Accepted: 3 November 2016

DOI: 10.1111/sms.12813

ORIGINAL ARTICLE

WILEY

Simultaneous validation of five activity monitors for use in adult populations

C. Powell[1] | **B. P. Carson**[1,2] | **K. P. Dowd**[3] | **A. E. Donnelly**[1,2]

[1]Department of Physical Education and Sport Sciences, Centre for Physical Activity and Health Research, University of Limerick, Limerick, Ireland

[2]Health Research Institute, University of Limerick, Limerick, Ireland

[3]Department of Sport and Health, Faculty of Science, Athlone Institute of Technology, Athlone, Westmeath, Ireland

Correspondence
Cormac Powell, Department of Physical Education and Sport Sciences, University of Limerick, Limerick, Ireland. Email: cormac.powell@ul.ie

Funding information
This research was supported by a University of Limerick Physical Education and Sport Sciences Postgraduate Scholarship (2013).

Numerous cut--points exist to measure physical activity by accelerometry. The ability to compare accelerometer findings from different devices from different locations may be advantageous to researchers. This study aimed to develop and validate cut-- points for 1.5, 3, and 6 METs in five activity monitors simultaneously. Fifty--six participants (mean age=39.9 [±11.5] years) performed six activities while wearing a CosMED K4b^2 and five activity monitors: activPAL3 Micro, activPAL, ActiGraph GT1M, ActiGraph wGT3X--BT, and GENEActiv. Receiver operating characteristic curves and analysis were used to develop and validate cut--points for the vertical axis counts (all activity monitors) and sum of the vector magnitude (ActiGraph wGT3X-- BT and GENEActiv) for 15 second (all devices) and 60 second (ActiGraph devices) epochs. A random coefficients statistical model was used to derive MET predictive equations for all activity monitors. Bland--Altman plots examined the variability in device error. No 1.5 MET cut--points were developed for the activPAL devices. All developed cut--points had high levels of sensitivity and specificity. When cross-- validated in an independent group, high levels of sensitivity and specificity remained (≥77.4%, monitor and intensity dependent). The mean bias based on the Bland-- Altman plots ranged from −0.03 METs to 0.35 METs (monitor dependent). This is the first study to develop and validate cut--points for five activity monitors simultaneously with high levels of sensitivity and specificity (≥77.4%). This is potentially a step toward cross--comparison/harmonization of data; however, further validation in a free--living environment is warranted.

KEYWORDS
accelerometer, cut-points, METs, physical activity, validation

1 | INTRODUCTION

Physical activity (PA) behaviors have been identified as key lifestyle variables which influence health. Increased partic-ipation in PA of specific intensities (ie, moderate to vigor-ous PA [MVPA]) positively impacts the onset, severity, and course of a range of chronic diseases and conditions, includ-ing diabetes mellitus, cancer, obesity, hypertension, bone and joint disease, and depression.[1–3] Additionally, emerging evi-dence now suggests that associations may exist between PA intensities at the lower end of the activity intensity continuum (such as sedentary time and light intensity [LI] PA) and indi-ces of health.[4]

Accelerometry has become the preferred method for objectively examining PA under free--living conditions, due to the portability, affordability, and convenience of the activity monitors and the abundance of information that can be obtained.[5,6] Accelerometers have made the measure-ment of PA more practical over extended periods, in non-- clinical settings and in large--scale epidemiological studies.[2,7] Considerable effort has been made in validating acceler-ometers as measures of PA intensity in adult populations.

TABLE 1 Characteristics of activity monitors

	activPAL3 Micro	activPAL	ActiGraph wGT3X-BT	ActiGraph GT1M	GENEActiv
Size (mm)	23.5 × 43 × 5	35 × 53 × 7	33 × 46 × 15	38 × 36 × 18	43 × 40 × 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 on each wrist
Application	Nitrile sleeve and waterproof dressing (Tegaderm)	Double-sided adhesive strip (PALstickie)	Elastic belt	Elastic belt	Strap
Range (g)	±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

However, this is a continuous task due to the large variety of accelerometers that are available to researchers and the constant updating of activity monitor models and software, thus creating noise and confusion in the literature. Due to the large number of individual activity monitor validations, it is not possible to cross--validate the results between differ-ent devices or directly compare the results from different studies.[8,9] The ability to compare data from different studies would be advantageous to researchers, as it would enable the harmonization of data from different large--scale studies with greater accuracy, providing a greater understanding of asso-ciations between PA behavior and health. A range of differ-ent activity monitors have been employed in the many large-- scale cross--sectional and surveillance studies of populations worldwide. Harmonizing activity measurements from these datasets is problematic since the majority of activity monitor validations have been undertaken independently, using differ-ent validation methods and activity protocols. The purpose of this study was to determine and validate count--to--activity cut--points for LIPA, MPA, and VPA in five commonly uti-lized PA monitors (activPAL, activPAL3 Micro, ActiGraph wGT3X--BT, ActiGraph GT1M, and GENEActiv). The find-ings of this study may potentially enable researchers to better compare data from studies that have employed any of the five different activity monitors.

2 | METHODS

2.1 | Participants

Participants were recruited from university staff and students and members of the local community. To be considered for inclusion, participants had to be male or female between the ages of 18--65 years with no injury or illness that prevented participation in PA. A recruitment email was circulated to identify potential participants, who were then screened to determine their suitability for inclusion. All participants returned signed consent forms prior to undertaking the study. All participants were allocated a number and a randomiza-tion table was used to assign each participant to either the Development Group or the Cross--validation Group. Ethics committee approval was granted by the University Research Ethics Committee (EHSREC 11--48 and EHSREC 10--26) in compliance with the Declaration of Helsinki (2008).

2.2 | Activity monitors

Five activity monitors were employed during the testing pro-tocol. The characteristics of the included monitors are shown in Table 1. Prior to testing, all monitors were initialized on the same computer, allowing the output from all of the activ-ity monitors to be synchronized. To reduce any potential interdevice differences, the same activity monitors were used throughout the study. The low--frequency extension filter was applied to all ActiGraph wGT3X--BT data, as this enables the more accurate detection of accelerations in lower intensity activities.[10,11]

2.3 | Metabolic measurement

Oxygen consumption and carbon dioxide output were meas-ured using the breath--by--breath function of a portable meta-bolic unit, the CosMED K4b^2 (CosMED, Rome, Italy). The K4b^2 has been shown to be a reliable measure of oxygen consumption.[12,13] The device was worn during rest and activ-ity measurements. Each participant was fitted with a rubber facemask (Hans Rudolph, Kansas City, USA) with a built--in seal to prevent air leaks. The device was calibrated in--line with the manufacturer's guidelines before each trial. The K4b^2 data were downloaded and stored on a PC after each individual trial. Resting metabolic rate (RMR) was measured to allow the activity intensities to be individualized for each participant, as use of the standard 3.5 mL kg min^{-1} has been shown to have limitations for calculating metabolic rate[14] and

