Development of raw acceleration cut-points for wrist and hip accelerometers to assess sedentary behaviour and physical activity in 5-7 year old children

Abstract:

The purpose of the study was to validate sedentary behaviour (SB), moderate-to-vigorous physical activity (MVPA) and vigorous physical activity (VPA) accelerometer cut-points for wrist and hip-worn ActiGraph devices in 5-7 year old children. Forty-nine (n=27 girls) 5-7-year-old children were recruited. Participants wore an ActiGraph GT9X accelerometer, recording data at 100Hz subsequently downloaded in 1s epochs, on both wrists and the right hip during a standardised protocol (10 tasks ranging from lying to running), and during recess. Cut-points were generated using ROC analysis using direct observation as a criterion reference in the cut-point generation group (n=22, 50% girls). Subsequently, cut-points were modified using Confidence intervals equivalency analysis until optimal cut-points were identified. Cut-points were then cross-validated using a cross-validation group (n=10, 60% girls). SB cut-points were 36mg (Sensitivity(Sn)=79.8%, Specificity(Sp)=56.8%) for non-dominant wrist, 39mg (Sn=75.4%, Sp=70.2%) for dominant wrist and 20mg (Sn=78%, Sp=50.1%) for hip. MVPA cut-points were 189mg (Sn=82.6%, Sp=78%) for non-dominant wrist, 181mg (Sn=79.1%, Sp=76%) for dominant wrist and 95mg (Sn=79.3%, Sp=75.6%) for hip. VPA cut-points were 536mg (Sn=75.1%, Sp=68.7%) for non-dominant wrist, 534mg (Sn=67.6%, Sp=95.6%) for dominant wrist and 325mg (Sn=78.2%, Sp=96.1%) for hip. All accelerometer placements demonstrated adequate levels of accuracy for SB and PA assessment.

Key Words: Accelerometry, validation, raw signal, objective measurement, criterion validity

Authors:

Mr. Matteo Crotti[1,2*] , Dr. Foweather Lawrence[2] , Dr. James R. Rudd[1] , Mrs. Liezel Hurter[2] , Mr. Sebastian Schwarz[1] , Dr. Lynne M. Boddy[2]

Affiliations:

1. Department of Sport Studies, Leisure and Nutrition, Liverpool John Moores University, UK
2. Physical activity Exchange, Research Institute for Sport and Exercise Science, Liverpool John Moores University. UK

Corresponding Author:

* Mr Matteo Crotti: No 5 Primrose Hill, Liverpool, L3 2EX, M.Crotti@2016.ljmu.ac.uk

Key Words: Accelerometry, validation, raw signal, objective measurement, criterion validity

Introduction

Accelerometers are the most widely used devices to assess physical activity (PA) and sedentary behaviours (SB) in children and have proved to be a feasible method to assess children on a large scale (1, 2). For many years, hip-worn accelerometers were the preferred devices for PA assessment (3). A major problem with hip-worn devices is poor compliance, which has been attributed to discomfort whilst wearing or forgetting to wear the devices after removal (4). However, it was reported that a 24h wear time protocol with hip monitors can lead to high levels of compliance (5). More recently, researchers have used wrist-worn accelerometers as they obtain better wear compliance (4, 6) and are suitable for 24-h per day recording, allowing sleep-time assessment (7, 8). A further advantage of wrist-worn accelerometers is that they are more sensitive to upper body movement, considered as a significant component of children's PA (4).

Traditionally, accelerometer output was reduced to proprietary units defined as "counts" (9). However, comparing PA and SB estimates across studies that have used different devices brands is problematic because of the brand specific data processing algorithms used (10). Consequently, a methodological harmonisation was recommended involving the use of raw acceleration signals rather than counts, regardless of the device brand (11). Raw signals consist of gravitational accelerations assessed at sample frequencies typically above 10Hz. The Euclidean Norm Minus One (ENMO), calculated using the R GGIR package, is emerging as the most frequently used metric when processing raw acceleration data generated from the most commonly used triaxial accelerometers (ActiGraph, GENEActiv and Axivity) (12, 13). The use of raw acceleration metrics such as ENMO have the potential to facilitate comparisons between different brands and wear sites (4) and to increase researchers' control over data processing. PA and SB intensity cut-points derived for use with ENMO data have been developed for the ActiGraph accelerometers for older children and adults (14, 15). Due to the characteristic intermittent nature of the movement behaviours during childhood and in view of the differences in movement dynamics observed in different age groups it is fundamental to create age specific cut-points (16, 17). However, to the best of our knowledge no calibration study has established raw acceleration cut-points for ActiGraph devices to assess PA or SB in 5-7 year old children.

The majority of previous calibration studies have been performed in laboratories and involved equipment such as treadmills or indirect calorimetry that could affect children's movement patterns and gait (18). Concerns have been raised about the ecological validity of such settings and it is has been recommended that future calibration studies should involve activities that are representative of free-living PA (19). Additionally, calibration studies should consider accelerometers' limitations in assessing SB based on the absence of or low levels of acceleration and distinguishing stationary activities such as standing stationary from SB (15, 20).

A further consideration in developing cut-points concerns the statistical techniques used to identify and validate intensity thresholds. Calibration studies have typically used Receiver Operating Characteristic (ROC) curve analysis for the calculation of SB and PA intensity cut-points from raw accelerometer data (18). Intensity thresholds were typically derived by coding and grouping all the accelerations recorded during the calibration protocol into binary indicator variables (0 or 1) based on the observed or measured activity level (18). However, the proportion of data from each activity level (e.g. SB, LPA, MPA and VPA) used in ROC analysis plays a key role in determining PA and SB cut-points and in some case could lead to low accuracy in SB and PA assessment. For example the presence of a high proportion of SB acceleration in the ROC analysis dataset could lead to LPA, MPA and VPA cut-points that are too low to accurately classify the behaviour (21). In light of this, alternative statistical procedures that could lead to increased diagnostic accuracy should be evaluated. The use of ‘pairs’ of activity levels in ROC analysis (e.g. SB versus LPA) rather grouped activities (i.e., SB versus LPA, MPA and VPA) has the potential to account for disproportions of data in different activity levels and might lead to improved diagnostic accuracy. However, to date, no study has evaluated the diagnostic accuracy of SB and PA cut-points calculated by ROC curve analysis using ‘pairs’ of activity levels.

