Title: Simultaneous Validation of Count-to-Activity Thresholds for Five Commonly Used Activity Monitors in Adolescent Research: A Step Towards Data Harmonisation

Published in: Journal for the Measurement of Physical Behaviour

Reference: Hayes, G., Dowd, K., MacDonncha, C. and Donnely, A., 2021. Simultaneous Validation of Count-to-Activity Thresholds for Five Commonly Used Activity Monitors in Adolescent Research: A Step Toward Data Harmonization. Journal for the Measurement of Physical Behaviour, 4(4), pp.333-342.

DOI: https://doi.org/10.1123/jmpb.2021-0023.

Abstract

Background: Multiple activity monitors are utilised for the estimation of moderate-to-vigorous intensity physical activity (MVPA) in youth. Due to differing methodological approaches, results are not comparable when developing thresholds for the determination of MVPA. This study aimed to develop and validate count-to-activity thresholds for 1.5, 3 and 6 METs in five of the most commonly used activity monitors in adolescent research. **Methods:** Fifty-two participants (mean age 16.1 ± 0.78 years) selected and performed activities of daily living while wearing a CosMED K4b^2 and five activity monitors; ActiGraph GT1M, ActiGraph wGT3X-BT, activPAL3 Micro, activPAL, and GENEActiv. Receiver operating characteristic analysis was used to examine the area under the curve and to define count-to-activity thresholds for the vertical axis (all monitors) and the sum of the vector magnitude (ActiGraph wGT3X-BT, activPAL3 Micro) for 15 second (all monitors) and 60 second (ActiGraph monitors) epochs. **Results:** All developed count-to-activity thresholds demonstrated high levels of sensitivity and specificity. When cross-validated in an independent group (N = 20), high levels of sensitivity and specificity generally remained (≥ 73.1%, intensity and monitor dependent).

Conclusion: This study provides researchers with the opportunity to analyse and cross-compare data from different studies that have not employed the same motion sensors.

Key Words

Accelerometer, count-to-activity thresholds, METs, physical activity, validation, adolescents.

Background

Accelerometer-based activity monitors have become the preferred method of choice for measuring activity behaviour in free-living populations (Dowd et al., 2018). Most accelerometer-based devices record and store raw accelerations, however, different companies vary on how they filter and process the data. Generally, once exported, the output from the accelerometer is presented in arbitrary units (referred to as activity or accelerometer counts) over a specified unit/epoch of time (e.g. 60 seconds). The accelerometer counts provide an indication of overall movement, but the fundamental research challenge is to determine how counts can be converted into more meaningful units. This challenge is usually addressed in calibration studies where the accelerometer counts are related to either energy expenditure, oxygen consumption or metabolic equivalents (METs) to give a more interpretable measure of physical activity (PA) intensity (Harrington et al., 2011).

As accelerometer-based monitors have evolved, so too has the amount of studies validating their use as measures of PA intensity. Controversy now exists when analysing PA data, as researchers must choose from multiple devices and multiple sets of count-to-activity thresholds for sedentary, light, moderate and vigorous intensity PA (Powell et al., 2017). The use of multiple devices, coupled with varying thresholds developed in specific populations (based on age and gender) and differing validation analysis techniques makes it difficult to cross-compare the results between devices or studies. For example, Van Hecke and colleagues highlight that across the youth PA literature, five different count-to-activity thresholds ranging from >1000 counts.min^{-1} to >3000 counts.min^{-1} were used to define MVPA measured with accelerometers in children (Van Hecke et al., 2016). Similarly, four different count-to-activity thresholds were used in adolescents ranging from >1500 counts.min^{-1} to >3000 counts.min^{-1}. The increased number of count-to-activity thresholds used within youth PA research is a major factor in the variation in reported levels of MVPA across youth populations (Van Hecke et al., 2016). Moreover, harmonising activity measurements from different studies is difficult since the majority of validation studies have been completed independently, using different validation methods and activity protocols (Powell et al., 2017).

Using the same validation methods, activity protocols and statistical analysis techniques, the aim of this study was to simultaneously develop and validate count-to-activity thresholds for 1.5, 3 and 6 MET activities for five of the most commonly used activity monitors in adolescent research (activPAL, ActiGraph wGT3x-BT, ActiGraph GT1M & GENEActiv) using VO_2 determined from a potable metabolic unit as the criterion. The thresholds developed for each of the included monitors will enable researchers to reprocess and compare data that was collected using different activity monitors potentially enabling the harmonisation of data with greater accuracy, thus providing greater knowledge of the relationship between activity behaviours and indices of health in adolescent populations.

Methods

Participant Eligibility and Recruitment

A convenience sample of 62 adolescent males and females were recruited from two second-level schools in the [redacted]. Eligible participants were required to be between the ages of 15-18 years and free from injury or any medical condition that would constrain their participation in PA or exercise. No exclusions were placed on the participant's level of fitness or other health behaviours. Once written informed consent was obtained from the school principals, all participants were provided with parental and participant information sheets, consent forms and a physical activity readiness questionnaire (PAR-Q). Participants were selected for inclusion once participant assent and parent consent were provided and they had successfully completed the PAR-Q. A lowest random numbers table was used to assign each participant to either an equation development group or a cross-validation group. Ethics committee approval was granted by the Faculty of Education and Health Sciences Research Ethics Committee at the [redacted].

Activity Measurement Devices

During testing procedures, participants were required to wear five activity monitors. The characteristics of each activity monitor are described in Table 1. To reduce any potential inter-device error, the same activity monitors were used throughout the entire testing period. The alignment of data was achieved by synchronising the internal clock of each activity monitor with the same computer. All monitors were set to record with 15 second epochs. To enable a more accurate detection of acceleration during the lower intensity activities (Migueles et al., 2017), the low-frequency extension filter was applied to the ActiGraph wGT3X-BT data.

Metabolic Testing

Breath-by-breath oxygen uptake (VO_2, $ml.kg.min^{-1}$) and carbon dioxide production (VCO_2, $ml.kg.min^{-1}$) was recorded in 30 second averages for each activity using the CosMed $K4b^2$ portable metabolic unit ($K4b^2$) (CosMed, Rome, Italy). The device has been shown to be a reliable measure of oxygen uptake over a range of exercise intensities (Bassett et al., 2012; McLaughlin et al., 2001). Briefly, the $K4b^2$ is a battery operated, portable, wireless device worn on the chest via a harness with a heart-rate receiver attached. Each participant was fitted with a rubber facemask (Hans-Rudolph, Kansas City, USA) affixed via a head harness. Prior to each testing session and following a standard 45-minute warm-up period, the $K4b^2$ was calibrated immediately in accordance with the specifications of the manufacturer. Flow control and gas calibration were performed using the $K4b^2$ automated calibration procedures and the CO_2 and O_2 analysers were calibrated against a reference gas of known concentrations (4.01% CO_2 and 16 % O_2) as well as room air. The output from each accelerometer was aligned with the steady state VO_2 for each activity performed. Resting metabolic rate (RMR) was measured for each participant, which enabled the intensity of each activity to be individualised per participant.

