

12-2004

Comparison of Two Approaches to Structured Physical Activity Surveys of Adolescents

Robert G. McMurray

Kimberly B. Ring

Margarita S. Treuth

Gregory J. Welk

Russell R. Pate

University of South Carolina - Columbia, rpate@mailbox.sc.edu

See next page for additional authors

Follow this and additional works at: https://scholarcommons.sc.edu/sph_physical_activity_public_health_facpub



Part of the [Public Health Commons](#)

Publication Info

Postprint version. Published in *Medicine and Science in Sports and Exercise*, Volume 36, Issue 12, 2004, pages 2135-2143.

This Article is brought to you by the Physical Activity and Public Health at Scholar Commons. It has been accepted for inclusion in Faculty Publications by an authorized administrator of Scholar Commons. For more information, please contact digres@mailbox.sc.edu.

Author(s)

Robert G. McMurray, Kimberly B. Ring, Margarita S. Treuth, Gregory J. Welk, Russell R. Pate, Kathryn H. Schmitz, Julie L. Pickrel, Vivian Gonzalez, Maria João C. A. Almeida, Deborah Rohm Young, and James F. Sallis

Published in final edited form as:

Med Sci Sports Exerc. 2004 December ; 36(12): 2135–2143.

Comparison of Two Approaches to Structured Physical Activity Surveys for Adolescents

ROBERT. G. MCMURRAY¹, KIMBERLY B. RING², MARGARITA S. TREUTH³, GREGORY J. WELK⁴, RUSSELL R. PATE⁵, KATHRYN H. SCHMITZ⁶, JULIE L. PICKREL⁷, VIVIAN GONZALEZ⁹, M. JAOA. C. A. ALMEDIA⁵, DEBORAH ROHM YOUNG¹⁰, and JAMES F. SALLIS⁸

¹ Department of Exercise and Sport Science, University of North Carolina, Chapel Hill, NC

² Collaborative Coordinating Center, University of North Carolina, Chapel Hill, NC

³ Center for Human Nutrition, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD

⁴ Department of Health and Human Performance, Iowa State University, Ames, IA

⁵ Department of Exercise Science, University of South Carolina, Columbia, SC

⁶ Division of Epidemiology, School of Public Health, University of Minnesota, Minneapolis, MN

⁷ SDSU Foundation, San Diego State University, San Diego, CA

⁸ Department of Psychology, San Diego State University, San Diego, CA

⁹ Department of Biostatistics, Tulane University School of Public Health, New Orleans, LA

¹⁰ Department of Kinesiology, University of Maryland, College Park, MD

Abstract

Purpose—To compare the test-retest reliability, convergent validity, and overall feasibility/usability of activity-based (AB) and time-based (TB) approaches for obtaining self-reported moderate-to-vigorous physical activity (MVPA) from adolescents.

Methods—Adolescents (206 females and 114 males) completed two 3-d physical activity recalls using the AB and TB surveys, which contained identical lists of physical activities. The participants wore an MTI Actigraph[®] accelerometer for the same period.

Results—The TB instrument took about 3 min longer to complete ($P = 0.022$). Overall 2-d test-retest correlations for MVPA were similar for the two surveys ($r = 0.676$ and 0.667), but the girls had higher reliability on the AB survey than the boys (girls: $r = 0.713$; boys: $r = 0.568$). The overall 3-d correlations for MVPA surveys and Actigraph counts varied by gender (girls: AB = 0.265 vs TB = 0.314; boys: AB = 0.340 vs TB = 0.277). Correlations for vigorous physical activity and Actigraph counts were higher for the AB than for the TB ($r = 0.281$ vs 0.162). As the interval between completing the surveys and the days being recalled increased, reliability and validity were lower, especially for the AB survey.

Conclusion—For both genders, either approach is acceptable for obtaining MVPA information on a single day, but the TB approach appears to be slightly favored over the AB approach for obtaining multiple days of MVPA. A 3-d recall period appears to be too long for accurate recall of MVPA

Address for Correspondence: Robert McMurray, Dept of Exercise and Sport Science, CB#8700, Fetzer Gym, University of North Carolina, Chapel Hill NC 27599–8700; Email: exphys@email.unc.edu.

A complete list of activities used for the survey instruments, as well as the instruments and their instructions, can be obtained by contacting the primary author, Robert G. McMurray.

information from either instrument. For both genders, the surveys overestimated activity levels; thus, self-reports should be supplemented with objective data.

Keywords

ACCELEROMETERS; RELIABILITY; VALIDITY; YOUTH; EXERCISE

There are a number of self-report questionnaires used to obtain information on physical activity patterns in adolescents. Some of these focus on habitual activity (6,8,12,24), and others examine the previous 1–7 d of activity (20,21,22,25,27). Surveys that examine daily physical activity patterns have used various cues to assist respondents in completing the forms. One approach is a time-based (TB) recall strategy. The TB approach, used by the Previous Day Physical Activity Recall (PDPAR) (25,29) and 3-Day Physical Activity Recall (3DPAR) (18), requires the respondent to report physical activity performed during each of the previous 3 d, beginning with the most recent day and working backward. Each day is divided into half-hour time blocks with the dominant activity for that period chosen from a list of common activities. The person is prompted to report the type and intensity of the activity using time queues as the day progressed.

An alternative approach for surveys is an activity-based (AB) self-report measure, an example being the Self-Administered Physical Activity Checklist (SAPAC) (8,20–22). With this approach, the instrument is structured around a list of activities with minimal cues about the time of day. The respondent reports minutes per day of the specific activity performed over the previous 3 d. These two approaches have similar reliabilities and validities, and have been used with moderate success in youth. However, to our knowledge, there are no published reports comparing the reliability and validity of these physical activity assessment strategies in adolescents. Does one method provide more accurate recalls than the other, and is one method easier for youth to complete than the other? Therefore, the purpose of this study was to compare the 1) test-retest reliability, 2) convergent validity, and 3) overall feasibility/utility of the AB and TB approaches for obtaining self-reports of physical activity from adolescents.

