Impact of ActiGraph Sampling Rate and Inter-Monitor Comparability on Measures of Physical Activity in Adults

Clevenger, K.A., Brønd, J.C., Arvidsson, D., Montoye, A.H.K., Mackintosh, K.A., McNarry, M.A. & Pfeiffer, K.A.

Date of Submission: 3/15/2021

Abstract

ActiGraph is a commonly used, research-grade accelerometer brand, but there is little information regarding inter-monitor comparability of newer models. Additionally, whilst sampling rate has been shown to influence accelerometer metrics, its influence on measures of free-living physical activity has not been directly studied. **Purpose:** To examine differences in physical activity metrics due to inter-monitor variability and chosen sampling rate. **Methods:** Adults (n=20) wore two hip-worn ActiGraph wGT3X-BT monitors for one week, with one accelerometer sampling at 30 Hz and the other at 100 Hz, which was downsampled to 30 Hz. Activity intensity was classified using vector magnitude (VM), Euclidean Norm Minus One (ENMO), and Mean Amplitude Deviation (MAD) cut-points. Equivalence testing compared outcomes. **Results:** There was a lack of inter-monitor equivalence for ENMO, time in sedentary/light- or moderate-intensity activity according to ENMO cut-points, and time in moderate-intensity activity according to MAD cut-points. Between sampling rates, differences existed for time in moderate-intensity activity according to VM, ENMO and MAD cut-points, and time in sedentary/light-intensity activity according to ENMO cut-points. While mean differences were small (0.1-1.7 percentage points), this would equate to differences in moderate-to-vigorous-intensity activity over a 10-h wear-day of 3.6 (MAD) to 10.8 (ENMO) $min \cdot day^{-1}$ for inter-monitor comparisons or 3.6 (VM) to 5.4 (ENMO) $min \cdot day^{-1}$ for sampling rate. **Conclusions:** Epoch-level inter-monitor differences were larger than differences due to sampling rate, but both may impact outcomes such as time spent in each activity intensity. ENMO was the least comparable metric between monitors or sampling rates.

keywords: accelerometry, reliability, adult, methodology

Introduction

Since the 1980s, accelerometers have been used to estimate free-living energy expenditure and physical activity levels (Wong et al., 1981). ActiGraph accelerometers are the most widely-used brand of research-grade monitors (Migueles et al., 2017; Montoye et al., 2016) and have been used in large-scale interventions (Stevens et al., 2005), national surveillance efforts, such as the National Health and Nutrition Examination Study (NHANES) (Troiano et al., 2008), and in clinical trials (US National Library of Medicine, 2021). ActiGraph monitors have historically measured, filtered, and rectified acceleration (in g's) to generate 'activity counts' that are intended to be a measure of physical activity intensity (Chen & Bassett, 2005; John & Freedson, 2012). In recent models, including the GT3X, GT3X+, wGT3X-BT, and GT9X (Link) monitors, both 'raw' acceleration and activity count data are stored, and the user is able to select the sampling rate, in 10 Hz increments, from 30 to 100 Hz (John & Freedson, 2012). Since these functionalities were introduced, several researchers have focused on the development of acceleration-based metrics, and they have used a variety of sampling rates (de Almeida Mendes et al., 2018; Migueles et al., 2017).

Recent research has suggested that ActiGraph sampling rate impacts the conversion of acceleration into activity counts (Brønd & Arvidsson, 2015; Clevenger et al., 2019). Specifically, a study in adults showed that an ActiGraph monitor using a sampling rate of 40 or 100 Hz resulted in the generation of additional activity counts compared to a second monitor collecting at 30 Hz during a semi-structured walking and running protocol (Brønd & Arvidsson, 2015). While Brønd et al. (2015) reported sampling rate was not an issue when using a multiple of 30 Hz, a recent review indicates that besides 30 Hz, users most often select a 100 Hz sampling rate (the maximum available for ActiGraph; Migueles et al., 2017). A limitation of prior research is that results could be, at least in part, attributable to inter-monitor variability introduced by the use of multiple monitors worn side-by-side. While there is evidence for inter-monitor comparability of older generations of ActiGraph devices (Aadland & Ylvisåker, 2015; Esliger & Tremblay, 2006; Jarrett et al., 2015; Ozemek et al., 2014; Santos-Lozano et al., 2013; Silva et al., 2010), there remain small differences in both acceleration (Montoye et al., 2018) and activity counts (Loprinzi & Smith, 2017; Ozemek et al., 2014) even in newer model monitors, potentially due to slight differences in monitor orientation or placement.

To account for the potential influence of using two monitors to examine the impact of sampling rate, a study in children utilized only one ActiGraph monitor that collected data at 100 Hz, which was later downsampled to 30 Hz (Clevenger et al., 2019). This study demonstrated that collected data, particularly activity counts, were still affected by sampling rate, even after eliminating inter-monitor differences (Clevenger et al., 2019). Specifically, it was estimated that approximately 15 minutes over the course of a 10-h wear-day could be classified as a different activity intensity when using a 100 Hz sampling rate compared to 30 Hz. While this difference would have a clear impact on the estimation of habitual physical activity, it is pertinent to note that this was an extrapolation based on a laboratory-based protocol involving a high level of moderate- or vigorous-intensity physical activity and a low level of sedentary time or light-intensity activity. Therefore, the actual impact of sampling rate on measures of habitual physical activity remains to be elucidated, particularly in adults. Understanding the effect of sampling rate on habitual physical activity measurement is important given that this information informs methodological decisions, comparability between studies using different sampling rates, and understanding of existing data, including national level physical activity data (Troiano et al., 2014).

While inter-monitor differences are generally considered an acceptable source of error, the free-living comparability of the ActiGraph wGT3X-BT in adults has not been established. Perhaps more importantly, the free-living comparability of the ActiGraph in general has not been well researched for measuring acceleration-based metrics, like Euclidean Norm Minus One (ENMO) (Bakrania et al., 2016; van Hees et al., 2014; van Hees et al., 2013) or Mean Amplitude Deviation (MAD) (Aittasalo et al., 2015; Bakrania et al., 2016; Vähä-Ypyä et al., 2015). In children and adults, the acceleration-based metric ENMO has demonstrated poorer reliability than MAD, vector magnitude (VM) counts, or VM acceleration (Clevenger et al., 2020a, 2020b). Therefore, more research is needed on inter-monitor comparability of recent ActiGraph models overall and particularly for acceleration-based metrics, as this may also impact measures of habitual physical activity, further compounding differences due to data collection decisions. The purpose of the present study was to partition the differences in habitual physical activity as measured by two monitors into differences attributable to inter-monitor variability vs. those resulting from the chosen sampling rate.

Methods

A convenience sample of college students was recruited for participation in this study by word of mouth and email after the University's Institutional Review Board approved this protocol. Following provision of written informed consent, an elastic belt was fitted around each participant's waist, with two ActiGraph wGT3X-BT accelerometers positioned over the right hip. To limit inter-monitor differences, only two pairs of accelerometers were used in this study (i.e., four monitors in total). The monitor pair assigned to the participant and the placement order (which monitor was medial or lateral) were randomized, and both monitors were worn for all waking hours for seven days except while sleeping, swimming, showering, or participating in other water-based activities.

Accelerometers (firmware 1.9.2) were initialized to collect acceleration data (in *g*'s), with one monitor randomly selected to sample at 30 Hz and the other at 100 Hz. Following data collection, data were downloaded as .gt3x files using ActiLife (version 6.13.3, ActiGraph, Pensacola, FL). The 100 Hz data were resampled to 30 Hz by converting the original .gt3x files to .wav files using Java software (Oracle Corp., Redwood Shores, CA) and then using the *resample* function available in MATLAB (MathWorks Inc., Natwick, MA). Once resampled, the 30 Hz files were converted back to .gt3x files using the Java program (Clevenger et al., 2019). Thus, there were three data files per participant: i) collected 30 Hz data; ii) collected 100 Hz data; and iii) downsampled 100 to 30 Hz data. This enabled the partitioning of differences between monitors collecting at 100 and 30 Hz in to inter-monitor differences (30 vs. 30 Hz data from the two monitors) and intra-monitor differences (100 vs. 30 Hz data from the monitor originally collecting 100 Hz data). All subsequent processing steps were conducted for all three of these '.gt3x' files.

Data were loaded into R (version 1.1.463; Vienna, Austria) as .csv files using the *AGread* package (version 0.2.0) (Hibbing, 2018). Acceleration data were auto-calibrated (van Hees et al., 2019; van Hees et al., 2014) and calibration information can be found in Supplementary Table 1. ENMO was calculated over 5-s epochs, in line with previous research (Migueles et al., 2019). ENMO was calculated as the square root of the sum of the squared values of the auto-calibrated acceleration signals in each axis, minus 1, with negative values rounded up to zero (van Hees et al., 2013). Activity intensity of each epoch was classified using Hildebrand et al. (2014) ENMO cut-points as sedentary/light, moderate, or vigorous. The *acc* package (version 1.3.3) was used to calculate MAD in 5-s epochs; MAD measures the typical distance between the square root of the sum of the squared values of the raw acceleration (not auto-calibrated) signals from each axis and the mean value for a given time period (Aittasalo et al., 2015; Bakrania et al., 2016; Vähä-Ypyä et al., 2015; Vähä-Ypyä et al., 2015). MAD values were classified as sedentary/light, moderate, or vigorous using the Vähä-Ypyä et al. (2015) cut-points. For activity count data, VM was calculated over a 60-s epoch as the square root of the sum of the squares of activity counts from each axis, and activity intensity was classified as sedentary/light, moderate, or vigorous, using cut-points developed by Sasaki et al. (2011). A 60-s epoch was used for VM as this is the most commonly used epoch for this metric (Migueles et al., 2017) and because a 60-s epoch was used for cut-point development (Sasaki et al., 2011). However, data were also analyzed using a 5-s epoch to be consistent with the epoch used for ENMO and MAD as exploratory analysis (data not shown). Only triaxial metrics were included in the present analysis to account for small potential differences in orientation between monitors that would impact single-axis metrics.