In view of the gaps in the literature presented above, this study aimed to develop and validate raw acceleration cut-points for the estimation of SB and PA in 5-7-year-old children using ActiGraph devices, and to compare different methods of cut-point calculation.

Methods

Design and Participants

The study received institutional research ethics committee approval (17/SLN/004). After school gatekeeper consent was obtained from the headteacher of a single primary school in a metropolitan city in North-West England, parent/carer consent and child assent forms were distributed to potential participants (n = 60) aged between 5 and 7 years old and taken home to parent/carer. As a result, 49 children agreed to take part in the study. Data collection for the study took place between November-December, 2017.

Data Collection and Procedures

All the participants were invited to take part in a standardised activity protocol and to be video-recorded during school recess. Data collection took place in the school gymnasium and playground to mimic free-living conditions and increase the ecological validity of the study protocol. Children’s stature (The Leicester Height Measure, Child Growth Foundation, Leicester, United Kingdom), sitting stature and waist circumference to the nearest 0.1cm together with mass to the nearest 0.1kg (model 760, Seca, Hamburg, Germany) were measured using standard procedures (22). All measurements were taken twice, with a third measurement taken if the first two differed by more than >1%. Body mass index (BMI) was calculated from stature and mass. Children self-reported their dominant hand and additionally they were asked to write their name on a paper so researchers could double check hand dominance.

Activity monitors

Participants were fitted with an ActiGraph GT9X Link on both wrists and on the right hip, and wore the devices throughout the data collection session. The GT9X was set to record at 100Hz and measured acceleration in a range of ±8*g* on x, y and z axes. Data were downloaded in 1s epochs.

Direct observation

Children's SB and PA were assessed using direct observation during the standardised activity protocol and during recess. Direct observation was chosen as the criterion reference for the classification of SB and PA levels as it is considered the most appropriate method to assess rapid changes in physical activity behaviours, typical of this age group, it does not involve equipment that might impair children's normal movements (17) and has been used for calibration purposes in previous studies (23, 24).

Calibration Protocol

The activity protocol lasted around 60 minutes in total, took place in the school hall during usual lesson time, and involved three participants at a time, rotating between 10 different tasks (Table 1). The selection of the tasks was informed by previous calibration studies in this age group, by observing children's typical recess play activities, and through consulting primary school teachers. Tasks were selected to encompass each activity intensity (SB, LPA, MPA and VPA) and were designed to simulate children's free-living PA and SB as accurately as possible. Four SB (Lying while watching TV, sitting while colouring, sitting and play with a tablet and playing with LEGO), one LPA (passive standing), two MPA (walking briskly together, throwing and catching) and three VPA (running, obstacle course run and hopping) activities were included in the protocol. The intensity of each activity in the protocol was classified using METs as reported in the youth compendium of physical activities (25). The most widely accepted intensity thresholds were used to classify the activities: SB (≤1.5METs), LPA (≥1.5–<3METs), MPA (≥3–<6 METs), VPA (≥ 6 METs) (26).

[TABLE 1 ABOUT HERE]

The activities were ordered into three different activity protocols and participants were randomised to one of the protocols. The three protocols were designed to allow three children to complete the protocol simultaneously. Children had 2 minutes rest after MPA and VPA tasks, while they were asked whether they needed more rest before starting each activity. Researchers independently conducted live direct observations of children through the protocol, which involved continuously instructing and supervising children to ensure they were 'on task', and recording the start time and end times of each activity.

Recess observation

Recess was included in the study protocol to capture children's behaviours during free-living conditions. Children were asked to participate in school recess as normal whilst wearing the devices. Each researcher video-recorded one child for a period of 10 minutes during either morning or lunchtime recess. Based on previous studies measuring activity levels during recess and previous observations of children's recess in the school involved, we expected children to spend the highest proportion of recess in LPA and a progressively lower amount of time in MPA, VPA respectively (27). Behaviours during recess were assessed and classified on a second-by-second basis (in order to match accelerometery 1s epochs) using the Youth compendium of physical activities (25). Before proceeding with the video analysis, the research team analysed three randomly selected video-recordings jointly in a single group session where behaviour classification was discussed until unanimous consensus was reached. Subsequently, one researcher classified children's recess behaviours second-by-second based on the activities and METs reported in the Youth compendium of physical activities (SB: ≤1.5METs, LPA: >1.5&<3METs, MPA: ≥3&<6 METs, or VPA: ≥ 6 METs) (25). Uncertainties with the classification of children's behaviours that emerged during analysis were discussed and resolved with the research team by consensus.

Data analysis

ActiGraph accelerations were downloaded and converted to .csv format data using Actilife software (ActiLife v6.13.3). Subsequently, the package GGIR version 1.11-0 from R software version 3.2.5 (R Foundation, www.r-project.org) was used to process raw data and calculate average ENMO accelerations for each 1 second epoch. As a result, csv documents presenting ENMO and related timestamps were produced. Acceleration data were then paired with SB and PA observation data. The first and last 15 seconds of each task in the activity protocol were deleted to account for possible start and end time imprecision, transition time delays, and irregular movement patterns, as well as to control for learning effect and fatigue. Only data from participants that completed both the standardised protocol and observation of recess were included in the final analysis. The final sample of participants was randomly divided into a cut-point generation (22 participants, n = 11 girls) and a cross-validation (10 participants, n = 6 girls) group for analysis. Shapiro Wilk test was performed to assess distribution normality of decimal age, height, weight, BMI both in participants included and excluded from the study. Subsequently, either independent samples t-test or Mann-Whitney test were performed to assess differences in decimal age, height, weight and BMI between participants in the two groups based on normality distribution test. Differences in the distribution of males and females between participants included and excluded was assessed using Chi-square test.

In this study we proposed a novel approach to cut-point calculation divided in 3 phases comprising 1) initial ROC analysis, 2) the use of equivalence testing to identify the likely optimum cut-points at the group level and 3) cross validation of the cut-points.

Phase 1. During the first phase cut-points were calculated using ROC curve analysis in the cut-point generation group. R package pROC was used to perform ROC and calculate SB, MVPA and VPA cut-points.