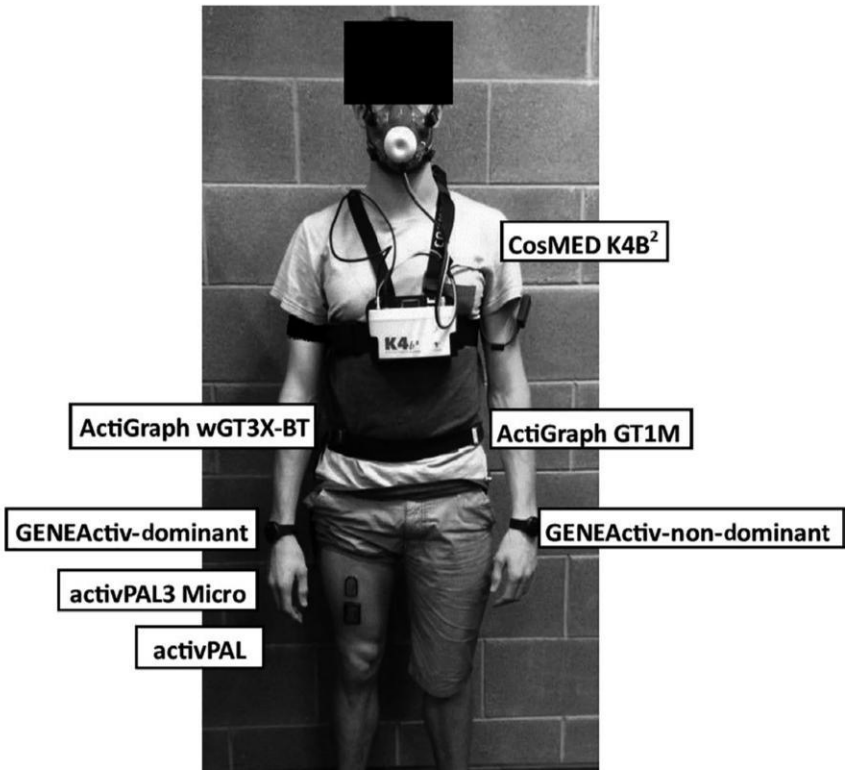

FIGURE 1 Wear location for each device during testing protocol

does not represent the RMR of the general population.[15] For this study, the measured value for each individual's RMR was deemed to be 1 MET, and all other measured values were expressed relative to this.

2.4 | Calibration activities

Activities included in a calibration protocol must be representative of the activities performed in regular daily activities by the intended population.[16,17] The calibration activities included in this study were sitting in a chair (5 minutes), standing still (5 minutes), dish handling (5 minutes), slow walking (7 minutes), brisk walking (7 minutes), and jogging (7 minutes). These activities have been used in previous validation studies[5,18–20] and were selected to mimic common everyday activities that the average person would engage in. The mean value of the last 2 minutes of each activity (exclud-ing RMR) was used in the data analysis. This was deemed appropriate as VO_2 reaches a steady state in healthy adults after 3 minutes during moderate intensity activity.[21]

2.5 | Testing protocol

As per the RMR measurement recommendations put forth by Compher et al.,[22] participants arrived at the testing facil-ity following an overnight fast (~12 hours), had abstained from alcohol and caffeine, and had not undertaken any unaccustomed activity or exercise during the same time period. Participants wore light shorts, t--shirt or vest, socks, and running shoes. Mass was measured without shoes to the nearest 0.1 kg using electronic scales (Seca model 770; Seca Ltd, Birmingham, UK) and height was measured without shoes to the nearest 0.25 cm using a stadiometer (Seca model 214; Seca Ltd). Body mass index (BMI) was calculated using the standard formula (kg m^{-2}). After anthropometric data had been collected, participants were fitted with the activity monitors and the $K4b^2$. The activity monitors were attached according to the manufacturer's guidelines; activ-PAL3 Micro to the anterior aspect of the right thigh using a nitrile sleeve and Tegaderm dressing (PAL Technologies Ltd, Glasgow, Scotland); activPAL to the anterior aspect of the right thigh using a PALstickie (PAL Technologies Ltd, Glasgow, Scotland); ActiGraph wGT3X--BT and ActiGraph GT1M on the right and left iliac crests, respectively, using elastic belts, and the GENEActiv on the left and right wrists using straps (manufacturer of the GENEActiv do not state if the monitor should be worn on the dominant or non--dominant side, so a monitor was worn on each side). Figure 1 depicts how all of the devices were worn simultaneously.

Following the initial setup, the activities were carried out in ascending intensity. A single observer instructed each participant when to start and stop a particular activity. The nature of the activity and the exact start and finish times were recorded by the observer. Resting VO_2 was measured for 15 minutes (after a 10--minute resting period), allowing the participant adequate time to reach a rested state. During this time, the participant was in prone position in a darkened room. For the sitting activity, participants sat while looking

forward, with their feet placed flat on the ground and hands on their knees. During the standing activity, participants again looked forward with their feet shoulder width apart and hands relaxed by their sides. Participants were instructed to refrain from talking during the sitting and standing activities.

Following this, participants undertook four other activi-ties, including dish handling, slow walking (2.5--4.5 km h^{-1}), brisk walking (4.5--6.5 km h^{-1}), and jogging (6.5--8.5 km h^{-1}). Participants were allowed to rest for 1 minute between activities or for longer if heart rate had not returned to <100 BPM after 1 minute. For the three locomotion activities, partici-pants walked at a self--selected speed within the speed bands outlined above. For each speed band, the upper and lower time limits required to complete one full circle of the prede-termined track (40 m in length) were calculated. These times were then used to compare the participant's actual walking speed with the prescribed speed. During the first minute of the measurement, the time it took to complete a full circle was recorded. If the time taken to complete the track was either too fast or too slow, the participant was instructed to slow down or speed up. Once the participant was comfortable at a specific speed and their times were consistent, the partici-pants were encouraged to maintain that pace for the reminder of the trial, thus giving them a consistent speed for the final 2 minutes that was comfortable for them. Participants per-formed each individual task for either 5 or 7 minutes, inten-sity dependent. The less metabolically demanding tasks were 5 minutes, with the more demanding tasks being 7 minutes long.[21] Again, participants were instructed to refrain from talking during the abovementioned activities. Once all the activities were completed, data from all the activity monitors were downloaded to the lead investigator's computer.

2.6 | Data processing

After all files were downloaded from the monitors, the respective proprietary software was used to process the data: activPAL3 Micro and activPAL (v 7.2.32), ActiGraph wGT3X--BT and ActiGraph GT1M (ActiLife v 6.11.4), and GENEActiv (v 2.2). For the activPAL and ActiGraph GT1M, accelerations in the vertical axis were extracted; for the activ-PAL3 Micro and ActiGraph wGT3X--BT, accelerations in the vertical axis and the sum of the vector magnitudes (SVM) were extracted; and for the GENEActiv, the SVM was extracted. While the ActiGraph wGT3X--BT and GENEActiv give SVM as one of their outputs, the SVM was computed for the activPAL3 Micro as $\sqrt{(X^2 + Y^2 + Z^2)}$.[23] The SVM from the GENEActiv is the gravity subtracted SVM. [24] Activity counts per 15 second epochs were developed for all devices using the respective proprietary software. In addition, activ-ity counts per 60 second epochs were also developed for out-puts from both ActiGraph devices. From the K4b^2, VO_2 data were averaged over 15 or 60 second periods. The average data over the final 2 minutes of each activity were selected and exported to SPSS (v 21, SPSS Inc., Chicago, USA) for further analysis.