METHODS

Study design

A test-retest design was used (Table 1). Test-retest reliability was determined by having half the subjects completed the same survey a second time, one day later. The other half of the participants completed the opposite survey the next day to assess which instrument had better recall capabilities over the previous 3 d. To obtain an objective measure of activity, the MTI Actigraph® accelerometer (MTI Health Systems, Fort Walton Beach, FL, U.S.) were placed on the participants, and data were collected over four consecutive days that overlapped with the days used for the self-report instruments. The counterbalanced research design ensured that there were no effects caused by the order in which the instruments were administered. Feasibility was assessed using those participants who completed both AB and TB instruments. These participants also completed a rating of satisfaction and preference when they completed the first survey, and their results were compared. To examine reliability, half the participants repeated the same instrument on two successive days, and the common days were compared. Validity was assessed by comparing the results from 3 d of recall with an objective measure of physical activity (MTI Actigraph® accelerometer) obtained over the same 3-d period. Table 1 provides a conceptual diagram that shows how the data were used to answer the specific research questions.

Participants

A convenience sample of 205 female and 116 male adolescents were recruited. A minimum of 30 girls and 14 boys were recruited from each of six sites in Arizona, California, Louisiana, Maryland, Minnesota, and South Carolina. To ensure a wide range of activity levels, the sample consisted of at least 10 girls per site who participated in sports teams or organized physical activity classes. The participants were randomly assigned to one of the four groups. Since this study was a substudy for a multifield site trial to reduce the decline in physical activity in adolescent girls, Trial of Activity for Adolescent Girls (TAAG), the lower recruitment of boys was by design. Before participating in the study, informed consent, using forms previously approved by each site's IRB, was obtained from the participant's parents, while assent was obtained from each participant.

Instrumentation

The TB approach was examined using the 3DPAR (18). The 3DPAR questionnaire required the person to recall the activities performed over the previous 3 d. To assist the respondent, each day was divided into 30-min segments or blocks. The respondent inserted in the block the "main activity" performed during that time period. The 3DPAR provided a list of activities with code numbers, arranged in categories (eating, sleeping, personal care, transportation, work/school, spare time, play/recreation, and exercise/workout). The respondent recorded the code number of the predominant activity that he or she performed during that 30-min block of time. Pate et al. (18) found that the 3DPAR was significantly, but moderately correlated with MTI Actigraph counts ($r = 0.28$ – 0.46). The 3DPAR was a modification of the PDPAR, which was previously validated in youth (25,26).

The AB approach was tested using a modification of the SAPAC. The original SAPAC consisted of 21 activities and additional queries on TV/video game/computer use. The instrument used simple timing cues: before, during, and after school. All activities were listed for all three time periods. For each activity done on the previous day or days, the respondent wrote the number of minutes in the blank for the appropriate time of day. The original SAPAC has been validated in younger children (22), and was moderately correlated with simultaneous accelerometer ($r = 0.33$ – 0.54) and heart-rate monitoring ($r = 0.30$). For the present study, the original SAPAC was modified to include the same list of 50 common activities as contained in the 3DPAR. The survey was further modified to assess activity over the previous 3 d. Also, two columns were added to the survey to determine where the activities took place (location), and with whom the activities were performed. This information is not reported as part of the present manuscript. The modified version of the SAPAC has not been used in the 3-d format, nor has it been validated in adolescents.

Participants completed a satisfaction survey after they finished their first physical activity recall. In a Likert scale format (1–5), the participant was asked six questions, including whether the instructions were easy to follow, whether the questionnaire was easy to complete, whether the questionnaire accurately described his/her activities, and whether the questionnaire took too long to finish. The questionnaire was not validated and presented here for descriptive purposes.

Physical activity was objectively measured in 30-s intervals using the Actigraph[®] accelerometer. The Actigraph is the most frequently used accelerometer for physical activity research, and has been shown to be a valid indicator of energy expenditure and activity levels in youth and adolescents (10,16,26). Correlations between Actigraph counts and measured energy expenditure were high for walking and running on treadmills: $r = 0.87$ (26). However, the correlations between Actigraph counts and heart-rate monitoring or direct observation were somewhat lower during uncontrolled, free-living conditions: $r = 0.45$ – 0.81 (5,10,14). The

accelerometer has limitations, and might not be a precise method for measuring energy expenditure in youth and adolescents (9,26). However, Welk et al. (27) have shown that accelerometry may be the most appropriate criterion to examine the validity of self-report instruments.

Procedures

Data were obtained over a 4-d period (Saturday through Tuesday) and coordinated to allow the self-report instruments and Actigraphs to assess activity on matched days (Table 1). On the Friday before the start of measurement, all participants had their heights and body masses measured using standardized procedures, and were fitted with a previously initialized Actigraph. They wore the Actigraph continuously until Wednesday morning, except during sleep, while bathing, or during activities in which they were not allowed to use them (e.g., football practice or soccer competition). On Wednesday, when the Actigraphs were removed, the participants were briefly interviewed to determine times they removed the monitors. The two physical activity surveys were administered in groups on Tuesday, and readministered on Wednesday. The time required to complete each survey was recorded. On Tuesday, groups 1 and 2 completed the TB instrument, while groups 3 and 4 completed the AB instrument. Upon completing the activity recall on the first day, the participants completed the satisfaction survey. On Wednesday, groups 1 and 3 completed the same physical activity survey as on Tuesday, while groups 2 and 4 completed the other PA survey. This design allowed the assessment of both instruments over the same days, while controlling any order effect.

Data processing

Because the two instruments created different outcome measures (blocks vs minutes), physical activity levels were calculated using specific published protocols. For the AB instrument, physical activity levels were calculated using 1-min increments of moderate-to-vigorous physical activity (MVPA ≥ 3.0 METs) or vigorous physical activity (VPA ≥ 6.0 METs), as previously described by Sallis et al. (22). The min for all activities ≥ 3.0 METs were summed for each day to obtain min of MVPA per day, while the min for all activities ≥ 6 METs were totaled daily for vigorous physical activity (VPA). If more than 12 h·d⁻¹ of MVPA were reported, that day was not included in the analyses ($N = 10$ out of 420 d). The number of activities each day that met the MVPA criteria was also reported. For the TB instrument, physical activity levels were calculated using the number of blocks (30-min segments) of MVPA or VPA. MVPA was defined by Pate et al. (18) as having MET values of ≥ 3.0 , while VPA was defined as ≥ 6.0 METs. The specific activities and the total number of blocks of activity that met the criteria were reported for each day. If more than 24 blocks of activity were reported for 1 d, that day was excluded from the analyses ($N = 3$ out of 456 d).