The measurement of individual RMR was appropriate as use of the standard 1 MET as 3.5 $ml.kg.min^{-1}$ has been shown to have limitations for calculating metabolic rate (Mansoubi et al., 2015) and it does not represent the RMR of younger populations (Butte et al., 2018). For this study, the measured value for each individual's RMR was assigned to be 1 MET, and all other measured values were expressed relative to this.

Calibration activities

The selection of the activities included in this research was predominantly informed by previous research that highlights best practice for wearable monitor calibration and validation studies (Bassett et al., 2012; Welk, 2005). Prior to testing, each participant was provided with a bank of 20 exercises that had been used in previously published accelerometer validation research that were categorised based on the compendium of PA and ranged in intensity from sedentary to vigorous (Ainsworth et al., 2011). The participants were required to rank in order how regularly they achieved each activity on a daily basis. Using this information, the main researcher assigned activities to each participant for their testing period. The random selection of the activities included aimed to mimic a more free-living natural environment. Each participant performed a maximum of eight standardised activities (see Supplementary Table 1 for activities) across four intensity categories (sedentary, light, moderate and vigorous). Sedentary and light intensity activities were performed for a duration of 5 minutes, while moderate- and vigorous-intensity activities were performed for 7 minutes. These durations were selected as VO_2 should reach steady state after 3 minutes for light intensity activities and after 3-5 minutes for more intense exercise (Poole et al., 1991). The attainment of steady state was confirmed by inspection of recorded HR and VO_2 values (Trost et al., 2005). The first 2 minutes (to allow for steady state) and the last minute (subjects stopped due to volitional fatigue) of each activity were eliminated. The mean value of the 3^{rd} and 4^{th} minute of the sedentary and light activity and the mean value of the 5^{th} and 6^{th} minute of the moderate and vigorous activities was used for this analysis as participants were at steady state energy expenditure during this time (Saint-Maurice et al., 2016).

Testing Protocol

Participants were asked to refrain from eating, consuming caffeine and smoking for a minimum of 3 hours prior to attending the testing centre. In the 12 hours preceding testing, participants were asked to refrain from any structured PA or exercise. Participants attended the centre wearing gym clothing and running shoes. Upon arrival, height was measured to the nearest 0.25 cm, using a portable stadiometer (Seca model 214; Seca Ltd, Birmingham, UK) and mass was measured to the nearest 0.1kg using a portable electronic scale (Seca model 770; Seca Ltd, Birmingham, UK). Body mass index (BMI) was recorded using the standard formula (Kg/m^{-2}).

Once anthropometric data were obtained, the K4 b^2 metabolic unit and all activity monitors were fitted on participants (Figure 1) in accordance with the manufacture's guidelines (Table 1). The GENEActiv devices were placed on both the right and left wrists as hand-dominance was/is not specified by the manufacturer. All activities were performed in ascending intensity. A single observer recorded and signalled to the participant when to start and stop each activity.

To initiate the study protocol participants were required to lie in a reclined position in a dimly lit, quiet room for 10 minutes before having their resting VO_2 measured for a 15-minute period. Participants were not permitted to sleep during this time and the 10-minute period was deemed an appropriate time frame to ensure that participants were in a fully rested state (Treuth et al., 2004). To determine individual RMR only minutes 5-15 of the 15-minute measurement period were included for analysis. The output from each accelerometer was aligned with the steady state VO_2 for each activity performed. To facilitate the assessment of the included ambulatory activities (i.e. walking and running), a 40-meter circular track was constructed within the testing centre. Participants were required to complete some of the ambulatory activities within specified time limits, but at a pace that was comfortable for them. The activities with specified time limits included; slow walking (2.5-4.5 $km.hr^{-1}$), brisk walking (4.5-6.5 $km.hr^{-1}$) and running (6.5-8.5$km.hr^{-1}$). Within each speed category, the upper- and lower-time limit required to complete one full circle of the 40-meter track was calculated. The time taken for a participant to complete one full circle of the track was used to estimate their actual speed. To standardise pace and movement, once the participant was comfortable at the self-selected speed and achieved consistent timing within the speed bands, they were cued to maintain that pace for the remainder of the activity and where possible a research student completed the activity with the participant. The self-selected pacing approach was used to mimic free-living activity and to avoid the clustering effect that set speeds may have during statistical analysis. A rest period of 2 minutes was permitted between the more metabolically demanding activities or longer if the heart rate had not returned to < 100 beats per minute after 1 minute. A brief description of the included ambulatory activities included can be observed in Supplementary Table 1.

Data Processing

After accelerometer files were downloaded, activity counts for each 15 second epoch were provided using each of the monitors respective proprietary software's (see Table 1 for proprietary software). Accelerations in the vertical axis were extracted for uniaxial accelerometers (activPAL and ActiGraph GT1M). Accelerations in the vertical axis and the sum of the vector magnitudes (SVM; combined value from the three orthogonal axis ($\sqrt{(X^2+ Y^2+ Z^2)}$) were extracted for the triaxial accelerometers. The start and stop times of each activity were synchronised between the $K4b^2$ and each of the devices, ensuring alignment of the steady state VO_2 and accelerometer output for each activity performed. Participants individual RMR was used to calculate the MET value at each activity intensity (VO_2 $ml.kg{\cdot}min^{-1}$/ Resting VO_2

ml.kg.min^{-1}). For each activity, the VO_2 data and the 15- or 60-second epoch data were collated, averaged and exported to SPSS (v 21, SPSS Inc., Chicago, USA) for further analysis.

Statistical Analysis

All outcome variables were tested to meet the assumptions for normal distribution though visual inspection of histograms and using the Shapiro-Wilks test of normality. Descriptive statistics for the study sample and outcome variables were calculated and presented as means and standard deviations (SD). Spearman rho correlation coefficients were calculated between the VO_2 values and the accelerometer output from each of the included devices. Independent samples t-tests were used to examine if differences existed between the participants of the development group and the participants of cross-validation group. Receiver operating characteristic (ROC) analysis was used to examine the Area under the Curve (AUC) and define count-to-activity thresholds for sedentary (≤1.5 MET's), light (> 1.5 - < 3 MET's), moderate ((≥ 3 - < 6 MET's) and vigorous (≥ 6 MET's) intensity activities with optimal levels of sensitivity (correctly identifying activities at or above the required intensity) and specificity (correctly excluding activities below the intensity threshold) (Dowd et al., 2012; Zweig & Campbell, 1993).Sensitivity, Specificity and AUC values were developed, examined and interpreted with optimal values for LIPA, MPA and VPA being identified for the development group. An AUC of 1 represents perfect classification. ROC-AUC values of ≥ 0.90 are considered excellent, 0.80-0.90 good, 0.70-0.79 fair and < 0.70 poor (Jago et al., 2007). The count-to-activity thresholds developed using the development group were then cross-validated in the independent group.