[TABLE 2 ABOUT HERE]

Consistent with previous studies, ROC analysis was initially performed including all the SB and PA levels (i.e. all recorded data across all activities). In contrast to previous research, and to reduce bias associated with unequal distributions of PA behaviours (28), ROC analysis was performed including pairs of activity levels, for example: SB versus LPA, MPA versus VPA (Table 2). To evaluate the effect of passive standing on the diagnostic accuracy of the cut-points, the acceleration signals collected during standing while watching TV were excluded from some of the conditions ROC analysis (Table 2). The Youden index and Distance method (selecting the point in the ROC curve that is closer to the left corner of the ROC curves plot) were used to calculate cut-points (29). The Area Under the ROC curve (AUC) and the related confidence interval (ciAUC) were calculated as a measure of a test's ability to discriminate between different conditions. Sensitivity and specificity were calculated. Agreement between the criterion method (direct observation) and accelerometer estimates generated using the cut-points was assessed using % of agreement (%Ag) and Cohen's Kappa (CK). CK values were considered poor when lower than 0.00, slight when between 0.00 and 0.20, fair when between 0.21 and 0.40, moderate when between 0.41 and 0.60, substantial when between 0.61 and 0.80 and almost perfect when between 0.81 and 1.00 (30). Lastly, equivalency analysis was used to assess the group-level equivalence between the observation and cut-point derived SB and PA estimates (31). Equivalency analysis compares an equivalence region derived from a criterion reference (e.g. observation) to the confidence interval for the difference in means between the criterion reference and a different method (e.g. accelerometery). The equivalence region is centred on the mean derived from the criterion reference while the confidence interval is centred on the mean obtained from the method to compare. Non-equivalence is rejected at the level α if 100(1-2α)% confidence interval for the difference in means lies entirely within the equivalence region. Based on previous research using equivalency testing to compare PA assessment methods, we used an equivalence region of ±10% the mean of the time spend in SB or PA activities assessed using the criterion method (observation) (32). Subsequently, we calculated the 90% confidence interval (as α was set at 0.05) for the difference in means between observed and cut-points derived time spent in SB and PA activities. Cut-point derived estimates were considered equivalent if the 90% confidence interval of the difference in means fell within the ±10% equivalence region.

Phase 2. Time spent in SB and PA levels derived from observation and ROC analysis generated cut-points were compared using equivalency. Subsequently, the most accurate cut-points were increased or decreased by 1m*g* progressively until cut-points providing the optimum estimates at the group-level (based on equivalency analysis) of SB, MVPA and VPA respectively were identified. Sensitivity, specificity, %Ag, and CK were re-examined for the revised cut-points and relative Bland Altman plots were produced (33).

Phase 3. In the third phase, the revised cut-points developed in phase 2 were applied to the cross-validation group. In this phase agreement and accuracy were calculated for SB, LPA, MPA, MVPA and VPA. Sensitivity, specificity, %Ag, CK were calculated and equivalency analysis was performed. Additionally, Mean absolute percent error (MAPE) was calculated as an individual-level measure of error and relative Bland Altman plots were produced.

Results

Forty-nine children (45% male;) agreed to take part in the study. Seventeen children did not complete the recess observation due to poor weather (heavy rain, icy conditions) and time constraints (data collection was restricted to December 2017). Thirty-two children (47% male;) completed all the assessments and were therefore included in the final analysis. The children who completed all the assessment included 12 children aged 5 years, 12 children aged 6 years and 8 children aged 7 years. Participant characteristics can be found in Table 3. No significant differences (p>0.05) were found between participants included and excluded from the analysis in terms of gender, decimal age, height, weight and BMI.

[TABLE 3 ABOUT HERE]

Children were video recorded during recess for an average of 7 minutes and 17 seconds (range: 3 minutes and 35 seconds to 10 minutes and 11 seconds). Table 4 presents mean ENMO, standard deviation and number of observations for each activity children engaged in during the standardised activity protocol and recess.

[TABLE 4 ABOUT HERE]

Phase 1: Cut-points calculated using the Youden and Distance methods are presented in Supplementary material 1 (see Supplementary Tables 1, 4 and 5). Most of the AUC were higher than 0.7 apart from “SB=1 and LPA=0” in the dominant wrist and hip placement with AUC equal to 0.611 and 0.689, respectively. The majority of cut-points presented higher sensitivity than specificity. Sensitivity ranged from 65.3% to 99.1% while specificity ranged from 61.8% to 96.5%. In terms of agreement, %Ag ranged from 71.5% to 95% while CK ranged from 0.43 to 0.82 representing moderate to substantial agreement.

Cut-points that included all the SB and PA levels in the ROC analysis generally presented higher AUC, higher sensitivity and lower specificity compared to the cut-points developed using pairs of activity levels. Moreover, the cut-points that included all SB and PA levels generally presented better agreement with observation for SB and lower agreement with observation for MVPA and VPA compared to cut-points developed using pairs of activity levels. Furthermore, excluding standing while watching TV from the ROC analysis resulted in an increase in AUC for SB and a decrease in the AUC for MPA and VPA ROC curves.

Based on the equivalency analysis (Figures 1-3) the cut-points developed using paired activity levels provided a better group-level estimate of time spent in SB, MVPA and VPA compared to cut-points developed using all the SB and PA levels (see CK and %Ag reported in Supplementary material 1: Supplementary Tables 1, 4 and 5). In general, Distance cut-points provided better estimates of SB, MVPA and VPA compared to Youden cut-points.

Phase 2: Results from phase 2 can be found in the Supplementary material 1 (Supplementary Tables 1-5). The cut-points providing the most comparable estimates of SB, MVPA and VPA were identified using equivalency testing (See Figures 1-3). Sensitivity, specificity, %Ag and CK observed in phase 2 cut-points were either similar or higher compared to the those observed in phase 1 meaning that cut-points developed in phase 2 obtained higher agreement with the criterion reference for SB and PA. SB cut-points demonstrated lower %Ag and CK compared to the MVPA and VPA cut-points. Based on equivalency analysis, the amount of time spent in SB, MVPA and VPA calculated using phase 2 cut-points was equivalent on average at the group level to the observed values with the exception of the SB hip accelerometer cut-point. LPA and MPA displayed lower agreement with the observed values in comparison to other PA levels. Wider limits of agreement where observed in Bland Altman plots for hip SB and LPA cut-points compared to wrist cut-points (see Supplementary material 2: Supplementary Figures 1-6). Furthermore, a linear relation between bias and average of the differences was observed in Bland Altman plots of SB (Supplementary material 2: Supplementary Figures 1-3) as children engaged in approximatively the same amount of SB (23min).

