## METHODS

Forty-four patients completed the IPAQ (short and long forms) twice. Test-retest reliability was assessed by Spearman correlation coefficients ($r$) and intraclass correlation coefficients. Additionally, standard error of measurement and minimal detectable change were calculated. Concurrent validity was determined by an accelerometer. Spearman correlation coefficients were calculated between IPAQ scores and accelerometer data. Bland-Altman analyses were performed for both reliability and validity.

## RESULTS

Fair to good correlation coefficients were found for test-retest reliability of the total activity scores ($r = 0.49-0.81$, intraclass correlation coefficient = 0.27-0.71). Standard error of measurement and minimal detectable change were large. For concurrent validity, weak to moderate correlation coefficients were found for total and activity scores ($r = -0.07$ to 0.54). Systematic bias was found between the IPAQ and accelerometer data, with higher scores on the IPAQ.

## CONCLUSION

Overall, the IPAQ showed fair reliability and weak concurrent validity. These results are in line with previous studies of the reliability and validity of the IPAQ. Due to systematic bias and large standard error of measurement and minimal detectable change, the IPAQ may only be suitable for intergroup comparisons. J Orthop Sports Phys Ther 2013;43(9):650-659. Epub 25 July 2013. doi:10.2519/jospt.2013.4422

**KEY WORDS:** hip replacement, knee replacement, older adults, osteoarthritis

---

The health benefits of regular physical activity are undeniable. As a result, in today's society, there is a growing awareness of the importance of staying physically active. Regular physical activity, as well as enhancing musculoskeletal fitness, have proven to be effective in the primary and secondary prevention of several chronic diseases and are associated with a reduction of all-cause mortality. It has also been proven that physical activity
improves muscle strength, balance, and coordination, all of which are beneficial to the prevention of falls. In the presence of an arthroplasty, this is even more relevant—after all, a periprosthetic fracture or dislocation resulting from a fall accounts for significant morbidity.

To stimulate regular physical activity, national and international physical activity guidelines recommend 30 minutes or more of moderate-intensity physical activity at least 5 days a week, or vigorous-intensity physical activity for 20 minutes at least 3 days a week. Assessment of physical activity, however, is challenging, as it ideally includes 3 dimensions: activity type, frequency, and intensity. To assess physical activity at a population level, self-report questionnaires are the most feasible option. However, as no questionnaire has yet been accepted nationally or internationally, comparison is difficult. The International Physical Activity Questionnaire (IPAQ), developed with support from the World Health Organization, represents an attempt to standardize assessment of physical activity internationally. The IPAQ is available in a short form and a long form, and is considered one of the most feasible instruments to assess physical activity by means of a questionnaire. A study by Naal et al reported a completion rate of 92% for the short form of the IPAQ. Both the short and long forms of the IPAQ were tested for reliability and validity in 12 countries among a general population (adults aged 18 to 65 years). The tests found acceptable reliability \( r = 0.66-0.91 \) and moderate validity \( r = 0.02-0.52 \) compared to an accelerometer. van Poppel et al studied the Dutch version of the IPAQ and also found good reliability (intraclass correlation coefficient [ICC] = 0.70-0.96) and moderate validity \( r = 0.36-0.49 \) of the IPAQ compared with an accelerometer. The IPAQ was also tested for reliability and validity in disease-specific settings (eg, diabetes, schizophrenia, and HIV), with results similar to those found in the general population.

Given the benefits of regular physical activity for health and musculoskeletal fitness, more insight is needed into the physical activity of patients after THA and TKA. The IPAQ can be of help for standardizing international comparison of assessment of physical activity in this expanding subset of the (mainly elderly) population. Hence, the objective of this study was to examine the test-retest reliability and concurrent validity of the short and long forms of the Dutch IPAQ in a THA and TKA population.

METHODS

Study Sample

Individuals who had undergone primary THA and/or TKA due to idiopathic osteoarthritis at University Medical Center Groningen (a university hospital) or Martini Hospital Groningen (a large teaching hospital) at least 1 year previously were consecutively selected from hospital records. Selected patients were contacted by phone and asked to participate in the study. If willing to participate, the patients received a confirmation letter by mail. The study took place from August 2011 to October 2011. Exclusion criteria were musculoskeletal and/or neurological disorders that could interfere strongly with daily physical activity, knee or hip joint procedures less than 1 year previously, and cognitive and/or language impairments that could interfere with the questionnaire survey. Data collection was performed to assess test-retest reliability and to determine concurrent validity of the IPAQ. The study was approved by the Medical Ethical Board of University Medical Center Groningen. All subjects gave written informed consent.

