable, the validity of the IPAQ-SF requires further investigation. Poor correlation between the IPAQ-LUS and accelerometer was significantly improved by increasing the sample size by adding its non-English counterpart,<sup>41</sup> but it demonstrated better criterion validity at  $<1000$  min/week of PA.<sup>46</sup>

Testing the ability of the SRPAQ to identify PA change in different subgroups of participants assesses its construct validity.<sup>16</sup> Expectedly, the IPAQ-L7S showed that age significantly affects an individual's PA; however, the presence of comorbidities and smoking did not.<sup>40</sup> This may be due to low numbers within these participant groups and lack of clarity over the degree to which the comorbidity affected their health status. The authors suggested that the unexpected weak associations between total PA, motivation and the ability to undertake PA,<sup>45</sup> were due to variability in the recording skills of participants, affecting data consistency, so repeated experimentation would be beneficial.

Encouragingly, it has been shown that higher IPAQ scores correlate with lower mortality rates and a reduced risk of cardiovascular disease, demonstrating its worth for public health assessment.<sup>52</sup> This supports its interpretability, which although not considered a measurement property, is a clinically important characteristic of any outcome measure. However, suggestion of inconsistent translation of the IPAQ may not allow the same accreditation to non-English versions, therefore investigation into each translation would be necessary.<sup>33</sup>

#### Recent Physical Activity Questionnaire

The development of the RPAQ with a 1 month recall period was to calculate the individual's PA in relation to home, work, transport and leisure time.<sup>37</sup> The RPAQ demonstrated high test–retest reliability and was strongly associated with the DLW technique for vigorous PA, giving the RPAQ criterion validity. This single study used different populations for the reliability and validity components of analyses, with the validity study consisting of a good sample size.<sup>37</sup> The stability of the PA between repeated measures was assumed and the testing environment was not described, allowing for the COSMIN methodological quality to be described as good. Interestingly, the authors evaluated the ICC of 0.62 for home activity as poor, whereas the other literature would have ranked



this more highly,<sup>53</sup> which reduces standardisation of the ICC and data synthesis across studies.

### PA Assessment Tool

The PAAT measures moderate and vigorous PA over the preceding 7 days, and is intended to be completed within 7 min.<sup>43</sup> This tool would be ideal for implementation in a healthcare professional's waiting area and determines whether this 7-day PA is similar to the patient's normal activity. A single study evaluating the PAAT demonstrated significant reliability for vigorous activity ( $r=0.618$ ;  $p<0.001$ ); however, by only using 67 participants, the strength of the evidence is reduced. Furthermore, participant interpretation of moderate and vigorous PA may introduce ambiguity of reporting, and cause inconsistency in results. This study evaluated a high proportion of female volunteers who, it has been stated, may demonstrate reduced PA in their lifetime due to pregnancy and child care. This could reduce the application of findings into male populations; however, limited research on female PA neither supports nor refutes this hypothesis.<sup>54</sup> Additionally, the PAAT quantifies the energy for moderate and vigorous PA only, and against potentially outdated Compendium values,<sup>55</sup> which may affect the applicability of the results found.

### Six-point scale

The six-point scale was devised to reduce time and financial constraints associated with lengthy questionnaires. PA is classified on a sliding scale with additional descriptors referring to the amount of perspiration and depth of breathing when examples of exercise are inappropriate.<sup>40</sup> This may introduce subjectivity in participant reporting and inconsistency between test subjects. Structural validity was assessed by comparing the six-point scale and HAP with the IPAQ-L7S,<sup>40</sup> demonstrating that a brief scale can assess PA adequately when compared to the longer SRPAQs on a heterogeneous sample recruited by verbal invitation. By subgrouping the participant group, they were able to discover that increasing age and the presence of comorbidities had an expected negative effect on PA. Smoking did not have an effect on the amount of PA completed, but a small sample size may not have enabled suitable analysis to be performed. It was concluded that the six-point scale can be used for assessing PA in large-scale heterogeneous sample epidemiological studies, but the poor COSMIN methodological quality, due to reduced description of the six-point scale and the non-directional hypotheses, limits confidence in this finding and more rigorous investigations are encouraged.<sup>37</sup>

### Human activity profile

The HAP measures PA levels to give an estimate of the individual's average energy expenditure.<sup>37</sup> The HAP poorly correlated with the IPAQ-L7S<sup>40</sup>; although using a different  $\kappa$  classification this would have been described as fair,<sup>56</sup> and deemed moderate with the six-point scale.

