"Cross-Validation of Two Accelerometers for Assessment of Physical Activity and Sedentary Time in Preschool Children"
by Alhassan S et al.
Pediatric Exercise Science

Cross-validation of two accelerometers for assessment of physical activity and sedentary time in preschool children

Brief running head: Actiwatch Activity Assessment in Preschoolers

Sofiya Alhassan[1], John R. Sirard[1], Laura B. F. Kurdziel[2],

Samantha Merrigan[3], Cory Greever[1] & Rebecca M. C. Spencer[2,3]

[1]Department of Kinesiology, University of Massachusetts, Amherst

[2]Neuroscience & Behavior Program, University of Massachusetts, Amherst

[3]Department of Psychological & Brain Sciences, University of Massachusetts, Amherst

Corresponding author:
Sofiya Alhassan, PhD
Associate Professor
University of Massachusetts
Department of Kinesiology
110 Totman Building
30 Eastman Lane
Amherst, MA 01003-9258
Phone: 1 - 413-545-3475
Fax: 1- 413-545-2906
Email: alhassan@kin.umass.edu

"Cross-Validation of Two Accelerometers for Assessment of Physical Activity and Sedentary Time in Preschool Children"
by Alhassan S et al.
Pediatric Exercise Science

Abstract

Purpose: The purpose of this study was to cross-validate previously developed Actiwatch (AW; Ekblom et al. 2012) and AcitGraph (AG; Sirard et al. 2005; AG-P, Pate et al. 2006) cut-point equations to categorize free-living physical activity (PA) of preschoolers using direct observation (DO) as the criterion measure. A secondary aim was to compare output from the AW and the AG from previously developed equations. **Methods**: Participants' (n=33; age=4.4±0.8 yrs; females, n=12) PA was directly observed for three 10-minute periods during the preschool-day while wearing the AW (non-dominant wrist) and AG (waist). Device specific cut-points were used to reduce the AW-E (Ekblom et al. 2012) and AG (AG-S, Sirard et al. 2005; AG-P, Pate et al. 2006) data into intensity categories. Spearman correlations (r_{sp}) and agreement statistics were used to assess associations between the DO intensity categories and device data. Mixed model regression was used to identify differences in times spent in activity intensity categories. **Results**: There was a significant correlation between AW and AG output across all data (r_{sp}=0.41, p<0.0001) and both were associated with the DO intensity categories (AW: r_{sp}=0.47, AG: r_{sp}=0.47; p<0.001). At the individual level, all devices demonstrated relatively low sensitivity but higher specificity. At the group level, AW-E and AG-P provided similar estimates of time spent in moderate-to-vigorous PA (MVPA, AW-E: 4.7±4.1, AG-P: 4.4±3.3), compared with DO (5.1±3.5). Conclusion: The AW-E and AG-P estimated times spent in MVPA were similar to DO, but the weak agreement statistics indicate that neither device cut-point equations provided accurate estimates at the individual level.

Keywords: wrist-worn Actiwatch, waist-worn ActiGraph, sleep, physical activity

Introduction

Sleep, sedentary behavior, and physical activity (PA) are all distinct behaviors with different physiological responses (38) and all play a role in pediatric health outcomes (2, 7, 12, 19, 21). In order to assess the combined effects of these behaviors on children’s health, it is critical for researchers to have accurate and reproducible methods for assessing the behaviors. For example, using a single device to collect PA and sleep data would be an important advance in differentiating their contributions to pediatric obesity and other health outcomes. Accelerometers, which detect accelerations in body position, are increasingly being recognized as a practical tool to objectively assess both PA and sleep, particularly for younger children who cannot validly or reliably self-report these behaviors. Currently different commercially-available accelerometers are used to assess PA and sleep (8, 15, 27, 28, 33, 35, 46).

Doubly labeled water can be used to assess energy expenditure in children over several days. However, due to the high cost and the inability of doubly labeled water to assess daily movement pattern, the criterion standard for the objective assessment of daily patterns of PA in preschoolers are waist-worn accelerometers, which are capable of detecting body acceleration at a person’s center of gravity (i.e., trunk) (11, 16, 29, 36). One of the most popular waist-worn accelerometers for assessing PA in children is the tri-axial ActiGraph (AG, ActiGraph, LLC, Pensacola, FL). Unfortunately, a limitation in data collection using waist-worn accelerometers is that participants are sometimes instructed to remove the device when sleeping, therefore only capturing movement during waking hours (16, 36, 46). In addition, due to their placement, waist-worn accelerometers are not able to capture upper body activities and could potentially underestimate the PA of young children (36, 46).

Although there have been algorithms for assessing sleep using hip-worn accelerometers, the most accurate placement of accelerometer devices for sleep assessment is on the wrist (1, 4, 14, 20, 42). The most widely used accelerometer for assessing sleep is the Actiwatch Spectrum (AW, Philips Healthcare, Andover, MA), a uniaxial accelerometer intended to be worn on the wrist of the non-dominant arm (4, 6, 8, 9, 16). In addition to assessing sleep, the placement of the AW on the wrist makes it possible to also assess arm movements during waking hours (8, 16). Previous studies have validated the use of several wrist-worn accelerometers in the assessment of PA in elementary school age children (6 – 11 years of age) (8, 26, 30-33, 37, 41). Currently, only one study has validated the use of a wrist-worn AG accelerometer to categorize PA in very young children (15 – 36 months of age) (15).