Questionnaire

The Dutch versions of the short and long forms of the IPAQ were used to assess physical activity. The design of the IPAQ makes it possible to test compliance with current national and international physical activity recommendations. Both the short (7 questions) and long (27 questions) forms of the IPAQ assess physical activity performed in the prior 7 days.

Questions in the long form of the IPAQ are prestructured into 5 subcategories: (1) job-related physical activity; (2) transportation-related physical activity; (3) housework, house maintenance, and caring for family; (4) recreation, sports, and leisure-time physical activity; and (5) time spent sitting. The long form of the IPAQ asks participants to report the frequency (days per week) and duration (time usually spent per day) of moderate-intensity physical activity, vigorous-intensity physical activity, walking activities, bicycling, and time spent sitting.

The short form of the IPAQ is derived from the long form, but its questions are not prestructured into subcategories. Participants are asked to report the frequency and duration of moderate-intensity physical activity, vigorous-intensity physical activity, walking activities, and time spent sitting. Intensity-related questions are accompanied by activity examples (eg, moderate-intensity physical activity is accompanied by cycling at a normal pace) in both forms.

Accelerometer

To determine the concurrent validity of the short and long forms of the IPAQ, physical activity was also assessed by means of an accelerometer (ActiGraph GT1M; ActiGraph, LLC, Pensacola, FL). This compact (3.8 × 3.7 × 1.8 cm) and lightweight (27 g) uniaxial accelerometer measures and stores vertical accelerations in a magnitude of approximately 0.05g to 2.1g. It is limited to a band frequency of 0.25 to 2.5 Hz, thus it detects only human motions and is not sensitive to the frequencies of other sources. The data from the accelerometer are reported in counts, which are the number of accelerations measured within a user-determined time frame (epoch). In this study, the epoch was set at 60 seconds, so the outcome was expressed as counts per minute. The accelerometer was tightly
fixed on a belt around the participant’s waist (sagittal line). Participants were instructed to wear the accelerometer all day, except when sleeping, showering, bathing, or swimming. Data from the accelerometer were processed by means of ActiGraph software (ActiLife Version 5.7.4.0; ActiGraph, LLC).

**Test-Retest Reliability**
To assess test-retest reliability, participants were assessed at an initial home visit and a second visit 9 to 12 days later. At the initial home visit, the researcher provided the participant with detailed verbal instructions on the IPAQ. Subsequently, the participant filled in the short form of the IPAQ, which was then collected by the researcher. Later that day, the researcher provided the participant with the long form of the IPAQ, which the participant completed and returned by mail, ensuring that answers to the short form could not be used to complete the long form. Participants who returned an incomplete long form were contacted by phone to retrieve the missing answers. Identical procedures were followed on the second home visit 9 to 12 days later.

**Concurrent Validity**
At the initial home visit, instructions on how to wear the accelerometer were provided by the researcher. The participants wore the accelerometer from that day until the second visit. Participants also kept a diary in which they noted when and why they were not wearing the accelerometer. Scores on the second-visit IPAQ were used to determine concurrent validity of the IPAQ. Hence, the accelerometer and the second IPAQ administration covered the same measurement period.

**Calculation of the Activity Score of the Accelerometer**
Activity score data collected from the short and long forms of the IPAQ were expressed in min·d⁻¹, so noted activity min·d⁻¹ values were multiplied by noted activity days per week. Using the *Guidelines for Data Processing and Analysis of the International Physical Activity Questionnaire*,²⁰ which uses the 2000 Ainsworth Compendium of Physical Activities,² activities were assigned a metabolic equivalent task (MET) value, 1 MET expressing the energy expenditure in rest per minute (sitting). Based on these guidelines, walking activity scored 3.3 METs, cycling 6.0 METs, moderate-intensity physical activity 4.0 METs, and vigorous-intensity physical activity 8.0 METs. Subsequently, activities were divided into moderate intensity (3–6 METs) or vigorous intensity (greater than 6 METs) by using the intervals of Freedson et al.²³ After following data-processing rules,²⁰ the resulting variables were (1) total activity score (MET-min·wk⁻¹), (2) total time (min·wk⁻¹) spent physically active, (3) time (min·wk⁻¹) spent on walking activity, (4) time (min·wk⁻¹) spent on moderate-intensity physical activity, (5) time (min·wk⁻¹) spent on vigorous-intensity physical activity, and (6) time (min·wk⁻¹) spent sitting.