[FIGURE 1 - 2 - 3 ABOUT HERE]

Phase 3: The final cut-points developed in phase 2 were applied to the cross-validation group and the results are presented in Table 5.

[TABLE 5 ABOUT HERE]

Consistent with phase 2, SB cut-points demonstrated lower %Ag and CK compared to MVPA and VPA cut-points. LPA and MPA displayed lower agreement with the observed values in comparison to other PA levels with sensitivity between 27.4%-39.8%, specificity between 78.5%- 94.3%, %Ag between 67.5%- 87.7% and CK between 0.06-0.36. Based on the equivalency analysis, estimates were equivalent on average at the group level for SB, and MVPA for non-dominant wrist cut-points, and for SB for the dominant wrist cut-points. No estimates were considered equivalent for the hip placement. Non-dominant wrist placement showed slightly higher CK and %Ag together with lower MAPE and better results in equivalency analysis compared to hip placement in SB and LPA classification (Figure 4). Similarly, non-dominant wrist placement showed higher CK and %Ag compared to dominant wrist placement in SB and LPA classification. Wider limits of agreement were observed in Bland Altman plots for hip SB and LPA cut-points (Supplementary material 2: Supplementary Figures 16-21) compared to wrist cut-points confirming results from equivalency analysis and MAPE. In line with what observed in phase 2, a linear relation between bias and average of the differences was observed in Bland Altman plots of SB (Supplementary material 2: Supplementary Figures 16-18).

[FIGURE 4 ABOUT HERE]

Discussion

This study developed raw acceleration SB and PA cut-points in 5-7 year old children for wrist and hip worn accelerometers. SB, MPA, MVPA and VPA cut-points demonstrated adequate levels of agreement (i.e. fair to substantial CK agreement, %Ag ≥ 73%) and error (MAPE ≤ 21.6%) with the criterion reference for all Accelerometer placements. LPA measurement presented lower agreement with the criterion method compared to SB, MPA, MVPA and VPA, in line with findings observed in previous studies (34) with higher levels of error reported in hip placement (MAPE = 51.9%) compared to non-dominant (MAPE = 19.6%) and dominant placement (MAPE = 18.6%). However, the %Ag observed in this study in LPA classification was higher than the one observed in previous literature (34) suggesting that the cut-points are adequate for the use in the field. Non-dominant wrist cut-points performed slightly better than other placements in assessing SB and LPA behaviours presenting higher levels of %Ag and CK compared to both dominant wrist and hip placement together with lower levels of MAPE, better agreement in equivalency analysis and smaller confidence interval in Bland Altman plots compared to hip placements for SB and LPA. Not surprisingly, SB cut-points presented lower agreement with the criterion reference compared to MVPA and VPA cut-points confirming the known limitations of accelerometers when aiming to distinguish SB from passive standing LPA (15). This study also demonstrated that combining equivalency analysis with ROC analysis could lead to more accurate cut-points than the ones derived from ROC analysis alone, based on the higher levels of agreement observed in Phase 2 compared to Phase 1 of the statistical analysis we reported.

SB cut-points were higher at the wrist than hip placement (36m*g*, 39m*g* and 20m*g* for non-dominant wrist, dominant wrist and hip placement respectively), in line with the majority of cut-points developed in previous literature (18). However, the opposite was reported by Hildebrand et al. (15) who created SB cut-points for ActiGraph accelerometers using ENMO in a similar older age group (7-11 years old). Hildebrand et al. (15) obtained higher cut-points for the hip placement compared to wrist placement (63.3m*g* and 35.6m*g* for hip and non-dominant wrist placement, respectively). Possible reasons behind this inconsistency in hip placement cut-points could be that Hildebrand et al. (15) utilised different activities in their protocol, used the Youden method alone in the ROC analysis to identify cut-points, and involved a different criterion reference (i.e. activPAL).

Interestingly, higher sensitivity than specificity values were observed in Hildebrand et al. (15) and in our study. Hildebrand et al. (15) argued that the lower levels of specificity might be due to the inclusion of standing as LPA in the study protocol. Passive standing might lead to the absence of registered accelerations or low accelerations similar to SB activities. Despite being classified as LPA based on energy expenditure and/or the posture, standing watching TV does not necessarily involve movement and therefore could be classified as passive standing (35). Previous research has demonstrated the limitations of accelerometers in distinguishing stationary behaviours such as passive standing from SB (20, 36). Another limitation of SB assessment using cut-points in is the lack of consideration of posture that is a key aspect of SB identification (37). This is confirmed by the results of our study where the mean acceleration during passive standing (Table 4) was below the SB cut-points.

SB raw acceleration cut-points have been developed by Schaefer et al. (34) and Duncan et al. (38) in GENEActiv devices for children aged between 5-7, though, rather than using ENMO these studies utilised different metrics to represent acceleration signals. SB cut-point presented in both Schaefer et al. (34) and Duncan et al. (38) studies were higher than SB cut-points developed in this study (36mg, 39mg, 20mg) with values of 190m*g* and 75m*g* (converted from time to independent unit m*g*) respectively. This is in line with previous studies where higher accelerations were observed in GENEActiv compared to ActiGraph when measuring the same participants simultaneously (39). However, key reasons for the disparity in cut-points is likely due to the different metrics that have been used to represent the acceleration meaning cut-points are not directly comparable (18).

Hildebrand et al. developed MVPA and VPA cut-points for ActiGraph using ENMO in 7-11 year old children. Their reported cut-points were higher for both wrist (MVPA: 201.4mg, VPA: 707.0mg) and hip (MVPA: 142.6mg, VPA: 464.6mg) placements compared to the ones in our study (MVPA: 189mg for non-dominant wrist, 181mg for dominant wrist and 95mg for hip; VPA: 536mg for non-dominant wrist, 534mg for dominant wrist and 325mg for hip) (Table 5). There are several potential reasons for the differences between the Hildebrand cut-points and the ones reported in the present study. For example, the difference in age range between the participants involved, the use of indirect calorimetry as criterion reference rather than observation, using linear regression for cut-points identification and the use of different activities within the study protocol (14). Van Loo et al. (40) assessed the accuracy of three sets of MVPA and VPA raw accelerometers cut-points developed by Hildebrand et al. (15) Philips et al. (41) and Schaefer et al. (34) for GENEActiv wrist mounted devices in 5-8 year old children and found that these cut-points led to considerable misclassification of PA levels. Interestingly, none of the cut-points examined by van Loo et al. (40) were originally developed from a sample of 5-8 years old children (15, 34, 41) and therefore it is possible that they were not adequate for the classification of MPA, MVPA and VPA in that age group.