Current accelerometer designs have enabled researchers to objectively assess both PA and sleep using similar accelerometer technology (i.e., tri-axial accelerometer) (16). Therefore, the measurement of both behaviors with a single device could improve our understanding of these behaviors on a range of health outcomes while also minimizing research cost and participant burden. The ability of the wrist-worn AW to validly register sleep makes it a viable option for measuring both sleep and PA. Currently, only one study has developed wrist worn PA cut-points for the AW in elementary school-age children (8) and it is unknown if the AW can accurately classify PA intensities in preschool-age children. Therefore, the purpose of this study was to cross-validate previously developed AW [Ekblom *et al.* (8)] and AG [Sirard *et al*. (35) and Pate *et al*. (25)] cut-point equations to categorize free-living physical activity (PA) of preschoolers using direct observation (DO) as the criterion measure. A secondary aim of this study was to compare output from the AW and the AG using those previously developed equations.

Methods

Participants

This study used data from preschool-age children (n = 74 children) who were participating in a larger study focused on sleep and cognition. Data were collected from October 2013 to February 2015. Families of preschool-age children (2.9 - 5 years of age) from 10 preschool centers within the greater Springfield, MA area were invited to participate in the study. Families were eligible if the child was enrolled in full-day preschool and had no sleep or learning disorders. The study was approved by the University of Massachusetts Amherst Institutional Review Board. Parents provided informed consent for their child's participation in the study. Children also provided spoken assent to wear both monitors and be observed on several occasions while at their preschool.

Procedures

On day one of the study, participants were asked to wear an AW (Actiwatch Spectrum (AW, Philips Healthcare, Andover, MA; sampling frequency = 32 Hz, dynamic range = 0.5 - 2 G) on their non-dominant wrist 24 hours per day for 16 days during which there were three 10-minute observation periods (30-mins total). During observation periods, participants were also asked to wear an AG (GT3X accelerometer; ActiGraph, LLC, Pensacola, FL; sampling frequency = 30 Hz, dynamic range = 6 G). The AG accelerometer was attached to an adjustable elastic belt and worn around participant's waists at the center of their lower back to be unobtrusive (40). The AW and the AG have been validated to assess sleep and PA, respectively, in preschool-age children (4, 23). Both monitors were programmed to store data at 15-second epochs and the AW was initialized to start collecting data at 7 am on day one of the study.

Observations were conducted using the Observational System for Recording Physical Activity in Children-Preschool Version (OSRAC-P) (5, 24). The three observation periods consisted of morning classroom time, afternoon classroom time, and outside or gym time. Trained observers categorized participants' PA into five intensity levels: stationary, limb movement, light, moderate, or vigorous. Activity was classified as stationary if the participant was in a resting or in a motionless state with no major limb movement (e.g., sitting down or standing quietly). Activity was classified as limb movement if the activity was stationary with easy movement of limbs(s) or trunk, or leg movements without movement of the entire body from one place to another (e.g., sitting down with trunk or limb movement). Activity was classified as light, moderate, or vigorous if the activity involved translocation (moving body from one location to another) at a slow and easy pace (e.g., walking), moderate pace (e.g., brisk walking), or fast pace (e.g., running), respectively. Any activity normally classified as limb movement, slow easy, or moderate could be "upgraded" to the next intensity category if the activity required more effort (e.g., carrying a heavy object would change slow walking from light to moderate intensity; pushing a swing would change from stationary with limb movement to light intensity). A modified version of the OSRAC-P observation (5 seconds) and recording session (25 seconds) procedures were used in the present study. In the present study, children were observed for 15-seconds and then during the subsequent 15-seconds, observers recorded the highest activity classification observed. This modified protocol was utilized to enable us to match the OSRAC-P observation time frame (15 seconds) to the epochs (15 seconds) used for the AW and AG accelerometers. Observations were timed using the rTimer app on an iPhone. All observers underwent training and were required to demonstrate high inter-observer reliability (>90% agreement) prior to assessing PA using the OSRAC-P. For descriptive purposes, standing height to the nearest millimeter (stadiometer) and body weight to the nearest

0.1 kg (digital scale) were assessed by trained data collectors and used to calculate body mass index (BMI; kg/m^2).

Data Cleaning

The time stamped counts from the AG and AW were temporally matched with the epochs for which direct observations were assessed. Only time points where participants had data for all three measurements (AG, AW, direct observation) were used for analysis. The Sirard *et al*. (AG-S; (35)) and Pate *et al*. (AG-P; (25)) 15-second epoch count cut-points were used to reduce the waist-worn AG data. Only counts from the vertical axis from the AG accelerometer were used for analysis. The age-specific, 15-second AG-S count cut-offs for 3, 4, and 5 year olds for the different activity intensities were sedentary ≤301, ≤363, ≤398; light 302-614, 364-811, 399-890; moderate 615-1230, 812-1234, 891-1254; and vigorous ≥1231, ≥1235, ≥1255, respectively (35). The AG-P 15-second count cut-offs were sedentary < 38, light 39-419, moderate 420-841, and vigorous ≥842 (13, 25, 39, 43). To reduce the AW data, the Ekblom *et al.* (AW-E; (8)) 15-second cut-points were used to differentiate activity intensity categories (sedentary ≤79, light 80-261, moderate 262-405, and vigorous ≥406). For direct observation data, activities classified as stationary and light limb movements were combined into a single category, sedentary, in order to mirror the cut-points for the two monitors.