2.7 | Statistical analysis

Independent t tests were used to examine characteristic differences between the Development Group and Cross--validation Group. Receiver operating characteristic (ROC) curves and analysis were used to calculate an area under the curve (AUC) and define cut--points for 1.5, 3, and 6 METs with optimal levels of sensitivity (correctly identified points at or above the activity intensity thresholds) and specificity (correctly excluded activities below the activity intensity thresholds).[5] This involves recoding the data to create binary indicator variables (0 or 1). Sensitivity, specificity, and AUC were examined and interpreted, with the optimal values for LIPA, MPA, and VPA being identified in the Development Group. An AUC of 1 represents perfect classification. AUC values of ≥0.90 are considered excellent, 0.80--0.89 good, 0.70--0.79 fair, and <0.70 poor.[25] The cut--points determined from the ROC analysis were subsequently cross--validated in the Cross--validation Group. For the activPAL and activPAL3 Micro, no cut--points were developed for 1.5 METs, as the proprietary algorithms within the software provide accurate estimates for sitting/lying and standing, which are conse-quently used in place of any 1.5 MET (sedentary) threshold. A random coefficients statistical model, which accounts for repeated measures taken from the same participants, was used to examine the relationship between MET values and accel-erometer counts for all of the included monitors. Regression equations for predicting activity METs were developed for all monitors. The concordance correlation coefficient (CCC) was used to assess the fit of the equation. The standard error estimate (SEE) was presented with the CCC. Rosenberger et al.[26] have previously used a combination of ROC analysis and mixed--model analysis to determine intensity cut--points and predictive equations, respectively. Bland--Altman plots were used to examine the mean bias and the upper and lower limit of agreement for all monitor--predicted METs in com-parison with the CosMED measured METs.[27]

3 | RESULTS

Seventy--two individuals responded to the initial recruitment email, with 65 being recruited for the study. Eight participants were unavailable for testing on test days, leaving 57 partici-pants to be tested. Fifty--six datasets were included in the final analysis. One dataset was excluded due to the K4b^2 failing to record the metabolic data. The characteristics of the included participants can be seen in Table 2. Independent t tests were used to examine whether differences existed between the

TABLE 2 Participant characteristics—mean (SD)

	All participants (n=56)	Development group (n=30)	Cross-validation group (n=26)	Between-group difference[a] (*P*)
Sex	25 Males/31 females	14 Males/16 females	9 Males/17 females	—
Age (years)	39.9 (11.5)	38.1 (11.2)	42.0 (11.6)	.212
Mass (kg)	73.7 (12.5)	75.8 (10.7)	71.3 (14.0)	.197
Height (m)	1.70 (0.09)	1.70 (0.09)	1.70 (0.09)	.406
BMI	25.0 (3.7)	25.5 (3.5)	24.5 (3.9)	.328
RMR (mL kg min^{-1})	3.27 (0.62)	3.25 (0.64)	3.29 (0.60)	.839

[a]Independent *t* tests used to examine between--group differences. Significance set at $P<.05$.

TABLE 3 Mean (SD) for all measures for all participants

Measure	Units	Sitting	Standing	Dish handling	Slow walking	Fast walking	Jogging
VO_2	mL kg min^{-1}	3.45 (1.00)	3.56 (0.94)	4.99(1.57)	10.26 (1.89)	13.55 (2.36)	30.99 (5.92)
Energy expenditure	METs	1.02 (0.14)	1.09 (0.14)	1.52(0.32)	3.14 (0.67)	4.07 (1.16)	9.72 (2.58)
ActiGraph wGT3X-BT	Counts. 15 s^{-1}	0 (2)	0 (0)	0(0)	547 (163)	940 (209)	2044 (458)
ActiGraph wGT3X-BT	Counts. 60 s^{-1}	1 (10)	0 (1)	0(1)	2222 (610)	3812 (741)	8406 (1937)
ActiGraph wGT3X--BT (SVM)	Counts. 15 s^{-1}	1 (3)	1 (3)	14(32)	728 (151)	1099 (189)	2233 (460)
ActiGraph wGT3X--BT (SVM)	Counts. 60 s^{-1}	0 (0)	0 (1)	57(129)	2924 (612)	4407 (766)	8936 (1844)
ActiGraph GT1M	Counts. 15 s^{-1}	0 (0)	0 (0)	0(0)	528 (155)	946 (191)	2093 (469)
ActiGraph GT1M	Counts. 60 s^{-1}	0 (0)	0 (0)	0(0)	2112 (619)	3786 (764)	8374 (875)
activPAL3 Micro	Counts. 15 s^{-1}	1 (12)	1 (7)	48(82)	6649 (206)	9765 (2493)	16845 (3582)
activPAL3 Micro (SVM)	Counts. 15 s^{-1}	1 (1)	1 (1)	159(251)	10239 (2705)	15560 (2818)	26855 (5825)
activPAL	Counts. 15 s^{-1}	0 (0)	0 (1)	1(11)	3606 (1206)	5403 (1140)	10994 (1288)
GENEActiv-dominant	g s^{-1}·15 s^{-1}	4 (2)	3 (1)	60(36)	70 (18)	104 (32)	394 (32)
GENEActiv-non-dominant	g s^{-1}·15 s^{-1}	3 (2)	3 (1)	54(25)	75 (24)	108 (33)	372 (76)
Speed	km h^{-1}	—	—	—	3.71 (0.32)	5.21 (0.34)	7.97 (0.47)

Development Group and Cross--validation Group. No significant differences were observed for age, mass, height, and BMI.

The mean (standard deviation) for all participants for K4b^2 measured VO_2, METs, accelerometer outputs for all activities, and speed during the locomotive activities are pre-sented in Table 3. As the activity intensity increased, the VO_2 and the MET values increased. Similarly, as the intensity of the activity increased, the outputs from all activity monitors increased (excluding sitting and standing).

3.1 | ROC analysis

Cut--points per 15 second epochs were developed in the Development Group for 1.5, 3, and 6 METs for all of the included activity monitors, using ROC analysis. Similarly, cut--points per 60 second epochs were developed for the ActiGraph monitors (ActiGraph wGT3X--BT and ActiGraph GT1M). Cut--points were developed for the vertical axis in the ActiGraph wGT3X--BT, ActiGraph GT1M, activPAL3 Micro, and activPAL, with cut--points also being developed using the SVM for the ActiGraph wGT3X--BT, activPAL3 Micro, and GENEActiv. For the GENEActiv, cut--points were developed for the activity monitor worn on both the dominant and non--dominant hand. The cut--points for all the activity monitors are shown in Table 4, along with the AUC, sensitivity, and specificity values for each developed cut--point.

For 1.5 METs, sensitivity and specificity ranged from 85.3%--90.9% and 89.9%--100%, respectively. For 3 METs, the ranges were 87.6%--94.5% and 87.6%--94.6%. Finally, the 6 METs ranges were 94.3%--98.2% and 94.3%--98.3%. Additionally, all of the developed cut--points had excellent AUC values (ie, >0.90).