Count and acceleration data from the same monitor were aligned based on timestamp, and non-wear-time was classified as continuous strings of 20 minutes of zero counts in the vertical axis using the *accelerometry* package (version 3.1.2) (Van Domelen & Pittard, 2014). Peeters et al. (2013) reported this non-wear classification resulted in the lowest amount of misclassification compared to self-report log books in adults. The three files per participant were then aligned based on timestamp, and only times classified as wear-time from all three files included. As the goal was not to produce estimates of habitual physical activity levels, no minimum wear-time per day was required, but participants were required to have at least 10 hours of wear data over the seven-day wear-period to be included in the subsequent analysis. This duration is in line with previous monitor comparison studies (Lee et al., 2013; Ried-Larsen et al., 2012; Vanhelst et al., 2012) and is longer than the protocols used in currently available studies regarding the impact of sampling rate (Brønd & Arvidsson, 2015; Clevenger et al., 2019).

At the epoch-level, Pearson's *r* correlation coefficients and mean absolute difference and percent difference were calculated between 100 Hz and downsampled 30 Hz data (intra-monitor) and between downsampled and collected 30 Hz data (inter-monitor). Correlation coefficients were classified as no (r<0.20), low (r=0.20-0.39), moderate (r=0.40-0.59), moderately high (r=0.60-0.79), or high ($r \geq 0.80$) relationship (Safrit & Wood, 1995). Bland Altman plots (1986) and bias were generated using the *blandr* package (version 0.5.1). Using the *irr* package (version 0.84.1) (Gamer et al., 2012), epoch-level agreement between activity intensities as classified using ENMO, MAD, and VM cut-points was assessed using weighted Kappa, which accounts for activity intensities being ordered, and percent agreement. Kappa coefficients were interpreted as no ($\kappa \leq 0.20$), minimal (κ=0.21-0.39), weak (κ=0.40-0.59), moderate (κ=0.60-0.79), strong (κ=0.80-0.90), or almost perfect (κ>0.90) agreement (McHugh, 2012). Confusion matrices were also used to compare activity intensity classification between datasets.

Mean ENMO, MAD, VM, and percent of wear-time spent in each physical activity intensity according to the ENMO, MAD, and VM metrics were calculated for each participant. Pearson's *r* correlation coefficient, mean absolute difference and percent difference were calculated for these collapsed data. Using the R package *TOSTER* (version 0.3.4) (Lakens, 2017), two, one-sided tests of equivalence (TOST) were used to compare mean VM, ENMO, MAD, and percent of wear-time spent in each activity intensity per participant across the three data files. In this method, 90% confidence intervals around the mean difference for each variable are constructed and if the confidence interval does not overlap or exceed the equivalence bounds, then the monitors are considered equivalent ($p<0.05$). Similar to prior research (Clevenger et al., 2020a, 2020b), equivalence bounds were initially set as 5% of the mean value for each variable. However, for percent of wear-time spent in moderate and vigorous activity, the equivalence bounds were modified to 0.5 percentage points, since using the 5% of the mean criterion resulted in extremely narrow bounds that have little practical meaning. Finally, mean absolute differences in percent time spent in each activity intensity were used to estimate inter- and intra-monitor differences in min·day^{-1} in each intensity based on a 10-h wear-day.

Results

Twenty adults (18-30 y of age) completed this study, with an average of 73.3 ± 23.2 hours of wear-time. Although not required, all participants had four or more wear days. Epoch level comparisons are shown in Table 1 (inter-monitor) and Table 2 (intra-monitor), while data collapsed to mean value per participant and percent time spent in each activity intensity are reported in Table 3 (inter-monitor) and Table 4 (intra-monitor). There were no notable differences in the results using VM at a 60-s or 5-s epoch, so only results using a 60-s epoch are reported (to align with prior research and the method in which the cut-points were developed).

At the epoch level, mean absolute percent differences ranged from 61.4% (VM) to 92.8% (ENMO) for inter-monitor differences and 38.3% (MAD) to 42.2% (ENMO) for intra-monitor differences. Correlations at the epoch level were classified as moderate-to-high for MAD (0.721-0.744) and ENMO (0.708-0.765), and high for VM (0.808-0.813) for both inter- and intra-monitor differences. Bland Altman plots are shown in Figure 1. Bias (lower, upper limits of agreement) for VM was 15.9 (-1709.3, 1741.1) counts·min^{-1} for the inter-monitor comparison and 46.6 (-1751.3, 1844.5) counts·min^{-1} for the intra-monitor comparison. Bias for ENMO was 2.8 (-86.5, 92.2) *mg* for the inter-monitor comparison and 0.3 (-81.0, 81.5) *mg* for the intra-monitor comparison. Bias for MAD was 2.0 (-137.7, 141.8) *mg* for the inter-monitor comparison and 2.3 (-133.0, 137.6) *mg* for the intra-monitor comparison.

The Kappa coefficient was classified as moderate for all metrics and comparisons (≥0.626). Confusion matrices for inter- and intra-monitor comparisons are shown in Tables 5 and 6, respectively. For both inter- and intra-monitor comparisons, the greatest agreement was for sedentary/light behavior, in which 95.9-98.3% of epochs were classified as sedentary/light by both datasets. For moderate- and vigorous-intensities, between 60.9-76.0% of epochs were classified identically between datasets.

When collapsed to mean values per participant, mean absolute percent differences ranged from 3.2% (VM) to 25.9% (ENMO) for inter-monitor differences (Table 3) and 5.8% (MAD) to 6.0% (VM) for intra-monitor differences (Table 4). Inter-monitor mean absolute percent differences in percent time spent in various activity intensities ranged from 0.6% (sedentary/light behavior according to VM and MAD cut-points) to 32.4% (vigorous activity according to ENMO cut-points; Table 4). Intra-monitor differences in percent time spent in various activity intensities ranged from 0.6% (sedentary/light behavior according to VM cut-points) to 30.9% (vigorous activity according to VM cut-points; Table 4). Correlation coefficients for the collapsed data were all classified as high (≥0.940; Tables 3 and 4), except inter-monitor comparisons for ENMO (r=0.468) and percent time spent in sedentary/light- (r=0.614) and moderate-intensity activity (r=0.605) according to ENMO cut-points, which were classified as moderate or moderate-to-high.

Results of the equivalence tests are shown in Tables 3 and 4. For inter-monitor comparisons, monitors were equivalent for all outcomes except ENMO, percent time spent in sedentary/light- or moderate-intensity activity according to ENMO cut-points, and percent time spent in moderate-intensity activity according to MAD cut-points. For intra-monitor comparisons, monitors were equivalent for all outcomes except percent time spent in moderate-intensity activity according to VM, ENMO, and MAD cut-points and percent time spent in sedentary/light-intensity activity according to ENMO cut-points.

When presented as min·day^{-1} (Tables 3 and 4), inter-monitor differences equated to 6.6 (MAD) to 30.0 (ENMO) min·day^{-1} across intensities. For moderate- to vigorous-intensity physical activity, specifically, differences would be 10.8 min·day^{-1} as classified by ENMO cut-points, compared to 3.6-4.8 min·day^{-1} for MAD or VM. Intra-monitor differences across all intensities were 7.2 (VM) to 10.8 (ENMO) min·day^{-1} or 3.6-5.4 min·day^{-1} of moderate- to vigorous-intensity physical activity when extrapolated to a 10-h wear-day.

Table 1. Mean absolute differences (± SD) and correlations between data types for epoch-level vector magnitude (VM), Euclidean Norm Minus One (ENMO), and Mean Amplitude Deviation (MAD) for the comparison between the collected 30 Hz data and the downsampled 30 Hz data (i.e., inter-monitor comparison)

	Mean Absolute Difference	**Mean Absolute Percent Difference**	**Pearson's *r***	**Kappa**	**Percent agreement**
VM (counts·min^{-1})	315.1 ± 822.0	61.4 ± 76.4	0.813	0.768	95.7
ENMO (*mg*)	21.2 ± 40.5	92.8 ± 73.3	0.708	0.626	92.8
MAD (*mg*)	22.9 ± 67.6	68.0 ± 74.2	0.721	0.650	92.1

Table 2. Mean absolute differences (± SD) and correlations between data types for epoch-level vector magnitude (VM), Euclidean Norm Minus One (ENMO), and Mean Amplitude Deviation (MAD) for the comparison between the collected 100 Hz data and the downsampled 30 Hz data (i.e., intra-monitor sampling rate comparison)

	Mean Absolute Difference	**Mean Absolute Percent Difference**	**Pearson's *r***	**Kappa**	**Percent agreement**
VM (counts·min^{-1})	274.3 ± 876.5	40.6 ± 71.9	0.808	0.788	96.2
ENMO (*mg*)	10.6 ± 40.1	42.2 ± 65.1	0.765	0.744	95.3
MAD (*mg*)	16.0 ± 67.2	38.3 ± 66.7	0.744	0.741	94.6