The Actigraph accelerometer data were first examined for compliance. Data from individuals wearing the monitor for <11.2 h on weekdays and <7.2 h on weekend days were excluded (23). The data were then analyzed separately for MVPA and VPA. The data were examined using two methods. The first method applied empirically derived thresholds for MVPA and VPA, based on the research of Treuth et al. (23), using a sample of similarly aged youth. The accelerometer counts (ct) representing the two thresholds of intensity were ≥ 1500 ct·30 s⁻¹ for MVPA and ≥ 2600 ct·30 s⁻¹ for VPA. Although these thresholds represented MVPA and VPA in middle school girls, these MET values (4.6 for MVPA and 6.5 for VPA) were higher than those previously used to define MVPA and VPA (7,9,15), and also higher than the thresholds used by the two survey methods (3 and 6 METs, respectively). Therefore, in the second method of analysis we redefined the Actigraph thresholds based on the standard 3- and 6-MET thresholds, which were 580 ct·30 s⁻¹ for MVPA and 2300 ct·30 s⁻¹ for VPA (7,9,18,25).

Data analyses

Feasibility and overall utility were assessed two ways. First, the time to complete each questionnaire (TB vs AB) was compared using a *t*-test. Second, the results on the Satisfaction Surveys were compared using chi-square. Test-retest reliability (stability) was determined over the two matched days (Sunday and Monday) for both the AB (group 2) and TB (group 1) instruments using a Pearson product-moment correlation. Correlations were computed separately for Sunday and Monday, as well as the 2 d combined. To test the reliability of reporting specific activities, the 10 most frequently reported activities for each instrument were determined, and the percent agreement in reporting the same activity on the same day was computed.

The validity of each instrument to measure MVPA and VPA was determined using temporally matched Actigraph accelerometer counts. Data from groups 3 and 4 provided the data for the AB instrument, while groups 1 and 2 provided the data on the TB instrument. For the AB instrument, the total number of min of reported MVPA and VPA were directly compared to the number of min for similar intensities of activity obtained from accelerometer counts. For the TB instrument, the primary outcome measure was the number of blocks (30 min) that were MVPA or VPA. The number of blocks of reported MVPA and VPA were compared to the number of min obtained from the Actigraph using correlation analyses. Correlations were computed for each day. Correlations for the 3-d overall results were computed by taking the average of the Fisher's Z transformation of each day's data. These computations were completed using the two Actigraph scoring methods described above. Because the two surveys used different outcome metrics, direct comparisons of the two instruments were not completed.

The ability of the two instruments to produce recalls of the same activity on the same day was assessed by computing the percent agreement for those participants who completed both instruments. Agreement between the two instruments was assessed for the 10 most frequently reported activities using data from groups 2 and 4. In addition, the percent agreement in reporting the same 10 activities on the same day (test-retest) was computed for each instrument separately, using information from groups 1 and 3.

RESULTS

The mean and standard deviation of the physical characteristics and reported physical activity levels of the participants are presented in Table 2. Two hundred and six girls and 114 boys from six sites participated. Representation from six sites was similar, varying by no more than nine participants between the sites. The racial/ethnic distribution was 60.9% Caucasian, 15.3% African American, 7.2% Hispanic, and 16.6% multiethnic. There were 154 sixth graders, 13 seventh graders, and 152 eighth graders. The participants represented a narrow age range, but a wide range of size (body mass index) and activity levels. In general, the participants reported a mean of $\sim 146 \text{ min}\cdot\text{d}^{-1}$ of MVPA using the AB survey, and approximately 5.6 blocks per day of MVPA on the TB survey. In contrast, the average number of minutes of MVPA determined from the Actigraph was much lower, approximately $28 \text{ min}\cdot\text{d}^{-1}$. The standard deviations for all three methods indicate a wide variation in activity levels across participants.

Feasibility/utility

The AB survey took significantly less time to complete than the TB survey (28 ± 8 vs 31 ± 10 min, respectively; $P = 0.022$); however, the actual difference in time was small. Analyses of the satisfaction survey indicated that 31% of the participants reported that both the TB and the AB took too long to complete. The instructions for the instruments were equally easy to follow. Eighty percent of the participants indicated that both instruments were easy to complete, while

65% agreed that the instruments accurately described the activities. There was no difference as to which instrument the participants preferred ($P > 0.15$).

Reliability

The test-retest correlations were computed for the 2 d combined (overall) and for each of the 2 d in common (Table 3). Because of poor compliance, data from seven participants completing the AB instrument and three participants completing the TB instrument were not included in the analysis. Overall (Monday + Sunday) test-retest correlations for MVPA were similar for the AB and TB instrument for the girls ($r = 0.713$ vs $r = 0.707$, respectively); however, for the boys, the correlation for the TB instrument was somewhat higher than for the AB instrument ($r = 0.673$ vs $r = 0.568$, respectively). For both genders, the ability to recall MVPA for the previous day (Monday) was better than the ability to recall 2 d earlier (Sunday). In addition, the data from Table 3 also suggest that the girls were better able to recall activity 2 d earlier (Sunday) using the TB instrument than the AB instrument, while the boys had better recall using the AB than the TB instrument.

The test-retest correlations for VPA are also presented in Table 3. The overall (Monday + Sunday) correlations were somewhat higher for the TB survey than for the AB survey for both genders (overall $r = 0.832$ vs $r = 0.627$, respectively). The ability to recall VPA for the previous day's activity (Monday) was slightly higher for the AB method, compared with the TB method; however, the correlations reflecting the recall of activities 2 d prior (Sunday) were generally much lower for the AB method than for the TB method.