3.2 | Cross--validation of developed cut-- points

The cut--points that were developed in the Development Group were subsequently cross--validated in the Cross--validation Group. The sensitivity and specificity of the cut--points were

TABLE 4 Cut--points for 1.5, 3, and 6 METs (including sensitivity and specificity values) for all the activity monitors developed using ROC analysis in the Development Group

Activity monitor	Axes	Unit	Epoch length (seconds)	AUC	Cut-point	Sensitivity	Specificity
1.5 METs							
ActiGraph wGT3X-BT	Vertical	Counts	15	0.926	0	0.866	0.909
ActiGraph wGT3X-BT	Vertical	Counts	60	0.926	1	0.866	0.909
ActiGraph wGT3X-BT	SVM	Counts	15	0.964	15	0.904	0.902
ActiGraph wGT3X-BT	SVM	Counts	60	0.964	61	0.904	0.902
ActiGraph GT1M	Vertical	Counts	15	0.927	0	0.853	1.000
ActiGraph GT1M	Vertical	Counts	60	0.927	1	0.853	1.000
activPAL3 Micro[a]	Vertical	Counts	15	—	—	—	—
activPAL3 Micro[a]	SVM	Counts	15	—	—	—	—
activPAL[a]	Vertical	Counts	15	—	—	—	—
GENEActiv-dominant	SVM	$g\ s^{-1}$	15	0.961	51	0.898	0.899
GENEActiv-non-dominant	SVM	$g\ s^{-1}$	15	0.972	47	0.909	0.911
3 METs							
ActiGraph wGT3X-BT	Vertical	Counts	15	0.990	397	0.942	0.944
ActiGraph wGT3X-BT	Vertical	Counts	60	0.991	1705	0.942	0.944
ActiGraph wGT3X-BT	SVM	Counts	15	0.991	627	0.942	0.944
ActiGraph wGT3X-BT	SVM	Counts	60	0.991	2504	0.942	0.944
ActiGraph GT1M	Vertical	Counts	15	0.989	427	0.943	0.944
ActiGraph GT1M	Vertical	Counts	60	0.989	1736	0.943	0.944
activPAL3 Micro	Vertical	Counts	15	0.991	5123	0.938	0.935
activPAL3 Micro	SVM	Counts	15	0.990	8873	0.937	0.938
activPAL	Vertical	Counts	15	0.994	3007	0.945	0.946
GENEActiv-dominant	SVM	$g\ s^{-1}$	15	0.959	68	0.876	0.876
GENEActiv-non-dominant	SVM	$g\ s^{-1}$	15	0.979	64	0.917	0.919
6 METs							
ActiGraph wGT3X-BT	Vertical	Counts	15	0.993	1028	0.943	0.943
ActiGraph wGT3X-BT	Vertical	Counts	60	0.997	4429	0.962	0.963
ActiGraph wGT3X-BT	SVM	Counts	15	0.996	1261	0.962	0.963
ActiGraph wGT3X-BT	SVM	Counts	60	0.996	5041	0.962	0.963
ActiGraph GT1M	Vertical	Counts	15	0.996	1084	0.964	0.960
ActiGraph GT1M	Vertical	Counts	60	0.996	4334	0.964	0.960
activPAL3 Micro	Vertical	Counts	15	0.999	12317	0.980	0.982
activPAL3 Micro	SVM	Counts	15	0.999	18791	0.982	0.983
activPAL	Vertical	Counts	15	0.999	6479	0.980	0.982
GENEActiv-dominant	SVM	$g\ s^{-1}$	15	0.999	142	0.980	0.982
GENEActiv-non-dominant	SVM	$g\ s^{-1}$	15	0.993	157	0.980	0.982

[a]No cut--points were developed for the activPAL3 Micro (vertical axis and SVM) and the activPAL (vertical axis) for 1.5 METs as the devices are able to differentiate between sitting and standing.

examined to determine their ability to correctly differentiate between PA intensities. This information is provided in Table 5.

The sensitivity and specificity values in the Cross--validation Group were all high (≥77.4%). 1.5 METs ranges were 86.3%--95.1% and 89.5%--100%%, 3 METs ranges were 90.3%--98.3% and 77.4%--91.2%, while 6 METs ranges were 93%--98% and 93.7%--100%. The developed cut--points for VPA appeared to be stronger than the other intensity domains.

3.3 | Mixed--model analysis and Bland--Altman plots

Regression equations for METs were created for all activity monitors, as well as the CCC and SEE for all equations.

TABLE 5 Cross--validation of the cut--points developed in the Development Group with sensitivity and specificity values in the Cross--validation Group

Activity monitor	Axes	Unit	Epoch length (seconds)	Sensitivity	Specificity
1.5 METs					
ActiGraph wGT3X-BT	Vertical	Counts	15	0.886	0.954
ActiGraph wGT3X-BT	Vertical	Counts	60	0.891	0.972
ActiGraph wGT3X-BT	SVM	Counts	15	0.943	0.954
ActiGraph wGT3X-BT	SVM	Counts	60	0.943	0.954
ActiGraph GT1M	Vertical	Counts	15	0.863	1.000
ActiGraph GT1M	Vertical	Counts	60	0.869	1.000
activPAL3 Micro[a]	Vertical	Counts	15	—	—
activPAL3 Micro[a]	SVM	Counts	15	—	—
activPAL[a]	Vertical	Counts	15	—	—
GENEActiv-dominant	SVM	$g\ s^{-1}$	15	0.940	0.909
GENEActiv-non-dominant	SVM	$g\ s^{-1}$	15	0.951	0.895
3.0 METs					
ActiGraph wGT3X-BT	Vertical	Counts	15	0.983	0.877
ActiGraph wGT3X-BT	Vertical	Counts	60	0.983	0.883
ActiGraph wGT3X-BT	SVM	Counts	15	0.983	0.896
ActiGraph wGT3X-BT	SVM	Counts	60	0.983	0.896
ActiGraph GT1M	Vertical	Counts	15	0.968	0.912
ActiGraph GT1M	Vertical	Counts	60	0.959	0.904
activPAL3 Micro	Vertical	Counts	15	0.947	0.890
activPAL3 Micro	SVM	Counts	15	0.955	0.896
activPAL	Vertical	Counts	15	0.903	0.897
GENEActiv-dominant	SVM	$g\ s^{-1}$	15	0.912	0.774
GENEActiv-non-dominant	SVM	$g\ s^{-1}$	15	0.947	0.787
6.0 METs					
ActiGraph wGT3X-BT	Vertical	Counts	15	0.979	0.937
ActiGraph wGT3X-BT	Vertical	Counts	60	0.979	0.954
ActiGraph wGT3X-BT	SVM	Counts	15	0.979	0.958
ActiGraph wGT3X-BT	SVM	Counts	60	0.979	0.954
ActiGraph GT1M	Vertical	Counts	15	0.979	0.949
ActiGraph GT1M	Vertical	Counts	60	0.980	0.950
activPAL3 Micro	Vertical	Counts	15	0.977	0.982
activPAL3 Micro	SVM	Counts	15	0.961	1.000
activPAL	Vertical	Counts	15	0.930	0.978
GENEActiv-dominant	SVM	$g\ s^{-1}$	15	0.977	0.951
GENEActiv-non-dominant	SVM	$g\ s^{-1}$	15	0.977	0.978

[a]No cut--points were validated for the activPAL3 Micro (vertical axis and SVM) and the activPAL (vertical axis) for 1.5 METs as the devices are able to differentiate between sitting and standing.