Data Analyses

Of the total sample (n = 74), participants' data were excluded when the AW reported "off-wrist" (n = 11), when the AG registered continuous zeros (n =15), or when a participant was absent during direct observation session (n = 15). All measures [AG-P, AG-S, AW-E, and direct observation (DO)] were available for 33 children (age = 4.4±0.8 yrs; females, n =12; weight =

18.4±4.4 kg; height = 105.7±7.3 cm; BMI = 16.4±2.4 kg/m^2). Across the 33 participants, there was a total of 1461 data points (44.4±17.7 direct observation data points per participant; approximately 22 minutes with 2 DO data points per minute). Descriptive statistics for the sample were calculated using means, standard deviations, and percentages, where applicable. The analyses were conducted to cross-validate AW-E, AG-P, and AG-S for each 15-second interval with corresponding intervals for DO (criterion validity). Non-parametric statistics were calculated since the AW-E and AG output data were not normally distributed. Also, each participant contributed a different number of data points; up to 60 data points across 30 minutes of observations (n = 1461 data points). Spearman correlations compared the DO classifications (ordinal data) with the 15-sec count values from the AW-E, AG-P and AG-S output for all participants and observations. Wilcoxon Rank Sum Tests were calculated to determine the ability of the AW-E, AG-P, and AG-S to differentiate among DO-determined PA intensity categories (sedentary, light, moderate, and vigorous). Overall agreement statistics (% agreement, sensitivity, specificity, Kappa, positive and negative predictive value) were calculated using a 4x4 table with minutes spent in the four intensity categories from DO and from AW-E, AG-P, and AG-S. In addition, agreement statistics between AW-E, AG-P, AG-S, and DO were calculated for each intensity separately by dichotomizing each data point (e.g., sedentary vs. not sedentary). Using MVPA as an example, positive predictive values (PPV) indicates that of all the data points classified by a device calibration equation as MVPA, the percent of those data points that were also classified as MVPA by DO. In addition, Bland-Altman plots were used to visually assess the agreement between AW-E, AG-P, AG-S and DO for MVPA. For these plots, the difference between the cut-points (AW-E, AG-P, or AG-S) was calculated as the device equation estimate minus DO. A positive or negative mean difference indicated that the utilized cut-point equation either overestimated or underestimated MVPA compared to DO, respectively.

Many field-based physical activity-related studies are interested in the ability of device specific equations to assess overall time spent in activity intensity categories, rather than the classification accuracy of each 15-second time interval. Therefore, a second set of analyses summed the time spent in each intensity category, for the DO, AW-E, AG-P, AG-S, across all observations for each child (N = 33). Estimates of total time spent in intensity categories among devices were compared using a mixed model regression. All analyses were calculated using SAS (ver. 9.2, Cary, NC).

Results

Overall, there were moderate positive associations (Spearman r) between DO intensity categories and counts/15-sec from the AW output ($r = 0.47$) and between DO and the AG output (also $r = 0.47$). The association between AW and AG output was of a similar magnitude ($r = 0.41$). The Wilcoxon Rank Sum tests indicate that, despite considerable variability, the AW-E, AG-P and AG-S were able to differentiate between the median values for adjacent DO intensity levels (all comparisons between adjacent intensity categories were $p \leq 0.03$).

When AW and AG output data are split into four PA intensity categories, the percent agreement with the DO for the AW-E, AG-P, and the AG-S were, 50.3% ($K = 0.13$; $K_w = 0.19$), 48.6% ($K = 0.23$; $K_w = 0.32$), and 44.1% ($K = 0.21$; $K_w = 0.32$), respectively. Agreement statistics calculated when the device outputs were dichotomized are provided in Table 1. While percent agreements were at least 60% across the three device equations and intensity levels, Kappa statistics indicate only "slight" ($K = 0.01$ to 0.20) or "fair" ($K = 0.21$ to 0.40) agreement with DO (17, 18). Except for sedentary time, sensitivity estimates (true positives) were relatively low (5.3%

to 62.3%) and specificity estimates (true negatives) were relatively high (50.7% to 98.5%) across all PA intensity categories. The results for PPV and NPV follow a similar pattern. Except for sedentary time, PPV was relatively low (17.1% to 51.6%) and NPV relatively high (73.2% to 91.0%). Using MVPA as an example, these results indicate that all three device equations were generally good to excellent at distinguishing when a data point was *not* MVPA (specificity and NPV) but only fair to good at correctly identifying actual MVPA (sensitivity and PPV), using DO as the criterion measure.

To determine the accuracy of the device equations based on metrics used in most studies of free-living children, we calculated the total minutes spent in activity intensity categories across all data points for each participant (Table 2). Using the Ekblom cutpoints (AW-E) to process the AW data resulted in a significant underestimation of sedentary time and overestimation of light and total PA time (i.e., light + moderate + vigorous PA); the estimate of MVPA time was similar to DO. In comparison, using the Sirard cutpoints (AG-S) to process the AG data resulted in a significant overestimation in minutes spent in sedentary time and underestimation of time spent in all other intensity categories, compared with DO. Using the Pate cutpoints (AG-P) to process the AG data produced similar estimates of time spent in sedentary, MVPA, and total PA, compared with DO.

The Bland-Altman plots for DO versus AW-E (Figure 1) and DO versus AG-P (Figure 2) reflect a similar level of agreement with both the AW-E and AG-P slightly underestimating minutes spent in MVPA compared with the DO estimates. The range of difference scores is slightly greater for the AW-E with little evidence of increased difference scores with increased mean levels of MVPA for either device methods. The Bland-Altman plot comparing AG-S and DO (Figure 3) reflects a substantial underestimation of MVPA by AG-S, compared with DO, and this underestimation was consistent across the range of mean MVPA values.