Table 3. Mean absolute differences and correlations between data types for individual-level vector magnitude (VM), Euclidean Norm Minus One (ENMO), and Mean Amplitude Deviation (MAD) for the comparison between the collected 30 Hz data and the downsampled 30 Hz data (i.e., inter-monitor comparison)

	Mean ± SD		**Mean Absolute Difference**	**Mean Absolute Difference in min·day^{-1}**	**Mean Absolute Percent Difference**	**Pearson's *r***	**Equivalence Bounds**	**Equivalence**
	30 Hz	**Downsampled**						
VM (counts·min^{-1})	761.7 ± 211.2	746.6 ± 220.2	22.6 ± 22.1	-	3.2 ± 3.5	0.992	-26.26, 3.99	Yes
Sedentary/ Light	91.3 ± 3.3	91.4 ± 3.4	0.6 ± 0.5	3.6 ± 3.0	0.6 ± 0.5	0.976	-0.41, 0.18	Yes
Moderate	7.1 ± 2.3	7.0 ± 2.3	0.6 ± 0.5	3.6 ± 3.0	8.7 ± 7.3	0.946	-0.21, 0.39	Yes
Vigorous	1.6 ± 1.6	1.5 ± 1.7	0.2 ± 0.2	1.2 ± 1.2	20.8 ± 40.0	0.985	-0.09, 0.15	Yes
ENMO (*mg*)	35.0 ± 9.8	33.0 ± 12.0	8.7 ± 7.4	-	25.9 ± 20.1	0.468	-6.51, 2.55	No
Sedentary/ Light	89.8 ± 3.3	89.9 ± 3.9	1.7 ± 2.7	10.2 ± 16.2	2.0 ± 3.1	0.614	-1.44, 1.11	No
Moderate	9.4 ± 3.0	9.2 ± 3.9	1.7 ± 2.7	10.2 ± 16.2	17.0 ± 22.7	0.605	-1.14, 1.38	No
Vigorous	0.9 ± 0.7	0.8 ± 0.7	0.1 ± 0.1	0.6 ± 0.6	32.4 ± 37.9	0.966	-0.03, 0.12	Yes
MAD (*mg*)	41.4 ± 10.5	39.5 ± 11.0	2.3 ± 3.6	-	5.8 ± 8.4	0.965	-3.07, -0.77	Yes
Sedentary/ Light	87.8 ± 3.6	88.2 ± 3.6	0.5 ± 0.6	3.0 ± 3.6	0.6 ± 2.7	0.984	-0.67, -0.15	Yes
Moderate	11.0 ± 3.3	10.7 ± 3.3	0.5 ± 0.5	3.0 ± 3.0	4.5 ± 6.1	0.984	0.11, 0.58	No
Vigorous	1.1 ± 0.9	1.1 ± 0.9	0.1 ± 0.1	0.6 ± 0.6	23.0 ± 30.5	0.986	-1.62, 0.12	Yes

VM classified using Sasaki et al. (2011) cut-points; ENMO classified using Hildebrand et al. (2014) cut-points; MAD classified using Vähä-Ypyä et al. (2015) cut-points; min·day^{-1} estimate based on 10-h wear-day

Table 4. Mean absolute differences (SD) and correlations between data types for individual-level vector magnitude (VM), Euclidean Norm Minus One (ENMO), and Mean Amplitude Deviation (MAD) for the comparison between the collected 100 Hz data and the downsampled 30 Hz data (i.e., intra-monitor comparison)

	Mean (SD)		**Mean Absolute Difference**	**Mean Absolute Difference in min·day^{-1}**	**Mean Absolute Percent Difference**	**Pearson's *r***	**Equivalence Bounds**	**Equivalence**
	100 Hz	**Downsampled**						
VM (counts·min^{-1})	790.8 ± 218.7	746.6 ± 220.2	44.2 ± 60.3	-	6.0 ± 7.9	0.962	-68.22, 20.25	Yes
Sedentary/ Light	90.9 ± 3.3	91.4 ± 3.4	0.6 ± 0.7	3.6 ± 4.2	0.6 ± 0.7	0.979	-0.84, -0.29	Yes
Moderate	7.4 ± 2.2	7.0 ± 2.3	0.4 ± 0.5	2.4 ± 3.0	5.9 ± 8.2	0.976	0.17, 0.57	No
Vigorous	1.7 ± 1.7	1.5 ± 1.7	0.2 ± 0.3	1.2 ± 1.8	30.9 ± 58.9	0.989	0.09, 0.29	Yes
ENMO (*mg*)	33.3 ± 12.0	33.0 ± 12.0	1.9 ± 2.7	-	5.9 ± 7.5	0.961	-1.66, 1.02	Yes
Sedentary/ Light	89.1 ± 5.0	89.9 ± 3.9	0.9 ± 1.9	5.4 ± 11.4	1.0 ± 2.4	0.943	-1.59, -0.11	No
Moderate	10.0 ± 5.1	9.2 ± 3.9	0.8 ± 1.9	4.8 ± 11.4	6.5 ± 10.4	0.948	0.02, 1.51	No
Vigorous	0.9 ± 0.7	0.8 ± 0.7	0.1 ± 0.1	0.6 ± 0.6	16.1 ± 34.3	0.983	0.03, 0.13	Yes
MAD (*mg*)	41.8 ± 11.2	39.5 ± 11.0	2.2 ± 2.6	-	5.8 ± 6.5	0.949	-3.74, -0.90	Yes
Sedentary/ Light	87.5 ± 3.9	88.2 ± 3.6	0.7 ± 1.3	4.2 ± 7.8	0.8 ± 1.5	0.940	-1.23, -0.19	Yes
Moderate	11.3 ± 3.5	10.7 ± 3.3	0.6 ± 1.1	3.6 ± 6.6	5.4 ± 9.5	0.944	0.13, 1.04	No
Vigorous	1.2 ± 1.0	1.1 ± 0.9	0.1 ± 0.2	0.6 ± 1.2	17.4 ± 32.8	0.982	0.05, 0.20	Yes

VM classified using Sasaki et al. (2011) cut-points; ENMO classified using Hildebrand et al. (2014) cut-points; MAD classified using Vähä-Ypyä et al. (2015) cut-points; min·day^{-1} estimate based on 10-h wear-day

Table 5. Confusion matrix showing agreement in activity intensity classifications between collected 30 Hz and downsampled 30 Hz data (inter-monitor comparison) based on Sasaki et al. (2011) vector magnitude cut-points in counts·min^{-1}, Hildebrand et al. (2014) Euclidean Norm Minus One (ENMO; *mg*) cut-points, and Vähä-Ypyä et al. (2015) mean amplitude deviation (MAD; *mg*) cut-points. The collected 30 Hz data served as the referent group and numbers represent percent of epochs within each activity intensity classified as that intensity according to the downsampled 30 Hz data.

	Downsampled Sasaki Classification		
30 Hz Sasaki Classification	Sedentary/Light	Moderate	Vigorous
Sedentary/Light	97.9	1.9	0.2
Moderate	26.5	71.7	1.8
Vigorous	15.1	8.9	76.0
	Downsampled Hildebrand Classification		
30 Hz Hildebrand Classification	Sedentary/Light	Moderate	Vigorous
Sedentary/Light	96.3	3.5	0.2
Moderate	37.3	61.6	1.1
Vigorous	26.6	12.5	60.9
	Downsampled Vähä-Ypyä Classification		
30 Hz Vähä-Ypyä Classification	Sedentary/Light	Moderate	Vigorous
Sedentary/Light	95.9	3.8	0.2
Moderate	33.9	64.9	1.3
Vigorous	25.9	12.1	60.9

Table 6. Confusion matrix showing agreement in activity intensity classifications between collected 100 Hz and downsampled 30 Hz data (intra-monitor sampling rate comparison) based on Sasaki et al. (2011) vector magnitude cut-points in counts·min^{-1}, Hildebrand et al. (2014) Euclidean Norm Minus One (ENMO; *mg*) cut-points, and Vähä-Ypyä et al. (2015) mean amplitude deviation (MAD; *mg*) cut-points. The collected 100 Hz data served as the referent group and numbers represent percent of epochs within each activity intensity classified as that intensity according to the downsampled 30 Hz data.

	Downsampled Sasaki Classification		
100 Hz Sasaki Classification	Sedentary/Light	Moderate	Vigorous
Sedentary/Light	98.3	1.5	0.2
Moderate	24.8	75.0	0.3
Vigorous	18.6	5.4	76.0
	Downsampled Hildebrand Classification		
100 Hz Hildebrand Classification	Sedentary/Light	Moderate	Vigorous
Sedentary/Light	97.9	1.9	0.2
Moderate	26.0	73.8	0.2
Vigorous	26.5	3.8	69.7
	Downsampled Vähä-Ypyä Classification		
100 Hz Vähä-Ypyä Classification	Sedentary/Light	Moderate	Vigorous
Sedentary/Light	97.4	2.4	0.2
Moderate	24.2	75.4	0.5
Vigorous	28.1	4.4	67.5

Discussion

The present study explored the impact of inter-monitor variability and intra-monitor differences due to chosen sampling rate of the ActiGraph wGT3X-BT on the estimation of free-living physical activity in adults. While we provide information on differences in epoch-level and mean VM, ENMO, and MAD, it is of particular interest to understand the impact on outcome measures like time spent being physically active. Inter- or intra-monitor variability resulted in differences in moderate-to-vigorous-intensity physical activity of less than 5 min·day^{-1} for VM and MAD, but 5.4-10.8 min·day^{-1} for ENMO, with the largest impact from inter-monitor variability. Whether this magnitude of difference is acceptable will likely depend on the study design and research questions, and potentially the population of interest. Previous activity-promoting interventions in healthy and older adults have demonstrated improvements of approximately 5-10 min·day^{-1} (Barone Gibbs et al., 2017; Napolitano et al., 2010). In clinical populations, a difference of this magnitude has been associated with changes in physical functioning and pain in those with or at risk of knee osteoarthritis (Dunlop et al., 2017; Liu et al., 2016) or lung function and quality of life for patients with interstitial lung disease (Hur et al., 2019). Whilst the present study only included healthy adults, there is no reason to expect that the intra- and inter-monitor differences would vary according to the population on which they are determined. Therefore, the current findings are likely to be applicable across the health spectrum.