Smoking did not significantly affect the PA, but expectedly, physicality of occupation did significantly affect the categorisation of PA by the HAP, which improves its construct validity. Disappointingly, this study did not use exploratory or factor analysis statistical methods and, consequently, methodological quality was poor,<sup>37</sup> reducing confidence in findings.

### Single-item measure

The single-item measure was investigated to observe whether the longer SRPAQs were, in fact, any more beneficial. The validity of the single-item measure<sup>48</sup> was tested against the accelerometer in a healthy population of good representation to the general public. This study reports significant correlation between the accelerometer, which is stronger when PA was taken at 10 min intervals. By analysing 10 min PA intervals, the sensitivity of testing was improved, although the specificity was better when testing total moderate-vigorous PA.<sup>57</sup> This lack of standardisation may limit the use of the single-item measure in clinical practice and research. However, the idea that a single question could be as reliable and valid as the longer SRPAQ is interesting and clinically relevant, so further investigation would be beneficial.

### G-S recall

The criterion validity of the G-S recall against the accelerometer resulted in moderate to strong correlation in varying PA. The authors of this study did note a greater variability with higher levels of PA and reported that the G-S recall has restrictions when used to assess individual PA levels, however, the G-S recall data were not significantly altered by their use in heterogeneous groups.<sup>49</sup> The ranking of this portion of the validity study was deemed as 'poor' by COSMIN, due to the lack of hypotheses prior to data collection. Furthermore, the statistical analysis was not optimal,<sup>16</sup> causing the inferences to clinical and research practice to be restricted. This may limit its use, so further investigation for use is encouraged.

### PAQs in research and clinical practice

To enable SRPAQ data to be equated to national guidelines,<sup>23</sup> units need to be comparable. This is a limitation for questionnaires assessing PAEE, as the guidelines focus on time, thereby reducing the clinical significance of SRPAQ data, the comparison between different SRPAQs due to ambiguity and transferability into practice.<sup>58</sup> Development of SRPAQs has reduced completion time, with the PAAT, six-point scale and single-item measure feasible for use in clinical practice, but further research is required to support their use across languages/populations. Conversely, the HAP is a lengthy questionnaire, thus increasing costs and reducing clinical usability.

The RPAQ demonstrates a strong correlation for assessing monthly vigorous PA energy expenditure, but not sedentary PA.<sup>47</sup> Likewise, the PAAT and single-item

measure only investigate moderate/vigorous PA over the preceding 7 days, questioning its use for all intensities of activity, as it may affect the categorisation of the individual as active or sedentary. The inconsistency of recording 10 and 30 min PA data for the single-item measure reduces standardisation, potentially leading to errors in data analysis.<sup>48</sup> The G-S recall did detect differences between gender and weight classifications, with no difference shown when correlating their data with the gold standard.<sup>49</sup> Similarly, there was moderate to strong correlation between moderate to vigorous PA collected comparable to the accelerometer. However, poor COSMIN methodological quality for its hypothesis testing sub-study<sup>16</sup> questions the inference that can be made when using the tool clinically or in research. Therefore, due to the ambiguity surrounding the PA intensity analysed, further scrutiny is required to improve consistency.

The IPAQ has the advantage of being researched extensively across languages, populations and demographic groups. It has been developed into multiple formats, making the IPAQ the most extensively used SRPAQ worldwide.<sup>51</sup> There is good correlation between the long and short forms via all modes of delivery; however, caution should be used if comparing the two, as the IPAQ long form (IPAQ-LF) can over-estimate PA to a greater extent than the IPAQ-SF.<sup>46</sup> The questionable validity of the IPAQ-SUS and time-consuming administration of the IPAQ-LF restrict their use clinically. The IPAQ-S7S is, at present, the preferred measure, owing to its excellent test-retest reliability over 7 days. Moreover, the IPAQ-S7S is supported across cultures within non-English speaking countries, assisting transferability of information, and this, along with its short format, makes it feasible for research and clinical practice. Owing to studies with low sample sizes demonstrating poor validity, a study investigating an adequate sample size of over 100 participants is required.