Results

A total of 62 participants were recruited for the study. Nine participants were unavailable to take part on the day of testing, leaving fifty-three participants to be tested. One data set was removed due to K4 b^2 malfunction. Participant characteristics are presented in Table 2.

Independent samples t-tests showed that there were no significant differences between participants in the development group and the cross-validation group for age, mass, height, BMI and RMR. Spearman rho correlation coefficients examining the strength of association between the accelerometer outputs and VO_2 achieved r_s-values that were ≥ 0.8 and thus were considered to be highly correlated. Table 3 presents the K4b^2 measured VO_2, MET values and accelerometer outputs for all of the included activities.

Receiver Operating Characteristic Analysis

By means of ROC analysis, count-to-activity thresholds *per 15 second epochs* were established for 1.5, 3 and 6 METs in the development group for all devices (3 and 6 METs for the activPAL devices only). Count-to-activity thresholds *per 60 second epochs* were established in the development group for 1.5, 3 and 6 METs for the ActiGraph accelerometers only. The count-to-activity thresholds for all of the included activity monitors along with the AUC, sensitivity and specificity values for each of the developed count-to-activity thresholds are presented in Table 4. Overall, the sensitivity and specificity values of the developed count-to-activity thresholds for 1.5 METs ranged from 97.4-98.1% and 97.2-99.3% respectively. For 3 METs, the sensitivity and specificity values of the developed thresholds ranged from 91.8-94.8% and 92.0-94.6% respectively. For 6 METs, the sensitivity and specificity values for the developed count-to-activity thresholds for all devices ranged from 85.1-95.7% and 84.5-95.8% respectively. All AUC values for the developed thresholds were ≥ 0.9, and were therefore considered to be excellent.

Cross-validation of developed count-to-activity thresholds

The count-to-activity thresholds that were established within the development group were then cross-validated in an independent sample (cross-validation group, N = 20). ROC analysis was used to determine if the count-to-activity thresholds optimized sensitivity and specificity (inclusion or exclusion of PA intensities at, above or below the defined count-to-activity threshold). The cross-validation information is presented in Table 4. The sensitivity and specificity values for 1.5 METs ranged between 87.4-94.8% and 78.0-98.8% respectively. The sensitivity and specificity values for 3 METs ranged from 85.7-95.7% and 92.5%98.1% respectively. The 6 METs sensitivity and specificity values ranged between 65.4-84.6% and 74.4-95.9% when cross-validated in an independent sample. The developed count-to-activity thresholds appeared stronger for 3 METs compared to the other activity intensities.

Discussion

This study presents a range of count-to-activity thresholds for some of the most commonly used activity monitors in youth PA research that have demonstrated fair to excellent levels of sensitivity and specificity for accurately detecting LIPA, MPA and VPA. While the field of PA measurement has significantly advanced in recent years, issues still exist. Differences among monitor types (makes and models), attachment and wear location, calibration methods and the proliferation in data processing procedures pose evolving challenges (Migueles et al., 2017). Furthermore, the large number of independent activity-monitor validation studies makes it impossible to directly compare results from different studies or to cross-validate results between different devices (Powell et al., 2017). The present study contributes to the literature by addressing these issues. It is the first study to simultaneously validate (using the same validation protocols and analysis methodologies) five activity monitors for the determination of LIPA, MPA and VPA in an adolescent population.

The developed count-to-activity thresholds for all devices demonstrated high levels of sensitivity and specificity for the determination of 1.5, 3 and 6 METs when cross-validated in the independent group. The AUC values from the current study (all $\geq$ 0.942) are similar to those reported by Romanzini et al. (2014) for 3 METs (0.99) and 6 METs respectively (0.93). Similar to the current study, the AUCs from Romanzini *et* al. were developed using daily and locomotive tasks (Romanzini et al., 2014). Based on the AUC standards defined by Metz and colleagues, this suggests that at least 94.2% of the time, the developed count-to-activity thresholds correctly identify the activity intensity (Metz, 1978). The highest AUC values were observed for 3 METs (MPA) which suggests that the free-living movements associated with this intensity were correctly classified more frequently and incorrectly misclassified less frequently than in other intensities. The practical significance of this finding is important, as it supports the use of the developed MPA threshold values within this population.

Since the activPAL is employed as a postural measurement system, and SB is reported based on posture (Dowd et al., 2012; Kozey-Keadle et al., 2011), there is no requirement to calculate sedentary time based on acceleration. Consequently, no cut-points were developed for 1.5 METs for the activPAL devices. To the authors' knowledge, no count-to activity thresholds have been previously developed for activPAL and activPAL 3 micro devices using a mixed adolescent population. When comparing the activPAL 3 micro SVM and vertical axis count-to-activity thresholds, for MPA (3 METs), the count-to-activity thresholds developed for the activPAL 3 micro SVM demonstrated higher sensitivity and specificity values compared to the vertical axis. For this reason, we propose using the newly developed SVM count-to-activity threshold for the activPAL 3 micro in this population, as it gave the lowest number of false positive and false negative classifications. This supports the view that count-to-activity thresholds developed from triaxial accelerometer data may provide better estimates of MPA compared to those developed based on the vertical axis of uniaxial measures only (Bassett et al., 2012).

Noteworthy is the finding that the developed thresholds for both activPAL devices demonstrated lower sensitivity and specificity values for VPA (6 METs) compared to both ActiGraph and GENEActiv devices. Using the same activity devices as the present study, Powell and colleagues (2017) observed higher sensitivity and specificity values for VPA for all the included devices in an adult population. However, the inclusion of only one 6 MET activity from one exercise domain (jogging) may be a limitation of their study design and thus may have influenced the results. The present study included a variety of 6 MET activities that had high acceleration, upper limb ambulation, and up and down movements. The decreased sensitivity values reported may lend support to the claim that at higher exercise intensities the activPAL device has some difficulty correctly identifying when a person is exercising at or above the 6 MET intensity band. It is likely that with the increased number of activities of this type, the precision of most accelerometer-based measurement tools will decrease, as the acceleration value of the activity may not accurately capture the metabolic demands of the activity (Lyden et al., 2012). Although it is beyond the scope of this paper to discuss the effects of varying activity types, these findings may suggest that the inclusion of a variety of free-living 6 MET activities have affected the accuracy and precision of the activPAL device to detect accelerations within this intensity band. Given that most adolescents only spend a limited proportion of their day in vigorous activities, our results suggest that a reasonably valid estimate of total daily PA can still be estimated by the activPAL device. Furthermore, the frequent combining of MPA and VPA (primarily for the determination of whether youths achieve the recommended amount of PA) will largely nullify this misclassification at higher intensities.

