Assessment of physical activity in youth

Kirsten Corder, Ulf Ekelund, Rebekah M. Steele, Nicholas J. Wareham, and Søren Brage
Medical Research Council Epidemiology Unit, Institute of Metabolic Science, Cambridge, United Kingdom

Submitted 29 January 2008; accepted in final form 17 July 2008

Corder K, Ekelund U, Steele RM, Wareham NJ, Brage S. Assessment of physical activity in youth. J Appl Physiol 105: 977–987, 2008. First published July 17, 2008; doi:10.1152/japplphysiol.00094.2008.—Despite much progress with physical activity assessment, the limitations concerning the accurate measurement of physical activity are often amplified in young people due to the cognitive, physiological, and biomechanical changes that occur during natural growth as well as a more intermittent pattern of habitual physical activity in youth compared with adults. This mini-review describes and compares methods to assess habitual physical activity in youth and discusses main issues regarding the use and interpretation of data collected with these techniques. Self-report instruments and movement sensing are currently the most frequently used methods for the assessment of physical activity in epidemiological research; others include heart rate monitoring and multisensor systems. Habitual energy expenditure can be estimated from these input measures with varying degree of uncertainty. Nonlinear modeling techniques, using accelerometry perhaps in combination with physiological parameters like heart rate or temperature, have the greatest potential for increasing the prediction accuracy of habitual physical activity energy expenditure. Although multisensor systems may be more accurate, this must be balanced against feasibility, a balance that shifts with technological and scientific advances and should be considered at the beginning of every new study.

measurement; self-report; pedometry; accelerometry; heart rate

THE NATURE OF THE RELATIONSHIP between habitual physical activity and specific health outcomes varies considerably. Although such heterogeneity often has true biological origins, alternative explanations include the challenges surrounding the accurate assessment of physical activity. Precise methods of measurement are necessary to more accurately establish the dose-response relationship with various health outcomes, to monitor the effect of interventions, determine temporal trends in physical activity, and to make cross-cultural comparisons (120).

Determining the relationship between physical activity and health in youth may be more difficult than in adults because the associations may be weaker (47). This may be due to higher activity levels in children (ceiling effects), and also a longer lifetime of exposure for adults, with more time for disease to develop.

Physical activity dimensions include intensity, frequency, and duration, which together make up the total volume of activity. Another important dimension of physical activity is type or mode, e.g., walking and cycling. For some research questions, it may also be relevant to know something about the setting in which the physical activity takes place, e.g., inside or outside, playing with peers, supervised by adults, but this dimension will not be considered in depth in this review. The research question being posed in any particular study is key to identifying the most appropriate assessment method(s) for physical activity; hence unique considerations are required for measurement in youth. Other factors affecting the choice of instrument include study size, budget, resources, and staff available (85, 120).

Habitual physical activity may be viewed as a latent, not directly observable, time series of activity type and intensity, which varies greatly over short periods of time (seconds, minutes, hours), and from which researchers can sample and summarize, for example, into average daily estimates. In this review, the term intensity is used broadly and covers a range of different phenomena, e.g., vertical acceleration of the body, the beating frequency of the heart, or the rate of chemical energy expended over and above that of the body’s basal metabolic requirements. The latter will be referred to as physical activity energy expenditure (PAEE). Intensity variables are expressed in relation to time (e.g., meters per second^2, beats per minute, energy expenditure per day). For the purpose of this review, we will use the prefixes “daily” and “instantaneous” to specify the length of the reference time interval. In both adults and children, it is not possible to directly measure instantaneous PAEE during free-living conditions over prolonged periods of time; therefore, it must be inferred, for example, from answers to an activity questionnaire, registration of body movement, physiological parameters, or a combination of these. There is a critical distinction between what is measured and a variable...
that is the result of processing the raw information. The uncertainty in the prediction of a physical activity outcome at the end of this process will depend on the accuracy by which the raw measurement was obtained and on the assumptions made during processing.

Despite much progress with physical activity assessment, the limitations concerning measurement accuracy are often amplified in young people due to the cognitive, physiological, and biomechanical changes that occur during natural growth and development (4, 21, 49, 96, 99, 107) as well as the more intermittent pattern (i.e., more variable time series) of habitual physical activity in youth compared with adults. Young people’s physical activity is very intermittent (6, 8, 11), with up to 96% of activity bouts shorter than 10 s, with the majority lasting between 3 and 22 s (8, 11). This has implications for all aspects of measurement, processing, and interpretation of physical activity data in youth, including what data sampling frequency and epoch length to use and where to place activity monitors. Although many methodological considerations can be generalized between adults and children, this is not universal. Measurement issues relevant to the assessment of physical activity in youth will be discussed in more detail below.

Physical activity assessment methods can be divided into subjective and objective methods, which assess different aspects of physical activity and may be combined in any study. A review of specific methods in use for assessing habitual physical activity in youth was published in 2000 (56). This mini-review gives a summary of recent self-report methods but mainly focuses on objective methods and discusses the main issues regarding data interpretation.

SUBJECTIVE METHODS

Subjective methods include questionnaires, interviews, activity diaries (logs), and direct observation. The latter is useful for the assessment of physical activity in controlled situations and as a validation criterion, but the substantial investigator burden and invasion of study participant privacy make it unsuitable for use during free living. Recent studies, from 1997 onward, reporting validity of self-report instruments for assessing physical activity in children and adolescents are shown in Table 1. We restricted this summary to studies comparing the self-report to an objective method.