Discussion

The purpose of this study was to cross-validate previously developed AW and AG cut-points to categorize free-living PA of preschool-age children using DO as the criterion measure. Secondary aim was to compare outputs from the AW and the AG using the same previously developed equations. For MVPA, there was a moderate, yet statistically significant correlation between the AW-E and DO and also similar estimates of time spent in MVPA. Our results are similar to what others have observed in slightly older children (8, 9). For example, Ekblom *et al.* examined the validity of the wrist-worn AW in assessing PA in elementary school children using energy expenditure as the criterion (8). Researchers reported a strong correlation (r = 0.80) between the AW output and energy expenditure data. It is possible that the higher correlation observed in the Ekblom *et al.* study compared to the current study (r_{Sp} = 0.47) could be due to differences in the criterion measure used. The Ekblom *et al.* study utilized energy expenditure as their criterion measure, and therefore required their participants to perform structured activities in a controlled laboratory setting (8). In the current study, the criterion measure was DO, which was collected under free-living conditions. Finn *et al.* examined the association between AW data output (activity counts) and directly observed PA in preschool-age children during free-living conditions and reported a strong correlation when looking across all intensity categories (r = 0.74) (9). Differences between the current study and the Finn *et al.* study could be due to differences in the DO protocol. In the Finn et al. study both DO and AW data were averaged over relatively long 3-minute epochs for analyses.

The AW output (AW-E) was also compared to the waist-worn AG output (AG-P and AG-S), a substantially validated and commonly used measure of PA in preschool children. We observed a significant positive associations among the AW-E, AG-P, and AG-S. In addition, the outputs (raw count values) from both devices increased with higher DO intensity categories. Overall, the percent agreements with DO were similar for the AW-E and AG-P. However, the results indicate that the agreement between DO and the intensity categorizations based on the monitor equations for each intensity level and individual data points was variable. For example, using the Sirard cutpoint to process the AG data (AG-S) resulted in the highest sensitivity (87.4%) for sedentary time but the lowest sensitivity for all other intensity categories, compared to AG-P and AW-E. The light intensity threshold cutpoint for the AG-S is higher than for the AG-P and this is likely a large factor in the differences among agreement statistics between these two data processing methods. For MVPA, the AW-E was able to correctly categorize data points as MVPA (according to DO) only 46.9% of the time (sensitivity); using AG-P resulted in similar sensitivity (39.8%) but much lower for AG-S (16.2%). In general, for the PA intensity categories, specificity (and NPV) were greater than sensitivity. For example, although the AG-S cutpoints only resulted in a 16.2% positive predictive rate for MVPA, specificity was 95.2%. Because this analysis dichotomized the data (e.g., MVPA or not MVPA), this high specificity indicates that of all the data points categorized as not MVPA by DO, 95.2% of the data processed with the AG-S correctly categorized these data points as not MVPA. This does not, however, provide any clarity on whether those data points were sedentary or light intensity PA. The relatively high specificity values are also driven by the distribution of the data, with only five minutes (about 23% of total time) spent in MVPA. With fewer “true” MVPA data points, there is less data to establish greater sensitivity rates for these higher intensities. Since the behaviors of the children were not scripted, the relatively low amount of MVPA represents typical behavior patterns of children in preschool settings.

Ekblom *et al.* tested the agreement between the wrist-worn AW and hip-worn AG counts in elementary school-aged children during free living and reported strong correlations (r = 0.67) between AW and AG counts (8). Routen *et al.*(31) examined if activity estimates from the AW placed on the hip was comparable to AW placed on the wrist in elementary school-age children. Researchers used the Ekblom *et al.* (8) and Puyau *et al.* (27, 28) cut-points to analyze the wrist and hip data, respectively. Similar to the current study, their results showed that time spent in moderate and vigorous activity categories was greater when measured at the wrist compared with the hip (31). In addition, the researchers observed that time categorized as sedentary behavior was greater at the hip than at the wrist (31). The observed differences between AW-E and AG-Puyau output classifications compared with DO are largely a result of the placement of the monitors. For example, it is possible that activities that isolate a part of the body on which the device is positioned, such as vigorous arm movement while sitting could cause high readings on one device (recorded as moderate intensity by the wrist-worn AW) but read stationary on the waist-worn AG. When using the OSRAC-P, the intensity of an activity can be upgraded to the next intensity code if it is performed while carrying a heavy object or if the activity requires more effort such as pushing a swing. In these types of activities, both the DO and the AW-E would register a higher intensity compared to the AG equations. Furthermore, studies have shown that there is greater signal acceleration at the wrist during sedentary activity (i.e., sitting and coloring) and short bursts of intermittent high intensity activities (which is characteristic of most preschool-age children's activity pattern) (3, 31).

Another major finding of the current study was the large variability for both the AW and the AG outputs for each DO-defined activity intensity categories. This could be due to the way in which DO data were collected. For DO, the protocol was to record the highest intensity activity during the 15-second observation interval. This meant that even if a participant was stationary for 13-seconds, but ran the last two seconds of the interval, the participant's activity would have been classified by the OSRAC-P as vigorous intensity for that 15 second epoch. In these instances, the counts from the AW and AG would be lower despite the higher direct observation category because the accelerometers record the total acceleration accumulated during the epoch. Comparatively, there were also times in which a participant was enthusiastically running or jumping for the entire 15-second epoch. Because of this discrepancy between the DO protocol and the data collected by the devices, some misclassification is expected. Exactly how much misclassification is due to this issue is unclear, warranting further research using different accelerometer data processing techniques and/or different DO protocols to better match the variables being measured by these methods. Despite this limitation, the data in Table 2 and the Bland-Altman plots support the use of the AW-E and the AG-P to assess MVPA trends at the group level, but not at the individual level.

