Is it possible to assess free-living physical activity and energy expenditure in young people by self-report?1–3

Kirsten Corder, Esther MF van Sluijs, Antony Wright, Peter Whincup, Nicholas J Wareham, and Ulf Ekelund

ABSTRACT
Background: It is unclear whether it is possible to accurately estimate physical activity energy expenditure (PAEE) by self-report in youth.
Objective: We assessed the validity and reliability of 4 self-reports to assess PAEE and time spent at moderate and vigorous intensity physical activity (MVPA) over the previous week in British young people between 4 and 17 y of age.
Design: PAEE and MVPA were derived from the Children’s Physical Activity Questionnaire, Youth Physical Activity Questionnaire, and Swedish Adolescent Physical Activity Questionnaire; a lifestyle score indicative of habitual activity was derived from the Child Heart and Health Study in England Questionnaire. These data were compared with criterion methods, PAEE, and MVPA derived from simultaneous measurements by doubly labeled water and accelerometry in 3 age groups: 4–5 y (n = 27), 12–13 y (n = 25), and 16–17 y (n = 24). Validity was assessed by using Spearman correlations and the Bland-Altman method, and reliability was assessed by using intraclass correlation coefficients.
Results: The strength of association between questionnaire and criterion methods varied (r = 0.09 to r = 0.46). Some questionnaires were able to accurately assess group-level PAEE and MVPA for some age groups, but the error was large for individual-level estimates throughout. Reliability of the Youth Physical Activity Questionnaire and Child Heart and Health Study in England Questionnaire was good (intraclass correlation coefficient: 0.64–0.92).
Conclusions: Absolute PAEE and MVPA estimated from these self-reports were not valid on an individual level in young people, although some questionnaires appeared to rank individuals accurately. Age (the outcome of interest) and whether individual or group-level estimates are necessary will influence the best choice of self-report method when assessing physical activity in youth. Am J Clin Nutr 2009;89:862–70.

INTRODUCTION
The observed magnitude of the relation between physical activity and health varies considerably, especially in children (1); this is largely due to the difficulties of making accurate assessments of physical activity energy expenditure (PAEE) and patterns of physical activity in large populations.

Despite their accuracy, gold-standard methods of energy expenditure (EE) assessment, such as the doubly labeled water method (DLW), are prohibitively expensive for use in large studies; however, they are useful for the validation of other methods potentially suitable for large-scale use (2, 3). The DLW method provides precise information on total daily EE (TEE) during daily living over a relatively long period and can be used to calculate PAEE by using measured or estimated REE. Accelerometry is probably the most widely used objective method of assessing physical activity and provides an assessment of frequency, intensity, and duration of physical activity. Recent large-scale studies have successfully used accelerometry to assess physical activity in youth (4, 5). However, self-report methods may still be the only feasible way to assess physical activity in many situations and are important for assessing aspects of physical activity not easily measured objectively, such as mode and domain (6). Some questionnaires can accurately determine the mode of activity and can be used to adequately rank, group, or categorize physical activity levels (7, 8) and possibly to assess some aspects of moderate and vigorous physical activity (MVPA) (9–11), but they are generally considered less accurate than objective methods for estimating PAEE and MVPA (12, 13). There are specific problems regarding the accuracy of questionnaire-derived physical activity data from young children because of the cognitive limitations of those aged <10 y (14, 15); therefore, proxy reports are often used (16). Despite questionnaires often being used to estimate EE, most recent questionnaire validation studies have used accelerometry as a criterion (6, 9, 10, 17); relatively few have used DLW (7, 8). Because most questionnaires are validated against only one criterion method, the full potential of questionnaires to assess different dimensions of physical activity is rarely explored, even though these dimensions may be of differential importance to a variety of health outcomes. To our knowledge, this is the first study to simultaneously assess the validity of estimated PAEE and MVPA from physical activity questionnaires against both DLW and accelerometry and that should allow a better assessment of the strengths and weaknesses of the self-reported physical activity than one criterion method alone. Additionally, this study compares 4 questionnaires in different age groups, which
should be relevant for those assessing physical activity in large-scale studies, including those that include multiple age groups of youth.

Therefore, the aim of this study was to determine the reliability and validity of 4 physical activity questionnaires compared with measurements of PAEE and MVPA obtained with the DLW technique and accelerometry in 3 different age groups of young people.