When considering previous studies that examined raw acceleration cut-points in 5-7 year old children, only Schaefer et al. (34), Hildebrand et al. (14) and Van Loo et al. (40) reported %Ag. Schaefer et al. (34) reported slightly higher %Ag for the SB cut-point (83.3%) but lower %Ag for LPA (29.4%), MPA (41%) and VPA (88.7%) compared to our study (%Ag in this study: SB between 73% and 78.5%, LPA between 67.5% and 62.5%, MPA between 88.7% and 88.2%, VPA between 92 and 93.8%). Similarly, Hildebrand et al. (14) and Van Loo et al. (40) obtained lower %Ag for MPA and VPA (%Ag for Hildebrand et. (14): MPA between 33% and 55%, VPA between 68% and 80%; %Ag for Val Loo et al. (40): MPA between 45.4% and 52%, VPA between 70% and 93.6%). In this study according to Cohen's Kappa values, LPA estimates demonstrated slight agreement, while MPA estimates showed fair agreement, and SB, MVPA and VPA moderate to substantial agreement. Given that no previous calibration studies in this age group have reported CK, we suggest that future studies should include this measure of reliability to account for chance agreements. Overall, the %Ag reported in this study is higher than those observed in previous studies applying raw acceleration cut-points in 5-7-year-old children, demonstrating that the cut-points proposed in this study could lead to improved accuracy in PA assessment.

A major strength of this calibration study was its high ecological validity as the protocol included direct observation of children's SB and PA during recess within the school playground and during a standardised protocol of activities performed in their physical education hall. Additionally, this is the first accelerometer calibration study in this age group to consider different methods of cut-point calculation, including: i) exploring the use of paired activity levels in ROC curve analysis, ii) examining the Youden and distance methods for cut-point development, and iii) using equivalency methods to identify and refine cut-points. Further strengths are the use of the ENMO metric, emerging as the most frequently used metric to process raw acceleration and generate thresholds for multiple accelerometer placements (42).

Despite the advantages of using direct observation as criterion reference for SB and PA assessment exposed in our methods section, we acknowledge that direct observation is not the gold standard for the measurements of energy expenditure and presents a level of subjectivity. Furthermore, because of time constraints and participants' availability, it was not possible for the all the initial 49 participants to complete the study protocol and to obtain a balanced number of children within each age group involved in the study (12 children aged 5 years, 12 children aged 6 years, and 8 children aged 7 years). We recognise that the limited number of children in the cut-point generation group together with the use of statistical analysis methods maximizing accuracy might lead to over fitting related problems. For future calibration studies, we suggest involving an equal number of participants in each age group to guarantee that each age is equally represented in the sample, together with a bigger sample size to guarantee a better representation of the population. In line with previous research, we encountered difficulties in the selection of standardised LPA activities for the testing protocol. Similar to previous studies (15, 40, 43), we classified slow walking and standing as LPA. Given that passive standing might lead to misclassification of SB and LPA, other activities that are representative of 5-7 years old children free-living LPA should be identified in the future. Moreover, future studies should examine methods to integrate postural aspects to the measurement to account for accelerometers limitations in classifying sedentary behaviours.

Conclusions

SB, LPA, MPA, MVPA and VPA cut-points demonstrated adequate accuracy in all accelerometer placements. Non-dominant accelerometer placement presented slightly better agreement with the criterion reference compared to the dominant wrist and hip placements for SB and LPA. However, no other differences were highlighted between the accelerometer placement. These findings can be used to inform the decisions made by researchers in relation to the assessment of young children's PA and SB. Furthermore, the study protocol, methods and analysis can inform the development of more rigorous calibration studies and subsequent analyses to determine cut-points in the future. In view of our results, we suggest that cut-points developed using Youden method involving all SB and PA levels in ROC analysis can lead to large misclassification of SB and PA levels. Future researchers should include paired activity levels analysis together with distance method in ROC analysis in combination with equivalency analysis and Cohen's Kappa statistic to select the most accurate SB and PA cut-points.

Acknowledgments

The authors would like to thank the school that agreed to take part in the study and provided the space where the data collection for this study took place. Furthermore, the authors would like to thank the children who agreed to take part in the study and the researchers who contributed in the data collection.

Conflict of interest

No conflict of interest was reported between authors and other people involved in the study.

References

1. Atkin AJ, Gorely T, Clemes SA, Yates T, Edwardson C, Brage S, et al. Methods of Measurement in epidemiology: sedentary Behaviour. Int J Epidemiol. 2012;41(5):1460-71.
2. Cain KL, Sallis JF, Conway TL, Van Dyck D, Calhoon L. Using accelerometers in youth physical activity studies: a review of methods. J Phys Act Health. 2013;10(3):437-50.
3. Trost SG, McIver KL, Pate RR. Conducting accelerometer-based activity assessments in field-based research. Med Sci Sports Exerc. 2005;37(11 Suppl):S531-43.
4. Fairclough SJ, Noonan R, Rowlands AV, Van Hees V, Knowles Z, Boddy LM. Wear Compliance and Activity in Children Wearing Wrist- and Hip-Mounted Accelerometers. Med Sci Sports Exerc. 2016;48(2):245-53.
5. Tudor-Locke C, Barreira TV, Schuna JM, Mire EF, Chaput J-P, Fogelholm M, et al. Improving wear time compliance with a 24-hour waist-worn accelerometer protocol in the International Study of Childhood Obesity, Lifestyle and the Environment (ISCOLE). International Journal of Behavioral Nutrition and Physical Activity. 2015;12(1):11.
6. McLellan G, Arthur R, Buchan DS. Wear compliance, sedentary behaviour and activity in free-living children from hip-and wrist-mounted ActiGraph GT3X+ accelerometers. J Sports Sci. 2018;36(21):2424-30.
7. Morgenthaler T, Alessi C, Friedman L, Owens J, Kapur V, Boehlecke B, et al. Practice parameters for the use of actigraphy in the assessment of sleep and sleep disorders: an update for 2007. Sleep. 2007;30(4):519-29.

8. Fairclough SJ, Dumuid D, Taylor S, Curry W, McGrane B, Stratton G, et al. Fitness, fatness and the reallocation of time between children's daily movement behaviours: an analysis of compositional data. international journal of behavioral nutrition and physical activity. 2017;14(1):64.