The accuracy of information collected by subjective instruments is influenced by the ability to accurately recall all relevant details retrospectively, but it may also be influenced by the opinion and perception of the participant, proxy reporter, or investigator. For young children, self-report methods hold the extra limitation that a child may be less able to recall their physical activity than an adolescent or adult. This may be because their activity pattern is more variable and harder to remember, but differences in cognitive and linguistic ability by age also play a role (96).

The age of the target population will influence the selection of self-report method. Proxy-reported methods are appropriate for young children. However, these methods are susceptible to additional problems because recall of children’s physical activity is difficult for adults (76). For example, a teacher is unlikely to be able to constantly monitor any one child for long periods, and the teacher may also be responsible for other children.

The validity of interview-administered methods is higher than self-administered methods (Table 1). If an interview cannot be undertaken in a study setting, having an adult to check completion of a child’s self-report questionnaire is a reasonable alternative. However, reports of validity using this approach should clarify how much assistance children received because this may impact not only the validity of the method but also its feasibility in the real world. Activity diaries need young people to report specific activities in a predetermined period of time (e.g., each 15 min). Consequently, there is a high participant burden, which limits the usefulness and may also affect behavior. Even a 15-min epoch may be too long to capture some short-term activities, but shortening the epoch renders the diary too burdensome (17). Activity diaries have been successfully used in adolescents but not younger children, who cope less well with this complex task.

The estimation of energy expenditure from self-reports in childhood generally uses adult-derived standard energy costs of specific activities (1). There are no comprehensive reference values specific for youth. Furthermore, there are possible differences between metabolic equivalents multiples for the same activities in adults and children (45, 104). In addition, resting energy expenditure expressed per kilogram body weight is higher in children than in adults. These factors may explain the poor validity of self-report methods for the estimation of energy expenditure in children (97). Nevertheless, self-report methods are valuable for the assessment of activity setting and mode of activity behaviors and its determinants, which may be more difficult to assess objectively.

OBJECTIVE METHODS

Objective methods for assessing physical activity involve the measurement of physiological or biomechanical parameters and use this information to estimate physical activity outcomes, such as instantaneous and daily PAEE. Figure 1 shows the essential elements of the staged process of moving from what an instrument actually measures to the estimated parameter that the scientific user might report. Obviously, there are uncertainties at every step.

Pedometry. Pedometers usually consist of a horizontal spring-suspended lever arm that moves with the vertical acceleration of the hips during ambulation (115). Pedometers count the number of times a certain acceleration threshold is exceeded (mechanical pedometers) or the number of zero crossings in the acceleration waveform (piezoelectric pedometers) and sum this to give an overall estimate of steps taken. They are generally cheaper than accelerometers and thus more feasible for use in large studies (101). However, many models only store the total number of steps (not the time series of step frequency) with no additional information on the time over which these were accumulated. Consequently, these pedometers cannot assess intensity, duration, or frequency of activity bouts but only provide a value of total ambulatory activity. Newer pedometers store a daily value for the last 7 days (106), and some even store information about the time when the sensor was in motion, allowing a little more insight into children’s physical activity behavior (10). Correlations between VO2 and step frequency assessed by a hip-mounted pedometer reach as high as 0.92 during unregulated play in 9-yr-old children (38). Despite this high correlation in a specific context, the relationship between step frequency and energy expenditure is variable between walking, running, and other biomechanically different activities. There have been few free-living validation studies of pedometers and only one in children against doubly labeled water. This study indicated that daily pedometer counts could explain 58% of the variance in daily PAEE per kilogram (82), which is comparable with results from accelerometer validation studies.

Outputs from different pedometer brands are not comparable. A recent comparative study of pedometers in children during different walking speeds showed that although accuracy is generally good above a walking speed of ~3.2 km/h, error generally increases below this speed, and it is variable by monitor brand (9). Similar issues occur with the estimation of distance from pedometers (27), because distance is the product of stride length and frequency (101). Pedometer data may therefore not be comparable across different age groups due to differences in stride length. For example, if a child walks 10,000 steps with a stride length of 50 cm, the total distance would be 5 km. However, another child with a stride length of 75 cm would walk 7.5
### Table 1. Self-report methods validated for assessing physical activity in youth