For sedentary behavior, inter- and intra-monitor variability resulted in differences of less than 5 min·day^{-1} for VM and MAD, but 5.4-10.2 min·day^{-1} for ENMO. While we were not able to separate sedentary behavior from light-intensity physical activity due to the cut-points used in the present study, prior intervention differences in sedentary behavior of adults were, on average, 22 min·day^{-1} according to a recent review (Martin et al., 2015), while another study reported a minimally important difference of over 100 min·day^{-1} for improvements in physical functioning (Gaskin et al., 2016). Thus, inter- and intra-monitor differences are relatively small for VM and MAD metrics, particularly for measuring sedentary behavior, but more research is needed on using ENMO cut-points for assessing moderate-to-vigorous-intensity physical activity. The magnitude of inter- and intra-monitor differences over longer wear periods may also be of interest due to growing interest in collecting 24-h wear data. While the present study did not include 24-h movement data, extrapolating our results suggests differences of 7.2-25.9 min·day^{-1} across intensities.

Whilst prior research has reported on inter-monitor comparability, it has largely focused on count-based metrics, whereas our study investigates the comparability of count- and acceleration-based activity metrics from the ActiGraph wGT3X-BT monitor, which has not previously been reported. In line with our findings, previous studies of adults in free-living settings wearing two GT3X+ or GT9X monitors at the right hip have reported strong intraclass correlation coefficients (0.97-0.99) for mean VM and time spent in various activity intensities based on the Sasaki et al. (2011) cut-points (Jarrett et al., 2015), and strong Pearson's *r* correlation coefficients (0.92-0.99) for mean VM (Aadland & Ylvisåker, 2015; Loprinzi & Smith, 2017). Similarly, in laboratory-based protocols, correlations for counts between monitors have been reported to range from 0.82 to 0.99, depending on the activity type (Ozemek et al., 2014). The magnitude of the differences between mean group-level VM in the present study (15.1 counts$\cdot$min^{-1}) was also similar to, or smaller than, previous research (e.g., 13.7 counts$\cdot$min^{-1} (Jarrett et al., 2015) and 31.0 counts$\cdot$min^{-1} (Loprinzi & Smith, 2017)). Thus, the inter-monitor comparability of the wGT3X-BT appears similar to that of other ActiGraph models.

There is less research on the comparability of ActiGraph devices for acceleration-based metrics, marking another important contribution of the present analysis. Initial research by Montoye et al. (2018) reported that, in contrast to strong correlations for VM counts, there were weaker correlations for mean acceleration between two ActiGraph models (GT9X and GT3X+) during a semi-structured, laboratory-based protocol in adults. However, the present study suggests that comparability is only an issue for ENMO, not MAD. This is supported by free-living research in children that indicated strong correlations between waist-worn wGT3X-BT and GT9X monitors for mean VM counts and MAD (r=0.996 for both), but a lower (albeit still classified as moderately high) correlation for mean ENMO (r=0.618) and lack of equivalence for mean ENMO between monitors (Clevenger et al., 2020b). The equivalence of the acceleration-based metric MAD in the present study is supported by prior research in free-living adults wearing a wGT3X-BT and GT9X at the hip (Clevenger et al., 2020a). While interest in acceleration-based metrics from ActiGraph monitors is growing, comparability of specific metrics should be considered before widespread implementation as current evidence supports inter-monitor comparability of only the MAD metric.

As may be expected, the largest differences between monitors or sampling rates was at the epoch-level. For example, inter-monitor mean absolute differences at the epoch-level (e.g., 315.1 counts·min^{-1}) were larger than differences between means (e.g., 22.6 counts·min^{-1}), indicating that greater caution should be taken when comparing estimates at epoch-level resolution. This difference was largest for the acceleration-based metric ENMO (92.8%) which is in line with prior research comparing two models of ActiGraph devices worn side-by-side in children (mean absolute percent difference in ENMO of 110.9%) (Clevenger et al., 2020b) and adults (80.9%) (Clevenger et al., 2020a). Conversely, the MAD metric had a lower percent difference (68.0%); it may be less impacted by epoch-level fluctuations because it is an indication of variability, not necessarily magnitude, of acceleration over the 5-s epoch. It has also been postulated that epoch-level inter-monitor differences may be due in part to misalignments in timing between devices. An example of the alignment of a sub-sample of one participant's data is found in Supplementary Figure 1. Although all monitors were started using the same computer, Steel et al. (2019) indicated there was time drift for ActiGraph monitors of approximately 5-s over a seven-day period. As VM is analyzed over a 60-s epoch, it may be less impacted by small misalignments in timing between monitors compared to ENMO, which uses a 5-s epoch. However, analysis of VM at a 5-s epoch resulted in minimal changes in outcomes (data not shown) and differences in ENMO were still larger than for VM or MAD. Thus, while future studies may account for time drift between monitors, the worse comparability of ENMO is likely not just due to time drift.

While differences due to sampling rate were also larger at the epoch level than when data were collapsed to mean per participant, differences were smaller than those due to inter-monitor comparability. No prior research has examined the impact of sampling rate on MAD or ENMO, but mean absolute percent difference for VM in the present study (6.0%) was identical to that found in children (Clevenger et al., 2019). Specifically, Clevenger et al. (2019) indicated that sampling rate had a greater impact on counts than acceleration. This is in line with the present study in which mean MAD was equally impacted by monitor comparability and sampling rate, mean ENMO was impacted by monitor comparability to a greater extent than sampling rate, and mean VM was impacted by sampling rate more so than inter-monitor comparability. This finding is due to the greater bias for intra-monitor differences in VM compared to inter-monitor differences (Figure 1). Thus, as in prior research, use of a 100 Hz sampling rate results in the recording of additional counts which leads to bias and impacts mean VM and, to a lesser extent, acceleration-based metrics.

Bias in the present study (15.9 VM counts·min^{-1}) was smaller than a previous study of adults during increasing speeds of locomotion, in which bias between monitors using a 100 Hz and 30 Hz sampling rate ranged from 47 to 1,238 vertical axis counts·min^{-1} (Brønd & Arvidsson, 2015). As the impact of sampling rate has been shown to increase with increasing intensity (Brønd & Arvidsson, 2015; Clevenger et al., 2019) and participants in the present study spent the majority of their time in sedentary and/or light intensity behaviors (>90% of time), it is not surprising that differences due to sampling rate were low compared to prior semi-structured protocols. In line with the idea that sampling rate differences are larger at higher intensities, we found that more active participants had greater differences due to sampling rate. While we did not formally test these differences due to the small sample size, some preliminary examples are provided in the supplementary material. For example, scatter plots between 100 Hz and downsampled data were less linear (Supplementary Figure 2) and confusion matrices included more mismatches (Supplementary Table 2) in participants who were generally more active. However, inter-monitor differences seemed consistent among participants, irrespective of activity levels (Supplementary Figure 3 and Supplementary Table 3). Future research may aim to consider the differential influence of sampling rate on the measurement of free-living physical activity of more active individuals.

These findings should be replicated, as this study is not without limitations, primarily the small sample size. However, a key strength of the present study was the use of two monitors, which allowed for the simultaneous evaluation of inter-monitor differences and the impact of sampling rate on accelerometer metrics. Moreover, matching wear-time between data files also enhances the quality of the present study. Previous studies in which participants wore two waist-worn monitors during free-living have reported small, unaccounted for, differences in wear-time (0.8-5.5 $min \cdot day^{-1}$), which could confound results if not addressed (Aadland & Ylvisåker, 2015; Jarrett et al., 2015). Finally, only two pairs of monitors were used in the present study, which may artificially limit inter-monitor differences, warranting further research on inter-monitor comparability of the ActiGraph wGT3X-BT.

Conclusions

When designing future physical activity studies, researchers have many decisions to make, including selecting a monitor, the sampling rate, and the metric used to classify time spent being physically active. We demonstrate that inter-monitor comparability had a larger impact on epoch-level metrics than sampling rate, but that sampling rate had a larger impact on collapsed data depending on the physical activity intensity performed, especially count data due to consistent bias of higher counts being recorded by the 100 Hz versus the 30 Hz monitor. While we support the comparability of the wGT3X-BT monitor for VM and MAD metrics and related outcomes, more research is needed on the comparability of ENMO during free-living as variation in ENMO due to sampling rate or inter-monitor comparability resulted in mean absolute differences in moderate-to-vigorous-intensity physical activity of 5.4-10.8 min·day^{-1}.

Practical Implications

- ActiGraph wGT3X-BT accelerometers demonstrate high comparability for VM counts and MAD, but only moderate comparability for ENMO
- Sampling rate had a smaller impact than inter-monitor comparability on epoch-level monitor output, but counts were impacted to the greatest extent

Acknowledgements

There was no funding for this project.