**Statistical Analysis**
The data were analyzed using PASW Version 18.0 (SPSS Inc, Chicago, IL). Descriptive statistics were used to describe the main characteristics of the study participants.

**Test-Retest Reliability**
Test-retest reliability of the short and long forms of the IPAQ was determined by calculating Spearman correlation coefficients (r) and ICCs (2-way mixed, absolute agreement) between activity variables from both administrations. The following benchmarks according to Domholdt²⁴ were used to interpret the Spearman values: 0.00 to 0.25, little if any correlation; 0.26 to 0.49, weak correlation; 0.50 to 0.69, moderate correlation; 0.70 to 0.89, strong correlation; and 0.90 to 1.00, very strong correlation. Benchmarks suggested by Fleiss²⁵ were used to interpret the ICC values: greater than 0.75, excellent correlation; 0.40 to 0.75, fair to good correlation; less than 0.40, poor correlation.

Absolute reliability was assessed using Bland and Altman 95% limits of agreement.* The mean difference between the first and second administrations of the IPAQ was calculated, with zero within the 95% confidence interval (CI) of the mean difference as the criterion for absolute reliability and zero outside the 95% CI indicating a bias in the measurements.

The standard error of measurement (SEM) and the minimal detectable change at the 95% confidence level (MDC₉₅) were calculated. The SEM reflects the amount of error associated with an individual patient score.²⁶ MDC₉₅ is an estimate of the lowest change value that can be detected beyond the random measurement error.²⁷ To calculate the SEM, the standard deviation of the first administration (population at baseline) was multiplied by √(1−r), where r represented the ICC.* To calculate the MDC₉₅ at the population level, the fol-
Concurrent Validity
To assess concurrent validity, Spearman correlation coefficients were calculated between mean total activity scores retrieved from the IPAQ (in MET-min·wk⁻¹) and mean counts·min⁻¹ retrieved from the accelerometer. For the moderate-intensity and vigorous-intensity physical activity categories, mean time assessed by the IPAQ was compared with mean time assessed with the accelerometer. Because the activity score for walking activity and the moderate-intensity activity score of the IPAQ represent the same MET interval (3-6 METs), these time scores (min·wk⁻¹) were summed. This transformed IPAQ moderate-intensity time score (min·wk⁻¹) was then compared with the moderate-intensity time score (min·wk⁻¹) of the accelerometer to determine concurrent validity on moderate-intensity activity. To assess the concurrent validity of the activity score for sitting, min·wk⁻¹ spent sitting retrieved by the IPAQ was compared with the period in which the accelerometer captured less than 101 counts.

A comparison was also made between the short and the long forms of the IPAQ. Concurrent validity was assessed by calculating the Spearman correlation coefficients between all collected variables of the IPAQ (eg, total activity score in MET-min·wk⁻¹, time [min·wk⁻¹] spent on moderate- and vigorous-intensity physical activity, and time [min·wk⁻¹] spent sitting). Previously mentioned benchmarks were used to interpret Spearman correlation coefficients. To determine the absolute concurrent validity, a Bland-Altman analysis was performed, including calculation of the mean difference, with a corresponding 95% CI, between the IPAQ and the accelerometer. If zero did not lie within the 95% CI, a bias between the results of the IPAQ and those of the accelerometer was indicated. 

### RESULTS

The reliability and validity study included 44 patients (19 with THA, 20 with TKA, and 5 with both THA and TKA) (TABLE 1). Two patients in the THA group were excluded from the validity study because they did not wear the accelerometer long enough to meet the criteria. TABLE 1 shows the demographic characteristics of the study population. The majority of the included patients were female (77.3%).

Time spent on the different categories of physical activity intensity was substantially higher as measured by the IPAQ compared to that measured by the accelerometer (TABLE 2). Compared to the TKA group, the THA group was more physically active. In both groups, most physical activity time spent was of moderate intensity. The mean accelerometer wearing period was 7 days (n = 40), with a minimum of 6 days (n = 2). The mean accelerometer wearing time was 848 min·d⁻¹ (approximately 14 hours), with a mean of 210 noted counts·min⁻¹ (TABLE 2).