9. Troiano RP, McClain JJ, Brychta RJ, Chen KY. Evolution of accelerometer methods for physical activity research. Br J Sports Med. 2014;48(13):1019-23.

10. Chen KY, Bassett DR, Jr. The technology of accelerometry-based activity monitors: current and future. Med Sci Sports Exerc. 2005;37(11 Suppl):S490-500.

11. van Hees VT, Thaler-Kall K, Wolf KH, Brond JC, Bonomi A, Schulze M, et al. Challenges and Opportunities for Harmonizing Research Methodology: Raw Accelerometry. Methods Inf Med. 2016;55(6):525-32.

12. Bakrania K, Yates T, Rowlands AV, Esliger DW, Bunnewell S, Sanders J, et al. Intensity Thresholds on Raw Acceleration Data: Euclidean Norm Minus One (ENMO) and Mean Amplitude Deviation (MAD) Approaches. PLoS One. 2016;11(10):e0164045.

13. Migueles JH, Rowlands AV, Huber F, Sabia S, van Hees VT. GGIR: A Research Community–Driven Open Source R Package for Generating Physical Activity and Sleep Outcomes From Multi-Day Raw Accelerometer Data. Journal for the Measurement of Physical Behaviour. 2019;2(3):188-96.

14. Hildebrand M, VT VANH, Hansen BH, Ekelund U. Age group comparability of raw accelerometer output from wrist- and hip-worn monitors. Med Sci Sports Exerc. 2014;46(9):1816-24.

15. Hildebrand M, Hansen BH, van Hees VT, Ekelund U. Evaluation of raw acceleration sedentary thresholds in children and adults. Scand J Med Sci Sports. 2017;27(12):1814-23.

16. Hausdorff J, Zemany L, Peng C-K, Goldberger A. Maturation of gait dynamics: stride-to-stride variability and its temporal organization in children. Journal of applied physiology. 1999;86(3):1040-7.

17. Bailey RC, Olson J, Pepper SL, Porszasz J, Barstow TJ, Cooper DM. The level and tempo of children's physical activities: an observational study. Med Sci Sports Exerc. 1995;27(7):1033-41.

18. de Almeida Mendes M, da Silva IC, Ramires VV, Reichert FF, Martins RC, Tomasi E. Calibration of raw accelerometer data to measure physical activity: A systematic review. Gait & posture. 2018;61:98-110.

19. Crouter SE, Flynn JI, Bassett DR, Jr. Estimating physical activity in youth using a wrist accelerometer. Med Sci Sports Exerc. 2015;47(5):944-51.

20. Aguilar-Farias N, Brown WJ, Peeters GM. ActiGraph GT3X+ cut-points for identifying sedentary behaviour in older adults in free-living environments. J Sci Med Sport. 2014;17(3):293-9.

21. Zhou X-H, Obuchowski NA, McClish DK. Statistical Methods in Diagnostic Medicine: John Wiley & Sons; 2011. 477 p.

22. Dettwyler KA. Anthropometric standardization reference manual, abridged edition. Edited by Timothy G. Lohman, Alex F. Roche, and Reynaldo Martoll. Champaign, Illinois: Human Kinetic Books. 1991. American Journal of Physical Anthropology. 1993;92(2):239-41.

23. Johansson E, Larisch LM, Marcus C, Hagstromer M. Calibration and Validation of a Wrist- and Hip-Worn Actigraph Accelerometer in 4-Year-Old Children. PLoS One. 2016;11(9):e0162436.

24. Mackintosh KA, Fairclough SJ, Stratton G, Ridgers ND. A calibration protocol for population-specific accelerometer cut-points in children. PLoS One. 2012;7(5):e36919.

25. Butte NF, Watson KB, Ridley K, Zakeri IF, McMurray RG, Pfeiffer KA, et al. A Youth Compendium of Physical Activities: Activity Codes and Metabolic Intensities. Med Sci Sports Exerc. 2018;50(2):246-56.

26. Saint-Maurice PF, Kim Y, Welk GJ, Gaesser GA. Kids are not little adults: what MET threshold captures sedentary behavior in children? Eur J Appl Physiol. 2016;116(1):29-38.

27. Baquet G, Ridgers ND, Blaes A, Aucouturier J, Van Praagh E, Berthoin S. Objectively assessed recess physical activity in girls and boys from high and low socioeconomic backgrounds. BMC Public Health. 2014;14:192.

28. Obuchowski NA, Bullen JA. Receiver operating characteristic (ROC) curves: review of methods with applications in diagnostic medicine. Phys Med Biol. 2018;63(7):07tr1.

29. Perkins Neil J, Schisterman Enrique F. The Inconsistency of “Optimal” Cut-points Using Two ROC Based Criteria. Am J Epidemiol. 2006;163:670-5.

30. Landis JR, Koch GG. The measurement of observer agreement for categorical data. biometrics. 1977:159-74.

31. Dixon PM, Saint-Maurice PF, Kim Y, Hibbing P, Bai Y, Welk GJ. A Primer on the Use of Equivalence Testing for Evaluating Measurement Agreement. Med Sci Sports Exerc. 2018;50(4):837-45.

32. DeShaw KJ, Ellingson L, Bai Y, Lansing J, Perez M, Welk G. Methods for activity monitor validation studies: an example with the fitbit charge. Journal for the Measurement of Physical Behaviour. 2018;1(3):130-5.

33. Bland JM, Altman D. Statistical methods for assessing agreement between two methods of clinical measurement. The lancet. 1986;327(8476):307-10.

34. Schaefer CA, Nigg CR, Hill JO, Brink LA, Browning RC. Establishing and evaluating wrist cutpoints for the GENEActiv accelerometer in youth. Med Sci Sports Exerc. 2014;46(4):826-33.

35. Tremblay MS, Aubert S, Barnes JD, Saunders TJ, Carson V, Latimer-Cheung AE, et al. Sedentary behavior research network (SBRN)–terminology consensus project process and outcome. International Journal of Behavioral Nutrition and Physical Activity. 2017;14(1):75.

36. Ridgers ND, Salmon J, Ridley K, O'Connell E, Arundell L, Timperio A. Agreement between activPAL and ActiGraph for assessing children's sedentary time. Int J Behav Nutr Phys Act. 2012;9:15.