SUBJECTS AND METHODS

Participants

This study was carried out in a convenience sample of 82 volunteers aged between 4 and 17 y and recruited from schools in Cambridgeshire, United Kingdom, or as siblings of those enrolled in an ongoing but unrelated infant cohort study. Those invited into the study were 2 full classes taught by contact teachers at each of 3 secondary schools, all of those aged 4 or 5 y at 2 primary schools, and all 4- and 5-y-old siblings of volunteers attending the infant study during the measurement period (January 2005 to December 2006). Those who had a medical condition that severely limited their normal physical activity were to be excluded, although none with such conditions volunteered for the study. All volunteers received an explanation of the study and all provided assent; a parent provided written informed consent for participation in the study, which was approved by the Cambridge local research ethics committee. Of 337 subjects invited, 86 attended a testing session, DLW data were obtained for 84 volunteers, and accelerometry data were obtained for 82 volunteers. Volunteers were recruited in 3 age groups: 4–5 y (n = 29), 12–13 y (n = 27), and 16–17 y (n = 26). Of these 82 volunteers, 79 completed the first questionnaire administered. Of these 79 volunteers, accelerometry data were not available for 3 because of monitor malfunction, which resulted in 76 volunteers with complete validation data: 4–5 y (n = 27), 12–13 y (n = 25), and 16–17 y (n = 24). An additional 14 volunteers did not complete the second administration for reliability purposes, which resulted in the following valid pairs for analysis: 4–5 y (n = 20 pairs), 12–13 y (n = 21 pairs), and 16–17 y (n = 20 pairs).

Anthropometric measurements

Anthropometric measurements were made on the day of DLW dosing for all volunteers while they were wearing light clothing and no shoes and socks (Figure 1). Weight was measured to the nearest 0.2 kg with calibrated scales (Tanita TBF-531; Tanita, Tokyo, Japan), and height was measured to the nearest 0.1 cm with a calibrated stadiometer (Chasmons Ltd, London, United Kingdom).

Criterion methods

Doubly labeled water (PAEE)

Total EE (TEE) was measured by using the DLW technique over 11 consecutive days. Each volunteer received a weighed dose equivalent to 0.174 g H\textsuperscript{18}O and 0.07 g H\textsubscript{2}O/kg body wt, after providing 2 predose urine samples: 1 on the day of dosing and 1 on the preceding day. Volunteers were instructed to provide the first postdose urine sample ≈24 h after dosing and then at a similar time, but not the first void of the day, for the next 10 d (Figure 1). DLW analysis was carried out by using isotope ratio mass spectrometry, as described previously (18). TEE was calculated by using standard equations (19) and Schoeller’s estimation of carbon dioxide production (20), which normalizes $\Delta^{2}H/\Delta^{18}O$ space ratios to 1.04/1.01 = 1.03 (21, 22). TEE was obtained from carbon dioxide production, assuming carbohydrate, fat, and protein substrate oxidation with a respiratory quotient of 0.85 (23). REE was measured for a minimum of 15 min by using indirect calorimetry after at least a 2-h fast and 10 min of supine rest. Oxygen consumption and carbon dioxide production were measured with an online system (Jaeger Oxycon Pro; Viasys Health Care, Warwick, United Kingdom). A ventilated hood was used for this measurement in the oldest age group, and a hollow facemask was used in the other groups. Data were averaged over a 15-s epoch; the 2 most extreme values in each interval were disregarded. EE was then computed by using the de Weir equation (24). REE was determined by averaging EE over minutes 5–15 of the resting test. PAEE was calculated as PAEE = 0.9 × TEE – REE and PAEE/kg as (0.9 × TEE – REE)/kg body wt.

Accelerometry (physical activity)

Physical activity was objectively assessed by using the Actigraph accelerometer (model 7164; Manufacturing Technologies Inc, Shalimar, FL). The Actigraph has been shown to accurately assess EE in European children during free-living conditions (17, 25). The Actigraph was worn for 11 d, concurrent with the DLW measurement, set to record at 60-s epochs and placed centrally on the hip with the side of placement randomly assigned. Volunteers were asked to wear the monitors during waking hours and to remove them while bathing, showering, and swimming.

Accelerometry data were analyzed by using a batch processing program (www.mrc-epid.cam.ac.uk/Research/PA/Downloads.html) and was used as minutes spent in MVPA (MVPA\textsubscript{Act}), derived by using 3000 accelerometer cpm as the lower limit of moderate activity (26). Analyses were also carried out by using 1592 cpm as the lower limit of MVPA (27) because of a lack of consensus on the most suitable cutoffs when translating accelerometer intensity into physiologic intensity. When 20 min of consecutive zeros were present in the accelerometer data, they were removed and it was assumed that the monitor was unworn at that time; consequently, all days consisting of >600 min of valid data were included in the analysis.

Questionnaire methods

Youth Physical Activity Questionnaire

The Youth Physical Activity Questionnaire (YPAQ) was administered to the 2 oldest groups and was based on the Children’s Leisure Activities Study Survey (CLASS) (10). The original intent of the self-reported CLASS questionnaire for 10–12-y-olds was to assess type, frequency, and intensity of physical activity over a usual week and was compared with accelerometer-derived MVPA and EE data (10). The YPAQ lists 47 different activities with...
participants requested to self-report the frequency and duration of each activity for both week and weekend days over the past 7 d. Therefore, the YPAQ assesses mode, frequency, and duration of physical activity and sedentary activities throughout all domains, including school time and leisure time over the past 7 d.