### Test-Retest Reliability

#### Long Form of the IPAQ
Good correlations were found between the total activity score of both administrations (r = 0.78, ICC = 0.65) (TABLE 3). Moderate to good correlations were also found for the separate categories of physical activity intensity (walking activity, moderate intensity, and vigorous intensity) (TABLE 3). Sitting scored moderately high on test-retest reliability (TABLE 3). According to the Bland-Altman analysis (TABLE 3), no systematic bias was found between test and retest of the long form of the IPAQ. The SEM for the total score and the activity scores ranged between 24 and 2668. MDC₉₅ ranged between 10 and 1115.

#### Short Form of the IPAQ
Moderate correlations were found between the total activity score of both administrations (r = 0.62, ICC = 0.51) (TABLE 3). Spearman correlation coefficients for activity subcategories ranged from 0.53 to 0.71 (TABLE 3), which were moderate to strong accord-
ing to the benchmarks. ICCs ranged from 0.27 to 0.38 (TABLE 3), which indicated a poor correlation. Sitting scored moderately high on test-retest reliability (TABLE 3). The Bland-Altman analysis (TABLE 3) showed no significant systematic bias between the 2 measurements. The SEM for the total score and the activity scores ranged between 248 and 2487. MDC, ranged between 103 and 1039.

Concurrent Validity

Long Form of the IPAQ The total activity score showed a weak correlation with mean counts-min⁻¹ assessed with the accelerometer (TABLE 2). Separate correlation coefficients for moderate- and vigorous-intensity physical activity were weak to moderate (TABLE 2). Concurrent validity for time spent sitting was poor (TABLE 2). In all categories of physical activity intensity, systematically higher physical activity scores were obtained with the IPAQ compared to those obtained with the accelerometer (TABLE 2). Bland-Altman analysis and plots (TABLE 2, FIGURE 1) illustrated the difference in mean time obtained by the IPAQ and accelerometer. Moreover, these plots showed linear data distribution, indicating that the difference between the physical activity scores obtained with the IPAQ and with the accelerometer increased when higher physical activity scores were reported (FIGURE 1).

Short Form of the IPAQ As with the long form of the IPAQ, the total activity score obtained with the short form correlated weakly with the score obtained with the accelerometer (TABLE 2). Weak correlations were found for moderate- and vigorous-intensity physical activity. Concurrent validity was poor for time spent sitting (TABLE 2). Similar to the long form, a large systematic bias was seen in the Bland-Altman analysis (TABLE 2), and a similar linear data-distribution pattern in the bias was noticed in the Bland-Altman plots (FIGURE 2).

Long Form Versus Short Form of the IPAQ Both forms of the IPAQ correlated strongly on all categories of physical activity intensity, except for the category time spent walking, which could be considered moderate (TABLE 4). Bland-Altman analysis indicated a systematic bias for the total activity score, total time spent physically active, and time spent in moderate-intensity physical activity (TABLE 4). In these categories of physical activity intensity, substantially higher scores were obtained with the long form of the IPAQ.

DISCUSSION

So far, little is known about the physical activity of THA and TKA populations. The overall results of this study show that the Dutch IPAQ—both the short and the long forms—possesses fair test-retest reliability and relatively weak concurrent validity for assessing the physical activity of THA and TKA patients.

To the best of our knowledge, this is the first study to examine test-retest reliability and concurrent validity of the IPAQ within a THA and TKA population. Results of the test-retest procedure to determine test-retest reliability of both the long and short forms of the IPAQ in THA and TKA patients showed fair reliability, with the long form of the IPAQ scoring higher than the short form of the IPAQ and no systematic bias between the first and second IPAQ administrations.