<table>
<thead>
<tr>
<th>Reference</th>
<th>Self-Report Method</th>
<th>Participants</th>
<th>Time Frame</th>
<th>What Is Assessed</th>
<th>Administration</th>
<th>Criterion Method</th>
<th>Validity</th>
</tr>
</thead>
<tbody>
<tr>
<td>109</td>
<td>Assessment of Young Children’s Activity Using Video Technology (ACTIVITY)</td>
<td>47 Children (7.7 yr)</td>
<td>Past day</td>
<td>Habitual and MVPA</td>
<td>Computerized Caltrac, HR</td>
<td>Caltrac: r = 0.40 (P &lt; 0.001) Minutes over 50% maximum HR; r = 0.50 (P &lt; 0.001)</td>
<td></td>
</tr>
<tr>
<td>121</td>
<td>Youth Media Campaign Longitudinal Survey</td>
<td>192 Children and adolescents (9–19 yr)</td>
<td>Past day and past 7 days</td>
<td>Bouts and minutes of MVPA</td>
<td>Interview (telephone) Actigraph</td>
<td>Activity Caltrac, HR: r = 0.53 and 0.37 for time and bouts of MVPA (previous day); r = 0.24 for total weekly PA; r = 0.31 for bouts of weekly PA</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Physical activity recall (PAR)</td>
<td>46 Girls (12 yr)</td>
<td>Past day</td>
<td>Estimation of EE</td>
<td>Interview administered HR, Caltrac</td>
<td>HR: r = 0.50 (P &lt; 0.01) Caltrac; r = 0.2 (P &lt; 0.01)</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Physical Activity Questionnaire for Adolescents (PAQA)</td>
<td>33 Adolescents (15.7 ± 0.4 yr)</td>
<td>Habitual week</td>
<td>PA during school, transport and leisure during habitual week</td>
<td>Interview administered DLW</td>
<td>r = 0.62 (P &lt; 0.001) Underestimated by 3.8 ± 1.7 MJ (P &lt; 0.001)</td>
<td></td>
</tr>
<tr>
<td>53</td>
<td>Habitual Activity Questionnaire (HAQ)</td>
<td>683 Children (9–10 yr)</td>
<td>Past year</td>
<td>Mode and frequency of sport and PA outside school</td>
<td>Interview administered Caltrac</td>
<td>r = 0.09 for both years (P &lt; 0.02)</td>
<td></td>
</tr>
<tr>
<td>103</td>
<td>Minnesota Leisure Time Physical Activity Questionnaire (MLTPAQ)</td>
<td>35 Adolescents (15 yr)</td>
<td>Past year</td>
<td>Leisure time physical activity TEE estimated</td>
<td>Interview administered DLW (TEE)</td>
<td>r = 0.49 (P &lt; 0.01) to r = 0.73 (P &lt; 0.01)</td>
<td></td>
</tr>
<tr>
<td>88</td>
<td>Activity diary</td>
<td>20 Children and adolescents (5.5 to 16 yr)</td>
<td>24 h</td>
<td>Estimated TEE</td>
<td>Parent and child reported, any discrepancies resolved with staff member HR-estimated EE</td>
<td>HR: R² = 0.72 no significant difference between estimate of EE</td>
<td></td>
</tr>
<tr>
<td>46</td>
<td>Physical activity questionnaire</td>
<td>62 Children (4–8 yr)</td>
<td>1 day</td>
<td>MVPA</td>
<td>Parent or teacher reported Accelerometer, HR</td>
<td>HR: r = 0.40–0.45 (P &lt; 0.01–0.001) Accelerometer: r = 0.53 (P &lt; 0.0001) r = 0.33 (P &lt; 0.001)</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>Checklist to record outdoor playtime Recall of outdoor playtime</td>
<td>250 Children, (3.7 yr)</td>
<td>3 days</td>
<td>Outdoor playtime Parentally reported</td>
<td>Parentally reported Accelerometer (RT3 Triaxial)</td>
<td>Parentally reported Accelerometer (RT3 Triaxial)</td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>Activity diary</td>
<td>30 Adolescents (15 yr)</td>
<td>3 days</td>
<td>TEE and time spent at different intensity levels Self-administered Self-administered HR</td>
<td>Self-administered Accelerometer (Caltrac)</td>
<td>Mean difference 3.4 ± 14.4 (not significant)</td>
<td></td>
</tr>
<tr>
<td>53</td>
<td>Activity diary</td>
<td>69 Children (9–10 yr)</td>
<td>3 days</td>
<td>Short-term physical activity Minutes of MVPA</td>
<td>Self-administered Accelerometer (Caltrac)</td>
<td>r = 0.2 to 0.25 for years 3–5 (P &lt; 0.0001)</td>
<td></td>
</tr>
<tr>
<td>57</td>
<td>Physical Activity Questionnaire for older Children (PAQ-C)</td>
<td>97 Children and adolescents (9–14 yr)</td>
<td>Past 7 days</td>
<td>Minute of MVPA</td>
<td>Self-administered Accelerometer (Caltrac)</td>
<td>r = 0.39</td>
<td></td>
</tr>
<tr>
<td>58</td>
<td>Physical Activity Questionnaire for Adolescents (PAQ-A)</td>
<td>85 Adolescents (13–20 yr)</td>
<td>Past 7 days</td>
<td>Total activity</td>
<td>Self-administered Accelerometer (Caltrac)</td>
<td>r = 0.33</td>
<td></td>
</tr>
<tr>
<td>77</td>
<td>3-Day Physical Activity Recall (3DPAR)</td>
<td>70 Adolescents (8th and 9th graders African-American)</td>
<td>Past 3 days</td>
<td>Overall, MVPA and Vigorous PA</td>
<td>Self-administered Accelerometer (CSA)</td>
<td>Over 7 days: r = 0.35–0.51, (P &lt; 0.01) Over 3 days: r = 0.27–0.46 (P &lt; 0.05)</td>
<td></td>
</tr>
<tr>
<td>110</td>
<td>GEMS Activity Questionnaire (GAQ)</td>
<td>68 African-American girls (8–9 yr)</td>
<td>Past day and “habitual”</td>
<td>Habitual PA</td>
<td>Self-administered Accelerometer (CSA)</td>
<td>Previous day: r = 0.27 (P &lt; 0.03) Habitual: r = 0.28 (P &lt; 0.02) CAS: r = 0.37 (P &lt; 0.002) Pedometer r = 0.47 (P &lt; 0.001)</td>
<td></td>
</tr>
<tr>
<td>110</td>
<td>Activitygram</td>
<td>68 African-American girls (8–9 yr)</td>
<td>Past 3 days</td>
<td>Mode, intensity and duration of activity or rest Self-administered Accelerometer (CSA), pedometer</td>
<td>Self-administered Accelerometer (Trirac)</td>
<td>r = 0.51 (P &lt; 0.0003) significantly overestimated EE No group-level significant difference</td>
<td></td>
</tr>
<tr>
<td>69</td>
<td>Computerized Activity Recall (CAR)</td>
<td>45 Children (11.8 ± 1.0 yr)</td>
<td>Past day</td>
<td>Total activity and TE Self-administered (computerized)</td>
<td>Accelerometer (Trirac) DLW</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>Activity diary</td>
<td>50 Adolescents (15 yr)</td>
<td>7 days</td>
<td>TEE and PA level Self-administered after tuition and followed by discussion</td>
<td>Accelerometer (CSA) DLW</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