References

Aadland, E., & Ylvisåker, E. (2015). Reliability of the Actigraph GT3X+ accelerometer in adults under free-living conditions. *PloS one*, *10*(8), e0134606.

Aittasalo, M., Vähä-Ypyä, H., Vasankari, T., Husu, P., Jussila, A.-M., & Sievänen, H. (2015). Mean amplitude deviation calculated from raw acceleration data: a novel method for classifying the intensity of adolescents' physical activity irrespective of accelerometer brand. *BMC Sports Science, Medicine and Rehabilitation*, *7*(1), 18.

Bakrania, K., Yates, T., Rowlands, A. V., Esliger, D. W., Bunnewell, S., Sanders, J., Davies, M., Khunti, K., & Edwardson, C. L. (2016). Intensity thresholds on raw acceleration data: euclidean norm minus one (ENMO) and mean amplitude deviation (MAD) approaches. *PloS One*, *11*(10), e0164045.

Barone Gibbs, B., Brach, J. S., Byard, T., Creasy, S., Davis, K. K., McCoy, S., Peluso, A., Rogers, R. J., Rupp, K., & Jakicic, J. M. (2017). Reducing sedentary behavior versus increasing moderate-to-vigorous intensity physical activity in older adults: a 12-week randomized, clinical trial. *Journal of Aging and Health*, *29*(2), 247-267.

Bland, J. M., & Altman, D. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet*, *327*(8476), 307-310.

Brønd, J. C., & Arvidsson, D. (2015). Sampling frequency affects the processing of Actigraph raw acceleration data to activity counts. *Journal of Applied Physiology*, *120*(3), 362-369.

Chen, K. Y., & Bassett, D. R. (2005). The technology of accelerometry-based activity monitors: current and future. *Medicine and Science in Sports and Exercise*, *37*(11), S490.

Clevenger, K. A., Pfeiffer, K. A., Mackintosh, K. A., McNarry, M. A., Brønd, J. C., Arvidsson, D., & Montoye, A. H. (2019). Effect of sampling rate on acceleration and counts of hip- and wrist-worn ActiGraph accelerometers in children. *Physiological Measurement*.

Clevenger, K. A., Pfeiffer, K. A., & Montoye, A. H. (2020a). Cross-generational comparability of hip-and wrist-worn ActiGraph GT3X+, wGT3X-BT, and GT9X accelerometers during free-living in adults. *Journal of Sports Sciences*, 1-9.

Clevenger, K. A., Pfeiffer, K. A., & Montoye, A. H. (2020b). Cross-Generational Comparability of Raw and Count-Based Metrics from ActiGraph GT9X and wGT3X-BT Accelerometers during Free-Living in Youth. *Measurement in Physical Education and Exercise Science*, 1-11.

de Almeida Mendes, M., da Silva, I. C., Ramires, V. V., Reichert, F. F., Martins, R. C., & Tomasi, E. (2018). Calibration of raw accelerometer data to measure physical activity: A systematic review. *Gait & Posture*, *61*, 98-110.

Dunlop, D. D., Song, J., Lee, J., Gilbert, A. L., Semanik, P. A., Ehrlich-Jones, L., Pellegrini, C. A., Pinto, D., Ainsworth, B., & Chang, R. W. (2017). Physical activity minimum threshold predicting improved function in adults with lower-extremity symptoms. *Arthritis Care & Research*, *69*(4), 475-483.

Esliger, D. W., & Tremblay, M. S. (2006). Technical reliability assessment of three accelerometer models in a mechanical setup. *Medicine and Science in Sports and Exercise*, *38*(12), 2173-2181.

Gamer, M., Lemon, J., Fellows, I., & Singh, P. (2012). irr: Various Coefficients of Interrater Reliability and Agreement. R package version 0.84. Available at *http://CRAN.R-project. org/package= irr*. Accessed June 2021

Gaskin, C. J., Craike, M., Mohebbi, M., Salmon, J., Courneya, K. S., Broadbent, S., & Livingston, P. M. (2016). Associations of objectively measured moderate-to-vigorous physical activity and sedentary behavior with quality of life and psychological well-being in prostate cancer survivors. *Cancer Causes & Control*, *27*(9), 1093-1103.

Hibbing, P. (2018). AGread: Read Data Files from ActiGraph Monitors. Available at https://cran.r-project.org/web/packages/AGread/AGread.pdf Accessed June 2021

Hildebrand, M., VT, V. A. N. H., Hansen, B. H., & Ekelund, U. (2014). Age group comparability of raw accelerometer output from wrist- and hip-worn monitors. *Medicine and Science in Sports and Exercise*, *46*(9), 1816-1824. https://doi.org/10.1249/mss.0000000000000289

Hur, S. A., Guler, S. A., Khalil, N., Camp, P. G., Guenette, J. A., Swigris, J. J., & Ryerson, C. J. (2019). Minimal important difference for physical activity and validity of the International Physical Activity Questionnaire in interstitial lung disease. *Annals of the American Thoracic Society*, *16*(1), 107-115.

Jarrett, H., Fitzgerald, L., & Routen, A. C. (2015). Interinstrument reliability of the ActiGraph GT3X+ ambulatory activity monitor during free-living conditions in adults. *Journal of Physical Activity and Health*, *12*(3), 382-387.

John, D., & Freedson, P. (2012). ActiGraph and Actical physical activity monitors: a peek under the hood. *Medicine and Science in Sports and Exercise*, *44*(1 Suppl 1), S86.

Lakens, D. (2017). TOSTER: Two one-sided tests (TOST) equivalence testing. R package version 0.2. 5. Available at https://cran.r-project.org/web/packages/TOSTER/TOSTER.pdf Accessed June 2021

Lee, K.-Y., Macfarlane, D. J., & Cerin, E. (2013). Comparison of three models of actigraph accelerometers during free living and controlled laboratory conditions. *European Journal of Sport Science*, *13*(3), 332-339.

Liu, S.-H., Driban, J. B., Eaton, C. B., McAlindon, T. E., Harrold, L. R., & Lapane, K. L. (2016). Objectively measured physical activity and symptoms change in knee osteoarthritis. *The American Journal of Medicine*, *129*(5), 497-505. e491.

Loprinzi, P. D., & Smith, B. (2017). Comparison between wrist-worn and waist-worn accelerometry. *Journal of Physical Activity and Health*, *14*(7), 539-545.

Martin, A., Fitzsimons, C., Jepson, R., Saunders, D. H., van der Ploeg, H. P., Teixeira, P. J., Gray, C. M., & Mutrie, N. (2015). Interventions with potential to reduce sedentary time in adults: systematic review and meta-analysis. *British Journal of Sports Medicine*, *49*(16), 1056-1063.

McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia Medica*, *22*(3), 276-282.

US National Library of Medicine (2021). *ClinicalTrials.gov*. https://www.clinicaltrials.gov/ct2/results?cond=&term=actigraph&cntry=&state=&city=&dist=

Migueles, J. H., Cadenas-Sanchez, C., Ekelund, U., Nyström, C. D., Mora-Gonzalez, J., Löf, M., Labayen, I., Ruiz, J. R., & Ortega, F. B. (2017). Accelerometer data collection and processing criteria to assess physical activity and other outcomes: a systematic review and practical considerations. *Sports Medicine*, *47*(9), 1821-1845.

Migueles, J. H., Cadenas-Sanchez, C., Rowlands, A. V., Henriksson, P., Shiroma, E. J., Acosta, F. M., Rodriguez-Ayllon, M., Esteban-Cornejo, I., Plaza-Florido, A., & Gil-Cosano, J. J.

(2019). Comparability of accelerometer signal aggregation metrics across placements and dominant wrist cut points for the assessment of physical activity in adults. *Scientific Reports*, *9*(1), 1-12.

Montoye, A. H., Moore, R. W., Bowles, H. R., Korycinski, R., & Pfeiffer, K. A. (2016). Reporting accelerometer methods in physical activity intervention studies: a systematic review and recommendations for authors. *British Journal of Sports Medicine*, bjsports-2015-095947.

Montoye, A. H., Nelson, M. B., Bock, J. M., Imboden, M. T., Kaminsky, L. A., Mackintosh, K. A., Mcnarry, M. A., & Pfeiffer, K. A. (2018). Raw and Count Data Comparability of Hip-Worn ActiGraph GT3X+ and Link Accelerometers. *Medicine and Science in Sports and Exercise*, *50*(5), 1103-1112.

Napolitano, M. A., Borradaile, K. E., Lewis, B. A., Whiteley, J. A., Longval, J. L., Parisi, A. F., Albrecht, A. E., Sciamanna, C. N., Jakicic, J. M., & Papandonatos, G. D. (2010). Accelerometer use in a physical activity intervention trial. *Contemporary Clinical Trials*, *31*(6), 514-523.

Ozemek, C., Kirschner, M. M., Wilkerson, B. S., Byun, W., & Kaminsky, L. A. (2014). Intermonitor reliability of the GT3X+ accelerometer at hip, wrist and ankle sites during activities of daily living. *Physiological Measurement*, *35*(2), 129.

Peeters, G., van Gellecum, Y., Ryde, G., Farías, N. A., & Brown, W. J. (2013). Is the pain of activity log-books worth the gain in precision when distinguishing wear and non-wear time for tri-axial accelerometers? *Journal of Science and Medicine in Sport*, *16*(6), 515-519.

Ried-Larsen, M., Brønd, J. C., Brage, S., Hansen, B. H., Grydeland, M., Andersen, L. B., & Møller, N. C. (2012). Mechanical and free living comparisons of four generations of the Actigraph activity monitor. *International Journal of Behavioral Nutrition and Physical Activity*, *9*(1), 113.