We utilized two available cut-points to reduce the AG data. Compared to DO data, the AG-P data processing method was better at predicting sedentary and light activity, whereas, AG-S data processing method overestimated sedentary and underestimated light intensity activities. The differences observed between the two AG cut-points could be due to factors such as the types of activities and the criterion measure used in the calibration studies (8, 10, 31, 45). For example, Ekblom *et al.* reported that the types of activities used in the calibration study can have an influence on the association between the derived activity counts and the criterion measure used (i.e., energy expenditure estimates or direct observation) (8). The Sirard *et al.* (34) calibration study utilized semi-structured activities and direct observation, while the Pate *et al.* (25, 35) study utilized structured activities and indirect calorimetry to derive their cut-points. Researchers have reported that the use of structured activities could impact children's natural movement pattern (44). Validating accelerometers with structured activities and/or indirect calorimetry may not be representative of free-living circumstances. Free-living conditions allow for much more variability in movements, which explains, to some extent, the large variability in the AG counts for each direct observation category.

The current study had several strengths. The data was collected in the participants' natural environment and were all free-living activities. We used DO in those free-living environments as a criterion measure to validate PA intensity categorizations, using a previously developed data processing equation, of a wrist-worn accelerometer (AW-E), which has been shown to increase participant compliance with wearing the device (8, 31). Limitations of the current study include the limited availability of data (10 minutes per DO session) used per participant. Although this limits our ability to estimate MVPA, our PA assessment protocol reflects real-life movement. Future studies should conduct longer DO of free-play session in order to capture enough activities to accurately estimate MVPA. A limitation of the DO protocol is that only the intensity of the highest movement category was recorded. Another limitation of the study is the utilization of 15-second intervals to assess both the directly observed PA (momentary time sampling) and accelerometer data. Studies have shown that a 15-second interval is too long to pick up short bursts of intermittent high intensity activities that are characteristic of preschool-age children's play patterns (3, 31, 34). However, since the participants were wearing the AW for 16 days, the 15-second interval was the shortest that could be used while maintaining battery life and data storage on the devices. Most newer accelerometers are capable of collecting and storing the raw acceleration signal (≥30 Hz), which may allow for a more accurate representation of the intermittent nature of preschool children's PA. Nonetheless, we should note that the occasional misclassification of very small duration bouts of activities might not be meaningful in examining the impact of PA on health. Another limitation of the study is applying cut-point equations derived from hip-worn accelerometers to accelerometers placed on the lower back.

In conclusion, the results show that there was a moderate correlation between previously developed AW-E cut-points and DO for time spent in MVPA. Based on these results, it is possible that the AW-E can be used to identify, at the group level, times spent in MVPA in preschool-age children. Processing the AG with the Pate cutpoints resulted in similar correlations and agreement statistics with DO, and absolute estimates of time spent in MVPA, compared with DO and AW-E. It is recommended that future studies video tape the direct observation so that focal sampling and duration coding can be used to more accurately assess activity intensity and also account for sudden changes in PA behavior that cannot be captured with a 15-second time sampling interval. This alternative type of direct observation coding relies on recording the behavior and intensity each time the participant changes their behavior (focal sampling) and then recording how much time the participant spends in the behavior (duration coding) (22).

Acknowledgements

This work was supported in part by NIH R01 HL111695 (to R.M.C.S.). We thank all the participating preschool centers.

Conflict of interest statement

All authors have no conflict of interest or financial interest to report.

References

1. Ancoli-Israel S, Cole R, Alessi C, Chambers M, Moorcroft W and Pollak CP. The role of actigraphy in the study of sleep and circadian rhythms. *Sleep*. 2003; 26(3):342-392.

2. Atkinson G and Davenne D. Relationships between sleep, physical activity and human health. *Physiol Behav*. 2007; 90(2-3):229-235.

3. Baquet G, Stratton G, Van Praagh E and Berthoin S. Improving physical activity assessment in prepubertal children with high-frequency accelerometry monitoring: a methodological issue. *Prev Med*. 2007; 44(2):143-147.

4. Belanger ME, Bernier A, Paquet J, Simard V and Carrier J. Validating actigraphy as a measure of sleep for preschool children. *J Clin Sleep Med*. 2013; 9(7):701-706.

5. Brown WH, Pfeiffer KA, McLver KL, Dowda M, Almeida MJ and Pate RR. Assessing preschool children's physical activity: the Observational System for Recording Physical Activity in children-preschool version. *Res Q Exerc Sport*. 2006; 77(2):167-176.

6. Chen X, Sekine M, Hamanishi S, et al. The validity of nursery teachers' report on the physical activity of young children. *J Epidemiol*. 2002; 12(5):367-374.

7. Colley RC, Wong SL, Garriguet D, Janssen I, Connor Gorber S and Tremblay MS. Physical activity, sedentary behaviour and sleep in Canadian children: parent-report versus direct measures and relative associations with health risk. *Health Rep*. 2012; 23(2):45-52.

8. Ekblom O, Nyberg G, Bak EE, Ekelund U and Marcus C. Validity and comparability of a wrist-worn accelerometer in children. *J Phys Act Health*. 2012; 9(3):389-393.

9. Finn KJ and Specker B. Comparison of Actiwatch activity monitor and Children's Activity Rating Scale in children. *Med Sci Sports Exerc*. 2000; 32(10):1794-1797.

10. Freedson P, Pober D and Janz KF. Calibration of accelerometer output for children. *Medicine & Science in Sports & Exercise*. 2005; 37(11 Suppl):S523-530.

11. Grydeland M, Hansen BH, Ried-Larsen M, Kolle E and Anderssen SA. Comparison of three generations of ActiGraph activity monitors under free-living conditions: do they provide comparable assessments of overall physical activity in 9-year old children? *BMC Sports Sci Med Rehabil*. 2014; 6:26.