The inclusion of the ActiGraph devices concurrently with the other commonly employed devices in this research was to allow for the development of comparable PA and SB estimates in young populations. The 15 second count-to-activity thresholds developed for the vertical axis of the ActiGraph wGT3X-BT and ActiGraph GT1M were similar for 1.5 METs (0), 3 METs (479 v 495) and 6 METs (1192 v 1212). The vertical axis thresholds developed for the 60 second epoch were also similar; 1.5 METs (1 v 3), 3 METs (1916 v 1981) and 6 METs (4767 v 4850) respectively. The observed similarities between the count-to-activity thresholds developed for the vertical axis of the two devices suggests that researchers using the tri-axial model may use the vertical axis count-to-activity thresholds established for earlier uniaxial models to estimate the time spent in different PA intensities (Robusto & Trost, 2012). One plausible reason for the slight differences between the accelerometer outputs may be the result of the participants wearing the ActiGraph devices on opposite hips. Previously Vaha-Ypya et al. (2015) observed marginal differences when similar accelerometers were placed on opposite hips. The observed differences between the monitors were attributed in part to pelvic tilting during low intensity locomotor activities (similar to soccer dribbling in the present study) and to the curved edge features of an indoor track (similar to the present study) where participants could only walk/run in a counter-clockwise direction (Vähä-Ypyä et al., 2015). The observations presented highlight how sensitive accelerometers can be in detecting differences in movements and provide evidence of the importance of keeping the placement of the monitor constant where possible.

Differences were also apparent when the vertical axis and the SVM of the ActiGraph wGT3X-BT were compared. The difference between the count-to-activity thresholds became more pronounced as the intensity of the activities increased for the 15 second epochs (1.5 METs; 0 v 14, 3 METs; 479 v 605, 6 METs; 1192 v 1470) and the 60 second epochs (1.5 METs; 1 v 55, 3 METs; 1916 v 2421, 6 METs; 4767 v 5879). The ActiGraph wGT3X-BT SVM thresholds generally indicated higher AUC values compared to the vertical axis results and the SVM better discriminated MPA compared to the vertical axis. This finding suggests that there may be value in using the SVM over the vertical axis of the ActiGraph wGT3X-BT when directly assessing MPA in an adolescent population. Given that multi-axial accelerometers are thought to improve measures of complex PA patterns (Smith et al., 2018), our findings for the SVM of both the activPAL 3 micro and ActiGraph wGT3X-BT suggest that the use of tri-axial accelerometers in youth research may be particularly advantageous given that children and adolescent's free-living movement patterns are generally accumulated in sporadic bursts and in multiple planes.

This is the first study to develop count-to-activity thresholds for adolescents aged between 15- 18 years for the GENEActiv accelerometer. The 15 second count-to-activity thresholds developed for the SVM of the GENEActiv-dominant and GENEActiv-non-dominant were different across intensities; 1.5 METs (14 v 11), 3 METs (69 v 67) and 6 METs (147 v 163). Notably, when the count-to-activity thresholds were cross-validated in the independent group, the sensitivity values for both the GENEActiv-dominant and GENEActiv-non-dominant were lower for 1.5 METs (0.874 v 0.888) and 6 METs (0.846 v 0.808) compared to 3 METs (0.920 v 0.956). The observed differences may be explained by the wear location of the GENEActiv device and/or by the activities that were included for the 1.5 and 6 MET intensity bands. For example, depending on the task (e.g., writing homework for 1.5 METs or running variation with limited upper limb movement for 6 METs), specific body parts may move more than others thereby producing more or less accelerations that can be disproportionate to the metabolic cost of the activity. In this respect, accelerometers attached to the wrist or upper body may not have the sensitivity to detect or accurately account for activities that are lower extremity dominant. Notwithstanding this, the excellent sensitivity and specificity values reported for the 3 MET intensity band highlight that the GENEActiv device can be employed to provide valid free-living information regarding MPA while further investigation is warranted for sedentary time and VPA.

While count-to-activity thresholds provide researchers with an acceptable method to analyse output, errors still exist when accelerometers assess free-living activity behaviour. The last decade has seen efforts to move away from the use of count-to activity thresholds, through the advent of machine learning techniques to classify PA intensity (Trost, 2007). New methodology recommendations supporting the use of more sophisticated analysis techniques such as Hidden Markov Modelling and artificial neural networking, that use features of the raw acceleration signal rather than average monitor output, have been proposed (Rowlands et al., 2018; Smith et al., 2018). While significant advances have occurred using these statistical approaches, more research is warranted to train and refine these advanced processing methods before application, especially within free-living settings. As more researchers continue to work with raw acceleration data and pattern recognition techniques, there is hope that this will facilitate the improved harmonisation of physical activity data gathered via accelerometry. However, these methods still struggle to predict many regular day-to-day behaviours, while accurately estimating activity intensity is still a struggle using these methodologies (Farrahi et al., 2019). Until this area within PA research is advanced and the analysis methods implemented with confidence, the developed count-to-activity thresholds provided are an alternative to analyse monitor output and to support the harmonisation/comparison of accelerometer data across studies.

Strength and Limitations

There are many strengths to this study including the relatively large sample size, the inclusion of both males and females and the use of an identical protocol for determining the count-to-activity thresholds for a range of activity monitors. Another significant strength of this study is the measurement of the individuals' RMR rather than using the standard adult RMR conversion of 3.5 $ml.kg.min^{-1}$ to 1 MET. The inclusion of a variety of free-living sedentary activities and a range of locomotor activity intensities (light, moderate and vigorous) enabled our research design to mimic a more free-living natural environment despite being in an indoor setting (Bassett et al., 2012; Trost et al., 2005). The inclusion of self-paced walking and running activities within specified speed ranges is a strength of the current study, it reduces the chance of a statistical clustering effect which may occur when using set/specific speeds (Bassett et al., 2012). The inclusion of ROC analysis is another strength that enabled the selection of count-to-activity thresholds that optimized sensitivity at the cost of specificity or visa-versa and has been recommended for use in validation research (Bassett et al., 2012). This is the first study to use the same validation activities and statistical analysis techniques to develop count-to-activity thresholds for five of the most commonly used activity monitors in a mixed adolescent population. Further, it is the first study to develop and cross-validate count-to-activity thresholds for the GENEActiv and activPAL devices in a mixed adolescent population. Therefore, we recommend that further studies will be necessary to compare the performance of the newly developed thresholds against existing that are commonly used in the literature and/or manufacturer proprietary algorithms/approaches. This will be essential to provide evidence that our newly developed thresholds actually improves the performance of predicting intensity levels using the activity monitors used in this study.