Safrit, M. J., & Wood, T. M. (1995). *Introduction to measurement in physical education and exercise science*. William C. Brown.

Santos-Lozano, A., Santin-Medeiros, F., Cardon, G., Torres-Luque, G., Bailon, R., Bergmeir, C., Ruiz, J., Lucia, A., & Garatachea, N. (2013). Actigraph GT3X: validation and determination of physical activity intensity cut points. *International Journal of Sports Medicine*, *34*(11), 975-982.

Sasaki, J. E., John, D., & Freedson, P. S. (2011). Validation and comparison of ActiGraph activity monitors. *Journal of Science and Medicine in Sport*, *14*(5), 411-416.

Silva, P., Mota, J., Esliger, D., & Welk, G. (2010). Technical reliability assessment of the Actigraph GT1M accelerometer. *Measurement in Physical Education and Exercise Science*, *14*(2), 79-91.

Steel, C., Bejarano, C., & Carlson, J. A. (2019). Time drift considerations when using GPS and accelerometers. *Journal for the Measurement of Physical Behaviour*, *2*(3), 203-207.

Stevens, J., Murray, D. M., Catellier, D. J., Hannan, P. J., Lytle, L. A., Elder, J. P., Young, D. R., Simons-Morton, D. G., & Webber, L. S. (2005). Design of the trial of activity in adolescent girls (TAAG). *Contemporary Clinical Trials*, *26*(2), 223-233.

Troiano, R. P., Berrigan, D., Dodd, K. W., Masse, L. C., Tilert, T., & McDowell, M. (2008). Physical activity in the United States measured by accelerometer. *Medicine and Science in Sports and Exercise*, *40*(1), 181-188.

Troiano, R. P., McClain, J. J., Brychta, R. J., & Chen, K. Y. (2014). Evolution of accelerometer methods for physical activity research. *British Journal of Sports Medicine*, *48*(13), 1019-1023. https://doi.org/10.1136/bjsports-2014-093546

Van Domelen, D. R., & Pittard, W. S. (2014). Flexible R functions for processing accelerometer data, with emphasis on NHANES 2003–2006. *R J*, *6*, 52.

van Hees, V., Fang, Z., Zhao, J., Heywood, J., Mirkes, E., Sabia, S., & Migueles, J. (2019). GGIR: Raw Accelerometer Data Analysis. https://doi.org/10.5281/zenodo.1051064

van Hees, V. T., Fang, Z., Langford, J., Assah, F., Mohammad, A., da Silva, I. C., Trenell, M. I., White, T., Wareham, N. J., & Brage, S. (2014). Auto-calibration of accelerometer data for free-living physical activity assessment using local gravity and temperature: an evaluation on four continents. *American Journal of Physiology-Heart and Circulatory Physiology*.

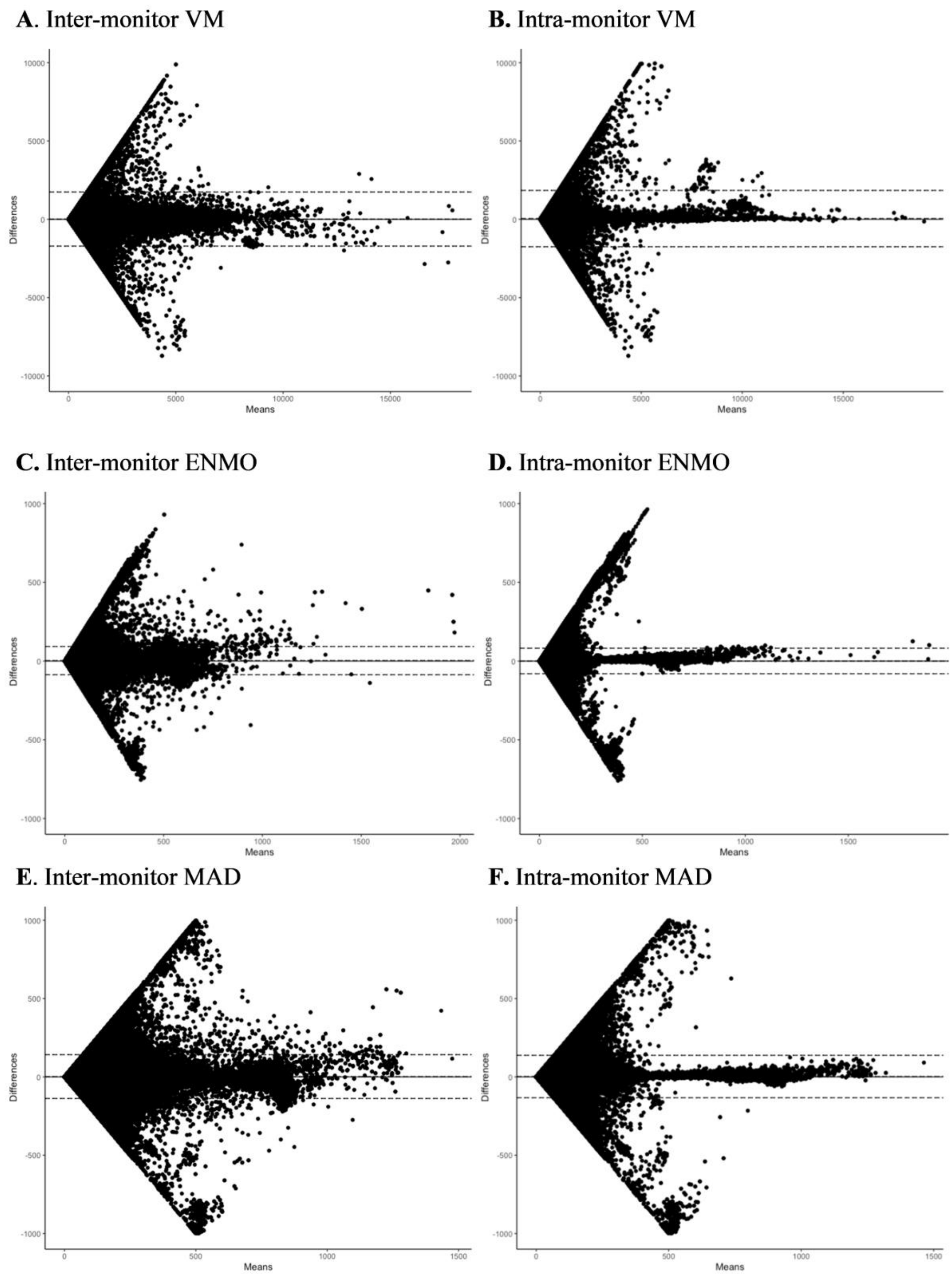
van Hees, V. T., Gorzelniak, L., Leon, E. C. D., Eder, M., Pias, M., Taherian, S., Ekelund, U., Renström, F., Franks, P. W., & Horsch, A. (2013). Separating movement and gravity components in an acceleration signal and implications for the assessment of human daily physical activity. *PloS One*, *8*(4), e61691.

Vanhelst, J., Mikulovic, J., Bui-Xuan, G., Dieu, O., Blondeau, T., Fardy, P., & Béghin, L. (2012). Comparison of two ActiGraph accelerometer generations in the assessment of physical activity in free living conditions. *BMC Research Notes*, *5*(1), 187.

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.

Vähä-Ypyä, H., Vasankari, T., Husu, P., Suni, J., & Sievänen, H. (2015). A universal, accurate intensity-based classification of different physical activities using raw data of accelerometer. *Clinical Physiology and Functional Imaging*, *35*(1), 64-70.

Wong, T. C., Webster, J. G., Montoye, H. J., & Washburn, R. (1981). Portable accelerometer device for measuring human energy expenditure. *IEEE Transactions on Biomedical Engineering*(6), 467-471.

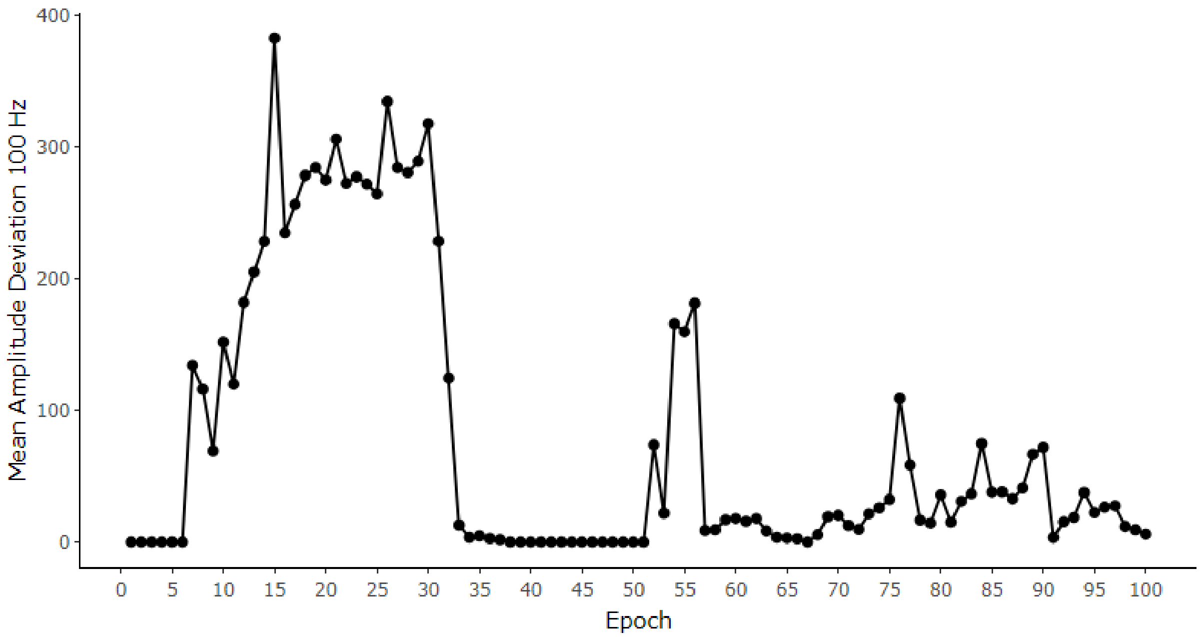

A. Inter-monitor VM
B. Intra-monitor VM
C. Inter-monitor ENMO
D. Intra-monitor ENMO
E. Inter-monitor MAD
F. Intra-monitor MAD
Differences
Means

Supplementary Table 1. Accelerometer calibration values for the four monitors used in the present study. Values were extracted using the ‘g.calibrate’ function in the GGIR package.