**continued**
km for the same amount of steps accumulated. Depending on which physical activity outcome is considered, the two children are either equally active (10,000 steps) or the second child is 50% more active than the first. Some pedometers may also give estimates of energy expenditure using estimates of stride length, sex, weight, or age (114).

Several large-scale studies have successfully used pedometers for the assessment of ambulatory physical activity in youth (29, 30, 61, 119). The relatively simple nature of pedometer output in steps per day makes it suitable for comparing walking levels between populations and studies because there is a limited number of data reduction techniques which can be used to summarize data. However, there may be potential confounding from certain activity types that are not measured by pedometry. Our recommendation for pedometers is to use them for large-scale studies when resource limitations prevent the use of more advanced objective methods or when a total volume of walking activity is the outcome of interest. Pedometer output should be expressed as steps per day without any further inference of distance or energy expenditure because the level of uncertainty in these predictions may be unacceptably high.

**Accelerometry.** Accelerometry is the most commonly used objective method of physical activity assessment in youth, and it has recently greatly increased in popularity relative to other objective methods in all age groups (91). A recent review of physical activity measurement in preschool children reported that 63% of habitual monitoring devices used were accelerometers, mainly of the same model (73). Accelerometers vary in size and cost and their use in youth has been described in detail elsewhere (28, 84, 91). There is not any one accelerometer that can be recommended for use over others because there have not been many direct comparisons of different accelerometers in children and adolescents, but these studies would be valuable for future research. A general recommendation would be that, for youth, an accelerometer should be small, able to measure acceleration accurately in the range of movement specific to young people, have high within- and between-instrument reliability, and store as highly time-resolved information as possible for the desired monitoring period.

An accelerometer quantifies one or more dimensions of movement of the body segment to which it is attached. Acceleration is a change in velocity with respect to time (SI unit: m/s²), enabling accelerometers to quantify the intensity of movement for that body part (20, 22, 40, 65). In many studies, however, accelerometry data are expressed as “movement counts,” an arbitrary value that is often not comparable between monitor brands (84). The translation of “counts” into an estimate of physiological activity intensity is contentious (8, 70, 80). This debate makes setting physical activity intensity thresholds difficult (44, 84, 91). Intensity thresholds for moderate and vigorous intensity physical activity, derived specifically for use in young people, vary widely from between 615 to 3,200 counts/min even with the same accelerometer model (8, 70, 80). Using these different thresholds on the same data, it is possible to show that the same group of adolescents is either inactive or sufficiently active, defined as accumulating at least 60 min of moderate-to-vigorous physical activity (MVPA) per day (Fig. 2, unpublished data). The thresholds for sedentary behavior are also variable, although there has been less examination of accelerometer-derived intensity thresholds to deter-
mine sedentary time, especially in youth (83, 84). This is, however, a growing area of research for which the intensity distribution of acceleration in the low range, and the sensitivity of different accelerometers to measure these low accelerations becomes important. Our recommendation is to move from the use of the somewhat arbitrary count-based cut points that may not be comparable between monitor brands and toward a more universally comparable approach of using acceleration (m/s²) to summarize accelerometry data.

Some researchers have attempted to go even further than categorizing activity into different intensities and to use accelerometer data as the basis for predicting energy expenditure equations. The key issue here is whether laboratory-derived prediction equations in children can be generalized to free living. This in turn is dependent on the range of activities included in the laboratory study that produced the prediction equations. The majority of studies have used sedentary activities and activities with mainly vertical movements to derive the prediction equations (23, 24, 67, 81, 113). This appears to be valid in laboratory settings but not necessarily in free-living (59, 71). A separate issue is whether or how the prediction equation deals with extreme data. As an illustration of this issue, Fig. 3 shows average daily PAEE estimates from doubly labeled water in a sample of 28 adolescents aged between 15 and 17 yr old (unpublished data) and the estimate from two PAEE prediction equations derived in youth (23, 113). These prediction equations were applied in their original form but also using a data filter for extreme values. The original equation derived using only flat treadmill activity (113) overestimated daily PAEE, and the equation based on lifestyle activities (23) underestimated daily PAEE. The overestimation from the treadmill-derived equation was dramatically reduced when the data filter was applied.