Overall, results were in line with those of other studies for test-retest reliability of the IPAQ in various languages and population settings. Craig et al. tested the IPAQ internationally in 12 countries in a normative adult population and found Spearman correlation coefficients ranging from 0.66 to 0.91 for the total activity score. van Poppel et al. studied the Dutch version of the IPAQ, also in a normative adult population, and found good test-retest reliability, with ICCs of 0.92 and 0.89 for the long and the short forms of the IPAQ, respectively (present study, 0.65 and 0.51, respectively). A previous study pointed out that the level of reliability is linked to the test-retest interval used. Reliability was higher when a short time interval was used (8 days or less), whereas a longer interval (2–3 weeks) accounted for a lower reliability.
With this in mind, van Poppel et al., as they pointed out themselves, might have overestimated reliability scores due to a 3-day measurement interval. It could be hypothesized that a highly educated, young test population may also increase the test-retest reliability of the instrument. Another recent study that solely sampled elderly subjects showed ICCs for the short form of the IPAQ that were comparable to those of the present study, with ICCs for total activity scores ranging from 0.50 to 0.65. The Short Questionnaire to Assess Health-enhancing physical activity, a physical activity questionnaire closely related to the IPAQ, found test-retest reliability comparable to that of the present study. Moreover, the test-retest reliability found in the present study is in line with that of other physical activity questionnaires; a review of 7 other physical activity questionnaires found Spearman correlation coefficients ranging from 0.34 to 0.89. The SEMs of both the short and long forms of the IPAQ were large. The MDC_{95} of both the short and long forms of the IPAQ were also rather large, but smaller than the (individual) SEM. The MDC_{95} represents the minimal amount of minutes

### TABLE 3

<table>
<thead>
<tr>
<th>Test</th>
<th>Retest</th>
<th>r</th>
<th>ICC</th>
<th>d*</th>
<th>SE d</th>
<th>95% CI</th>
<th>SEM</th>
<th>MDC_{95}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Long form of the IPAQ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total score</td>
<td>5181 ± 4509</td>
<td>4555 ± 3876</td>
<td>0.78</td>
<td>0.65</td>
<td>626 ± 3524</td>
<td>531</td>
<td>-446,1697</td>
<td>2668</td>
</tr>
<tr>
<td>Time vigorous</td>
<td>18 ± 45</td>
<td>19 ± 57</td>
<td>0.78</td>
<td>0.71</td>
<td>-1 ± 40</td>
<td>6</td>
<td>-13,11</td>
<td>24</td>
</tr>
<tr>
<td>Time moderate</td>
<td>815 ± 769</td>
<td>713 ± 687</td>
<td>0.81</td>
<td>0.70</td>
<td>102 ± 561</td>
<td>85</td>
<td>-69,273</td>
<td>421</td>
</tr>
<tr>
<td>Time walking</td>
<td>460 ± 609</td>
<td>395 ± 441</td>
<td>0.68</td>
<td>0.55</td>
<td>65 ± 504</td>
<td>76</td>
<td>-88,218</td>
<td>409</td>
</tr>
<tr>
<td>Time sitting</td>
<td>2963 ± 1229</td>
<td>3260 ± 1158</td>
<td>0.49</td>
<td>0.46</td>
<td>-296 ± 1237</td>
<td>186</td>
<td>-672,80</td>
<td>903</td>
</tr>
<tr>
<td>Short form of the IPAQ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total score</td>
<td>3524 ± 3553</td>
<td>3324 ± 3705</td>
<td>0.62</td>
<td>0.51</td>
<td>200 ± 3602</td>
<td>543</td>
<td>-895,1295</td>
<td>2487</td>
</tr>
<tr>
<td>Time vigorous</td>
<td>117 ± 307</td>
<td>44 ± 127</td>
<td>0.71</td>
<td>0.35</td>
<td>73 ± 265</td>
<td>40</td>
<td>-8,153</td>
<td>248</td>
</tr>
<tr>
<td>Time moderate</td>
<td>359 ± 375</td>
<td>407 ± 613</td>
<td>0.53</td>
<td>0.38</td>
<td>-48 ± 568</td>
<td>86</td>
<td>-220,125</td>
<td>295</td>
</tr>
<tr>
<td>Time walking</td>
<td>350 ± 337</td>
<td>407 ± 605</td>
<td>0.58</td>
<td>0.27</td>
<td>-58 ± 591</td>
<td>89</td>
<td>-237,122</td>
<td>288</td>
</tr>
<tr>
<td>Time sitting</td>
<td>3055 ± 1269</td>
<td>3124 ± 1253</td>
<td>0.59</td>
<td>0.60</td>
<td>-69 ± 1129</td>
<td>170</td>
<td>-412,274</td>
<td>803</td>
</tr>
</tbody>
</table>

**Abbreviations:** CI, confidence interval; d, difference between test and retest; ICC, intraclass correlation coefficient; IPAQ, International Physical Activity Questionnaire; MDC_{95}, minimal detectable change; r, Spearman correlation coefficient; SE, standard error; SEM, standard error of measurement.