To assess the instruments' ability to assess the full range of types of activities, we counted the number of specific MVPA activities the youth reported. On the initial survey, using the AB instrument, the girls reported 3.2 ± 2.5 different MVPA activities per day that were completed in $146 \text{ min} \cdot \text{d}^{-1}$, while the boys reported 2.9 ± 2.3 MVPA activities per day, completed in $147 \text{ min} \cdot \text{d}^{-1}$. Using the TB instrument, the girls named an average of 1.9 ± 1.5 activities per day, completed in ~ 5.5 blocks per day (Table 2). The boys reported an average of 1.7 ± 1.2 activities per day, completed in ~ 5.9 blocks per day. The number of MVPA activities reported during the retest was 16–38% less for the girls and 31–43% less for the boys, regardless of the instrument.

Validity

Overall, 77.1% of the participants met the minimum criteria for adherence to Actigraph monitoring. Adherence dropped slightly as the time period progressed, with 80.5% adherence on Saturday, and 73.1% adherence on Monday. Table 4 presents the correlations between the survey instruments and Actigraph counts using the empirically derived Actigraph thresholds (23). The overall correlations for MVPA (3 d combined) were weak for both the AB survey and the TB survey ($r = 0.239$ vs $r = 0.279$, respectively). Gender variability was noted in the day-to-day correlations for MVPA. For the girls, the overall correlation for MVPA between the Actigraph and the AB survey were weaker than for the TB survey; however, the reverse was true for the boys. The overall correlations for VPA were better with the AB method than with the TB method, with the correlations generally becoming lower as days of separation between completing the survey and the day of recall increased (e.g., previous day vs 2 or 3 d previous). The girls had higher correlations for VPA using the AB method than the TB method.

Using the standard thresholds for MVPA and VPA for the Actigraph (3 and 6 METs) and computing the correlations with the survey methods generally resulted in slightly lower overall correlations (Table 5). The use of the standard thresholds appeared to reduce the correlations for MVPA or VPA as the days of separation between completing the survey and the day of recall increased.

Comparison of the two instruments for reporting specific activities

Table 6 presents the top 10 activities reported on both instruments by gender. For the girls, considerable differences existed in the top five activities reported when comparing the two surveys. Seven of the top 10 activities were reported on both surveys, although the order varied slightly. However, calisthenics, playing catch, and yard work were reported only on the AB survey, while PE class, playground games, and “Other” were reported only for TB. Averaged across activities, girls reported the same activities on test and retest only $42 \pm 15\%$ of the time on the AB survey, and $47 \pm 18\%$ of the time for the TB survey.

Similar disparities in the top 10 activities existed for the boys. The AB survey included walking for exercise, playing catch, running/jogging, and yard work, whereas the TB survey included bicycling, PE class, swimming, and wrestling. The ability of the boys to recall the activities the second time was lower for the AB survey than for the TB survey (34 ± 20 vs $51 \pm 29\%$, respectively). When the genders were combined, consistent recall of the same activities occurred $38 \pm 18\%$ for the AB survey and $49 \pm 23\%$ of the time for the TB survey.

DISCUSSION

Both the AB and TB physical activity recall methods were shown to have similar utility, required about the same amount of time to complete, were similarly valid, and were equally acceptable to adolescents. Thus, both methods appear to be useful measures of physical activity among boys and girls in early adolescence. The TB instrument had slightly higher reliability and validity than the AB instrument for assessing 3 d of MVPA, whereas the AB approach had higher validity, but not reliability, for obtaining 3 d of VPA information. The TB approach had slightly better reliability for obtaining multiple days of self-reported physical activity information than the AB approach. Thus, the TB approach appears to be marginally favored over the AB approach.

The correlations between the Actigraph counts and MVPA and VPA for both instruments were low and inconsistent (Table 4). The low correlations between the instruments and the Actigraph counts were somewhat anticipated. Studies by Sallis et al. (21,22) using 4th and 5th grade children reported correlations of 0.15–0.33 between an AB survey and Caltrac[®] accelerometry. The original validation of the 3DPAR (18) resulted in similar correlations between 3-d accelerometry and survey ($r = 0.28$ – 0.46). Trost et al. (25) also reported similar correlations (0.19–0.23) between MVPA using the 1-d PDPAR (a TB instrument) and CSA accelerometers. Conversely, Pate et al. (18), using 7-d accelerometry counts and 3DPAR, found correlations of 0.38–0.51 in 13–15-yr-old girls. Other researchers have found higher correlations between physical activity surveys and heart rates (10,11,29), accelerometry (13,29), or oxygen uptake (1,26). These studies have used isolated activities (1,26), a single day of activity (29), or more generalized questionnaires (11,13), reducing their comparability with the present study. Thus, we have confidence in our current results, which indicate limited validity of adolescents' physical activity self-reports.

Table 2 suggests that the adolescents reported on the surveys an average over 2.5 h of MVPA, but the Actigraph counts indicated less than $40 \text{ min} \cdot \text{d}^{-1}$. These discrepancies are of some concern, as the survey instruments indicate that the youth are meeting recommended guidelines of $60 \text{ min} \cdot \text{d}^{-1}$ (4), whereas the Actigraph results do not confirm this. The Actigraph does not accurately record activities such as cycling or weightlifting, and cannot be worn during swimming (27), so some of the activities reported on the surveys may not have been detected by the accelerometers. Table 6 suggests that, other than cycling, the most common MVPA activities reported were detectable by the Actigraph. Thus, both survey instruments appear to overestimate MVPA. The most likely explanation for the differences is the nature and type of activity performed by these adolescents. Adolescents tend to perform activities more

intermittently or sporadically than adults (29); they rarely engage in continuous activities for 30 min at a time, unless they are participating in an organized sport (19). In addition, most sports activities are intermittent in nature, and may involve significant breaks or rest periods. Thus, adolescents may report a sports activity for 30 min, yet the time spent in actual movement is much less. The use of time blocks might help youth remember the events of previous days, but, as reported here, activity blocks using the TB approach should not be misconstrued as representing an entire 30 min of activity.