The limitations of this study should be considered. This study specifically targeted healthy male and female adolescents aged between 15-18 years and so the results cannot be generalised to young children or adults. The developed thresholds may not be applicable to adolescents with chronic illnesses, thus, population specific thresholds may be more appropriate for clinical populations. The count-to-activity thresholds from the development group were cross-validated in an independent group that performed the same activities. The inclusion of the same activities for cross-validation may have the potential for bias and subsequently exaggerate the accuracy of the developed count-to-activity thresholds (Powell et al., 2017).

Conclusions

Using the same analysis methods and study protocols, this is the first study to develop and cross-validate count-to-activity thresholds for a range of the most commonly used activity monitors in adolescent populations. This study expands the current PA measurement literature by providing age specific count-to-activity thresholds for the determination of light, moderate and vigorous intensity activity using the vertical axis and/or SVM of the included devices. Furthermore, it provides researchers with the opportunity to analyse and cross-compare data from different studies that have not employed the same motion sensors. The potential to cross-compare data from different studies should enable PA researchers to; (i) draw more powerful conclusions regarding movement behaviours and parameters of health, and; (ii) facilitate the interpretation and application of data to address important PA research questions.

Bibliography:

Ainsworth, B. E., Haskell, W. L., Herrmann, S. D., Meckes, N., Bassett Jr, D. R., Tudor-Locke, C., Greer, J. L., Vezina, J., Whitt-Glover, M. C., & Leon, A. S. (2011). 2011 Compendium of Physical Activities: a second update of codes and MET values. *Medicine & Science in Sports & Exercise, 43*(8), 1575-1581.

Bassett, D. R., Jr., Rowlands, A., & Trost, S. G. (2012, Jan). Calibration and validation of wearable monitors. *Med Sci Sports Exerc, 44*(1 Suppl 1), S32-38. https://doi.org/10.1249/MSS.0b013e3182399cf7

Butte, N. F., Watson, K. B., Ridley, K., Zakeri, I. F., McMurray, R. G., Pfeiffer, K. A., Crouter, S. E., Herrmann, S. D., Bassett, D. R., & Long, A. (2018). A youth compendium of physical activities: activity codes and metabolic intensities. *Medicine and science in sports and exercise, 50*(2), 246.

Dowd, K. P., Harrington, D. M., & Donnelly, A. E. (2012). Criterion and concurrent validity of the activPAL professional physical activity monitor in adolescent females. *PLOS ONE, 7*(10), e47633. https://doi.org/10.1371/journal.pone.0047633

Dowd, K. P., Szeklicki, R., Minetto, M. A., Murphy, M. H., Polito, A., Ghigo, E., van der Ploeg, H., Ekelund, U., Maciaszek, J., & Stemplewski, R. (2018). A systematic literature review of reviews on techniques for physical activity measurement in adults: a DEDIPAC study. *International journal of behavioral nutrition and physical activity, 15*(1), 15.

Farrahi, V., Niemelä, M., Kangas, M., Korpelainen, R., & Jämsä, T. (2019, 2019/02/01/). Calibration and validation of accelerometer-based activity monitors: A systematic review of machine-learning approaches. *Gait & Posture, 68*, 285-299. https://doi.org/https://doi.org/10.1016/j.gaitpost.2018.12.003

Harrington, D. M., Welk, G. J., & Donnelly, A. E. (2011, Mar). Validation of MET estimates and step measurement using the ActivPAL physical activity logger. *J Sports Sci, 29*(6), 627-633. https://doi.org/10.1080/02640414.2010.549499

Jago, R., Zakeri, I., Baranowski, T., & Watson, K. (2007, Jun). Decision boundaries and receiver operating characteristic curves: new methods for determining accelerometer cutpoints. *J Sports Sci*, *25*(8), 937-944. https://doi.org/10.1080/02640410600908027

Kozey-Keadle, S., Libertine, A., Lyden, K., Staudenmayer, J., & Freedson, P. S. (2011, Aug). Validation of wearable monitors for assessing sedentary behavior. *Med Sci Sports Exerc*, *43*(8), 1561-1567. https://doi.org/10.1249/MSS.0b013e31820ce174

Lyden, K., Kozey Keadle, S. L., Staudenmayer, J. W., & Freedson, P. S. (2012, Nov). Validity of two wearable monitors to estimate breaks from sedentary time. *Med Sci Sports Exerc*, *44*(11), 2243-2252. https://doi.org/10.1249/MSS.0b013e318260c477

Mansoubi, M., Pearson, N., Clemes, S. A., Biddle, S. J., Bodicoat, D. H., Tolfrey, K., Edwardson, C. L., & Yates, T. (2015). Energy expenditure during common sitting and standing tasks: examining the 1.5 MET definition of sedentary behaviour. *BMC public health*, *15*(1), 516.

McLaughlin, J. E., King, G. A., Howley, E. T., Bassett, D. R., Jr., & Ainsworth, B. E. (2001, May). Validation of the COSMED K4 b2 portable metabolic system. *Int J Sports Med*, *22*(4), 280-284. https://doi.org/10.1055/s-2001-13816

Metz, C. E. (1978). Basic principles of ROC analysis. Seminars in nuclear medicine,

Migueles, J. H., Cadenas-Sanchez, C., Ekelund, U., Delisle Nystrom, C., Mora-Gonzalez, J., Lof, M., Labayen, I., Ruiz, J. R., & Ortega, F. B. (2017, Sep). Accelerometer Data Collection and Processing Criteria to Assess Physical Activity and Other Outcomes: A Systematic Review and Practical Considerations. *Sports Med*, *47*(9), 1821-1845. https://doi.org/10.1007/s40279-017-0716-0

Poole, D. C., Schaffartzik, W., Knight, D. R., Derion, T., Kennedy, B., Guy, H. J., Prediletto, R., & Wagner, P. D. (1991). Contribution of excising legs to the slow component of oxygen uptake kinetics in humans. *Journal of applied physiology*, *71*(4), 1245-1260.