Prediction equations for free-living PAEE derived on the basis of doubly labeled water measurements may be more accurate than those using laboratory activities, but these equations can only be used for daily PAEE prediction and not instantaneous PAEE prediction. Alternative analytical strategies have also been developed for instantaneous PAEE prediction; these include decision trees and artificial neural networks (22, 26, 74, 89, 124, 125) and addition of physiological parameters (12, 24, 41, 42, 50, 54, 123). More recently, activity-type classification schemes for interpretation of accelerometry data have emerged that aim to detect a multidimensional movement signature and assign membership to a (limited) set of activity types (7, 37, 48, 75, 79, 122). These methods have not been fully investigated in youth, although they should theoretically be valuable for the assessment of the unique physical activity patterns of youth. It has yet to be shown whether the more complex computations required for these analyses can be justified in large-scale epidemiological studies compared with more traditional approaches, because processing may be prohibitively slower.

**HR monitoring.** Heart rate (HR) monitoring may be used to assess PAEE in children, both in controlled and free-living environments (38, 62). However, it has several important potential limitations. The relationship between HR and instantaneous PAEE is relatively poor at low intensity activity and a large interindividual variation in HR at different activity intensities means that some form of individual calibration of the HR-PAEE [oxygen uptake (VO₂)] relationship (32, 90, 93) is required as depicted in Fig. 1. Earlier studies used universal HR intensity cut points to assess physical activity, irrespective of the large interindividual variation in HR in young people (3), but using HR to estimate PAEE in youth without some form of individual calibration is less accurate. HR monitoring is also susceptible to poor pick-up or interference, especially during free-living conditions, although novel electrode technologies may help overcome some of these issues (52, 102). Nonetheless, the necessity for skin contact impacts on the feasibility of the method in young people, due to their more sensitive skin with a potentially higher risk for allergic reactions. HR monitoring is useful in older children and adolescents but may be more difficult in young children. This is because they may be unsettled during calibration tests in unfamiliar environments, especially when a face mask is required while measuring VO₂. This may alter their HR and the measured HR-PAEE relationship. Additionally, the sporadic nature of young children’s physical activity, combined with the lag of the HR response to activity, may affect assessment of the intermittent pattern of activity (92).

The relationship between HR and instantaneous PAEE is approximately linear during moderate to vigorous activity intensities but not at low intensity activity and while sedentary, therefore a “flex point” is often used to define the point above which the HR-PAEE relationship becomes linear (62). Below the flex point, the HR-PAEE relationship is more uncertain because heart rate can be raised by environmental conditions such as anxiety, stress, or increased temperature without a corresponding increase in energy expenditure (12). The flex HR method generally assumes that instantaneous energy expenditure is equal to rest below the flex point (62). However, this is unlikely to always be the case, and this assumption will therefore introduce negative error into the prediction of PAEE. In contrast, setting the flex point too low will introduce positive bias. There is no consensus regarding the definition of flex HR, which is often defined as the average of the lowest HR during exercise and the highest HR
during rest (116). Some researchers have defined flex HR as the average between resting and exercising HR +5 or even +10 beats/min (31, 39). The flex HR point may also be calculated differently between studies using different types or intensities of exercise, or it may be estimated on the basis of sleeping HR. Although the flex HR method appears to be reasonably valid for estimating average daily energy expenditure on a group but not individual level in children (62), activity estimates using different applications of the flex HR method may not necessarily be comparable (63).

The individual HR-PAEE (V̇O₂) relationship is usually derived in a laboratory environment and then subsequently applied to predict PAEE in a free-living situation. This individual calibration can be time consuming and still may not be sufficiently precise to assess free-living PAEE on the individual level. The choice of activities in the calibration protocol, and in particular their intensity will affect the accuracy of the prediction, even though HR displays a more universal relationship with instantaneous PAEE across different activities than accelerometry (63). When it is not possible to carry out indirect calorimetry during a calibration procedure, simple step tests with a standard workload for even just sleeping HR in combination with age and sex can be used for simplified individual calibration, with some reduction in precision (13, 23, 24, 63).

To avoid the problems at low activity intensities, it has been suggested that HR monitoring should only be used to assess time spent in moderate and vigorous activity with HR above a certain level in children (86). The use of HR rather than accelerometry for the determination of MVPA has not received much attention. However, the use of HR reserve (HRR), the difference between maximum HR and resting HR, allows estimation of time spent in MVPA and incorporates some level of individual calibration by the use of resting HR and possibly age. Figure 2 includes an estimate of time spent in MVPA using 50% of HRR in adolescents aged 12–13 yr old, with the HRR defined as [200 beats/min – sleeping HR]. This estimate of MVPA is comparable to the estimates from uniaxial accelerometry in the same individuals when using the highest cut points (Fig. 2). HR information generally has less variation than accelerometric information in aspects of measurement, on-board processing, and data storage, in that it uses agreed standards (heart period in milliseconds, HR in beats per minute) to express the information; consequently HR data from different studies may be more comparable when used for the same purpose. As for accelerometry, however, there are also benefits of higher data resolution, since measurement noise is more efficiently handled with as much information as possible (105).