The regression equations, CCC, and SEE for all the activ-ity monitors are shown in Table 6. From the Bland--Altman plots, the mean bias and limits of agreement between activity monitor--predicted METs (based on the regression equations) and the K4b^2 measured METs are also presented in Table 6.

Using the above developed regression equations for each device to determine the MET value associated with a particular accelerometer output, the predicted MET values were compared to the K4b^2 criterion measure. All predicted METs (excluding the activPAL) had a positive mean bias, ranging from 0.16 to 0.35 METs. The limits of agreement for all of the devices ranged from −2.56 to 3.00 METs. The Bland--Altman plots are presented in Figures 2 (all ActiGraph related outputs) and 3 (all other devices). The middle line represents the mean bias, with the top and bottom lines depicting the limits of agreement (upper and lower limit).

TABLE 6 Regression equations (developed using the Development Group), concordance correlation coefficient (CCC), standard error estimate (SEE), mean bias, and limits of agreement for predicted METs for all activity monitors

Device	Axes	Unit	Regression equation	CCC	SEE	Mean bias	Upper limit	Lower limit
ActiGraph wGT3X-BT	Vertical	Counts	$1.177671 + (0.004271 \times \text{counts}.\ 15\ s^{-1})$	0.92	1.21	0.30	2.77	−2.16
ActiGraph wGT3X-BT	Vertical	Counts	$1.177770 + (0.000988 \times \text{counts}.\ 60\ s^{-1})$	0.95	0.98	0.12	2.45	−2.22
ActiGraph wGT3X-BT	SVM	Counts	$1.054381 + (0.003644 \times \text{counts}.\ 15\ s^{-1})$	0.95	0.90	0.13	2.41	−2.15
ActiGraph wGT3X-BT	SVM	Counts	$1.055121 + (0.000911 \times \text{counts}.\ 60\ s^{-1})$	0.95	0.90	0.14	2.44	−2.16
ActiGraph GT1M	Vertical	Counts	$1.195210 + (0.003995 \times \text{counts}.\ 15\ s^{-1})$	0.95	0.98	0.12	2.38	−2.14
ActiGraph GT1M	Vertical	Counts	$1.189174 + (0.001000 \times \text{counts}.\ 60\ s^{-1})$	0.95	0.98	0.11	2.38	−2.15
activPAL3 Micro	Vertical	Counts	$0.991163 + (0.000457 \times \text{counts}.\ 15\ s^{-1})$	0.94	0.99	0.16	2.36	−2.04
activPAL3 Micro	SVM	Counts	$0.961574 + (0.000290 \times \text{counts}.\ 15\ s^{-1})$	0.94	0.99	0.14	2.43	−2.15
activPAL	Vertical	Counts	$1.073557 + (0.000718 \times \text{counts}.\ 15\ s^{-1})$	0.97	0.76	−0.03	2.49	−2.56
GENEActiv-D	SVM	$g\ s^{-1}$	$1.145753 + (0.022043 \times g\ s^{-1} \cdot 15\ s^{-1})$	0.95	0.90	0.35	3.00	−2.31
GENEActiv-ND	SVM	$g\ s^{-1}$	$1.154510 + (0.022261 \times g\ s^{-1} \cdot 15\ s^{-1})$	0.94	1.04	0.20	2.47	−2.07

4 | DISCUSSION

The primary aim of this study was to develop and validate cut--points from five different activity monitors for LIPA, MPA, and VPA in an adult population using the same study protocol and analysis methodology. This study pre-sents a collection of cut--points for some of the most com-monly employed activity monitors in the measurement of PA, which have high levels of sensitivity and specificity for accurately detecting activities of LIPA, MPA, and VPA intensities. The use of accelerometry to measure and classify PA is an ever--expanding field. The number of activity monitors, cut--points, device wear locations, etc., available to researchers makes it difficult to compare data from different studies.[8,9] The cut--points developed in this study may offer researchers a means of comparing data from previous studies that have utilized different activity monitors. This could potentially be done by allowing researchers to reanalyze preexisting data that were collected with one of the included activity monitors using the cut--points developed in this study. This could allow researchers to cross--compare data collected using different devices. The collection of raw acceleration signal and the development of pattern recognition and machine learning are proposed as solutions to the aforementioned issues. While different machine learning/pattern recognition approaches have been developed, with some showing good performances, widespread application has been difficult due to the high cost and the need for multiple sensors to be attached to the body.[28]

When validated in the Cross--validation Group, the developed cut--points all showed high levels of sensitivity and specificity for 1.5, 3, and 6 METs, respectively. No cut-- points were developed for 1.5 METs for the activPAL3 Micro (vertical axis and SVM) and the activPAL (vertical axis) due to activPAL devices being able to differentiate between sitting, standing, and moving. As the activPAL devices are able to accurately quantify the amount of time spent sitting,[29] there is no need to develop a cut--point for determining sed-entary time. Upon examining the devices in closer detail, the ActiGraph wGT3X--BT and the ActiGraph GT1M had the same or similar 15 second epoch vertical axis cut--points for 1.5 METs (both 0), 3 METs (397 vs 427), and 6 METs (1028 vs 1084). The 60 second epoch cut--points for these two devices were also very similar; 1 vs 1, 1705 vs 1736, and 4429 vs 4334. The count output from the vertical axis of the ActiGraph wGT3X--BT and ActiGraph GT1M should be similar, but not the exact same as the devices use differ-ent sensors/accelerometers (Personal correspondence with ActiGraph). Additionally, as the two ActiGraph devices were worn on opposite hips, there is the potential for small differences to be observed. Vaha--Ypya et al. observed small differences between two similar accelerometers placed on opposite hips during a continuous locomotive test, due in part to the participants walking/running in a counter--clockwise direction (similar to the present study) throughout the test.[8,9] The differences between the vertical axis cut--points and the SVM cut--points for the ActiGraph wGT3X--BT became less pronounced as the activity intensity increased for both the 15 second epochs (0 vs 15, 397 vs 627, and 1028 vs 1261) and the 60 second epochs (1 vs 61, 1705 vs 2504, and 4429 vs 5041).