Powell, C., Carson, B. P., Dowd, K. P., & Donnelly, A. E. (2017, Dec). Simultaneous validation of five activity monitors for use in adult populations. *Scand J Med Sci Sports*, *27*(12), 1881-1892. https://doi.org/10.1111/sms.12813

Robusto, K. M., & Trost, S. G. (2012). Comparison of three generations of ActiGraph™ activity monitors in children and adolescents. *Journal of sports sciences, 30*(13), 1429-1435.

Romanzini, M., Petroski, E. L., Ohara, D., Dourado, A. C., & Reichert, F. F. (2014, 2014/01/02). Calibration of ActiGraph GT3X, Actical and RT3 accelerometers in adolescents. *European Journal of Sport Science, 14*(1), 91-99. https://doi.org/10.1080/17461391.2012.732614

Rowlands, A. V., Mirkes, E. M., Yates, T., Clemes, S., Davies, M., Khunti, K., & Edwardson, C. L. (2018, Feb). Accelerometer-assessed Physical Activity in Epidemiology: Are Monitors Equivalent? *Med Sci Sports Exerc, 50*(2), 257-265. https://doi.org/10.1249/mss.0000000000001435

Saint-Maurice, P. F., Kim, Y., Welk, G. J., & Gaesser, G. A. (2016, Jan). Kids are not little adults: what MET threshold captures sedentary behavior in children? *Eur J Appl Physiol, 116*(1), 29-38. https://doi.org/10.1007/s00421-015-3238-1

Smith, M. P., Horsch, A., Standl, M., Heinrich, J., & Schulz, H. (2018, 2018/10/10). Uni- and triaxial accelerometric signals agree during daily routine, but show differences between sports. *Scientific Reports, 8*(1), 15055. https://doi.org/10.1038/s41598-018-33288-z

Treuth, M. S., Schmitz, K., Catellier, D. J., McMurray, R. G., Murray, D. M., Almeida, M. J., Going, S., Norman, J. E., & Pate, R. (2004, Jul). Defining accelerometer thresholds for activity intensities in adolescent girls. *Med Sci Sports Exerc, 36*(7), 1259-1266.

Trost, S. G. (2007). State of the art reviews: measurement of physical activity in children and adolescents. *American Journal of lifestyle medicine, 1*(4), 299-314.

Trost, S. G., McIver, K. L., & Pate, R. R. (2005, Nov). Conducting accelerometer-based activity assessments in field-based research. *Med Sci Sports Exerc, 37*(11 Suppl), S531-543.

Vähä-Ypyä, H., Vasankari, T., Husu, P., Mänttäri, A., Vuorimaa, T., Suni, J., & Sievänen, H. (2015). Validation of Cut-Points for Evaluating the Intensity of Physical Activity with Accelerometry-Based Mean Amplitude Deviation (MAD). *PLOS ONE, 10*(8), e0134813. https://doi.org/10.1371/journal.pone.0134813

Vaha-Ypya, H., Vasankari, T., Husu, P., Suni, J., & Sievanen, H. (2015, Jan). A universal, accurate intensity-based classification of different physical activities using raw data of accelerometer. *Clin Physiol Funct Imaging*, *35*(1), 64-70. https://doi.org/10.1111/cpf.12127

Van Hecke, L., Loyen, A., Verloigne, M., van der Ploeg, H. P., Lakerveld, J., Brug, J., De Bourdeaudhuij, I., Ekelund, U., Donnelly, A., Hendriksen, I., Deforche, B., & on behalf of the, D. c. (2016, 2016/06/28). Variation in population levels of physical activity in European children and adolescents according to cross-European studies: a systematic literature review within DEDIPAC. *International journal of behavioral nutrition and physical activity*, *13*(1), 70. https://doi.org/10.1186/s12966-016-0396-4

Welk, G. J. (2005, Nov). Principles of design and analyses for the calibration of accelerometry-based activity monitors. *Med Sci Sports Exerc*, *37*(11 Suppl), S501-511.

Zweig, M. H., & Campbell, G. (1993, Apr). Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin Chem*, *39*(4), 561-577.