The total number of 30-min blocks of MVPA from the TB instrument varied from 0 to 18 blocks of MVPA per day (Table 2). If the number of blocks were interpreted as 30 min of MVPA, they represent a range of response of 0–540 min·d⁻¹. For reasons previously mentioned, these results should not be interpreted literally as “minutes” of activity. Since the AB approach also showed variability, with a range of 0–573 min·d⁻¹ of MVPA, we cannot conclude which method gives a more accurate estimate of actual amounts of physical activity. The AB approach allowed for the recall of activities that lasted a shorter period of time than the 30-min segments used by the TB approach. For example, consider a person who bikes to the basketball court for 5 min, plays basketball for 20 min, and bikes home for 5 min. If the person fills out an AB survey, he or she may submit biking for a total of 10 min, and basketball for 20 min. However, if the person completes a TB survey, which asks for the major activity that took place within that 30-min block, he or she may only submit basketball. Thus, a potential limitation to the TB approach is that short bouts of activity, which are characteristic of youth activity patterns, may not be as apparent as with the AB approach. This limitation is also supported by the smaller number of activities reported in the TB survey compared with the AB survey.

A consistent finding of the present study was that as the interval between the physical activity and day of recall increased, the recall became less reliable and valid. Our data (Tables 3 and 4) suggest that adolescents may be able to recall the previous 2 d of activity, but beyond 2 d, limitations in memory reduce their ability to recall the information. This pattern has been reported previously (20). Based on our results, we recommend that physical activity surveys for early adolescents limit their recall span to the previous 2 d. It may be preferable to estimate habitual physical activity by collecting 1- or 2-d recalls on multiple occasions.

One of the suggested benefits of the self-report measures is the ability to collect data on participation in specific activities. However, the present results (Table 6) document that adolescents have a limited ability to recall specific activities for specific days, even over short intervals. Attempting to recall the same activity from 2 d earlier resulted in less than 50% recall accuracy, regardless of recall approach. Better results might have been obtained if we had asked the respondents if they had performed the activity in the last week, without specifying the day. However, the benefits of this approach are speculation. This finding is somewhat troubling, as it suggests that these young people had some difficulty recalling the details of their physical activities.

Studies have shown that boys report more MVPA than girls (2,3). In contrast, boys in our study reported similar MVPA blocks, or min, as girls, and the standard deviations and ranges of scores were similar for both genders. In our study, we purposefully oversampled girls who were regularly participating in physical activity (1/3 of sample), and obtained a convenient sample of boys without regard to physical activity levels. Thus, our results may have been an artifact of the sampling. However, the intended purpose of the study was not to focus on gender differences in MVPA, but on the ability to recall MVPA using different information recall formats. It is interesting to note that even though the surveys suggested similar amounts of self-reported MVPA, the average number of min from the Actigraph for boys was 17 min·d⁻¹ higher than for girls.

We had anticipated that there would be considerable similarity between the specific activities reported by the boys and girls. Five of the top 10 activities were similar for girls and boys on the TB survey, while 8 of the top 10 were similar between the sexes using the AB instrument. The main differences were that the boys reported football and wrestling, whereas the girls reported dance and playground games. These findings are in agreement with previous research on 12–14-yr-old boys and girls (2). Since both instruments had the same list of activities and recalled the same days, we cannot explain the differences in specific activities reported between our two survey methods (5/10 vs 8/10). One would expect a similar agreement between the two surveys in the top 10 activities.

The present study highlights the limitations of self-report with adolescents. Validity of reports of activity levels on specific days was generally low, though overall results tended to be somewhat better for girls. Self-reports continue to be used primarily because of their low cost and convenience. Of particular concern were the generally low correlations with the physical activity measured by motion sensors. The large difference between self-reports and accelerometry-measured MVPA min is another indication that adolescent reports of physical activity duration should not be interpreted literally (17). All of these criticisms combined suggest that surveys probably should not be used to estimate energy expenditure in youth, and that more objective measures, such as doubly labeled water or indirect calorimetry, should be used whenever possible (28). Given the limitations of the self-reports, they could be useful for gathering information on the specific activities that youth are engaged in, and determining how those activities change with increasing age or in response to an intervention. Self-reports could also be useful in obtaining general activity levels so that youth could be potentially categorized into sedentary or active.

It is important to note that accelerometers measure ambulatory movement (7,9,23,26). Activities that do not involve ambulation, such as cycling, rowing, or weight lifting, may be recorded as little or no activity. Thus, survey methodology might be needed to assess nonambulatory activities. Although more research is needed to define the best combination of measures for various populations and study purposes, it appears that multiple methods of measurement may be required to maximize the accuracy of physical activity assessments.

In this study Actigraph thresholds, or cut points for MVPA and VPA, surfaced as another point of concern. Based on data from adults, Freedson et al. (7) suggested that 1952 and 5725 $\text{ct}\cdot\text{min}^{-1}$ (976 and 2862 $\text{ct}\cdot 30\text{ s}^{-1}$) were the thresholds for MVPA and VPA, respectively. Handelman et al. (9) noted that 2191 and 6893 $\text{ct}\cdot\text{min}^{-1}$ (1096 and 3447 $\text{ct}\cdot 30\text{ s}^{-1}$) were the thresholds for MVPA and VPA, respectively. However, these data were determined on adults. In a study of 6–16-yr-old youth, Puyau et al. (16) found considerable differences between the Freedson's adult cut points and proposed cut points for MVPA and VPA in youth of 3200 and 8200 $\text{ct}\cdot\text{min}^{-1}$ (1600 and 4100 $\text{ct}\cdot 30\text{ s}^{-1}$), respectively. Given such a wide variation, we chose to develop our own thresholds, based on counts from our study of similarly aged girls, and using NIH-derived reference activities (16) for defining MVPA and VPA (23). This method seemed most appropriate, since it was developed specifically on our target population. Since this method resulted in higher MET values than were used for the survey instruments being assessed, we also evaluated the accelerometry data based on the standard thresholds of 3 and 6 METs. In general, the correlations between surveys and Actigraph counts were higher for our empirical method, suggesting that our thresholds may have been more appropriate for adolescents than the standard thresholds of 3 and 6 METs.