37. Rowlands AV, Yates T, Olds TS, Davies M, Khunti K, Edwardson CL. Wrist-Worn Accelerometer-Brand Independent Posture Classification. Med Sci Sports Exerc. 2016;48(4):748-54.

38. Duncan MJ, Wilson S, Tallis J, Eyre E. Validation of the Phillips et al. GENEActiv accelerometer wrist cut-points in children aged 5–8 years old. European journal of pediatrics. 2016;175(12):2019-21.

39. Rowlands AV, Mirkes EM, Yates T, Clemes S, Davies M, Khunti K, et al. Accelerometer-assessed Physical Activity in Epidemiology: Are Monitors Equivalent? Med Sci Sports Exerc. 2018;50(2):257-65.

40. van Loo CM, Okely AD, Batterham MJ, Hinkley T, Ekelund U, Brage S, et al. Wrist acceleration cut-points for moderate-to-vigorous physical activity in youth. Med Sci Sports Exerc. 2018;50(3):609.

41. Phillips LR, Parfitt G, Rowlands AV. Calibration of the GENEA accelerometer for assessment of physical activity intensity in children. J Sci Med Sport. 2013;16(2):124-8.

42. Welk GJ. Harmonizing Monitor-and Report-Based Estimates of Physical Activity Through Calibration. Kinesiology Review. 2019;8(1):16-24.

43. van Loo CM, Okely AD, Batterham MJ, Hinkley T, Ekelund U, Brage S, et al. Wrist accelerometer cut-points for classifying sedentary behavior in children. Med Sci Sports Exerc. 2017;49(4):813.

Table 1: Standardised activity protocol

Sedentary behaviours	
Lying while watching TV	Lie comfortably on a mat while watching an age appropriate television programme or movie for 10 minutes.
Sitting while colouring	Colouring exercise while sitting at a table for 5 minutes.
Sitting playing with a tablet	Play games on a tablet while sitting on a chair for 5 minutes.
Playing with LEGO	Sit or lie on the floor while playing with Lego for 5 minutes.
Light physical activity	
Standing while watching TV	Stand and watch a video for 5 minutes.
Moderate physical activity	
Walking briskly self-paced	Walk briskly for 2 minutes, at a self-selected pace around a designated track or circuit. A researcher walked with the child encouraging him/her to maintain the pace.
Throwing and catching	Child and researcher passed the ball to each other continuously for 2 minutes.
Vigorous physical activity	
Running	Run for 2 minutes, at a self-selected pace around a designated track or circuit.
Obstacle course	Run for 2 minutes on a course around cones. This course was designed to mimic typical run/chase type activities and involved slalom, dodging tasks and fast changes of direction.
Hopping	Complete a hopscotch course for 2 minutes.

Table 2. Dichotomization of the data for the ROC analysis

Sedentary	
"1"	**"0"**
SB	LPA, MPA, VPA.
SB	LPA excluding standing while watching TV, MPA, VPA
SB	LPA
SB	LPA excluding standing while watching TV
Moderate physical activity	
"1"	**"0"**
MPA,VPA	SB, LPA
MPA,VPA	SB, LPA excluding standing watching TV
MPA	LPA
MPA	LPA excluding standing watching TV
Vigorous physical activity	
"1"	**"0"**
VPA	SB, LPA, MPA.
VPA	SB, LPA excluding standing watching TV, MPA
VPA	MPA

Scored "**1**" when the condition is present; Scored "**0**" when the condition is absent; **SB:** Sedentary behaviours; **LPA:** Light physical activity; **MPA:** Moderate physical activity; **VPA:** Vigorous physical activity.

Table 3. Participants' descriptive data

Initial group (n=49)	Males (n=22)		Females (n=27)	
	Mean	**SD**	**Mean**	**SD**
Decimal age (years)	6.5	0.8	6.5	0.7
Height (cm)	120.2	6.7	120.4	9.0
Weight (Kg)	23.6	3.9	24.4	6.1
BMI (Kg/m^2)	16.3	1.8	16.6	2.1

Final group (n=32)	Males (n=15)		Females (n=17)	
	Mean	**SD**	**Mean**	**SD**
Decimal age (years)	6.4	0.8	6.4	0.7
Height (cm)	119.4	6.3	120.2	9.5
Weight (Kg)	23.3	4.2	24.2	7.0
BMI (Kg/m^2)	16.2	2.0	16.5	2.5

Table 4. Accelerations observed in each SB and PA level recorded

				Non-		**Dominant**		**Hip**	
Intensity (MET)	**Standardised Protocol**	**MET**	**Obs (s)**	**Mean (m*g*)**	**SD (m*g*)**	**Mean (m*g*)**	**SD (m*g*)**	**Mean (m*g*)**	**SD (m*g*)**
Sedentary	Lying while watching TV	1.2	18155	17	37	15	37	12	14
	Sitting while colouring	1.6	8640	20	47	37	65	11	13
	Sitting and playing with a tablet	1.4	8640	11	21	23	28	9	12
	Playing with LEGO	1.5	8640	52	48	51	47	11	12
Light	Standing	1.7	8640	20	39	12	27	9	13
Moderate	Walking briskly self-paced	4.6	2880	294	289	255	271	178	100
	Throw and catch	4.9	2790	444	370	432	374	83	88
Vigorous	Running	7.8	2865	1071	581	1115	601	607	179
	Obstacle course	7.2	2880	744	424	719	396	446	165
	Hopping	6.3	2563	844	552	762	491	452	241
	Recess								
Sedentary	Sitting down	1.4	51	64	64	67	80	18	27
Light	Standing	1.7	3007	103	165	117	210	45	88
	Walk slow	2.5	6164	204	249	207	266	120	128
Moderate	Walk brisk	4.6	665	528	397	473	398	336	196
	Jog slow	5.5	1364	652	459	644	537	434	259
	Dancing	3.6	13	654	557	347	340	162	126
	Ball games	6.0	23	773	337	652	379	379	189

	Jumping-jack	5.9	107	931	463	1081	449	281	247
Vigorous	Jog fast	6.8	1178	1103	632	1032	688	599	290
	Running	7.8	510	1772	894	1766	999	808	254
	Hopping	6.3	437	883	537	782	575	528	259
	Jump rope	6.9	577	801	390	1140	456	649	241
	Ball games	6.1	75	1663	696	1347	633	604	204

Obs: Number of observation of each behaviours where each observation corresponds to 1 second spent in the activity observed.