*Values are mean ± SD.

†Total activity score in metabolic equivalent task-min wk⁻¹.

‡Time spent in min wk⁻¹.

FIGURE 1. Bland-Altman plot of the total time assessed with the long form of the IPAQ (n = 44). Difference between total time derived from the IPAQ (second administration) and accelerometer. The x-axis is mean total time: (IPAQ time + accelerometer time)/2 (min wk⁻¹). The y-axis is difference in total time: IPAQ time – accelerometer time (min wk⁻¹). Abbreviation: CI, confidence interval; IPAQ, International Physical Activity Questionnaire.
that total activity scores of the IPAQ correlated weakly with mean activity counts·min⁻¹ measured by the accelerometer. The concurrent validity found in our study is in line with that reported in previous studies on the concurrent validity of the IPAQ, determined by comparisons with accelerometers. For the total activity score, Craig et al. found Spearman correlation coefficients ranging from 0.02 to 0.52, with a median of 0.30. van Poppel et al. found similar concurrent validity, with correlation coefficients for the total activity score of 0.40 for the short form and 0.25 for the long form. A recent review that included 23 studies showed correlation coefficients ranging from 0.25 to 0.33 for the short form of the IPAQ.

In this study, concurrent validity was determined between the accelerometer and the second IPAQ administration. The reason for this was that they covered the same measurement period. However, it can also be argued to use the first IPAQ administration to avoid possible bias from (a) increased physical activity awareness by wearing the accelerometer and (b) a learning effect. Using the first administration of the IPAQ resulted in higher, moderate correlation coefficients for total activity of 0.63 for the short version and 0.52 for the long version. From the study of Craig et al., which tested the reliability and validity of the IPAQ in 12 countries, it was not clear whether the first or second administration of the IPAQ was used.

When looking at the Spearman correlation coefficients of the separate scores of physical activity intensity, moderate-intensity physical activity did not score as high and was not as stable as vigorous-intensity physical activity (moderate-intensity long and short forms: \( r = 0.18 \) and 0.09; vigorous-intensity long and short forms: \( r = 0.54 \) and 0.38). This finding is not unique, as shown by another study. A possible explanation could be that moderate-intensity physical activity happens so often that it goes unnoticed, in contrast with vigorous-intensity physical activity, which is often deliberately planned.

With respect to concurrent validity, the results of the present study showed that total activity scores of the IPAQ correlated weakly with mean activity counts·min⁻¹ measured by the accelerometer. The concurrent validity found in our study is in line with that reported in previous studies on the concurrent validity of the IPAQ, determined by comparisons with accelerometers. For the total activity score, Craig et al. found Spearman correlation coefficients ranging from 0.02 to 0.52, with a median of 0.30. van Poppel et al. found similar concurrent validity, with correlation coefficients for the total activity score of 0.40 for the short form and 0.25 for the long form. A recent review that included 23 studies showed correlation coefficients ranging from 0.25 to 0.33 for the short form of the IPAQ.

In this study, concurrent validity was determined between the accelerometer and the second IPAQ administration. The reason for this was that they covered the same measurement period. However, it can also be argued to use the first IPAQ administration to avoid possible bias from (a) increased physical activity awareness by wearing the accelerometer and (b) a learning effect. Using the first administration of the IPAQ resulted in higher, moderate correlation coefficients for total activity of 0.63 for the short version and 0.52 for the long version. From the study of Craig et al., which tested the reliability and validity of the IPAQ in 12 countries, it was not clear whether the first or second administration of the IPAQ was used.

When looking at the Spearman correlation coefficients of the separate scores of physical activity intensity, moderate-intensity physical activity did not score as high and was not as stable as vigorous-intensity physical activity (moderate-intensity long and short forms: \( r = 0.18 \) and 0.09; vigorous-intensity long and short forms: \( r = 0.54 \) and 0.38). This finding is not unique, as shown by another study. A possible explanation could be that moderate-intensity physical activity happens so often that it goes unnoticed, in contrast with vigorous-intensity physical activity, which is often deliberately planned.