12. Iglowstein I, Jenni OG, Molinari L and Largo RH. Sleep duration from infancy to adolescence: reference values and generational trends. *Pediatrics*. 2003; 111(2):302-307.

13. Janssen X, Cliff DP, Reilly JJ, et al. Predictive validity and classification accuracy of ActiGraph energy expenditure equations and cut-points in young children. *PLoS One*. 2013; 8(11):e79124.

14. Jean-Louis G, von Gizycki H, Zizi F, Spielman A, Hauri P and Taub H. The actigraph data analysis software: II. A novel approach to scoring and interpreting sleep-wake activity. *Percept Mot Skills*. 1997; 85(1):219-226.

15. Johansson E, Ekelund U, Nero H, Marcus C and Hagstromer M. Calibration and cross-validation of a wrist-worn Actigraph in young preschoolers. *Pediatr Obes*. 2015; 10(1):1-6.

16. Lambiase MJ, Gabriel KP, Chang YF, Kuller LH and Matthews KA. Utility of actiwatch sleep monitor to assess waking movement behavior in older women. *Med Sci Sports Exerc*. 2014; 46(12):2301-2307.

17. Landis JR and Koch GG. An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics*. 1977; 33(2):363-374.

18. Landis JR and Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977; 33(1):159-174.

19. Liou YM, Liou TH and Chang LC. Obesity among adolescents: sedentary leisure time and sleeping as determinants. *J Adv Nurs*. 2010; 66(6):1246-1256.

20. Littner M, Kushida CA, Anderson WM, et al. Practice parameters for the role of actigraphy in the study of sleep and circadian rhythms: an update for 2002. *Sleep*. 2003; 26(3):337-341.

21. Lowry R, Eaton DK, Foti K, McKnight-Eily L, Perry G and Galuska DA. Association of Sleep Duration with Obesity among US High School Students. *J Obes*. 2012; 2012:476914.

22. Lyden K, Petruski N, Staudenmayer J and Freedson P. Direct observation is a valid criterion for estimating physical activity and sedentary behavior. *J Phys Act Health*. 2014; 11(4):860-863.

23. Pate R, Almeida MJ, McIver KL, Pfeiffer KA and Dowda M. Validation and calibration of an accelerometer in preschool children. *Obesity*. 2006; 14(11):2000-2006.

24. Pate R, McIver K, Dowda M, Brown WH and Addy C. Directly observed physical activity levels in preschool children. *J Sch Health*. 2008; 78(8):438-444.

25. Pate RR, Almeida MJ, McIver KL, Pfeiffer KA and Dowda M. Validation and calibration of an accelerometer in preschool children. *Obesity (Silver Spring)*. 2006; 14(11):2000-2006.

26. Phillips LR, Parfitt G and Rowlands AV. Calibration of the GENEA accelerometer for assessment of physical activity intensity in children. *J Sci Med Sport*. 2013; 16(2):124-128.

27. Puyau MR, Adolph AL, Vohra FA and Butte NF. Validation and Calibration of Physical Activity Monitors in Children. *Obesity Research*. 2002; 10(3):150-157.

28. Puyau MR, Adolph AL, Vohra FA, Zakeri I and Butte NF. Prediction of activity energy expenditure using accelerometers in children. *Med Sci Sports Exerc*. 2004; 36(9):1625-1631.

29. Reilly J, Kelly LA, Montgomery C, et al. Validation of Actigraph accelerometer estimates of total energy expenditure in young children. *International Journal of Pediatric Obesity*. 2006; 1(3):161-167.

30. Rosenberger ME, Haskell WL, Albinali F, Mota S, Nawyn J and Intille S. Estimating activity and sedentary behavior from an accelerometer on the hip or wrist. *Med Sci Sports Exerc*. 2013; 45(5):964-975.

31. Routen AC, Upton D, Edwards MG and Peters DM. Discrepancies in accelerometer-measured physical activity in children due to cut-point non-equivalence and placement site. *J Sports Sci*. 2012; 30(12):1303-1310.

32. Rowlands AV, Rennie K, Kozarski R, et al. Children's physical activity assessed with wrist- and hip-worn accelerometers. *Med Sci Sports Exerc*. 2014; 46(12):2308-2316.

33. Schaefer CA, Nace H and Browning R. Establishing wrist-based cutpoints for the Actical accelerometer in elementary school-aged children. *J Phys Act Health*. 2014; 11(3):604-613.

34. Sedentary Behaviour Research N. Letter to the editor: standardized use of the terms "sedentary" and "sedentary behaviours". *Appl Physiol Nutr Metab*. 2012; 37(3):540-542.

35. Sirard J, Trost S, Pfeiffer KA, Dowda M and Pate RR. Calibration and evaluation of an objective measure of physical activity in preschool children. *J Phys Act Health*. 2005; 2(3):345-357.

36. Sirard JR and Pate RR. Physical activity assessment in children and adolescents. *Sports Medicine*. 2001; 31(6):439-454.

37. Tracy DJ, Xu Z, Choi L, Acra S, Chen KY and Buchowski MS. Separating bedtime rest from activity using waist or wrist-worn accelerometers in youth. *PLoS One*. 2014; 9(4):e92512.

38. Tremblay MS, Colley RC, Saunders TJ, Healy GN and Owen N. Physiological and health implications of a sedentary lifestyle. *Appl Physiol Nutr Metab*. 2010; 35(6):725-740.

39. Trost SG, Fees BS, Haar SJ, Murray AD and Crowe LK. Identification and validity of accelerometer cut-points for toddlers. *Obesity (Silver Spring)*. 2012; 20(11):2317-2319.