Specific thresholds for Actigraph counts presented one problem, but the definition for MET values in adolescents provided an additional conundrum. Different MET values were used to define the intensity levels on the questionnaires compared with on the Actigraph. Presently, definitions of exercise intensity for youth and adolescents are not well established, and there

are few studies to guide this decision. The questionnaires adopted the adult definitions of moderate (3–5.9 METs) and vigorous (≥ 6 METs) because they were the best available at the time the questionnaires were developed. Recent studies conducted as part of TAAG (23) indicated that a higher MET value was needed to define the activity intensities, at least when using the Actigraph. Presently, there is no way to reconcile these discrepancies. Nonetheless, the validity correlations reported were similar to those in previous studies, suggesting that the differences in definitions did not alter the findings dramatically. This suggests that more evidence is needed to develop standard definitions of moderate and vigorous physical activity for youth of various ages.

Each survey method had strengths. Overall, the TB approach seems to be marginally better than the AB approach for recalling multiple days of physical activity. This was evident when comparing the reliability correlations for recalling 2-d previous MVPA or VPA using the TB approach, compared to the AB approach (Table 3). Conversely, the TB approach was not as effective as AB approach in assessing all the activities performed by the adolescents. Using the AB approach, the adolescents named more than twice the number of different MVPA activities (not minutes) reported by the TB approach. Yet, both surveys had the same activity list. Moreover, the AB approach used min as the convenient unit of measure and, thus, should be more sensitive to change for longitudinal studies. However, the AB approach overreports actual minutes of activity. The format of dividing the day into 30-min segments appears to help the adolescent with recall tasks 2 d previous to completing the survey, but greatly overestimates total physical activity. The 30-min blocks could be used to determine the chronological activity patterns, which might be useful for some studies. However, the 30-min blocks might make this format less sensitive to change for prepost study designs, unless multiple activities could be reported within each time block.

From the results of this study, the following conclusions can be ascertained. First, early adolescents appear unable to provide valid physical activity recalls over a 3-d period. Compared to estimates of MVPA time from accelerometer monitoring, it appeared that both self-report instruments produced grossly inflated estimates. However, accelerometry failed to record *all* activities, since the adolescents were required to remove it during some activities (e.g., competitive sports and swimming), and the intensities of other activities were not recorded or underestimated (e.g., cycling and weightlifting). Furthermore, middle-school-aged youth showed a relatively poor ability to recall specific physical activities they had performed as recently as 24 h previously. Thus, information obtained from any self-report might be a crude estimate of activity levels and participation in specific activities. Second, for a single day's recall of physical activity, either approach—TB or AB—may be used. Third, to obtain information on multiple days of MVPA, the TB appears to be a slightly better approach. The half-hour blocks of time might make it easier for the adolescent to recall the day's events. Fourth, the AB approach appears to be slightly favored for obtaining information on VPA. Although both self-report measures could be judged to have similarly weak reliability and validity, and although in general the data suggest that the TB and AB recall strategies were similar in their performance, on balance it appears that the TB might perform slightly better for middle-school-aged youth. A final implication is that, whenever possible, objective measures such as accelerometers should be used to measure physical activity levels in adolescents (28).

Acknowledgements

This study was supported by grants from National Heart, Lung, Blood Institute of NIH, Grant # U01HL66845, HL66852, HL66853, HL66855, HL66856, HL66857, & HL66858, and from a Career Development Award from the Johns Hopkins Center for Adolescent Health Promotion and Disease Prevention.

We would like to thank the participants, the Project Coordinators, and the members of the TAAG Steering Committee: Larry Webber, Ph.D., Tulane University; John Elder, Ph.D., San Diego State University; Timothy Lohman, Ph.D.,

University of Arizona; Leslie Lytle, Ph.D., University of Minnesota; Deborah Rohm Young, Ph.D., University of Maryland College Park; Russell Pate, Ph.D., University of South Carolina; June Stevens, Ph.D., The University of North Carolina at Chapel Hill; and Charlotte A. Pratt, Ph.D., National Heart, Lung, and Blood Institute.

References

1. Bassett DR, Ainsworth BE, Swartz AM, Strath SJ, O'Brien WL, King GA. Validity of four motion sensors in measuring moderate intensity physical activity. *Med Sci Sports Exerc* 2000;32:S471–S480. [PubMed: 10993417]
2. Bradley CB, McMurray RG, Harrell JS, Deng S. Changes in common activities of third through tenth graders. *Med Sci Sports Exerc* 2000;32:2071–2078. [PubMed: 11128854]
3. Caspersen CJ, Pereira MA, Curran KM. Changes in physical activity patterns in the United States by sex and cross-sectional age. *Med Sci Sports Exerc* 2000;32:1601–1609. [PubMed: 10994912]
4. Cavill N, Biddle S, Sallis JF. Health enhancing physical activity for young people: Statement of the United Kingdom expert consensus conference. *Pediatr Exerc Sci* 2001;13:12–25.
5. Coe D, Pivarnik JM. Validation of the CSA accelerometer in adolescent boys during basketball practice. *Pediatr Exerc Sci* 2001;3:373–379.
6. Crocker PRE, Bailey DA, Faulkner RA, Kowalski KC, McGrath R. Measuring general levels of physical activity: preliminary evidence for the Physical Activity Questionnaire for Older Children. *Med Sci Sports Exerc* 1997;29:1344–1349. [PubMed: 9346166]
7. Freedson PS, Melanson E, Sirard J. Calibration of the Computer Science and Applications, Inc. accelerometer. *Med Sci Sports Exerc* 1998;30:777–781. [PubMed: 9588623]
8. Gilmer MJ, Speck BJ, Bradley C, Harrell JS, Belyea M. The Youth Health Survey: Reliability and validity for assessing cardiovascular health habits in adolescents. *J School Health* 1996;66:106–111. [PubMed: 8857159]
9. Handelman D, Miller K, Baggett C, Debold E, Freedson P. Validity of accelerometry for the assessment of moderate intensity physical activity in the field. *Med Sci Sports Exerc* 2000;32:S442–S449. [PubMed: 10993413]
10. Janz KF. Validation of the CSA accelerometer for assessing children's physical activity. *Med Sci Sports Exerc* 1994;26:369–375. [PubMed: 8183103]
11. Janz KF, Witt J, Mahoney LT. The stability of children's physical activity as measured by accelerometry and self-report. *Med Sci Sports Exerc* 1995;28:1326–1332. [PubMed: 8531633]
12. Koo MM, Rohan TE. Comparison of four habitual physical activity questionnaires in girls aged 7–15 yr. *Med Sci Sports Exerc* 1999;31:421–427. [PubMed: 10188747]
13. Kowalski KC, Crocker PRE, Faulkner RA. Validation of the Physical Activity Questionnaire for Older Children. *Pediatr Exerc Sci* 1997;9:174–186.
14. Leenders NJM, Sherman WM, Nagaraja HN, Kien CL. Evaluation of methods to assess physical activity in free-living conditions. *Med Sci Sports Exerc* 2001;33:1233–40. [PubMed: 11445774]
15. National Institute of Health. Physical Activity and Cardiovascular Health: NIH Consensus Statement 1995;13:1–33.
16. Puyau MR, Adolph AL, Vohra FA, Butte NF. Validation and calibration of physical activity monitors in children. *Obes Res* 2002;10:150–157. [PubMed: 11886937]
17. Pate RR, Freedson PS, Sallis JF, et al. Compliance with physical activity guidelines: Prevalence in a population of children and youth. *Ann Epidemiol* 2002;12:303–308. [PubMed: 12062916]
18. Pate RR, Ross R, Dowda M, Trost SG, Sirard JR. Validation of a 3-day physical activity recall instrument in female youth. *Pediatr Exerc Sci* 2003;15:257–265.
19. Pate, RR.; Trost, SG.; Williams, C. Critique of existing guidelines for physical activity in young people. In: Biddle, S.; Sallis, J.; Cavill, N., editors. *Young and Active? Young People and Health-Enhancing Physical Activity—Evidence and Implications*. London: Health Education Authority; 1998. p. 162-176.
20. Sallis JF, Buono MJ, Roby JJ, Micale FG, Nelson JA. Seven-day recall and other physical activity self-reports in children and adolescents. *Med Sci Sports Exerc* 1993;25:99–108. [PubMed: 8423762]
21. Sallis JF, Condon SA, Goggin KJ, Roby JJ, Kolody B, Alcaraz JE. The development of self-administered physical activity survey for 4th grade students. *Res Quart Exerc Sport* 1993;64:24–31.