With respect to concurrent validity, the results of the present study showed that total activity scores of the IPAQ correlated weakly with mean activity counts·min⁻¹ measured by the accelerometer. The concurrent validity found in our study is in line with that reported in previous studies on the concurrent validity of the IPAQ, determined by comparisons with accelerometers. For the total activity score, Craig et al. found Spearman correlation coefficients ranging from 0.02 to 0.52, with a median of 0.30. van Poppel et al. found similar concurrent validity, with correlation coefficients for the total activity score of 0.40 for the short form and 0.25 for the long form. A recent review that included 23 studies showed correlation coefficients ranging from 0.25 to 0.33 for the short form of the IPAQ.

In this study, concurrent validity was determined between the accelerometer and the second IPAQ administration. The reason for this was that they covered the same measurement period. However, it can also be argued to use the first IPAQ administration to avoid possible bias from (a) increased physical activity awareness by wearing the accelerometer and (b) a learning effect. Using the first administration of the IPAQ resulted in higher, moderate correlation coefficients for total activity of 0.63 for the short version and 0.52 for the long version. From the study of Craig et al., which tested the reliability and validity of the IPAQ in 12 countries, it was not clear whether the first or second administration of the IPAQ was used.

When looking at the Spearman correlation coefficients of the separate scores of physical activity intensity, moderate-intensity physical activity did not score as high and was not as stable as vigorous-intensity physical activity (moderate-intensity long and short forms: \( r = 0.18 \) and 0.09; vigorous-intensity long and short forms: \( r = 0.54 \) and 0.38). This finding is not unique, as shown by another study. A possible explanation could be that moderate-intensity physical activity happens so often that it goes unnoticed, in contrast with vigorous-intensity physical activity, which is often deliberately planned.

With respect to concurrent validity, the results of the present study showed that total activity scores of the IPAQ correlated weakly with mean activity counts·min⁻¹ measured by the accelerometer. The concurrent validity found in our study is in line with that reported in previous studies on the concurrent validity of the IPAQ, determined by comparisons with accelerometers. For the total activity score, Craig et al. found Spearman correlation coefficients ranging from 0.02 to 0.52, with a median of 0.30. van Poppel et al. found similar concurrent validity, with correlation coefficients for the total activity score of 0.40 for the short form and 0.25 for the long form. A recent review that included 23 studies showed correlation coefficients ranging from 0.25 to 0.33 for the short form of the IPAQ.

In this study, concurrent validity was determined between the accelerometer and the second IPAQ administration. The reason for this was that they covered the same measurement period. However, it can also be argued to use the first IPAQ administration to avoid possible bias from (a) increased physical activity awareness by wearing the accelerometer and (b) a learning effect. Using the first administration of the IPAQ resulted in higher, moderate correlation coefficients for total activity of 0.63 for the short version and 0.52 for the long version. From the study of Craig et al., which tested the reliability and validity of the IPAQ in 12 countries, it was not clear whether the first or second administration of the IPAQ was used.

When looking at the Spearman correlation coefficients of the separate scores of physical activity intensity, moderate-intensity physical activity did not score as high and was not as stable as vigorous-intensity physical activity (moderate-intensity long and short forms: \( r = 0.18 \) and 0.09; vigorous-intensity long and short forms: \( r = 0.54 \) and 0.38). This finding is not unique, as shown by another study. A possible explanation could be that moderate-intensity physical activity happens so often that it goes unnoticed, in contrast with vigorous-intensity physical activity, which is often deliberately planned.
A systematic bias between the scores of the IPAQ and the accelerometer was noticed, with higher physical activity scores on the IPAQ compared to the accelerometer. This systematic bias tended to increase with increased reported physical activity on the IPAQ: the more physically active a person reported he or she was, the greater the bias of the accelerometer data. This bias may be the result of an overestimated physical activity level by the IPAQ or the result of an underestimated physical activity level by the accelerometer. Overestimation is an intrinsic property of self-assessment questionnaires. Bias because of socially desirable answers on the IPAQ may have also contributed to overestimation. This phenomenon is reported in a comparable study in which the IPAQ was also validated by means of accelerometry. This could have occurred in the present study if the participants felt pressure to report higher physical activity levels. Underestimation of physical activity levels by the accelerometer might also have taken place, because the accelerometer is uniaxial and is therefore insensitive to horizontal displacement and activity that requires little vertical displacement (eg, cycling, fitness). Moreover, the waist configuration of the accelerometer makes it impossible to notice movements from the upper body. Another reason for underestimation of physical activity measured by the accelerometer could have been due to the use of the Freedson equation for data processing. This equation has been validated in a young adult population (mean age, 23 years). But elderly persons with a THA or TKA may avoid high-impact lower-body movements, thus a large amount of moderate-intensity physical activity might not have been recorded by the accelerometer.