When comparing the activPAL3 Micro and activPAL vertical axis cut--points, the activPAL3 Micro cut--points for 3 and 6 METs were almost double those of the activPAL cut-- points, but they both had similar sensitivity and specificity values. Again, while the SVM cut--points of the activPAL3 Micro were higher than the vertical axis cut--points, the

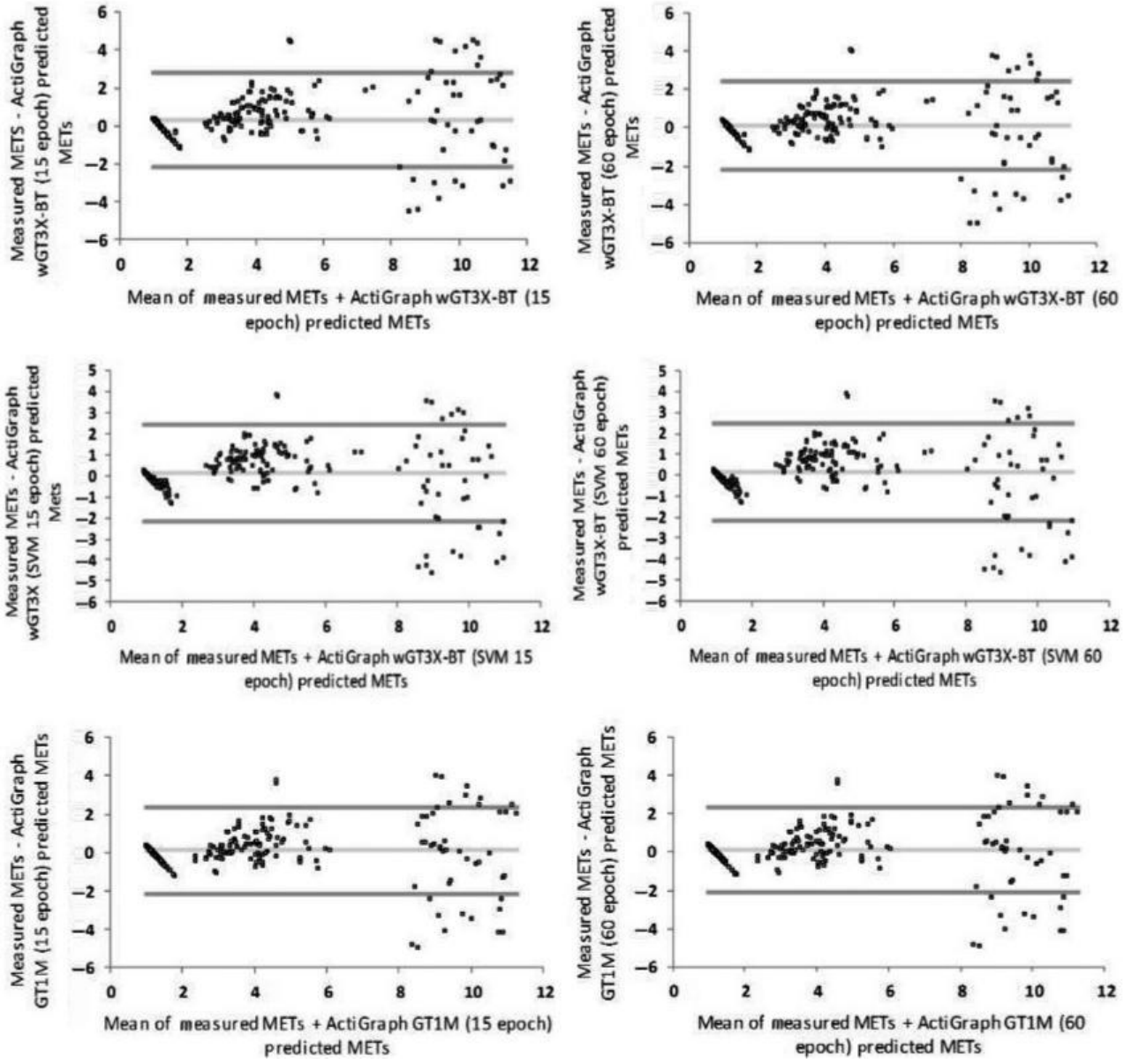

FIGURE 2 Bland--Altman plots for all ActiGraph predicted MET values based on the developed predictive equations

sensitivity and specificity values were almost identical. This suggests that cut--points for the activPAL devices are device specific. For the GENEActiv devices, the GENEActiv--dominant and GENEActiv--non--dominant produced differing cut--points across the different intensity domains (51 vs 47, 68 vs 64 and 142 vs 157). The largest noticeable difference for the GENEActiv devices was seen in the specificity values for 3 METs (0.774 and 0.787). Compared to the specificity val-ues for 1.5 METs (0.909 and 0.895) and 6 METs (0.951 and 0.978), the lower specificity values seen at 3 METs may sug-gest that the GENEActiv devices are less capable of detecting when a 3 MET intensity has not been reached, which may result from a large amount of upper limb movement with a disproportional metabolic cost.

A noticeable trend is the higher sensitivity and specificity values seen for VPA activities across all of the devices. Activities of 6 METs or more generally have greater acceler-ations or movements, thus enabling the activity monitors to correctly identify such intense movements more easily. Of the activities included in this study, jogging was by far the most intense activity (average MET value of 9.72). As the energy expenditure for jogging was more than double that of the next closest activity (fast walking, 4.07 METs), the devices had no difficulty differentiating between MPA and VPA. If there had been an intermediary activity eliciting approximately 6 METs, potentially it would have been harder to separate the different activity intensities. As the majority of researchers combine MPA and VPA (MVPA) together, potential issues in differentiating between the two intensities may not be problematic.

The differences in cut--points across the activity monitors (and between different activity monitor models from the same