**Children’s Physical Activity Questionnaire**

The Children’s Physical Activity Questionnaire (CPAQ) was administered to the youngest group and is also based on the CLASS questionnaire, very similar to the YPAQ but parentally reported and includes activities specific to young children, such as “playing in a playhouse.” The original intent of the proxy-reported CLASS questionnaire for 5–6-y-olds was to assess type, frequency, and intensity of physical activity over a usual week and was compared with accelerometer-derived MVPA and EE data (10). In this study, the CPAQ was parentally reported for the youngest group of volunteers over the past 7 d. The mother of the child completed all administrations of the questionnaire for all participants. Therefore, the CPAQ assesses mode, frequency, and duration of physical activity and sedentary activities throughout all domains over the past 7 d.

**Child Heart and Health Study in England Questionnaire**

This questionnaire is used in the Child Heart and Health Study in England (CHASE) Study, which is examining the health of ≈5000 9–10-y-old primary school children living in the United Kingdom (www.chasestudy.ac.uk) and is currently unpublished. This questionnaire consists of 25 questions and addresses the mode and frequency of physical activity and sedentary activities throughout all domains, including school time and leisure time, and also includes multiple-choice questions regarding lifestyle activities. Therefore, the CHASE questionnaire solicits semi-quantitative estimates of the duration spent in broad categories of physical activities and weekly frequency of discrete activities. It was administered to the 12–13-y-old group.

**Swedish Adolescent Physical Activity Questionnaire**

The Swedish Adolescent Physical Activity Questionnaire (SWAPAQ) is a translation of the self-reported past 7-d physical activity questionnaire that has been successfully validated against accelerometry in Swedish adolescents (28). The original questionnaire assessed frequency, duration, and intensity of activity during school, transport, and leisure time over the past 7 d, and data were used as the duration of MVPA and MET (metabolic equivalent)-minutes of physical activity calculated as duration × frequency × MET intensity (28). This version of the questionnaire consists of 25 questions and addresses the mode, frequency, and duration (in broad categories) of physical activity throughout all domains and was administered to the 16–17-y-old group.

**Questionnaire administration**

The questionnaires are available as “Supplemental Data” in the online issue. The questionnaires were either mailed to the parents of volunteers (4–5-y-olds) or distributed by school teachers (to 12–13- and 16–17-y-olds), and no instructions were given over that written on the questionnaires. For all volunteers, the first administration of the questionnaire was given on day 11 of the DLW measurement, referring to days 4–10 and then 7 d later referring to the same period, as shown in Figure 1. The first administration was used to assess validity; the second administration was used only to assess reliability. When a volunteer received 2 different questionnaires simultaneously, the YPAQ and CHASE for the 12–13-y-olds and the YPAQ and SWAPAQ for the 16–17-y-olds, the order in which the questionnaires were administered was randomly determined, however, in the same order for both administrations.

**Questionnaire data processing**

For the YPAQ and CPAQ, frequency and duration of listed physical activities were reported; these activities were assigned a MET value according to published values (29). Because only summary items were assessed in the SWAPAQ (eg, How much time on average did you spend doing vigorous physical activities?), it was not possible to assign individual MET values to specific activities. Consequently, values of 4 and 6 METs were assumed for moderate and vigorous physical activity, respectively, and MET-minutes were calculated as follows: duration × frequency × MET intensity.

Estimates of PAEE were derived from the YPAQ, CPAQ, and SWAPAQ by using a method similar to that described previously (2). It was assumed that 1 MET is equivalent to an oxygen consumption rate of 4.00 mL·kg⁻¹·min⁻¹ for 16–17-y-old adolescents and of 4.58 mL·kg⁻¹·min⁻¹ for 12–13-y-olds (30). For the 4–5-y-old children, the published values for the older age groups (30) were extrapolated down for 5-y-old children, which resulted in an estimate of 7.0 mL·kg⁻¹·min⁻¹ as the MET equivalent. The oxygen energy equivalent was assumed to be 0.0209 kJ/mL, and the formula used to estimate daily PAEE from the questionnaire data (PAEEQ) was as follows: PAEEQ (kJ·kg⁻¹·d⁻¹) = 1440 × [(0.0209 × MET) × (total MET-min/total time frame)]. Self-reported minutes of MVPA per week were also summed from the YPAQ, CPAQ, and SWAPAQ (MVPAAcc) for direct comparison with MVPADLW.

The CHASE questionnaire did not include information on activity duration, so MVPAQ and PAEEQ could not be calculated. Four multiple choice questions regarding active transport, school break activities, activity outside school, and the amount and frequency of "exercise that makes you out of breath" were summed into a lifestyle score to represent habitual physical activity. The multiple-choice answers to each question were ranked by using consecutive integers with the least active choice given a score of 1 and the most active option given the highest score (eg, 5). The numbers corresponding to each answer were then summed for each volunteer; the scores for each question are available from the corresponding author.