Although it has not been as widely used as accelerometry (36), HR monitoring has the potential, especially with new technologies, to be valuable for the assessment of both PAEE and MVPA. With current monitor designs it is less feasible for use in very young children. In older children and adolescents, this method could be considered, especially in populations with varied activity profiles, such as where cycling or hill walking are common activities, because these are captured less accurately by accelerometry. Combined sensors. Newer sensors combining one or more physiological measures with movement sensing have been used in adolescents (23, 24, 123). These new combined devices, including combined HR and movement sensors, combined movement and temperature sensors, and multisensor devices to determine the motion of multiple body segments, are still prohibitively expensive for use in large epidemiological studies. Nevertheless, many of these new methods provide interesting information on diverse characteristics of physical activity, which may be particularly applicable to children and certain clinical populations.

Prediction equations for instantaneous PAEE from separate accelerometry and HR equations combined by branched modeling give different weightings to HR and acceleration dependent on the level of HR and physical movement (12). Branched modeling reduces the error of energy expenditure prediction in adults (12, 25, 108) and children in controlled environments (23, 24). However, because the accuracy of methods generally tends to be greater in controlled environments, free-living studies are necessary to determine whether the accuracy of combined HR and movement sensing for prediction of daily PAEE is increased enough to outweigh the greater cost of measuring both movement and HR.

**ISSUES REGARDING PHYSICAL ACTIVITY ASSESSMENT IN YOUTH**

Table 2 gives a summary of the main features of the physical activity assessment methods discussed, but there are five major issues that influence the choice of method for any particular study.

“Hawthorne effect.” It is a general scientific fact that the process of observation alters the phenomenon being observed. In the field of physical activity measurement, this is a greater problem for objective than subjective methods, except activity diaries. Pedometers that display the numbers of steps may alter behavior. However, studies in children blinded to the step counts show that this effect is small (92, 118). Accelerometer counts are 3% higher during the first day of measurement than subsequent days in 11-yr-old children (68), suggesting an initial awareness of observation. Because the problem is not apparent on subsequent days, one solution may be to discount the first day entirely or to scale it.

**Placement, multiple sensors, and multidirectional monitoring.** This is an issue that is mainly relevant to movement sensors. Accelerometer and pedometer placement has not been systematically assessed in youth; however, outputs from monitors placed on the hip and waist do not differ substantially (72). These are also the most commonly used sites, possibly due to practical considerations because it may be intrusive for children to wear an ankle monitor. Most accelerometers used in the field are uniaxial, and pedometers are also only sensitive to movement in one axis. Some accelerometers are also sensitive to anteroposterior and/or bilateral movement and are thus biaxial or triaxial. The sporadic nature of young children’s physical activity may be more accurately measured by either multidirectional or multiple sensors, but this has not been investigated fully in children (91, 111). Triaxial accelerometer may be more accurate than uniaxial accelerometer to assess a range of activities in children, but it is uncertain whether this warrants the extra cost and size (38). Similarly, using multiple monitors attached to different body sites may explain more variance in activity than one alone, but the extra burden and cost may not be worth only marginal increases in accuracy (111).

**Sampling frequency and epoch length.** The frequency by which the raw signal is captured and the period over which it is summarized (the epoch) are critical issues. It is particularly important in young children, because of the short duration of their activity bouts (8, 11). Some accelerometers have memory capacity allowing data storage for over a week of continuous data collection in 5-s epochs, which captures more time spent at MVPA intensity than a 60-s epoch (72). The epoch length used should ideally be as short as possible, because data can always be reintegrated into a longer time frame but not vice versa.

**Representativeness of the monitored period to “usual” activity.** There is some evidence that between 4 and 9 full days of monitoring including 2 weekend days is required for a reliable estimate in youth (111). Seven days of continuous monitoring seem logical, but because protocol adherence tends to decrease with days of wear, it may be more feasible to opt for four full
days with at least 1 weekend day, as is often done in large
studies (87). The trade-off between feasibility and accuracy
was demonstrated in a large recent study of 11-yr-old children
that examined the number of days of monitoring required to
achieve reliability coefficients of 0.7, 0.8, and 0.9 (68). Three
days of monitoring were required to achieve a coefficient of
0.7, irrespective of whether a valid day was counted as ≥420
or ≥600 registered min/day. However, 5 days were required
for a coefficient of 0.8 and 11 days were needed to achieve a
reliability coefficient of 0.9. A total of 5,601 and 4,760 children
had ≥3 and ≥5 valid days (≥600 registered min/day) of data,
respectively, with numbers decreasing by a further 11 and 5%
with the additional requirement of both weekday and weekend
representation. There also appears to be an age effect, with
younger children having less day-to-day variability than older
children with regards to MVPA as it has been shown that 4–5
days of monitoring achieves a reliability coefficient of 0.80 in
young children, but between 8 and 9 days may be required for
adolescents (112). Because of the differences between week
and weekend days and the balance between feasibility and
validity, we suggest that as a minimum, studies in both children
and adolescents aim for at least 4 full days of monitoring
including one weekend day.