40. Trost SG, McIver KL and Pate RR. Conducting accelerometer-based activity assessments in field-based research. *Medicine & Science in Sports & Exercise*. 2005; 37(11 Suppl):S531-S543.

41. Trost SG, Zheng Y and Wong WK. Machine learning for activity recognition: hip versus wrist data. *Physiol Meas*. 2014; 35(11):2183-2189.

42. Tudor-Locke C, Barreira TV, Schuna JM, Jr., Mire EF and Katzmarzyk PT. Fully automated waist-worn accelerometer algorithm for detecting children's sleep-period time separate from 24-h physical activity or sedentary behaviors. *Appl Physiol Nutr Metab*. 2014; 39(1):53-57.

43. Van Cauwenberghe E, Gubbels J, De Bourdeaudhuij I and Cardon G. Feasibility and validity of accelerometer measurements to assess physical activity in toddlers. *Int J Behav Nutr Phys Act*. 2011; 8:67.

44. Welk G, Corbin C and Dale D. Measurement issues in assessment of physical activity in children. *Research Quarterly for Exercise and Sport*. 2000; 71(2):59-72.

45. Welk GJ. Principles of design and analyses for the calibration of accelerometry-based activity monitors. *Med Sci Sports Exerc*. 2005; 37(11 Suppl):S501-511.

46. Welk GJ, Corbin CB and Dale D. Measurement issues in the assessment of physical activity in children. *Res Q Exerc Sport*. 2000; 71(2 Suppl):S59-73.

"Cross-Validation of Two Accelerometers for Assessment of Physical Activity and Sedentary Time in Preschool Children"
by Alhassan S et al.
Pediatric Exercise Science

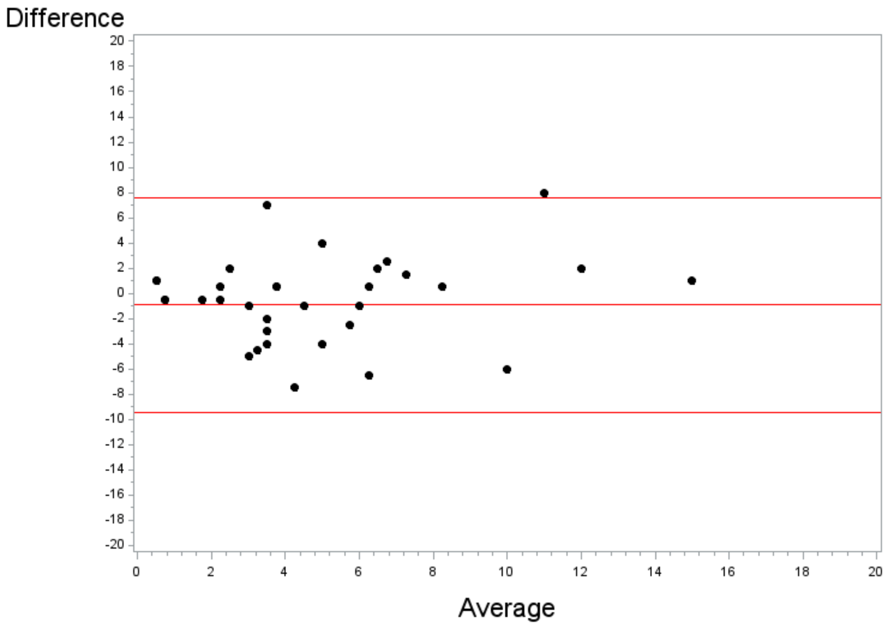

Figure 1. Bland-Altman plot of AW-E vs DO for minutes spent in MVPA. Difference = AW-E – DO.

"Cross-Validation of Two Accelerometers for Assessment of Physical Activity and Sedentary Time in Preschool Children"
by Alhassan S et al.
Pediatric Exercise Science

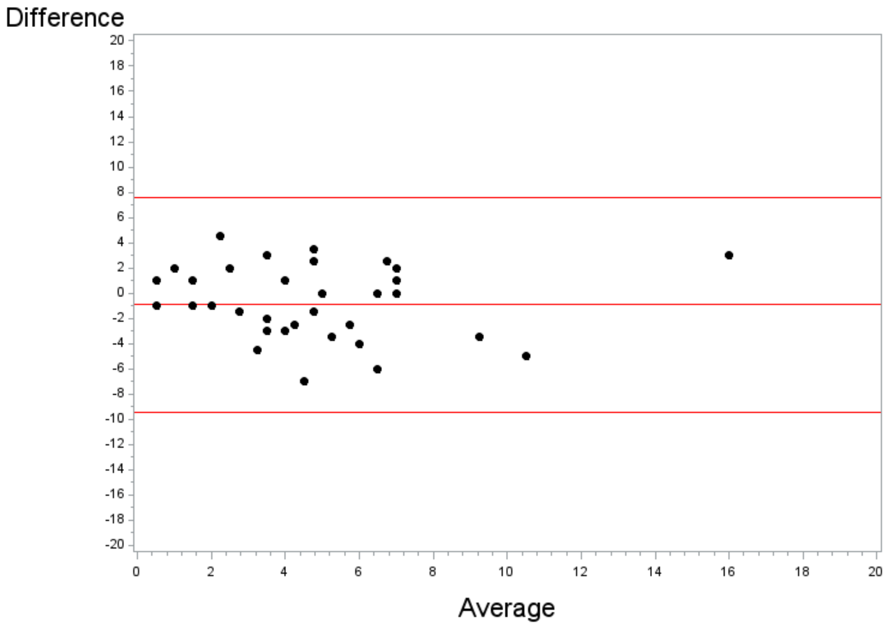

Figure 2. Bland-Altman plot of AG-P vs DO for minutes spent in MVPA. Difference = AG-P – DO.