**Statistical analysis**

Stata 10.0 (Statacorp, College Station, TX) was used for all analyses. Differences in activity levels between age groups were assessed by using linear or logistic regression, depending on the nature of the variable. Spearman correlations were used to determine the ability of the questionnaires to rank physical activity and EE summary variables. Those values most directly comparable were used for this analysis; therefore, PAEEQ was correlated with PAEE_DLW and MVPAQ correlated with MVPA_Acc for the CPAQ, YPAQ, and SWAPAQ. However, the lifestyle score from the CHASE questionnaire was correlated with both MVPA_Acc and PAEE_DLW.
VALIDITY OF 4 PAQs IN YOUTH

RESULTS

Descriptive characteristics and objectively measured physical activity levels of the volunteers are displayed in Table 1; there was a wide range of weights and BMIs between and within groups. The sample was predominantly white; only 3 volunteers were from another ethnic group. On the basis of accelerometer with ≥1952 cpm as the lower threshold of MVPA, the oldest group was significantly less active than were the other groups (all P < 0.01). However, when ≥3000 cpm was used as the lower threshold of MVPA, there were no significant differences between minutes spent at MVPA in each age group. However, PAEE per kilogram body weight differed significantly between age groups. Spearman correlations between PAEE and MVPA assessed by accelerometry using 1952 counts as the lower threshold of MVPA (27) was also used to assess the degree of agreement between methods. The difference (estimation error) between predicted and criterion values was calculated (predicted − criterion) and plotted against the criterion for each questionnaire and age group. The mean difference and direction of any systematic bias were examined, and the extent of any heteroscedasticity was determined by using the Breusch-Pagan/Cook-Weisburg test. In case of any heteroscedasticity, the ratio limits of agreement were calculated on a log scale (32).

The reliability of the CPAQ, YPAQ, and SWAPAQ to assess PAEE and MVPA and the CHASE questionnaire to assess the lifestyle score was determined by using intraclass correlation coefficients (ICCs), derived by using one-factor analysis of variance. Test-retest reliability was assessed to establish how much of the total variation in questionnaire variables was between subject variation, therefore determining the ability of one administration of the questionnaire to reliably assess physical activity.

TABLE 1
Participant characteristics by age group

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of subjects</td>
<td>27</td>
<td>25</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>Age (y)</td>
<td>4.9 ± 0.7a</td>
<td>13.1 ± 0.3</td>
<td>17.1 ± 0.6</td>
<td>a &lt; b &lt; c (P &lt; 0.001)</td>
</tr>
<tr>
<td>Height (cm)</td>
<td>110.1 ± 8.4</td>
<td>161.4 ± 7.5</td>
<td>169.5 ± 8.8</td>
<td>a &lt; b &lt; c (P &lt; 0.001)</td>
</tr>
<tr>
<td>Weight (kg)</td>
<td>20.2 ± 4.1</td>
<td>50.6 ± 9.8</td>
<td>63.3 ± 9.7</td>
<td>a &lt; b &lt; c (P &lt; 0.001)</td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td>16.5 ± 1.7</td>
<td>19.6 ± 1.3</td>
<td>22.0 ± 2.5</td>
<td>a &lt; b &lt; c (P &lt; 0.001)</td>
</tr>
<tr>
<td>Overweight (%)</td>
<td>28</td>
<td>21</td>
<td>19</td>
<td>NSD</td>
</tr>
<tr>
<td>Sex (% male)</td>
<td>62</td>
<td>70</td>
<td>30</td>
<td>a &gt; c (P = 0.02); b &gt; c (P = 0.003)</td>
</tr>
<tr>
<td>TEE_DLW (kJ/d)</td>
<td>6535 ± 1114</td>
<td>11,759 ± 2227</td>
<td>12,058 ± 2990</td>
<td>a &lt; b &lt; c (P &lt; 0.001)</td>
</tr>
<tr>
<td>TEE_DLW (kJ/d)</td>
<td>6647 (4415–7753)</td>
<td>11,566 (8198–14,184)</td>
<td>11,681 (7984–20,359)</td>
<td>NA</td>
</tr>
<tr>
<td>PAL (TEE/REE)</td>
<td>1.67 ± 0.2</td>
<td>1.86 ± 0.3</td>
<td>1.78 ± 0.3</td>
<td>a &lt; b &lt; c (P = 0.006)</td>
</tr>
<tr>
<td>PAL (TEE/REE)</td>
<td>1.67 (1.39–1.92)</td>
<td>1.78 (1.53–2.07)</td>
<td>1.70 (1.37–2.12)</td>
<td>NA</td>
</tr>
<tr>
<td>PAEE_kW (kJ · kg⁻¹ · d⁻¹)</td>
<td>97.7 ± 28.0</td>
<td>83.0 ± 29.6</td>
<td>65.6 ± 27.7</td>
<td>a &gt; b &gt; c (all P &lt; 0.01)</td>
</tr>
<tr>
<td>PAEE_kW (kJ · kg⁻¹ · d⁻¹)</td>
<td>95.8 (52.8–176.0)</td>
<td>72.2 (49.2–179.6)</td>
<td>52.1 (26.3–153.2)</td>
<td>NA</td>
</tr>
<tr>
<td>MVPAAcc1952 (min/wk)</td>
<td>564.7 ± 200.8</td>
<td>547.2 ± 195.7</td>
<td>371.2 ± 194.3</td>
<td>a, b &gt; c (all P &lt; 0.01)</td>
</tr>
<tr>
<td>MVPAAcc1952 (min/wk)</td>
<td>540.8 (187.8–1072.8)</td>
<td>533.8 (249.5–994.0)</td>
<td>310.3 (120.4–889.0)</td>
<td>NA</td>
</tr>
<tr>
<td>MVPAAcc3000 (min/wk)</td>
<td>250.1 ± 133.3</td>
<td>272.0 ± 148.9</td>
<td>232.2 ± 165.5</td>
<td>NSD</td>
</tr>
<tr>
<td>MVPAAcc3000 (min/wk)</td>
<td>171.1 (42.0–742.0)</td>
<td>244.4 (113.2–693.0)</td>
<td>211.6 (47.6–168.7)</td>
<td>NSD</td>
</tr>
<tr>
<td>Accelerometer (total counts/d)</td>
<td>547,714 ± 130,090</td>
<td>495,900 ± 136,806</td>
<td>345,823 ± 119,886</td>
<td>a, b &gt; c (P &lt; 0.001)</td>
</tr>
<tr>
<td>Accelerometer wear time (d)</td>
<td>8.0 ± 2.0</td>
<td>8.5 ± 2.7</td>
<td>7.8 ± 3.0</td>
<td>NSD</td>
</tr>
</tbody>
</table>