22. Sallis JF, Strikemiller PK, Harsha DW, et al. Validation of interviewer- and self-administered physical activity checklist for fifth grade students. *Med Sci Sports Exerc* 1996;28:840–851. [PubMed: 8832538]
23. Treuth MS, Schmitz K, Catellier DJ, et al. Defining accelerometer thresholds for physical activity intensities in adolescent girls. *Med Sci Sports Exerc* 2004;36:1259–1266. [PubMed: 15235335]
24. Treuth MS, Sherwood NE, Butte NF, et al. Validity and reliability of activity measures in African American Girls from GEMS. *Med Sci Sports Exerc* 2003;35:532–539. [PubMed: 12618587]
25. Trost SG, Ward DS, McGraw B, Pate RR. Validity of the Previous Day Physical Activity Recall (PDPAR) in fifth-grade children. *Pediatr Exerc Sci* 1999;11:341–348.
26. Trost SG, Ward DS, Moorehead SM, Watson PD, Riner W, Burke JR. Validation of the computer science and applications (CSA) activity monitor in children. *Med Sci Sports Exerc* 1998;30:629–633. [PubMed: 9565947]
27. Welk, GJ. Use of Accelerometry-Based Activity Monitors for the Assessment of Physical Activity. In: Welk, GJ., editor. *Physical Activity Assessments in Health Related Research*. Human Kinetics Publishers; Champaign: 2002. p. 125-142.
28. Welk GJ, Corbin CB, Dale D. Measurement issues for the assessment of physical activity in children. *Res Quart Exerc Sport* 2000;71(2 Suppl):S59–S73.
29. Weston AT, Petosa R, Pate RR. Validation of an instrument for measurement of physical activity in youth. *Med Sci Sports Exerc* 1997;29:138–143. [PubMed: 9000167]

TABLE 1
Conceptual diagram of research design used to assess feasibility, reliability, and validity.

	Fri	Sat	Sun	Mon	Tues	Wed	Use of the Data to Address Research Questions
	Don Actigraph	Wear Actigraph			Recall Sat, Sun, Mon Time-based recall	Remove Actigraph Recall Sun, Mon, Tues	
Group 1 (N = 65)		*	*	*			Validity information on Sat, Sun, & Mon for time-based recall Test-retest information for time-based recall of Sun & Mon
Group 2 (N = 92)			*	*	Time-based recall	Time-based recall	Validity information on Sat, Sun, & Mon for time-based recall Feasibility information using most proximal day of recall for both instruments
Group 3 (N = 84)			*	*	Activity-based recall	Activity-based recall	Validity information on Sat, Sun, & Mon for activity-based recall Test-retest information for activity-based recall of Sun & Mon
Group 4 (N = 70)			*	*	Activity-based recall	Time-based recall	Validity information on Sat, Sun, & Mon for activity-based recall Feasibility information using most proximal day of recall for both instruments

* Indicates days used for test-retest comparisons.

TABLE 2
Physical characteristics and physical activity survey results of the participants presented by gender and genders combined.

	<i>N</i>	Girls (Mean ± SD)	<i>N</i>	Boys (Mean ± SD)	<i>N</i>	Combined (Mean ± SD)
Age (yr)	206	12.5 ± 1.2	114	12.6 ± 1.1	320	12.5 ± 1.1
Stature (cm)	206	157 ± 8	114	158 ± 11	320	158 ± 9
Body mass (kg)	206	54.7 ± 14.0	114	56.1 ± 21.8	320	55.2 ± 17.2
BMI (BMMDNM _{kg} ⁻² , SMDNM _m ⁻²)	206	22.0 ± 4.8	114	22.2 ± 7.8	320	22.1 ± 6.0
TB MVPA (blocks) ^a	101	5.5 ± 2.7	51	5.9 ± 3.2	152	5.6 ± 2.9
AB MVPA (min) ^b	97	146 ± 120	43	147 ± 119	140	146 ± 119
Actigraph MVPA (min)	194	20.3 ± 13.3	109	37.7 ± 22.3	303	26.6 ± 19.0

^aTB, Time-based number of 30-min segments (blocks) of MVPA per day. Data obtained from groups 1 and 2 first recall only; group 4, second recall.