### Limitations

Limited narrative comparison of the SRPAQs was achievable due to their individual aims to report different PA intensities. For example, the IPAQ, HAP and RPAQ investigate a wide variety of PA, whereas the six-point scale and single-item measure are brief overviews of daily activity or exercise, and the PAAT concentrates on moderate/vigorous activity. The individual studies did not define their understanding of PA; and the SRPAQs did not indicate the difference between PA intensities. Furthermore, the single-item measure<sup>57</sup> may compromise breadth and depth of recorded data for speed of administration.<sup>49</sup> The main methodological flaws encountered were from the use of limited statistical tests and low subject numbers. Using the COSMIN scoring system, successive methodological limitations will reduce the rank from excellent methodological quality through to good, fair or poor, reflecting the confidence we can have in study findings. Additionally, within the reliability studies, it was generally assumed that no change in PA occurred rather

than statistically analysing PA stability. Furthermore, the Spearman correlation coefficient indicates association rather than true agreement.<sup>40</sup> Therefore, caution should be taken when accepting validity with only this supporting statistical test, which is true for the majority of included studies. Moreover, there were discrepancies in ranking the numerical outcome of the statistical tests between studies, which should be noted during data synthesis for each SRPAQ.

The ambiguity of common terminology used within PA research may have affected the electronic database search for this systematic review, possibly contributing to omissions in the included studies. In addition, as only English language SRPAQs were included, the sample sizes within the Craig *et al*<sup>41</sup> study were greatly reduced, diminishing the significance of their results. Furthermore, the aim of this systematic review was to comment on the measurement properties of SRPAQs, but, unfortunately, only studies investigating test-retest reliability, internal consistency and criterion, construct and structural validity, were identified, reflecting the limited research to date on measurement properties.

### CONCLUSION

Ambiguity in PA terminology, patient reporting of PA, and the variable nature of activity across the seasons and 7 days, makes daily activity difficult to assess using structured SRPAQs. Further inconsistencies within the PA assessed by each SRPAQ, limited measurement properties being evaluated and poor methodological quality of many studies, contributed to difficult data synthesis across studies. Consequently, the optimum SRPAQ has not been reported. The IPAQ-L7S is the most investigated SRPAQ, with all versions reported to have good test-retest reliability but limited criterion validity, potentially due to limited methodological quality. Based on current data, the IPAQ-S7S is the most appropriate outcome measure for clinical and research use, as it has excellent reliability and moderate correlation with accelerometry. The short version makes it efficient for clinicians, also making it more cost-effective. Future research should continue to assess SRPAQs against the 'gold standard' PA performance-based measures, in diverse socioeconomic groups worldwide. Assessment of responsiveness and interpretability would be valuable to relate the levels of PA required to reduce the risk of chronic disease for the individual. This is fundamental for disease prevention and therefore essential for the promotion of public health.

**Acknowledgements** The authors acknowledge Mr Gethin Lynch, affiliated to School of Sport, Exercise and Rehabilitation Sciences, University of Birmingham, UK, for assisting with article selection.

**Contributors** ZS was involved in finding and appraising all articles, completing COSMIN analysis, data extraction, analysis and synthesis, and wrote the article. This included documenting the supporting evidence for the need for the systematic review, adapting pre-defined protocol for its completion, and recording and synthesising the results obtained with inferences for clinical practice and research. RG was involved in the selection of articles, COSMIN analysis, and data extraction and synthesis for this



manuscript. AR was the mediating third reviewer, academic supervisor of the research and peer reviewed the submitted article.

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

**Competing interests** None declared.

**Ethics approval** Ethical approval was obtained from the University of Birmingham.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data sharing statement** No additional data are available.