1 NSD, no significant difference; NA, not applicable; REE, resting energy expenditure; DLW, doubly labeled water; TEE_kW, total energy expenditure assessed by DLW; MVPAAcc1952, moderate and vigorous physical activity assessed by accelerometer using 1952 counts as the lower threshold of MVPA (27); MVPAAcc3000, MVP assessed by accelerometer using 3000 counts as the lower threshold of MVPA (26); PAEE_DLW, physical activity energy expenditure assessed by DLW; PAL, physical activity level.

2 Differences between groups were calculated by using linear regression, except for overweight and sex (logistic regression). The Cole et al (33) threshold was used to define overweight.

3 Mean ± SD (all such values).

4 Median; range in parentheses (all such values).
lower MVPA threshold (≥1952 cpm) was used, the mean difference between MVPAQ and MVPAAcc was nonsignificant for all questionnaires, as indicated by the 95% CI (Table 4). The SWAPAQ had the largest CIs, which indicated the least accurate predictions of MVPAQ. Strong correlations were present in most Bland-Altman plots (Figure 2), which indicated that the degree of questionnaire error was dependent on activity level and manifested as underreporting at higher activity levels for the CPAQ and SWAPAQ. Additionally, MVPAQ estimates from the YPAQ (16–17-y group) showed heteroscedasticity, which indicated an increased measurement error as activity level increased. The ratio limits of agreement are also shown in Table 4. These limits indicate that the mean bias ranged from an underestimation of 49% to an overestimation of 43%, and any individual estimate may have differed by between 4.3 and 11.6 times the true value, depending on the questionnaire and age group. When the higher threshold of MVPA (≥3000 cpm) was used, the mean difference between MVPAQ and MVPAAcc was nonsignificant, as indicated by the 95% CI (Table 4) for the YPAQ (16–17-y-old group) and the SWAPAQ. However, estimates from the CPAQ and YPAQ (12–13-y-old group) differed significantly from the criterion. The YPAQ (12–13-y-old group) had the largest CIs, which indicated the greatest overestimation of MVPAQ. Strong correlations were present in some Bland-Altman plots (Figure 2), which indicated that the degree of questionnaire error was dependent on activity level and was manifested as underreporting at higher activity levels for the YPAQ (16–17-y-old group) and SWAPAQ. Additionally, MVPAQ estimates from the YPAQ (12–13-y-old group) and the SWAPAQ showed heteroscedasticity, which indicated increased measurement error as activity level increased. The ratio limits of agreement are also shown in Table 4. These limits indicate that the mean bias ranged from an overestimation of 3% to 201%, and any individual estimate may have differed by between 2.2 and 3.0 times the true value, depending on the questionnaire and age group. Modified Bland-Altman plots of the criterion (PAEE<sub>DLW</sub>) and differences between PAEEQ and PAEE<sub>DLW</sub> are shown in Figure 3.