New insights into physical activity give indications that sedentary behavior is becoming increasingly important as a parameter of health status. Sedentary behavior involves predominantly sitting, which represents little to no physical activity (1 to 1.5 times rest metabolism). The good test-retest reliability results for sitting, as assessed by means of the IPAQ, were in line with results found in previous studies. However, relative and absolute concurrent validity for sitting was poor, which makes the IPAQ not usable for obtaining time spent sitting. This result is not unique, as a review showed coefficients ranging from 0.07 to 0.61 (present study, r = –0.07 to 0.10).

There are some limitations to the present study. First of all, the study population was recruited by phone, which could have resulted in selection bias. However, in a reliability and validity study of this nature, this may be considered a nonissue, because the study was not designed to obtain a precise degree of physical activity. Second, the study population may be considered small for reliability assessment. A recent checklist proposed a minimum of 50 participants, and the present study included 44 participants. Furthermore, data collection for each individual took place on 2 different dates, so weather differences could have influenced physical activity behavior and, subsequently, test-retest reliability and validity. This, however, was unavoidable, and no large differences in weather conditions between measurements were observed. Finally, to assess concurrent validity, we used accelerometers. The highest level of evidence for validity would be obtained by comparing the IPAQ questionnaires with a gold standard, an instrument that could measure the same construct with perfect reliability and validity (criterion validity). However, there is no gold standard for the assessment of physical activity. The doubly labeled water method (DLWM) is often considered a gold standard for assessing total daily energy expenditure. However, considerations of cost aside, DLWM is not suitable for this type of study because the total daily energy expenditure, as measured by DLWM, is caused not only by physical activity but also by the basal metabolic rate and the thermic effect of food. More importantly, with DLWM, no information on duration, frequency, and intensity of physical activity can be obtained. Therefore, accelerometers are considered to be one of the best suitable instruments for validating physical activity questionnaires and are often used for this purpose.

**CONCLUSION**

Based on the results of this study, it can be concluded that both the long and the short forms of the IPAQ possess a fair test-retest reliability and weak concurrent validity for determining physical activity of patients following THA or TKA. These results are in line with previous studies of the reliability and validity of the IPAQ and with studies of reliability and validity of other self-reported physical activity questionnaires.

Based on the present results, it may be concluded that, due to a large systematic bias, absolute concurrent validity with an accelerometer is somewhat limited. In addition, the SEM and MDCₙ₀ of the IPAQ were large. For that reason, the IPAQ may not be applicable to determine precise amounts of individual physical activity. As a consequence, national and international physical activity guidelines cannot be tested on an individual basis. This conclusion is supported by previous research into the concurrent validity of the IPAQ, which also found that the IPAQ cannot be used on an individual basis but is suitable for intergroup comparison of physical activity. Use of the long IPAQ is preferred because it scored higher and more stable on test-retest reliability than the short form.

**KEY POINTS**

**FINDINGS:** The long and the short forms of the IPAQ possess a fair test-retest reliability and weak relative concurrent validity for determining the physical activity of patients following THA or TKA.
CAUTION: Due to a large systematic bias, absolute concurrent validity is somewhat limited. In addition, the SEM and MDC50 of the IPAQ were large. For that reason, the IPAQ may not be applicable to determining precise amounts of individual physical activity. Finally, the research design used in this study was not able to establish responsiveness to change over time.

IMPLICATIONS: The IPAQ seems to lack adequate precision for use in individuals; however, it may be suitable for group-level comparisons of physical activity. Use of the long IPAQ is preferred because it scored higher and more stable on test-retest reliability than the short form.

REFERENCES
37. Terwee CB, Bousmeester W, van Eijsland SL, de Vett HC, Dekker J. Instruments to assess...


43. Warburton DE, Nicol CW, Bredin SS. Health benefits of physical activity: the evidence. CMAJ.