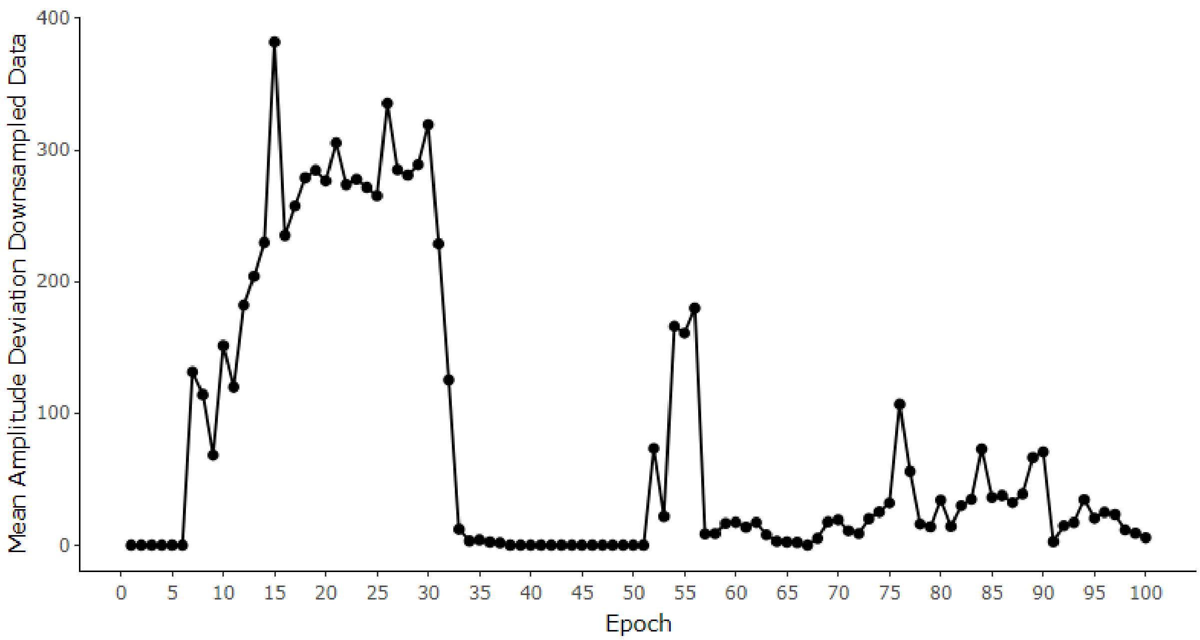
	Monitor A	Monitor B	Monitor C	Monitor D
Prior calibration error	0.019 ± 0.008	0.013 ± 0.008	0.017 ± 0.005	0.017 ± 0.005
Post calibration error	0.004 ± 0.002	0.004 ± 0.002	0.004 ± 0.002	0.004 ± 0.002
Offset x-axis	-0.002 ± 0.007	0.002 ± 0.002	0.002 ± 0.007	0.003 ± 0.002
Offset y-axis	-0.003 ± 0.005	0.003 ± 0.009	-0.005 ± 0.008	0.004 ± 0.007
Offset z-axis	0.002 ± 0.020	0.007 ± 0.011	0.006 ± 0.016	0.012 ± 0.014
Scale x-axis	0.999 ± 0.016	1.003 ± 0.016	0.985 ± 0.012	0.975 ± 0.016
Scale y-axis	0.993 ± 0.014	1.006 ± 0.016	0.993 ± 0.011	0.993 ± 0.022
Scale z-axis	0.994 ± 0.013	0.997 ± 0.011	0.997 ± 0.019	1.004 ± 0.020

Supplementary Figure 1. Example of Mean Amplitude Deviation in 5-s epochs for (a) a sub-sample of 100 Hz, (b) downsampled 30 Hz, and (c) collected 30 Hz data from one participant

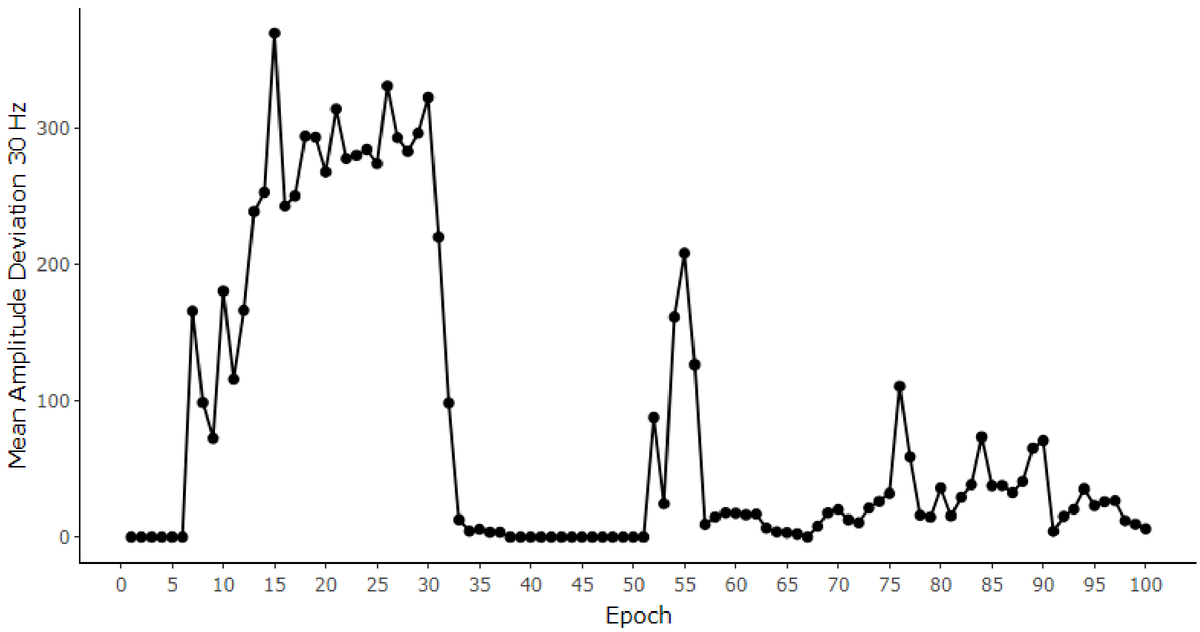
(a)



(b)

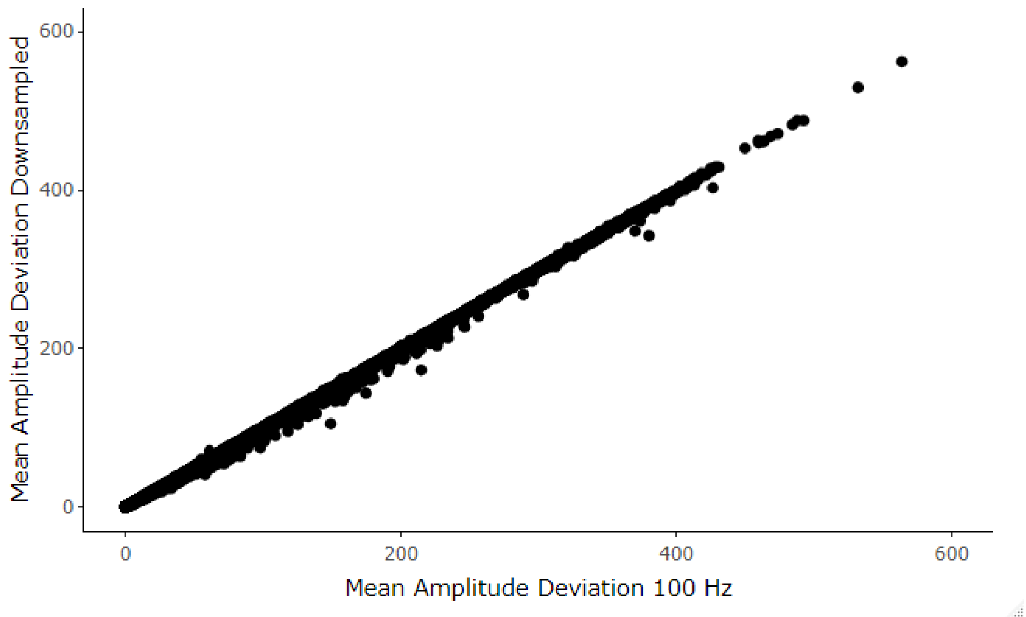


(c)