Mean differences, correlation of the error plots, and ratio limits of agreement are shown in Table 4. All questionnaires (PAEEQ) underestimated PAEE<sub>DLW</sub> and estimates from the YPAQ (16–17-y-old group) differed significantly from PAEE<sub>DLW</sub>. All other group-level estimates of PAEEQ were underestimated with nonsignificant but relatively wide CIs. There was a negative correlation between PAEE<sub>DLW</sub> and the difference (PAEEQ − PAEE<sub>DLW</sub>) in all Bland-Altman plots, which indicated an increasing underestimation of PAEE<sub>DLW</sub> by PAEEQ at higher activity levels. The ratio limits of agreement were calculated on a log scale and showed that these questionnaires underestimated PAEEQ and had wide ratio limits of agreement. Group-level PAEEQ estimates from these questionnaires were likely to underestimate PAEE<sub>DLW</sub> by between 56% and 24%, with any individual estimate differing by between 3.1 and 8.5 times the true value.

The test-retest reliability of all questionnaires is shown in Table 5. The summary variables calculated from the YPAQ showed high reliability between administrations. Reliability was somewhat lower for MVPAQ estimated from the CPAQ. The CHASE questionnaire showed significant reliability for the lifestyle score, but the SWAPAQ showed low reliability.

### DISCUSSION

This study assessed the validity and reliability of 4 physical activity questionnaires at assessing PAEE and MVPA (measured by DLW and accelerometry, respectively) in young persons of different ages. All questionnaires were valid at ranking either PAEE or MVPA, depending on age group and questionnaire. Some questionnaires were able to accurately assess group-level PAEE and MVPA, but error was large for individual estimates. A summary of the validity and reliability of all questionnaires is presented in Table 6.

The CPAQ and YPAQ (12–13-y-olds) were valid at ranking MVPA, and the YPAQ (16–17-y-old group) and SWAPAQ were valid at ranking PAEE. The lifestyle score from the CHASE

<table>
<thead>
<tr>
<th>Questionnaire summary variable</th>
<th>CPAQ (n = 27)</th>
<th>YPAQ, 12–13-y olds (n = 25)</th>
<th>YPAQ, 16–17-y-olds (n = 24)</th>
<th>SWAPAQ (n = 24)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MVPA&lt;sub&gt;Q&lt;/sub&gt; MVPA&lt;sub&gt;Acc&lt;/sub&gt;</td>
<td>486 (75–1540)</td>
<td>646 (21–2381)</td>
<td>332 (52–1169)</td>
<td>217 (0–750)</td>
</tr>
<tr>
<td>PAEEQ, PAEE&lt;sub&gt;DLW&lt;/sub&gt;</td>
<td>83 (27–229)</td>
<td>62 (2–227)</td>
<td>28 (4–98)</td>
<td>48 (0–275)</td>
</tr>
</tbody>
</table>

1 All values are medians; ranges in parentheses. MVPAQ, moderate and vigorous physical activity assessed by questionnaire; PAEEQ, physical activity energy expenditure assessed by questionnaire; CPAQ, Children’s Physical Activity Questionnaire; YPAQ, Youth Physical Activity Questionnaire; SWAPAQ, Swedish Adolescent Physical Activity Questionnaire.

<table>
<thead>
<tr>
<th>Questionnaire summary variable</th>
<th>CPAQ (n = 27)</th>
<th>CHASE (n = 25)</th>
<th>YPAQ, 12–13-y-olds (n = 25)</th>
<th>YPAQ, 16–17-y-olds (n = 24)</th>
<th>SWAPAQ (n = 24)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MVPA&lt;sub&gt;Q&lt;/sub&gt; MVPA&lt;sub&gt;Acc&lt;/sub&gt;</td>
<td>0.42 (P = 0.04)</td>
<td>—</td>
<td>0.42 (P = 0.04)</td>
<td>0.11 (P = 0.61)</td>
<td>0.23 (P = 0.27)</td>
</tr>
<tr>
<td>PAEEQ, PAEE&lt;sub&gt;DLW&lt;/sub&gt;</td>
<td>0.22 (P = 0.28)</td>
<td>—</td>
<td>0.09 (P = 0.67)</td>
<td>0.46 (P = 0.05)</td>
<td>0.40 (P = 0.04)</td>
</tr>
<tr>
<td>Lifestyle score: PAEE&lt;sub&gt;DLW&lt;/sub&gt;</td>
<td>0.45 (P = 0.02)</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>Lifestyle score: MVPA&lt;sub&gt;Acc&lt;/sub&gt;</td>
<td>—</td>
<td>0.12 (P = 0.57)</td>
<td>—</td>
<td>—</td>
<td></td>
</tr>
</tbody>
</table>

1 Values are Spearman correlation coefficients. MVPAQ, moderate and vigorous physical activity assessed by questionnaire (min/wk); PAEEQ, physical activity energy expenditure assessed by questionnaire (kJ · kg<sup>-1</sup> · d<sup>-1</sup>); CPAQ, Children’s Physical Activity Questionnaire; YPAQ, Youth Physical Activity Questionnaire; SWAPAQ, Swedish Adolescent Physical Activity Questionnaire; CHASE, Child Heart and Health Study in England; MVPA<sub>Acc</sub>, MVPA assessed by accelerometry (min/wk) using 3000 counts as the lower threshold (26); PAEE<sub>DLW</sub>, PAEE assessed by doubly labeled water (kJ · kg<sup>-1</sup> · d<sup>-1</sup>).
questionnaire was able to rank overall activity level in comparison with PAEE.