**Open Access** This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

## REFERENCES

- Haskell W, Blair S, Hill J. Physical activity: health outcomes and importance for public health policy. *Prev Med* 2009;49:280–2.
- WHO—World Health Organisation. *Physical activity recommendations for health*. Geneva, Switzerland: WHO Press, 2010. [http://www.who.int/dietphysicalactivity/physical\\_activity-recommendations-18-64years.pdf](http://www.who.int/dietphysicalactivity/physical_activity-recommendations-18-64years.pdf) (accessed 27 Feb 2015).
- Blair S. Physical inactivity: the biggest public health problem of the 21st century. *Br J Sports Med* 2009;43:1–2.
- Wittink H, Engelbert R, Takken T. The dangers of inactivity, exercise and inactivity physiology for the manual therapist. *Man Ther* 2011;16:209–16.
- Bize R, Johnson J, Plotnikoff R. Physical activity level and health-related quality of life in the general adult population: a systematic review. *Prev Med* 2007;45:401–15.
- Pedersen B, Saltin B. Evidence for prescribing exercise as therapy in chronic disease. *Scand J Med Sci Sports* 2006;16(Suppl 1):3–63.
- Haskell W. Physical activity by self-report: a brief history and future issues. *J Phys Act Health* 2012;9(Suppl 1):S5–10.
- Moore A, Jull G. Physical activity/exercise—what does it mean to students, clinicians, patients and researchers. *Man Ther* 2012;17:99.
- WHO—World Health Organisation. *The International Classification of Functioning, Disease and Health*. Geneva, Switzerland: World Health Organisation, 2001. <http://www.who.int> (accessed 5 May 2012).
- Bossmann T, Kirchberger I, Glaessel A, et al. Validation of the comprehensive ICF core set for osteoarthritis: the perspective of physical therapists. *Physiotherapy* 2011;97:3–16.
- Van Poppel M, Chinapaw M, Mokkink L, et al. Physical activity questionnaires for adults: a systematic review of measurement properties. *Sports Med* 2010;40:565–600.
- Jette A, Haley S. Contemporary measurement techniques for rehabilitation outcomes assessment. *J Rehabil Med* 2005;37:339–45.
- Bassett D. Validity and reliability issues in objective monitoring of physical activity. *Res Q Exerc Sport* 2000;71:30–6.
- Masse L, de Niet J. Sources of validity evidence needed with self-report measures of physical activity. *J Phys Act Health* 2012;9(Suppl 1):S44–55.
- Gibbs G. Analyzing qualitative data. London: SAGE Publications Ltd, 2007:90–104.
- Mokkink L, Terwee C, Patrick D, et al. The COSMIN Checklist Manual 2012 [online]. <http://www.cosmin.nl> (accessed 19 Sep 2012).
- World Health Organisation. Global Strategy on Diet, Physical Activity & Health 2010. <http://www.who.int/dietphysicalactivity/strategy/en/> (accessed 4 Dec 2014).
- Fairhall N, Sherrington C, Kurrle S, et al. ICF participation restriction is common in frail community-dwelling older people: an observational cross-sectional study. *Physiotherapy* 2011;97:26–32.
- Forsén L, Loland N, Vuillemin A, et al. Self-administered Physical Activity Questionnaires for the elderly: a systematic review of measurement properties. *Sports Med* 2010;40:601–23.
- Higgins J, Green S, eds. *Cochrane Handbook for Systematic Reviews of Interventions*. The Cochrane Collaboration, 2012. <http://www.cochrane-handbook.org> (accessed 5 May 2012).
- CRD—Centre for Reviews and Dissemination. *Systematic reviews: CRD's guidance for undertaking reviews in healthcare*. 3rd edn. CRD University of York: York Publishing Services, Ltd., 2009.
- Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ* 2009;339:b2535.
- NICE—National Institute for Clinical Excellence. *Four commonly used methods to increase physical activity: brief interventions in primary care, exercise referral schemes, pedometers and community-based exercise programmes for walking and cycling*. London: National Institute for Clinical Excellence, 2006. <http://www.nice.org.uk/PHI002> (accessed 11 Jan 2012).
- Pereira M, FitzerGerald S, Gregg E, et al. A collection of Physical Activity Questionnaires for health-related research. *Med Sci Sports Exerc* 1997;29(Suppl 6):S1–205.
- Yore M, Ham S, Ainsworth B, et al. Reliability and validity of the instrument used in BRFSS to assess physical activity. *Med Sci Sports Exerc* 2007;39:1267–74.
- Mokkink L, Terwee C, Patrick D, et al. The COSMIN study reached international consensus on taxonomy, terminology and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol* 2010;63:737–45.
- Pettee Gabriel K, McClain J, Lee C, et al. Evaluation of physical activity measures used in middle-aged women. *Med Sci Sports Exerc* 2009;41:1403–12.
- Rosenberg D, Bull F, Marshall A, et al. Assessment of sedentary behavior with the international Physical Activity Questionnaire. *J Phys Act Health* 2008;5(Suppl 1):S30–44.
- Brown D, Brown D, Heath G, et al. Associations between physical activity dose and health-related quality of life. *Med Sci Sports Exerc* 2004;36:890–6.