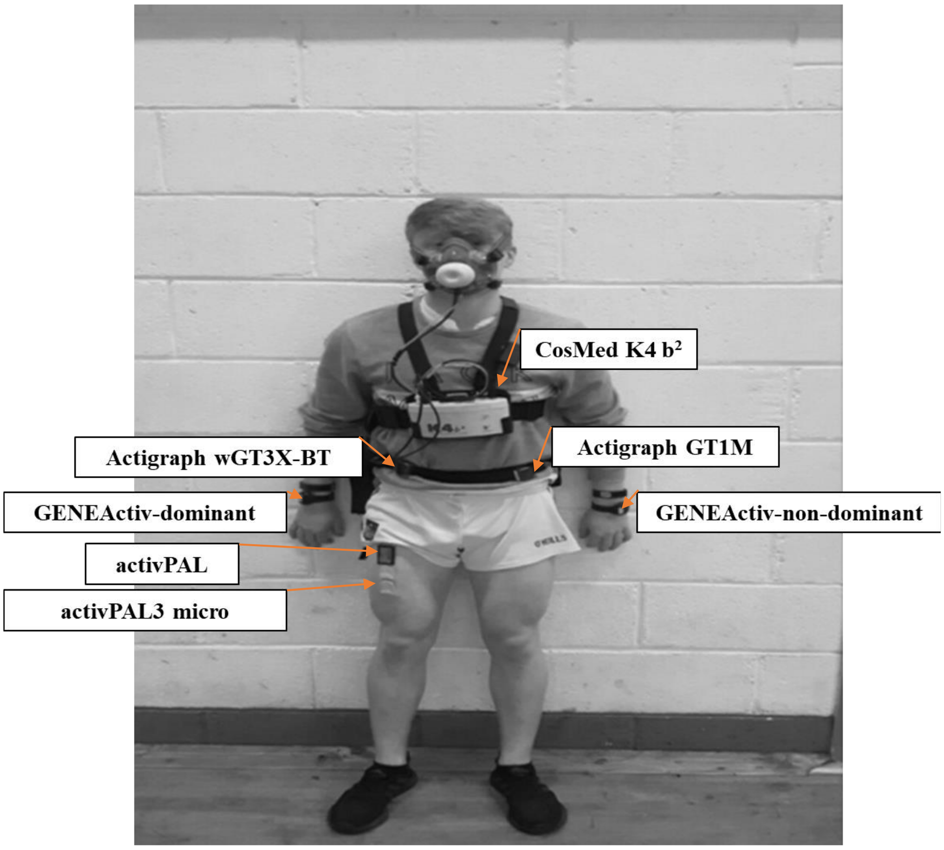

Figure 1: Wear Location of Devices

Table 1: Characteristics and specifications of all included devices

	activPAL 3 micro	activPAL	ActiGraph wGT3X-BT	ActiGraph GT1M	GENEActiv
Size (mm)	23.5 x 43 x 5	35 x 53 x 7	33 x 46 x 15	38 x 36 x 18	43 x 40 x 13
Mass (g)	10	15	19	27	16
Axes	3	1	3	2	3
Placement	Midpoint of anterior right thigh	Midpoint of anterior right thigh	Right iliac crest	Left iliac crest	One one each wrist
Application	Waterproofed with nitrile sleeve and tegaderm dressing	Waterproofed with nitrile sleeve and tegaderm dressing	Elastic Belt	Elastic Belt	Wrist strap
Range	±2	0-1.5	±8	±5	±8
Sample Frequency (Hz)	20	10	30	30	30
Epoch Length (seconds)	15	15	15 and 60	15 and 60	15
Software	activPAL v 7.2.32	activPAL v 7.2.32	ActiLife v 6.11.4	ActiLife v 6.11.4	GENEActiv v 2.2

Table 2: Summary of Participant Characteristics by total sample and by group (Mean (SD)).

	All Participants (n=52)	**Development group (n=32)**	**Cross-Validation group (n=20)**	**Between-group Differences[a] *(P)***
Sex	25 Males/ 27 Females	16 Males/ 16 Females	9 Males/ 11 Females	
Age (years)	16.1 (0.78)	16.0 (0.80)	16.2 (0.77)	0.543
Mass (kg)	67.2 (13.12)	66.2 (13.21)	66.2 (13.14)	0.479
Height (m)	1.7 (0.09)	1.7 (0.09)	1.7 (0.09)	0.862
BMI ($kg.m^{-2}$)	23.0 (3.79)	22.7 (3.48)	23.6 (4.27)	0.390
RMR ($ml.kg.min^{-1}$)	4.36 (0.75)	4.50 (0.76)	4.15 (0.72)	0.100

[a] Independent samples *t*-tests used to examine between-group differences. Statistical significance set at $P < 0.05$.

Table 3: Summary of K4 b^2 measured VO_2, MET values and accelerometer outputs for all of the included activities (Mean (SD)).

Measure	Units	Sitting	Screen Viewing	Standing	Slow walking	Brisk Walk	Dribble football	Jogging	Jogging Variation
VO_2	ml.kg.min^{-1}	5.15 (1.18)	4.88 (0.70)	5.10 (1.06)	15.02 (3.06)	20.03 (3.00)	19.00 (3.91)	34.06 (6.70)	38.31 (4.57)
Energy Expenditure	METs	1.20 (0.23)	1.12 (0.20)	1.16 (0.18)	3.50 (0.74)	4.66 (0.90)	5.00 (1.40)	7.78 (1.60)	8.57 (1.28)
ActiGraph wGT3X-BT	Counts.15s^{-1}	0 (1)	0 (0)	0 (0)	606 (196)	1071 (225)	704 (238)	2011 (531)	1552 (297)
ActiGraph wGT3X-BT	Counts.60s^{-1}	1 (3)	0 (0)	0 (0)	2425 (783)	4285 (901)	2815 (952)	8045 (2124)	6208 (1188)
ActiGraph wGT3X-BT (SVM)	Counts.15s^{-1}	3 (9)	0 (0)	1 (5)	768 (248)	1215 (231)	1030 (247)	2155 (522)	2123 (387)
ActiGraph wGT3X-BT (SVM)	Counts.60s^{-1}	11 (35)	0 (0)	4 (20)	3070 (994)	4859 (926)	4112 (990)	8620 (2088)	8516 (1550)
ActiGraph GT1M	Counts.15s^{-1}	0 (1)	0 (1)	0 (0)	640 (189)	1118 (220)	720 (263)	2056 (507)	1614 (375)
ActiGraph GT1M	Counts.60s^{-1}	1 (3)	1 (2)	0 (0)	2562 (756)	4473 (879)	2880 (1051)	8223 (2026)	6454 (1501)
activPAL 3micro	Counts.15s^{-1}	5 (15)	0 (1)	5 (32)	7020 (1691)	10636 (1755)	8129 (2285)	16025 (4571)	11980 (1244)
activPAL 3micro (SVM)	Counts.15s^{-1}	10 (31)	1 (1)	15 (74)	10581 (1972)	16505 (2518)	12865 (3117)	26012 (7349)	19234 (1721)
activPAL	Counts.15s^{-1}	10 (95)	36 (169)	2 (17)	3212 (853)	5041 (1018)	4140 (1117)	9882 (2825)	6734 (1391)
GENEActiv-dominant	g.s^{-1}.15.s^{-1}	5 (4)	4 (1)	4 (2)	75 (16)	119 (22)	146 (63)	351 (107)	330 (72)
GENEActiv- non-dominant	g.s^{-1}.15.s^{-1}	4 (3)	3 (1)	4 (2)	77 (17)	125 (29)	151 (69)	368 (102)	402 (84)

Table 4: Development group count-to-activity thresholds, AUC, sensitivity and specificity values for 1.5, 3 and 6 METs developed using ROC analysis in the development group.