There was inconsistency in correlations between age groups and between estimates from the same questionnaire and the 2 different criterion methods; validity of the CPAQ and YPAQ differed by age group despite being similar questionnaires. For the CPAQ, this may be explained by the proxy-reported nature of this questionnaire. The difference in validity between the self-
reported YPAQ administered to the 2 different age groups is likely to be due to a combination of factors, including differences between the criterion methods, which measure different aspects of physical activity. However, these discrepancies may be partly due to differences in activity levels and activity profiles between groups. Depending on the threshold used, the 12–13-y-olds carried out significantly more MVPA than did the older group \( (P < 0.001) \), and higher-intensity activities were more accurately reported than were lighter-intensity activities \( (7) \). Age is also likely to influence validity because older children provide more accurate self-report data \( (14, 15) \). One could hypothesize that this may especially affect light-intensity activities, which are already more difficult to report accurately, but that may heavily influence PAEE. The thresholds used to determine MVPA also influenced the absolute validity of the questionnaires (Table 4). Because there is no consensus on the most suitable cutoffs for time spent at different intensity levels, the results based on 2 different MVPA intensity thresholds were included and were essentially unchanged between intensity thresholds.

As described in Subjects and Methods, various assumptions had to be made for the estimation of PAEE \( Q \) because there is no standardized reference for youth to parallel the 1 MET equivalent of 3.5 mL \( \cdot \text{kg}^{-1} \cdot \text{min}^{-1} \) for adults. Nonetheless, this error is much lower than that which would have resulted if adult standards were used (data not shown). Estimations of EE from self-reports in youth generally use adult-derived standard energy costs of specific activities \( (29) \). This is because there are no comprehensive reference values specific to youth, despite possible differences between MET multiples for the same activities in adults and children \( (34) \). However, it should be possible to use adult-derived values to estimate EE in youth if adjusted for the higher REE of children \( (30) \). These factors may unavoidably

### FIGURE 3

Modified Bland-Altman plots for the sum of physical activity energy expenditure assessed by questionnaire (PAEE\( Q \)) compared with PAEE assessed by doubly labeled water (PAEE\( DLW \)) for each questionnaire and age-group. Mean differences (PAEE\( Q \) – PAEE\( DLW \)) plotted against the mean of PAEE\( Q \) and PAEE\( DLW \) for the Children’s Physical Activity Questionnaire (CPAQ) in the 4–5-y-old group \( (A; n = 27) \), the Youth Physical Activity Questionnaire (YPAQ) in the 12–13-y-old group \( (B; n = 25) \), the YPAQ in the 16–17-y-old group \( (n = 24) \), and the Swedish Adolescent Physical Activity Questionnaire (SWAPAQ) in the 16–17-y-old group \( (n = 24) \).
Table 6
Summary of the validity and reliability of all 4 questionnaires

<table>
<thead>
<tr>
<th>Questionnaire</th>
<th>Age group</th>
<th>Able to rank MVPA?</th>
<th>Able to rank PAEE?</th>
<th>Able to rank lifestyle score?</th>
<th>Group-level prediction of MVPA_{Acc1952}?</th>
<th>Group-level prediction of MVPA_{Acc3000}?</th>
<th>Group level prediction of PAEE?</th>
<th>Reliable estimates?</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPAQ</td>
<td>4–5 y</td>
<td>Yes</td>
<td>No</td>
<td>—</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes (MVPA)</td>
</tr>
<tr>
<td>YPAQ</td>
<td>12–13 y</td>
<td>Yes</td>
<td>No</td>
<td>—</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes (MVPA)</td>
</tr>
<tr>
<td>CPAQ</td>
<td>16–17 y</td>
<td>No</td>
<td>Yes</td>
<td>—</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>CHASE</td>
<td>12–13 y</td>
<td>—</td>
<td>Yes</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>Yes</td>
<td>Yes (MVPA)</td>
</tr>
<tr>
<td>SWAPAQ</td>
<td>16–17 y</td>
<td>No</td>
<td>Yes</td>
<td>—</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>

1 MVPA, moderate and vigorous physical activity; PAEE, physical activity energy expenditure; CPAQ, Children’s Physical Activity Questionnaire; YPAQ, Youth Physical Activity Questionnaire; SWAPAQ, Swedish Adolescent Physical Activity Questionnaire; CHASE, Child Heart and Health Study in England; MVPA_{Acc1952}, assessed by accelerometer (min/wk) using 1952 counts as the lower threshold of MVPA (26); MVPA_{Acc3000}, assessed by accelerometer (min/wk) using 3000 counts as the lower threshold of MVPA (26).

2 Compared with PAEE (kJ · kg^{-1} · d^{-1}).

Contribute to the limited validity of self-reports for the estimation of EE in children in this and other studies (35).