MET: Metabolic equivalent (1 MET equals the oxygen uptake of $3.5 mL \cdot Kg^{-1} \cdot min^{-1}$)

Table 5. Cut-points performance in cross-validation group

							Equivalency analysis derived mean and confidence interval	
	Cut-point (m*g*)	**Sn (%)**	**Sp (%)**	**CK (a.u)**	**%Ag (%)**	**MAPE (%)**	**Obs (min)**	**Cut-point (min)**
Non-dominant wrist								
SB	<36	79.8	56.8	0.57	78.5	9.3	23.0±2.3	22.8±1.4
LPA	≥36&<189	38.4	81.9	0.20	72.5	19.6	9.1±0.9	9.5±1.2
MPA	≥189&<536	39.0	93.7	0.34	87.7	19.0	4.7±0.5	4.2±0.6
MVPA	≥189	82.6	78.0	0.78	92.0	9.0	10.2±1.0	10±0.8
VPA	≥536	75.1	68.7	0.69	92.7	12.9	5.5±0.6	5.9±0.5
Dominant wrist								
SB	<39	75.4	70.2	0.46	73.0	10.1	23.0±2.3	23.1±1.7
LPA	≥39&<181	27.4	78.4	0.06	67.5	18.7	9.1±0.9	9.6±1.2
MPA	≥181&<534	39.8	93.5	0.35	87.7	14.4	4.7±0.5	4.3±0.5
MVPA	≥181	79.1	76.0	0.76	91.4	13.5	10.2±1.0	9.5±1.0
VPA	≥534	67.6	95.6	0.64	92.0	16.2	5.5±0.6	5.3±0.7
Hip								
SB	<20	78.0	50.1	0.50	75.3	21.2	23.0±2.3	23.3±3.1
LPA	≥20&<95	30.0	80.2	0.10	69.4	51.9	9.1±0.9	9.3±3.0
MPA	≥95&<325	39.1	94.3	0.36	88.2	21.6	4.7±0.5	4±0.7
MVPA	≥95	79.3	75.6	0.76	91.2	13.2	10.2±1.0	9.7±1.0
VPA	≥325	78.2	96.1	0.73	93.8	11.3	5.5±0.6	5.7±0.4

SB: Sedentary behaviours; **LPA:** Light physical activity; **MPA:** Moderate physical activity; **MVPA:** moderate to vigorous physical activity; **VPA:** Vigorous physical activity; **Sn:** Sensitivity; **Sp:** Specificity; **CK:** Cohen's Kappa; **%Ag:** Percentage of agreement. **MAPE:** mean absolute percent error; **a.u.:** Arbitrary units; **Obs:** Concerns the mean time spent in SB and PA levels obtained by observation ±10% of the mean time spent in a specific activity level derived from observation; **Cut-point:** Concerns the mean of the cut-points derived SB and PA levels and the related 90% confidence interval of the difference between observed and cut-point derived minutes spent in a specific activity level.

Figure 1. Non-dominant wrist equivalency analysis in Cut-point generation group (Phase 1-2)

[FIGURE 1 ABOUT HERE]

*: the cut-points marked with a * were calculated using ROC analysis Youden method.
#: the cut-points marked with a # were calculated using ROC analysis Distance method.
Phase 2: the cut points in Phase 2 was calculated using equivalency analysis method.
Solid line: The solid line concerns the 90% confidence interval of the difference between observed and cut-point derived minutes spent in a specific activity level. The confidence interval is centred on the mean of the cut-point derived time estimate of the activity level taken into consideration (i.e. SB, MVPA, VPA).
Dashed line: The dashed line concerns the ±10% interval of the mean time estimate of a specific activity level calculated using observation. The ±10% interval is centred on the mean of the observation derived time estimate of the activity level taken into consideration (i.e. SB, MVPA, VPA).

Figure 2. Dominant wrist equivalency analysis in Cut-point generation group (Phase 1-2)

[FIGURE 2 ABOUT HERE]

*: the cut-points marked with a * were calculated using ROC analysis Youden method.
#: the cut-points marked with a # were calculated using ROC analysis Distance method.
Phase 2: the cut points in Phase 2 was calculated using equivalency analysis method.
Solid line: The solid line concerns the 90% confidence interval of the difference between observed and cut-point derived minutes spent in a specific activity level. The confidence interval is centred on the mean of the cut-point derived time estimate of the activity level taken into consideration (i.e. SB, MVPA, VPA).
Dashed line: The dashed line concerns the ±10% interval of the mean time estimate of a specific activity level calculated using observation. The ±10% interval is centred on the mean of the observation derived time estimate of the activity level taken into consideration (i.e. SB, MVPA, VPA).

Figure 3. Hip equivalency analysis in Cut-point generation group (Phase 1-2)

[FIGURE 3 ABOUT HERE]

*: the cut-points marked with a * were calculated using ROC analysis Youden method.
#: the cut-points marked with a # were calculated using ROC analysis Distance method.
Phase 2: the cut points in Phase 2 was calculated using equivalency analysis method.
Solid line: The solid line concerns the 90% confidence interval of the difference between observed and cut-point derived minutes spent in a specific activity level. The confidence interval is centred on the mean of the cut-point derived time estimate of the activity level taken into consideration (i.e. SB, MVPA, VPA).
Dashed line: The dashed line concerns the ±10% interval of the mean time estimate of a specific activity level calculated using observation. The ±10% interval is centred on the mean of the observation derived time estimate of the activity level taken into consideration (i.e. SB, MVPA, VPA).

Figure 4. Standard confidence interval test in cross validation group (Phase 3)

[FIGURE 4 ABOUT HERE]

SB: Sedentary behaviours; **LPA:** Light physical activity; **MPA:** Moderate physical activity; **MVPA:** moderate to vigorous physical activity; **VPA:** Vigorous physical activity.
Solid line: The solid line concerns the 90% confidence interval of the difference between observed and cut-point derived minutes spent in a specific activity level. The confidence interval is centred on the mean of the cut-point derived time estimate of the activity level taken into consideration (i.e. SB, LPA, MPA, MVPA, VPA).
Dashed line: The dashed line concerns the ±10% interval of the mean time estimate of a specific activity level calculated using observation. The ±10% interval is centred on the mean of the observation derived time estimate of the activity level taken into consideration (i.e. SB, LPA, MPA, MVPA, VPA).