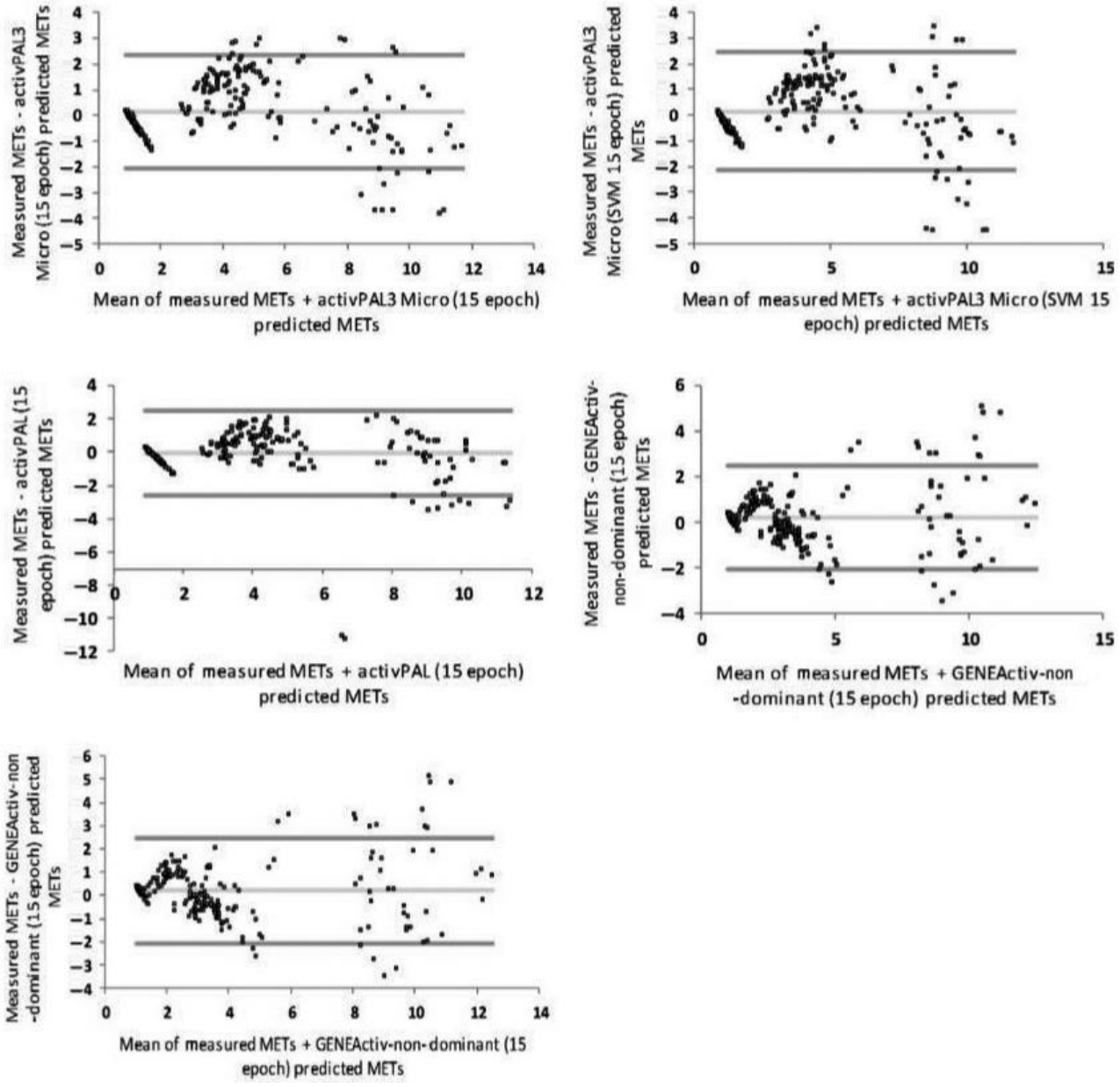

FIGURE 3 Bland--Altman plots for all other device predicted MET values based on the developed predictive equations

manufacture) could be partially due to the different sensors, accelerometers, and proprietary software that the activity monitors utilize. For the activity monitors in which cut--points for 15 and 60 second epochs were developed (both ActiGraph devices), the 60 second epoch was almost always four times that of the 15 second epoch. The slight differences observed could be due to how the proprietary algorithms of the soft-ware (ActiLife v6.11.4) extract and sort the accelerometry data. Wear location may also explain some of the differences seen between the included activity monitors. Depending on the task, specific body parts move more than others, thus pro-ducing more accelerations, and therefore have a higher count value.

Based on the AUC standards described by Metz, [25] all of the cut--points that were developed for the included activity monitors had AUC values of 0.926 or higher, suggesting that at least 92.6% of the time, the developed cut--points are able to correctly identify the activity intensity. The highest AUC values were seen for 6 METs. As alluded to previously, the more vigorous movements associated with this intensity make it easier to determine. Compared to ROC values produced by Vaha-Ypya et al.,[8,9] the AUCs from the current study are similar to those reported for 3 METs (0.971) and 6 METs (0.995), respectively. The AUCs from the Vaha--Ypya et al. study were developed based on a continuous locomotion test, while the AUCs from the current study included daily and locomotive tasks. In addition, the AUCs were developed across the intensity spectrum (LIPA--VPA).

Previous studies have developed and/or validated cut--points for some of the monitors included in this study:

ActiGraph GT3X SVM in adults,[23] GENEActiv in adults,[24] and activPAL in adolescents.[5] The previous cut--points developed for the ActiGraph GT3X and the GENEActiv were substantially higher than those developed here. This could be in part due to the use of the standard 3.5 mL kg min^{-1} which is equal to 1 MET for classifying activity intensity,[23,24] only including locomotive activities[23] or including a large range of hand dominant tasks.[24] No such cut--points have previously been developed for activPAL activity monitors for use in adult populations. The development of cut--points for the activPAL activity monitors may be important as it enable researchers to extract more activity intensity information from the activPAL and activPAL3 Micro data. While the software for both activity monitors allows the researcher to view the data based on events (sitting/lying, standing and stepping), it also estimates the metabolic cost of the activity that the person is engaging in. This is done using an algorithm based on step count. However, this has been shown to be problematic,[30,31] and the proposed use of cut--points will allow users of the activPAL activity monitors to quantify time spent in each physical activity band using similar methodology to that employed for other activity monitors.

Machine learning and pattern recognition may be a measurement option of PA in the future, as recent work by Miao and colleagues has highlighted that using the built--in sensors of smartphones to recognize typical physical activities without any firm body attachment may be feasible.[28] While the use of non--fixed attachments may become feasible to recognize different activities, it is critical that researchers continue to examine the associations between physical activity behaviors and health. Although the use of cut--points has its limitations,[32] it is currently the most valid and reliable method of examining accelerometer output available to researchers.[23] Until pattern recognition and machine learning are at the level of identifying a much greater number of activities, the use of cut--points will likely continue.

The main strength of this study is the simultaneous validation of five PA monitors in a single study. The comparison of validation studies is difficult due to the different methodologies employed by researchers. The use of an identical protocol for determining the cut--points for all of the activity monitors should improve the accuracy of cross--comparisons between datasets that have employed different activity -monitors. The measurement of individuals' RMR, rather than using the standard conversion of 3.5 mL kg min^{-1} as 1 MET, is another strength of this study. Both everyday tasks and locomotive tasks were included, to reflect the daily living habits of the general population. ROC analysis was chosen as the method to develop the cut--points, as it has the advantage of allowing the researcher to select cut--points that maximize sensitivity at the cost of specificity or vice versa.[16] As many newer accelerometers are triaxial, this study also developed cut--points using the SVM where possible. It is likely that triaxial accelerometers are superior to uniaxial accelerometers for predicting energy expenditure.[33] The greater sensitivity of the triaxial accelerometers may be advantageous at the lower end of the PA spectrum. The limitations of the study must be acknowledged. As the cut--points from the Development Group were cross--validated in the Cross--validation Group that performed the same activities, there is the potential for bias which could exaggerate the accuracy of the cut--points.[34] Also, the participants included in this study may not reflect the general population, as the majority had BMIs that were in the normal classification. For researchers working with overweight/obese cohorts, the development of population spe-cific cut--points may improve accuracy for those populations. Additionally, these cut--points were developed for an adult population and are therefore not suitable for children and/ or adolescents. While the inclusion of more activities (both sedentary and non--sedentary) may have further improved the study, the activities included were selected to reflect activities that individuals would perform on a daily basis. For example, walking is one of the most common PAs undertaken by adults in Europe, with 66.8% of people reported walking for at least 10 minutes consecutively on five or more occasions a week.[35]

4.1 | Perspective

This is the first study to develop and validate cut--points for a combination of five PA monitors (including use of activity monitors on dominant and non--dominant hands) using the same protocol and analysis methods. This study may provide researchers with the potential to cross--compare findings from different studies that have not employed the same activity monitoring devices. This potential cross--comparison of data may enable researchers to draw more powerful conclusions between PA behaviors and indices of health. The cut--points reported here may also have the potential to be used in pro-spective data collections where different activity monitors have been employed.

ACKNOWLEDGEMENTS

The authors would like to thank Ms. Aine O'Sullivan, Ms. Celia O'Driscoll, Ms. Colleen Molloy, Dr Christina Dillon, Ms. Grainne Hayes, Ms. Hannah McCormack, and Ms. Una Delahunt for their contribution to the data collection process.