^bAB, Activity-based mean (±SD) number of minutes of MVPA per day. Data obtained from groups 2 and 3 first recall only; group 4, second recall.

TABLE 3

Test-retest correlations of moderate-to-vigorous and vigorous physical activity for the activity-based and time-based for both days overall and each of the two days for the boys and girls.

Questionnaire	Overall	Monday	Sunday
Moderate-Vigorous PA			
Activity-Based			
Girls (<i>N</i> = 47)	0.713 [*]	0.702 [*]	0.455 [†]
Boys (<i>N</i> = 19)	0.568 [‡]	0.463	0.701 [#]
Both (<i>N</i> = 66)	0.667 [*]	0.657 [*]	0.490 [*]
Time-Based			
Girls (<i>N</i> = 48)	0.707 [*]	0.664 [*]	0.709 [*]
Boys (<i>N</i> = 23)	0.673 [#]	0.777 [*]	0.535 [†]
Both (<i>N</i> = 71)	0.676 [*]	0.710 [*]	0.630 [*]
Vigorous PA			
Activity-Based			
Girls (<i>N</i> = 47)	0.642 [*]	0.808 [*]	0.215
Boys (<i>N</i> = 19)	0.811 [*]	0.726 [#]	0.798 [*]
Both (<i>N</i> = 66)	0.627 [*]	0.727 [*]	0.369 [†]
Time-Based			
Girls (<i>N</i> = 48)	0.771 [*]	0.640 [*]	0.778 [*]
Boys (<i>N</i> = 23)	0.837 [*]	0.702 [#]	0.810 [*]
Both (<i>N</i> = 71)	0.832 [*]	0.688 [*]	0.823 [*]

^{*} $P \leq 0.0001$;

[#] $P \leq 0.001$;

[†] $P < 0.01$.

TABLE 4

Overall and day-by-day correlations between the moderate-to-vigorous physical activity and vigorous physical activity reported on each of the two instruments and Actigraph counts, defining MVPA ≥ 1500 ct·30 s⁻¹; VPA ≥ 2600 ct·30 s⁻¹.

Questionnaire	Overall	Monday	Sunday	Saturday
Moderate-Vigorous PA				
Activity-Based				
Girls (N = 75)	0.265†	0.371*	0.093	0.321 [#]
Boys (N = 32)	0.340†	0.268	0.512 [#]	0.218
Both (N = 107)	0.239 [#]	0.307 [#]	0.177	0.229†
Time-Based				
Girls (N = 86)	0.314 [#]	0.529*	0.157	0.224†
Boys (N = 44)	0.277	0.419 [#]	0.433 [#]	-0.056
Both (N = 130)	0.279 [#]	0.487*	0.264 [#]	0.060
Vigorous PA				
Activity-Based				
Girls (N = 75)	0.312 [#]	0.503*	0.030	0.366 [#]
Boys (N = 32)	0.289	0.386†	0.309	0.163
Both (N = 107)	0.281 [#]	0.441*	0.148	0.233†
Time-Based				
Girls (N = 86)	0.092	0.195	-0.005	0.084
Boys (N = 32)	0.265	0.525*	0.316	-0.094
Both (N = 130)	0.162	0.310 [#]	0.147	0.023

*
P ≤ 0.001

[#]
P ≤ 0.01

†
P ≤ 0.05

TABLE 5

Overall and day-by-day correlations between the moderate-to-vigorous physical activity and vigorous physical activity reported on each instrument and Actigraph results using the limits of 3 METs = 579 ct·30 s⁻¹ and 6 METs = 2330 ct·30 s⁻¹.

Questionnaire	Overall	Monday	Sunday	Saturday
Moderate-Vigorous PA				
Activity-Based				
Girls	0.208	0.253 [†]	0.162	0.208
Boys	0.307	0.351	0.420 [†]	0.138
Both	0.198	0.240 [†]	0.192	0.162
Time-Based				
Girls	0.280	0.467 [*]	0.100	0.239 [†]
Boys	0.391 [†]	0.504 [#]	0.570 [*]	0.039
Both	0.313 [*]	0.497 [*]	0.282 [#]	0.136
Vigorous PA				
Activity-Based				
Girls	0.316 [*]	0.493 [*]	0.064	0.359 [#]
Boys	0.281	0.364	0.319	0.154
Both	0.286 [#]	0.422 [*]	0.183	0.242 [#]
Time-Based				
Girls	0.104	0.209	0.012	0.090
Boys	0.262	0.478 [#]	0.327 [†]	-0.055
Both	0.187	0.323 [*]	0.185 [†]	0.044

* $P \leq 0.001$

$P \leq 0.01$

[†] $P \leq 0.05$

TABLE 6

The top 10 activities reported for both instruments by gender and the percentage that reported participation in the same activity for both the test and retest.

Gender	Activity	Time-Based		Activity-Based	
		%Reported	Test-Retest%	%Reported	Test-Retest %
Girls	Walking for travel	35	50	47	55
	PE class	24	75	40	35
	Basketball	16	69	38	42
	Running/Jogging	16	31	37	28
	Other	16	44	36	47
	Walking for exercise	12	42	32	75
	Bicycling	11	64	31	29
	Playing with children	11	45	21	48
	Playground games	11	18	19	26
	Dance	11	36	17	35
	Mean \pm SD		47 \pm 18		42 \pm 15
	Basketball	42	71	50	70
Boys	Walking for travel	26	62	35	7
	PE Class	24	75	33	23
	Bicycling	22	73	33	15
	Football	14	43	30	58
	Playing with children	12	0	28	45
	Swimming	10	60	25	20
	Softball/Baseball	8	75	23	33
	Wrestling	8	50	23	22
	Calisthenics/Exercises	6	0	23	44
	Mean \pm SD		51 \pm 29		34 \pm 20
	Basketball	42	71	50	70
	Running/Jogging	26	62	35	7
	Football	24	75	33	23