Previous studies examining the validity of the CLASS questionnaire, on which the YPAQ and CPAQ were based, suggested limited validity when compared with accelerometer data when parentally reported for 5–6 y-olds (r = −0.06–0.05) and when self-reported for 10–12-y-olds (r = −0.04 to r = 0.06) (10). The higher correlation coefficients observed in this study may have been partly due to our older sample of adolescents, because others have suggested an age threshold of 10 y as the cutoff for providing acceptable questionnaire data (14). The CLASS questionnaire assessed a “usual” week of activity, but the YPAQ and CPAQ assessed the past week. For validity purposes, data from the questionnaire and criterion are more directly comparable when referring to the same time period (ie, past week), and it may also be easier to report physical activity from the previous week than from a “usual” week. The original Swedish SWAPAQ showed better validity than did the current study: r = 0.51 for total accelerometer- assessed physical activity (28). This difference should not be due to age, because these British adolescents are slightly older than those originally assessed but may be explained by differences in sample sizes and activity levels between groups. The Swedish adolescents (20) were ∼15% more active than were those who participated in the present study. There are no published validation studies of the CHASE questionnaire with which to compare.

The small sample size and sex differences in each age group were limitations of this study but were unavoidable, primarily because of the high cost of the DLW method and difficulties with recruitment. However, because these questionnaires were administered with no verbal instructions from the investigators, these results should indicate the realistic reliability and validity of these questionnaires if they were to be used in a large epidemiologic study in a similar population. It was deemed necessary to use different questionnaires in the 3 age groups because cognitive abilities and activity profiles of different age groups vary substantially. This is supported by the fact that the validity of the CPAQ and YPAQ differ by age group, despite being very similar questionnaires. These differences indicate that data from the same questionnaire may not be comparable across different age groups. Although the validity of all 4 questionnaires is therefore not directly comparable, the evaluation of 4 age-specific questionnaires should be useful to those carrying out research on physical activity in different age groups of youth. Another limitation is that the 3 methods applied in this study—an accelerometer, DLW, and questionnaires—assess somewhat different dimensions of physical activity, and these results must therefore be interpreted with that in consideration. However, the associated strength of this design was the use of 2 different criterion methods, which allowed the simultaneous assessment of questionnaire validity to measure both total volume of physical activity (PAEE) and physical activity intensity and duration (MVPA). Additionally, use of only the summary variables considered here (MVPA_{Q} and PAEE_{Q}) somewhat limited the scope of these questionnaires, which do hold more comprehensive information on specific physical activities, including sedentary behavior. An additional subjective method of measurement, such as a detailed interview, diary, or direct observation, would be required to assess the validity of these sections.

The YPAQ was the most reliable questionnaire, which showed significant intraclass correlations for PAEE_{Q} and MVPA_{Q}. The lifestyle score (CHASE) and MVPA_{Q} from the CPAQ also showed acceptable reliability. The SWAPAQ had poor reliability, despite being completed by the oldest group; the reliability of this questionnaire was not assessed in the previous study and thus cannot be compared (28). The structured nature of the YPAQ, including detailed activity lists, compared with the open nature of the SWAPAQ questionnaire could account for this finding. The proxy-report nature of the CPAQ could explain the lower reliability because recall of the children’s physical activity is difficult for adults (36). Proxy reporting may lead to limitations in questionnaire validity, which also may apply to reliability because an adult is unable to constantly monitor a child and may also be in charge of other children.

For reliability purposes, the 2 administrations of all questionnaires were answered in reference to the same 7 d. The advantage of this is that the differences between the 2 administrations should only consist of reporting error, with no variation due to real differences in physical activity levels over time. The subsequent limitation of this method is the recall of a period further back in time; consequently, reporting error is likely to be greater with the second administration. Additionally, the short time interval between administrations may mean that the volunteers remembered their answers from the first administration.

Absolute PAEE and MVPA estimated from these self-reports were not valid on an individual level in young persons, although some questionnaires may rank individuals accurately for MVPA (CPAQ and YPAQ in 12–13-y-olds), PAEE (YPAQ in 16–17-y-olds and SWAPAQ), and overall physical activity (CHASE). Age,
the outcome variable of interest, and whether individual or group-level estimates are necessary will influence the best choice of measurement method when assessing physical activity in youth.

We thank Alice Tompson, Kate Westgate, Charlotte Ridgway, Paul Roberts, Rosalind Stevenson, and all of the field epidemiology and WTCRF teams for assistance with testing. We are also grateful to Mark Hemings for development of the Actigraph processing software and Les Black for valuable comments on the manuscript.

The authors' responsibilities were as follows—KC: collected all of the data, conducted the data analyses, and drafted the manuscript; EMFvS, PW, and UE: compiled the questionnaires; NJW and UE: supervised the data collection; and AW: analyzed and interpreted the doubly labeled water data. All of the authors provided critical input on the data analyses and on all versions of the manuscript and approved the final version. None of the authors had any conflicts of interest.

REFERENCES