"Cross-Validation of Two Accelerometers for Assessment of Physical Activity and Sedentary Time in Preschool Children"
by Alhassan S et al.
Pediatric Exercise Science

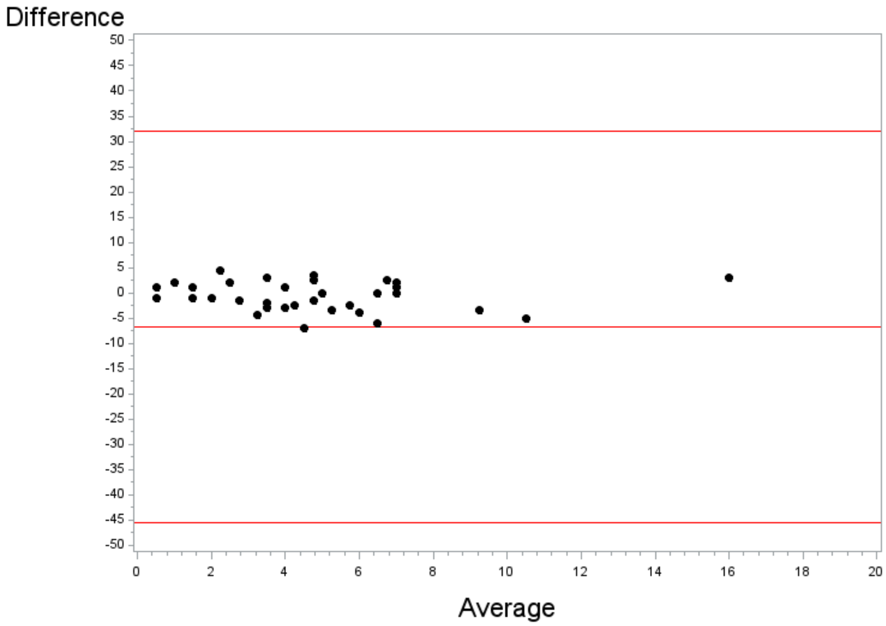

Figure 3. Bland-Altman plot of AG-S vs DO for minutes spent in MVPA. Difference = AG-S – DO.

"Cross-Validation of Two Accelerometers for Assessment of Physical Activity and Sedentary Time in Preschool Children"
by Alhassan S et al.
Pediatric Exercise Science

Table 1: Agreement Statistics for dichotomized output from AW-E, AG-S, and AG-P with DO (criterion)

	% Agreement	Kappa	Sensitivity	Specificity	PPV	NPV
Sedentary						
AW-E	65.4	0.31	41.5	89.1	79.1	60.5
AG-S	62.7	0.26	87.4	38.2	58.4	75.3
AG-P	61.5	0.40	60.1	79.7	74.7	66.8
Light PA						
AW-E	53.8	0.10	62.3	50.7	31.8	78.4
AG-S	65.0	0.01	18.5	82.2	27.8	73.2
AG-P	60.7	0.13	51.1	64.3	34.6	78.1
Moderate PA						
AW-E	82.8	0.11	22.0	89.7	19.6	91.0
AG-S	86.3	0.06	10.0	95.1	18.8	90.3
AG-P	80.3	0.09	24.0	86.8	17.1	90.9
Vigorous PA						
AW-E	86.3	0.31	33.9	94.0	45.7	90.6
AG-S	86.5	0.06	5.3	98.5	34.5	87.5
AG-P	86.1	0.19	18.0	96.2	41.5	88.8
MVPA						
AW-E	77.5	0.35	46.9	86.8	51.6	84.4
AG-S	76.9	0.15	16.2	95.2	50.5	79.0
AG-P	75.3	0.27	39.8	86.0	46.2	82.6

AG-S = ActiGraph data processed with Sirard et al. cutpoints; AG-P ActiGraph data processed with Pate et al. cutpoints; AW-E= Actiwatch data processed with Ekblom et al. cutpoints; DO = Direct Observation; PA = Physical Activity; MVPA = Moderate+Vigorous Physical Activity; PPV = Positive Predictive Value; NPV = Negative Predictive Value

Table 2: Minutes spent in intensity categories by device; Mean (SD)

	DO	AG-S †	AG-P	AW-E
Sedentary	11.06 (5.13)	16.54 (6.83) **	8.91 (5.13) ‡	5.80 (3.67) **
Light	5.98 (3.65)	3.98 (2.32) *	8.85 (4.85) * ‡	11.71 (4.36) **
Moderate	2.27 (1.91)	1.21 (1.39) *	3.18 (2.06) *	2.55 (2.00)
Vigorous [a]	2.86 (2.49)	0.44 (1.11) **	1.24 (2.17) **	2.12 (2.52)
MVPA [a]	5.14 (3.54)	1.65 (2.20) **	4.42 (3.34)	4.67 (4.05)
Total PA	11.12 (6.23)	5.64 (3.72) **	13.27 (6.27) ‡	16.38 (6.25) **

AG-S = ActiGraph data processed with Sirard et al. cutpoints; AG-P ActiGraph data processed with Pate et al. cutpoints; AW-E= Actiwatch data processed with Ekblom et al. cutpoints; DO = Direct Observation; Total PA = Light+Moderate+Vigorous PA.

Significantly different from DO; * $p \leq 0.05$, ** $p < 0.001$

AW-E significantly different from AG-S; † $p \leq 0.004$ all comparisons

AW-E significantly different from AG-P; ‡ $p \leq 0.03$

[a] Data were skewed for these variables. Analyses performed on log-transformed values. Non-transformed values presented in table.