Few studies have examined this issue for pedometer or HR
data but it is as important for these methods as for accelerom-
etry because it relates primarily to the variability of the behav-
ior, and less to the measurement method. There is some
evidence that 5 days of wear provide reliable pedometer data
with a reliability coefficient above 0.8 in 10- to 14-yr-old girls
(106). In the absence of a detailed literature on required days of
wear for pedometer and HR data, it seems logical to follow the
suggestions for accelerometry.

Seasonality, due to school terms, school holidays, and cli-
mate, is another important factor to consider for estimating
habitual physical activity in youth. For all methods of mea-
surement, seasonal variation means that a single measurement
of a week may not adequately reflect children’s habitual
physical activity because there may be considerable annual
intraindividual variation (43, 66). If a habitual estimate of
activity, defined as an annual average, is required, measure-
ment over more than one season would be preferable. Ac-
knowledging the limitations of a particular measurement pro-
tocol and carrying out longitudinal measurements over the
same time of year may be a more practical approach in some
situations. Alternatively, statistical methods may be used to

### Table 2. Summary characteristics of objective assessment methods for free-living physical activity in youth

<table>
<thead>
<tr>
<th>Method</th>
<th>Pedometry</th>
<th>Accelerometry</th>
<th>HR Monitoring</th>
<th>Combined HR and Movement Sensing</th>
<th>Multiple Site Movement Sensing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Invasiveness</td>
<td>Not invasive</td>
<td>Not invasive</td>
<td>Can be bulky</td>
<td>Not invasive</td>
<td>Relatively invasive</td>
</tr>
<tr>
<td>Cost</td>
<td>Inexpensive</td>
<td>Moderate</td>
<td>Moderate</td>
<td>Moderate to expensive</td>
<td>Expensive</td>
</tr>
<tr>
<td>Dimensions assessed</td>
<td>Total steps: total daily ambulatory PA</td>
<td>Activity counts: total daily PA, intensity, frequency, duration</td>
<td>Total daily PA, intensity, frequency, duration</td>
<td>Total daily PA, intensity, frequency, duration</td>
<td></td>
</tr>
<tr>
<td>Epoch length</td>
<td>Not applicable</td>
<td>As short as possible, preferably ≤5 s</td>
<td>As short as possible, preferably ≤15 s</td>
<td>As short as possible, preferably ≤15 s</td>
<td></td>
</tr>
<tr>
<td>Length of measurement</td>
<td>Recommended at least 4 full days of recording with 1 weekend day</td>
<td>Recommended at least 4 full days of recording with 1 weekend day</td>
<td>Recommended at least 4 full days of recording with 1 weekend day</td>
<td>Recommended at least 4 full days of recording with 1 weekend day</td>
<td></td>
</tr>
<tr>
<td>Recommended output</td>
<td>Total steps per day</td>
<td>Acceleration, time spent at different movement intensities</td>
<td>Net HR, time spent at different physiological intensities</td>
<td>Acceleration and net HR, time spent at different physiological and movement intensities</td>
<td></td>
</tr>
<tr>
<td>Application</td>
<td>Large scale studies assessing ambulatory PA</td>
<td>Medium to large studies assessing overall PA, intensity frequency, or duration</td>
<td>Medium to large studies assessing overall PA, intensity frequency, or duration</td>
<td>Moderate-size studies requiring an assessment of PAEE</td>
<td></td>
</tr>
<tr>
<td>Validated during free living in youth?</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No, only in controlled environments</td>
<td></td>
</tr>
<tr>
<td>Accuracy</td>
<td>Accurate for assessing steps during most walking speeds</td>
<td>Accurate estimation of instantaneous PAEE during flat locomotion and sedentary activities, variable accuracy in other activities (heavily dependent of inference model)</td>
<td>Accurate estimation of instantaneous PAEE at higher activity intensities, especially with good individual calibration. In general, poor accuracy at low activity intensities</td>
<td>Accurate estimation of instantaneous PAEE across all intensity levels</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Accurate classification of activity type and estimation of instantaneous PAEE across all intensity levels</td>
</tr>
</tbody>
</table>

PAEE, physical activity energy expenditure.
overcome what is essentially an issue of undersampling of the behavior under study (59, 60).

**Summarizing physical activity data and handling missing data.** Non-wear time needs to be identified in data records from activity monitors, because they are not always worn continuously over an entire measurement period. Following this, there needs to be decisions about how to deal with this missing information. Various methods have been used to address this issue, including comparison with a diary of wear time, assumption of sleep time, or identifying data segments of continuous inactivity surpassing a length beyond which it is deemed unlikely that the monitor could have been worn, e.g., 10–30 min. This length may also depend on the sensitivity of the monitor. Different techniques have been used to deal with missing data. These range from exclusion of the time period (equivalent to imputation with the average of remaining data for an individual) to replacement by time-of-day specific values estimated from the remaining data on an individual or group level and various averaging schemes using scaling for weekdays and weekend days (19, 59). The choices made can have dramatic effects on the activity estimate. For example, one study in European children showed more than a doubling of 6-yr tracking coefficients of activity between different data reduction techniques (60). Another study compared a laboratory-derived accelerometry equation with a free-living (doubly labeled water) derived equation for estimation of PAEE in children and found the laboratory equation to overestimate PAEE by 83% with potential non-wear time identified and assumed to equal the average of remaining data. The same equation underestimated PAEE by only 3%, when sleep time was identified and PAEE set to zero in these segments (71).