Activity Monitor	Axes	Unit	Epoch (seconds)	AUC	Cut-point	Sensitivity	Specificity
			1.5 METs				
ActiGraph wGT3X-BT	Vertical	Counts	15	0.988	0	0.976	0.993
ActiGraph wGT3X-BT	Vertical	Counts	60	0.988	1	0.976	0.993
ActiGraph wGT3X-BT	SVM	Counts	15	0.966	14	0.981	0.980
ActiGraph wGT3X-BT	SVM	Counts	60	0.966	55	0.981	0.980
ActiGraph GT1M	Vertical	Counts	15	0.987	1	0.976	0.980
ActiGraph GT1M	Vertical	Counts	60	0.987	3	0.976	0.980
GENEActiv-dominant	SVM	Counts	15	0.996	14	0.974	0.972
GENEActiv-non-dominant	SVM	Counts	15	0.993	11	0.986	0.987
			3 METs				
ActiGraph wGT3X-BT	Vertical	Counts	15	0.985	479	0.918	0.920
ActiGraph wGT3X-BT	Vertical	Counts	60	0.985	1916	0.918	0.920
ActiGraph wGT3X-BT	SVM	Counts	15	0.989	605	0.942	0.943
ActiGraph wGT3X-BT	SVM	Counts	60	0.989	2421	0.942	0.943
ActiGraph GT1M	Vertical	Counts	15	0.984	495	0.918	0.920
ActiGraph GT1M	Vertical	Counts	60	0.984	1981	0.918	0.920
ActivPAL 3 micro	Vertical	Counts	15	0.982	5934	0.929	0.925
ActivPAL 3 micro	SVM	Counts	15	0.984	9286	0.934	0.934
ActivPAL	Vertical	Counts	15	0.984	2899	0.940	0.940
GENEActiv-dominant	SVM	Counts	15	0.990	69	0.948	0.946
GENEActiv-non-dominant	SVM	Counts	15	0.989	67	0.942	0.943
			6 METs				
ActiGraph wGT3X-BT	Vertical	Counts	15	0.982	1192	0.936	0.935
ActiGraph wGT3X-BT	Vertical	Counts	60	0.982	4767	0.936	0.935
ActiGraph wGT3X-BT	SVM	Counts	15	0.985	1470	0.957	0.955
ActiGraph wGT3X-BT	SVM	Counts	60	0.985	5879	0.957	0.955
ActiGraph GT1M	Vertical	Counts	15	0.983	1212	0.926	0.926
ActiGraph GT1M	Vertical	Counts	60	0.983	4850	0.926	0.926
ActivPAL 3micro	Vertical	Counts	15	0.942	10226	0.851	0.845
ActivPAL 3micro	SVM	Counts	15	0.955	16100	0.872	0.874
ActivPAL	Vertical	Counts	15	0.975	5372	0.915	0.922
GENEActiv-dominant	SVM	Counts	15	0.991	147	0.953	0.953
GENEActiv-non-dominant	SVM	Counts	15	0.993	163	0.957	0.958

Table 5: Cross-validation of the count-to-activity thresholds developed in the development group. Sensitivity and specificity values for 1.5, 3 and 6 METs reported for the cross-validation group.

Activity Monitor	Axes	Unit	Epoch (seconds)	Sensitivity	Specificity
		1.5 METs			
ActiGraph wGT3X-BT	Vertical	Counts	15	0.897	0.951
ActiGraph wGT3X-BT	Vertical	Counts	60	0.897	0.951
ActiGraph wGT3X-BT	SVM	Counts	15	0.897	0.939
ActiGraph wGT3X-BT	SVM	Counts	60	0.948	0.780
ActiGraph GT1M	Vertical	Counts	15	0.888	0.854
ActiGraph GT1M	Vertical	Counts	60	0.888	0.854
GENEActiv-dominant	SVM	Counts	15	0.874	0.988
GENEActiv-non-dominant	SVM	Counts	15	0.888	0.988
		3 METs			
ActiGraph wGT3X-BT	Vertical	Counts	15	0.934	0.963
ActiGraph wGT3X-BT	Vertical	Counts	60	0.934	0.962
ActiGraph wGT3X-BT	SVM	Counts	15	0.956	0.944
ActiGraph wGT3X-BT	SVM	Counts	60	0.956	0.944
ActiGraph GT1M	Vertical	Counts	15	0.957	0.944
ActiGraph GT1M	Vertical	Counts	60	0.956	0.944
ActivPAL 3micro	Vertical	Counts	15	0.890	0.925
ActivPAL 3micro	SVM	Counts	15	0.912	0.935
ActivPAL	Vertical	Counts	15	0.857	0.953
GENEActiv-dominant	SVM	Counts	15	0.920	0.972
GENEActiv-non-dominant	SVM	Counts	15	0.956	0.981
		6 METs			
ActiGraph wGT3X-BT	Vertical	Counts	15	0.808	0.942
ActiGraph wGT3X-BT	Vertical	Counts	60	0.808	0.944
ActiGraph wGT3X-BT	SVM	Counts	15	0.654	0.942
ActiGraph wGT3X-BT	SVM	Counts	60	0.654	0.942
ActiGraph GT1M	Vertical	Counts	15	0.769	0.936
ActiGraph GT1M	Vertical	Counts	60	0.769	0.936
ActivPAL 3micro	Vertical	Counts	15	0.846	0.849
ActivPAL 3micro	SVM	Counts	15	0.769	0.860
ActivPAL	Vertical	Counts	15	0.731	0.936
GENEActiv-dominant	SVM	Counts	15	0.846	0.958
GENEActiv-non-dominant	SVM	Counts	15	0.808	0.959

Supplementary Table 1: Bank of study exercises, ranked by participants and included in their study protocol

Activities	Sedentary Activities
Sitting Still	Participants were asked to sit upright with their back placed firmly to the chair and feet placed on the ground.
Standing Still	Participants asked to stand upright unassisted.
Standing doing homework	Participants asked to stand and work at a desk typing on a laptop.
Sitting doing homework	Participants were seated at a desk and were given a passage from a text book and asked to write it onto an A4 pad.
Sitting playing on Phone	Participants were seated and asked to play with their phone. All app use was permitted however, no calls, texts or taking photos were permitted
Screen Viewing	Participants seated in a chair and asked to watch a short video.
	Light Intensity Activities
Slow Walking (2.5- 4.5 km.hr)	Participants were asked to briskly walk around a track at self-selected pace.
Cleaning Bedroom	A series of objects were scattered around a section of the hall. Participants were asked to tidy up a certain section of the hall as if they were at home. They were asked to pick up books, DVD's, bottles, 2 shirts from the floor and place the shirts on a hanger, move 3 boxes singly from one chair to another (each weighing no more than 3kgs).
Sweeping	Cleaning, sweeping, slow, light effort
Football	Participants were asked kick a soccer ball over and back
Loading and Unloading Boxes	Participants were required to move 3 boxes stacked on the top of a chair to another chair one at a time.
	Moderate Intensity Activities
Brisk Walking (4.5-6.5 km.hr)	Participants were asked to briskly walk around a track at self-selected pace.
Aerobic Exercise	Participants were asked to perform squats, lunges, knee raises, heel flicks, sit-ups and press-ups (mimic exercise routine). They performed each exercise for 1 minute each. Modifications were
Dribbling a basketball	Participants were asked to walk dribbling a basketball around a rectangular space
Dribbling a soccer ball	Participants were asked to walk dribbling a soccer ball around a rectangular space.
	Vigorous Intensity Activities
Jogging (6.5 – 8.5 km.hr)	Participants were asked to briskly walk around a track at a set pace.
Running Variation	Markers/ cones were placed around the track, at which participants hand to perform different movements. Participants were asked to jump on one leg into the air and raise both arms (to mimic them catching a football). Participants were asked to stop at the second cone and touch their ankles 5 times (standing straight between each one).
Aerobic Steps	6-8 inch step, participants were asked to step up on the step and step back down.

Exercise to music instruction	Participants were asked to perform exercise to music watching a video and asked to mimic the actions.