- Peterson B, Rosenthal D. The ICF: a primer for rehabilitation educators. *Rehabil Educ* 2005;19:81–94.
- Feldman L, Kaneva P, Demyttenaere S, et al. Validation of a Physical Activity Questionnaire (CHAMPS) as an indicator of postoperative recovery after laparoscopic cholecystectomy. *Surgery* 2009;146:31–9.
- Blair S, Dowda M, Pate R, et al. Reliability of long-term recall of participation in physical activity by middle-aged men and women. *Am J Epidemiol* 1991;133:266–75.
- Bauman A, Bull F, Chey T, et al. The IPS Group. The international prevalence study on physical activity: results from 20 countries. *Int J Behav Nutr Phys Act* 2009;6:21.
- Bauman A, Phongsavan P, Schoeppe S, et al. Physical activity measurement: a primer for health promotion. *Promot Educ* 2006;13:92–103.
- Frei A, Williams K, Vetsch A, et al. A comprehensive systematic review of the development process of 104 patient-reported outcomes (PROs) for physical activity in chronically ill and elderly people. *Health Qual Life Outcomes* 2011;9:116–59.
- Mokkink L, Terwee C, Knol D, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. *BMC Med Res Methodol* 2010;10:22.
- Terwee C, Mokkink L, Knol D, et al. *Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist*. Quality of Life Research, 2011. <http://www.comin.nl> (accessed 6 Jul 2012).
- Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas* 1960;20:37–46.
- De Vries H, Reneman M, Groothoff J, et al. Factors promoting staying at work in people with chronic nonspecific musculoskeletal pain: a systematic review. *Disabil Rehabil* 2012;34:443–58.
- Webster S, Khan A, Nitz C. A brief questionnaire is able to measure population physical activity levels accurately: a comparative validation study. *J Clin Gerontol Geriatr* 2011;2:83–7.
- Craig C, Marshall A, Sjostrom M, et al. International Physical Activity Questionnaire: 12-Country Reliability and Validity. *Med Sci Sports Exerc* 2003;35:1381–95.
- Van der Ploeg H, Tudor-Locke C, Marshall A, et al. Reliability and validity of the international Physical Activity Questionnaire for assessing walking. *Res Q Exerc Sport* 2010;81:97–101.
- Meriwether R, McMahon P, Islam N, et al. Physical activity assessment: validation of a clinical assessment tool. *Am J Prev Med* 2006;31:484–91.
- Maddison R, Mhurchu C, Jiang Y, et al. International Physical Activity Questionnaire (IPAQ) and New Zealand Physical Activity Questionnaire (NZPAQ): a doubly labelled water validation. *Int J Behav Nutr Phys Act* 2007;4:62.
- Levy S, Readdy R. Reliability of the international Physical Activity Questionnaire in research settings: last 7-day self-administered long form. *Meas Phys Educ Exerc Sci* 2009;13:191–205.
- Boon R, Hamlin M, Steel G, 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;44:741–6.
47. Besson H, Brage S, Jakes R, *et al.* Estimating physical activity energy expenditure, sedentary time, and physical activity intensity by self-report in adults. *Am J Clin Nutr* 2010;91:106–14.
  48. Milton K, Clemes S, Bull F. Can a single question provide an accurate measure of physical activity? *Br J Sports Med* 2013;47:44–8.
  49. Sirard J, Hannan P, Cutler G, *et al.* Evaluation of two self-report measures of physical activity with accelerometry in young adults. *J Phys Act Health* 2013;10:85–96.
  50. Papathanasiou G, Georgoudis G, Georgakopoulos D, *et al.* Criterion-related validity of the short international Physical Activity Questionnaire against exercise capacity in young adults. *Eur J Cardiovasc Prev Rehabil* 2010;17:380–6.
  51. Lee P, Macfarlane D, Lam T, *et al.* Validity of the International Physical Activity Questionnaire Short Form (IPAQ-SF): a systematic review. *Int J Behav Nutr Phys Act* 2011; 8:115.
  52. Pitsavos C, Kavouras S, Panagiotakos D, *et al.* GREECS Study Investigators. Physical activity status and acute coronary syndromes survival. *J Am Coll Cardiol* 2008;51:2034–9.
  53. Munro B. Correlation. In: Munro B. ed. *Statistical methods for health care research*. 5th edn. Philadelphia: Lippincott Williams & Wilkins, 2005:239–58. Chapter 10.
  54. Harrison C, Thompson R, Teede H, *et al.* Measuring physical activity during pregnancy. *Int J Behav Nutr Phys Act* 2011;8:19.
  55. Ainsworth B, Haskell W, Whitt M, *et al.* Compendium of physical activities: an update of activity codes and MET intensities. *Med Sci Sports Exerc* 2000;32(Suppl 9):S498–504.
  56. Hicks C. *Research methods for clinical therapists: applied project design & analysis*. 5th edn. Edinburgh: Churchill Livingstone, 2009.
  57. Milton K, Bull F, Bauman A. Reliability and validity testing of a single-item physical activity measure. *Br J Sports Med* 2011;45:203–8.
  58. Bauman A, Ainsworth B, Bull F, *et al.* Progress and pitfalls in the use of the International Physical Activity Questionnaire (IPAQ) for adult physical activity surveillance. *J Phys Act Health* 2009; 6(Suppl 1):S5–8.