Determination of non-wear time is commonly cited as a problem for accelerometry but is also important for other objective methods. There is no single accepted criterion for the identification of how much wear time is necessary to constitute a valid day of measurement. Approximately 10 h (600 min) of minimal daily wear time are often used for youth (33, 64, 87), which have been shown to maximize reliability in 5-yr-old children (78). When ≥420 and ≥600 min/day of data were used as the minimum recording length requirements in a study of 7,159 British 11-yr-old children, reliability was not substantially different. Because of the employment of many different strategies for identifying and dealing with missing or “non-wear time” in accelerometer data, comparability between studies is often difficult (91). Consequently, it has been suggested that accelerometers should be worn for 24 h a day (91), but this may not be practical due to discomfort during sleep, for example from monitors positioned around the waist. It is important that volunteers are encouraged to follow protocol, because if volunteers are instructed to wear monitors for 24 h a day but some remove the monitor at night, the data may not be comparable if expressed as a mean daily value. It may be most practical for volunteers to wear monitors during all waking hours and be strongly encouraged to follow protocol, whatever that may be.

Some combined HR and movement sensors are worn for 24 h a day, which makes assumptions of non-wear time less of an issue. However, the identification and subsequent handling of potentially noisy HR data, due to poor electrode-to-skin contact or electrical interference should, but does not necessarily, reduce the error in daily PAEE estimates. For example, if all erroneous data are removed, the problem does not disappear, but instantaneous PAEE in those segments is assumed to correspond to average instantaneous PAEE of the remaining data. This may in some cases be justified but alternative methods include more sophisticated noise classification schemes followed by inference of the latent HR trace (105). The continued development and subsequent evaluation of postprocessing algorithms are important for the field of habitual activity monitoring and for the estimation of PAEE during free living.

**FEASIBILITY VS. VALIDITY**

Generally, there is an inverse association between feasibility and validity of physical activity assessment methods for use in youth to measure or estimate different aspects of physical activity. Feasibility relates to the population and setting of a study as well as to the specific instrument(s) used to collect the raw information. Validity also relates to population and setting but is in principle device independent because it relates only to the raw information and how this is processed through all stages of complete inference (Fig. 1). Validity is ultimately defined by the uncertainty in the final result of the process of obtaining the raw data and processing it into an estimate of physical activity. The optimal balance between feasibility and validity is therefore dependent on which activity variable is considered in the study hypothesis. The doubly labeled water technique combined with a measure of resting metabolic rate offers a relatively direct and accurate measure of PAEE over several days but with no information about the activity intensity profile or which activity types were performed, and the method is still prohibitively expensive for large studies. Similarly, a multisensor device may be the most accurate method of activity-type classification and intensity estimation, but it may either cost too much or not be practical for use in large studies or the relevant age group. Technological advances and development of novel estimation methods constantly shift the balance between validity and feasibility. This balance will also differ by age and cultural setting. Although adolescents may be more similar to adults, they may be less willing to wear an activity monitor if it clashes with their idea of style and fashion, for example.

It is not possible to define when a child becomes an adult in terms of physical activity measurement, because this would depend on interactions between age, developmental stage, and sociocultural setting, and possibly with wide individual variability. In terms of growth and development, certain gait characteristics like contralateral movement of the upper limbs develop already around the age of 3–4 yr (107), whereas stiffness of the muscle-tendon complexes only increases from puberty onward (21, 99, 100). Along with the natural weight gain, the resulting decrease in mass-specific elasticity until puberty dictates a steady decrease in step frequency with age when walking or running at a given speed. Because mass-specific resting metabolic rate generally declines from the age of 2 yr (49) and motor skills generally improve with maturity (16), the relationship between net physiological energy expenditure and biomechanical measures of work rate varies with age and body size (14, 15, 34, 94). In addition to these biological changes, there are behavioral changes as a child grows older, some if not all of which may differ by culture and
era (43, 55, 95, 98, 117). For example, the inability of a hip-mounted accelerometer to estimate PAEE of cycling is irrelevant in populations where nobody cycles. Together, all these factors impact to a varying degree on both validity and feasibility of specific methods.

Ultimately, the choice of assessment method is a trade-off between feasibility and validity and all aspects should be thoroughly considered before use in a study. In the face of constant technological and scientific advances; however, this choice is time-dependent, and has to be balanced against resources available at the beginning of a study. If there is uncertainty with regards to feasibility of a specific data acquisition method in a specific study setting, this should be tested before commitment to the full study, and ideally with a contingency plan which can be promptly executed if feasibility proves unacceptable.

CONCLUSION

Self-report instruments and movement sensing are currently the most commonly used methods for the assessment of physical activity in epidemiological research. Habitual energy expenditure is often not measured directly, but rather it is estimated with varying degrees of uncertainty. Nonlinear modeling techniques to estimate energy expenditure and activity mode, using accelerometry with perhaps physiological parameters like HR or temperature have the greatest potential for increasing the accuracy of energy expenditure prediction of habitual physical activity in youth. Such methods in particular may benefit from increased time resolution of the recorded information that technological advances now allow. A move towards more investigator-controlled data handling strategies and away from monitor-based processing should maximize flexibility and aid comparability in the field. This is necessary to better understand the role of physical activity in the health and well-being of youth.

REFERENCES