CONFLICT OF INTERESTS

The authors declare no conflict of interest.

REFERENCES

1. Haskell WL, Blair SN, Hill JO. Physical activity: health outcomes and importance for public health policy. *Prev Med*. 2009;49:280–282.

2. Lee IM, Shiroma EJ. Using accelerometers to measure physical activity in large--scale epidemiological studies: issues and challenges. *Br J Sports Med*. 2014;48:197–201.
3. Warburton DE, Nicol CW, Bredin SS. Health benefits of physical activity: the evidence. *CMAJ*. 2006;174:801–809.
4. Healy GN, Dunstan DW, Shaw JE, Zimmet PZ, Owen N. Objectively mea-sured sedentary time and light--intensity physical activity are independently associated with components of the metabolic syndrome: the AusDiab study. *Diabetologia*. 2007;50:S67–S68.
5. Dowd KP, Harrington DM, Donnelly AE. Criterion and concurrent validity of the activPAL professional physical activity monitor in adolescent females. *PLoS One*. 2012;7:e47633.
6. Warren JM, Ekelund U, Besson H, Mezzani A, Geladas N, Vanhees L. Assessment of physical activity—a review of methodologies with reference to epidemiological research: a report of the exercise physiology section of the European Association of Cardiovascular Prevention and Rehabilitation. *Eur J Cardiovasc Prev Rehabil*. 2010;17:127–139.
7. Pulakka A, Cheung YB, Ashorn U, et al. Feasibility and validity of the ActiGraph GT3X accelerometer in measuring physical activity of Malawian toddlers. *Acta Paediatr*. 2013;102:1192–1198.
8. Vaha-Ypya H, Vasankari T, Husu P, et al. Validation of cut--points for eval-uating the intensity of physical activity with accelerometry--based mean amplitude deviation (MAD). *PLoS One*. 2015;10:e0134813.
9. Vaha-Ypya H, Vasankari T, Husu P, Suni J, Sievanen H. A universal, accu-rate intensity--based classification of different physical activities using raw data of accelerometer. *Clin Physiol Funct Imaging*. 2015;35:64–70.
10. Cain KL, Conway TL, Adams MA, Husak LE, Sallis JF. Comparison of older and newer generations of ActiGraph accelerometers with the normal filter and the low frequency extension. *Int J Behav Nutr Phys Act*. 2013;10:51.
11. Wanner M, Martin BW, Meier F, Probst-Hensch N, Kriemler S. Effects of filter choice in GT3X accelerometer assessments of free--living activity. *Med Sci Sports Exerc*. 2013;45:170–177.
12. McLaughlin JE, King GA, Howley ET, Bassett DR Jr, Ainsworth BE. Validation of the COSMED K4 b2 portable metabolic system. *Int J Sports Med*. 2001;22:280–284.
13. Schrack JA, Simonsick EM, Ferrucci L. Comparison of the Cosmed K4b(2) portable metabolic system in measuring steady--state walking energy expen-diture. *PLoS One*. 2010;5:e9292.
14. Mansoubi M, Pearson N, Clemes SA, et al. Energy expenditure during com-mon sitting and standing tasks: examining the 1.5 MET definition of seden-tary behaviour. *BMC Public Health*. 2015;15:516.
15. Kozey S, Lyden K, Staudenmayer J, Freedson P. Errors in MET estimates of physical activities using 3.5 ml × kg(−1) × min(−1) as the baseline oxygen consumption. *J Phys Act Health*. 2010;7:508–516.
16. Bassett DR Jr, Rowlands A, Trost SG. Calibration and validation of wearable monitors. *Med Sci Sports Exerc*. 2012;44:S32–S38.
17. Welk GJ. Principles of design and analyses for the calibration of accelerometry--based activity monitors. *Med Sci Sports Exerc*. 2005;37:S501–S511.
18. Calabro M, Lee JM, Saint-Maurice PF, Yoo H, Welk GJ. Validity of physical activity monitors for assessing lower intensity activity in adults. *Int J Behav Nutr Phys Act*. 2014;11:119.
19. Grant PM, Ryan CG, Tigbe WW, Granat MH. The validation of a novel activity monitor in the measurement of posture and motion during everyday activities. *Br J Sports Med*. 2006;40:992–997.
20. Stanton R, Guertler D, Duncan MJ, Vandelanotte C. Validation of a pouch-- mounted activPAL3 accelerometer. *Gait Posture*. 2014;40:688–693.
21. Kilding AE, Jones AM. V02 'overshoot' during moderate--intensity exercise in endurance--trained athletes: the influence of exercise modality. *Respir Physiol Neurobiol*. 2008;160:139–146.
22. Compher C, Frankenfield D, Keim N, Roth-Yousey L. Best practice methods to apply to measurement of resting metabolic rate in adults: a systematic review. *J Am Diet Assoc*. 2006;106:881–903.
23. Sasaki JE, John D, Freedson PS. Validation and comparison of ActiGraph activity monitors. *J Sci Med Sport*. 2011;14:411–416.
24. Esliger DW, Rowlands AV, Hurst TL, Catt M, Murray P, Eston RG. Validation of the GENEA Accelerometer. *Med Sci Sports Exerc*. 2011;43:1085–1093.
25. Metz CE. Basic principles of ROC analysis. *Semin Nucl Med*. 1978;8: 283–298.
26. Rosenberger ME, Haskell WL, Albinali F, Mota S, Nawyn J, Intille S. Estimating activity and sedentary behavior from an accelerometer on the hip or wrist. *Med Sci Sports Exerc*. 2013;45:964–975.
27. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;1:307–310.
28. Miao F, He Y, Liu J, Li Y, Ayoola I. Identifying typical physical activity on smartphone with varying positions and orientations. *Biomed Eng Online*. 2015;14:32.
29. Kozey-Keadle S, Libertine A, Lyden K, Staudenmayer J, Freedson PS. Validation of wearable monitors for assessing sedentary behavior. *Med Sci Sports Exerc*. 2011;43:1561–1567.
30. Harrington DM, Welk GJ, Donnelly AE. Validation of MET estimates and step measurement using the ActivPAL physical activity logger. *J Sports Sci*. 2011;29:627–633.
31. Powell C, Carson BP, Dowd KP, Donnelly AE. The accuracy of the SenseWear Pro3 and the activPAL3 Micro devices for measurement of energy expenditure. *Physiol Meas*. 2016;37:1715–1727.
32. Trost SG, Loprinzi PD, Moore R, Pfeiffer KA. Comparison of accelerometer cut points for predicting activity intensity in youth. *Med Sci Sports Exerc*. 2011;43:1360–1368.
33. Hanggi JM, Phillips LR, Rowlands AV. Validation of the GT3X ActiGraph in children and comparison with the GT1M ActiGraph. *J Sci Med Sport*. 2013;16:40–44.
34. Crouter SE, Horton M, Bassett DR Jr. Validity of ActiGraph child--specific equations during various physical activities. *Med Sci Sports Exerc*. 2013;45:1403–1409.
35. Hallal PC, Andersen LB, Bull FC, Guthold R, Haskell W, Ekelund U. Global physical activity levels: surveillance progress, pitfalls, and prospects. *Lancet*. 2012;380:247–257.