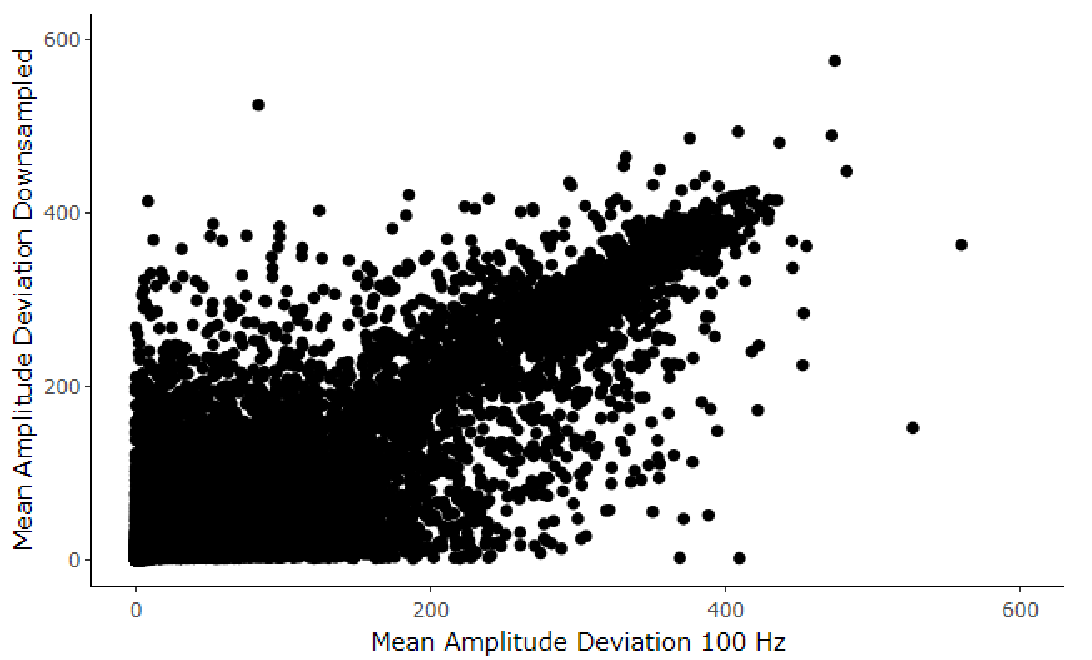

Mean Amplitude Deviation 30 Hz
0
100
200
300
0
5
10
15
20
25
30
35
40
45
50
55
60
65
70
75
80
85
90
95
100
Epoch

Supplementary Figure 2. Scatter plot between downsampled and 100 Hz data for two participants (panels a and b, participant A and B, respectively). Average Mean Amplitude Deviation (MAD) was ~27-28 *mg* for participant A and ~36-38 *mg* for participant B.

(a)



(b)



Supplementary Table 2. Confusion matrices showing agreement in activity intensity classifications using Vähä-Ypyä et al. (2015) mean amplitude deviation (MAD; *mg*) cut-points between collected 100 Hz and downsampled 30 Hz data (intra-monitor comparison) for the same two participants (panels a and b, participant A and B, respectively) shown in Supplementary Figure 3. The collected 100 Hz data served as the referent group and numbers represent percent of epochs within each activity intensity classified as that intensity according to the downsampled 30 Hz data. Average MAD was ~27-28 *mg* for participant A and ~36-38 *mg* for participant B.

(a)

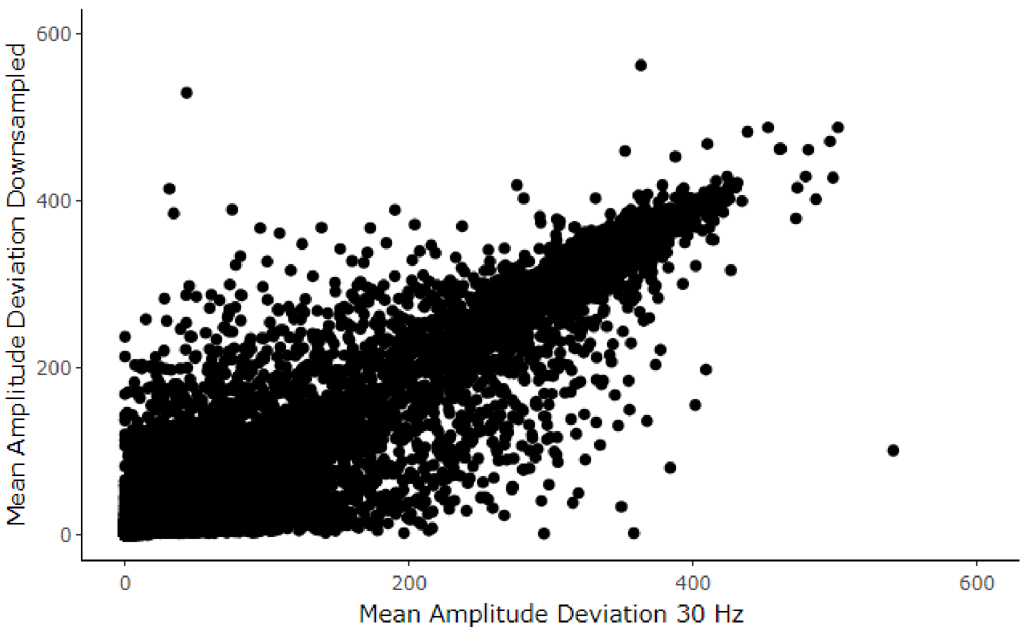
	Downsampled Vähä-Ypyä Classification		
100 Hz Vähä-Ypyä Classification	Sedentary/Light	Moderate	Vigorous
Sedentary/Light	100.0	0.0	0.0
Moderate	0.9	99.1	0.0
Vigorous	0.0	11.1	88.9

(b)

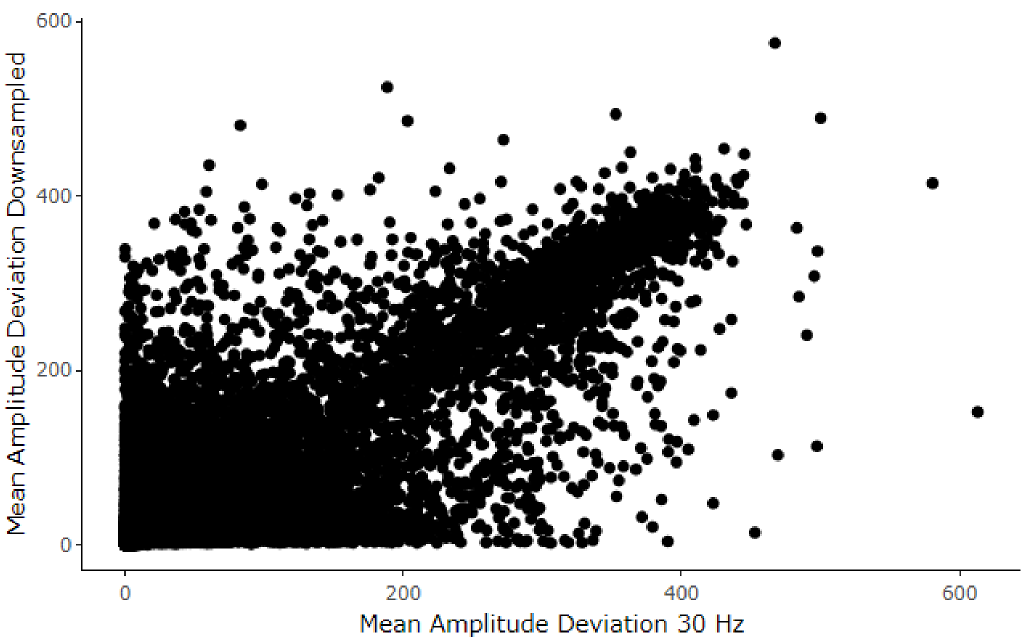
	Downsampled Vähä-Ypyä Classification		
100 Hz Vähä-Ypyä Classification	Sedentary/Light	Moderate	Vigorous
Sedentary/Light	96.2	3.8	0.0
Moderate	28.6	70.8	0.6
Vigorous	0.0	29.0	71.0

Supplementary Figure 3. Scatter plot between downsampled and 30 Hz data for two participants (a and b). Average Mean Amplitude Deviation (MAD) was ~27-28 *mg* for participant A and ~36-38 *mg* for participant B.

(a)



(b)



Supplementary Table 3. Confusion matrices showing agreement in activity intensity classifications using Vähä-Ypyä et al. (2015) mean amplitude deviation (MAD; *mg*) cut-points between collected 30 Hz and downsampled 30 Hz data (inter-monitor comparison) for the same two participants (a and b) shown in Supplementary Figure 3. The collected 30 Hz data served as the referent group and numbers represent percent of epochs within each activity intensity classified as that intensity according to the downsampled 30 Hz data. Average MAD was ~27-28 *mg* for participant A and ~36-38 *mg* for participant B.

(a)

	Downsampled Vähä-Ypyä Classification		
30 Hz Vähä-Ypyä Classification	Sedentary/Light	Moderate	Vigorous
Sedentary/Light	98.6	1.4	0.0
Moderate	18.2	81.6	0.2
Vigorous	0.0	59.0	41.0

(b)

	Downsampled Vähä-Ypyä Classification		
30 Hz Vähä-Ypyä Classification	Sedentary/Light	Moderate	Vigorous
Sedentary/Light	95.8	4.2	0.0
Moderate	33.8	65.7	0.5
Vigorous	3.6	18.2	78.2