Comparability of Postural and Physical Activity Metrics from Different Accelerometer Brands Worn on the Thigh: Data Harmonization Possibilities

Charlotte L Edwardsona,b, Benjamin D Maylorc, Nathan P Dawkinsd, Tatiana Plekhanova, and Alex V Rowlandsa,b

aAssessment of Movement Behaviours (Amber) Group, Leicester Lifestyle and Health Research Group, Diabetes Research Centre, University of Leicester, Leicester General Hospital, Leicester UK; bNational Institute for Health Research (NIHR) Leicester Biomedical Research Centre (BRC), University Hospitals of Leicester NHS Trust and the University of Leicester, Leicester, UK

ABSTRACT

The aim was to establish which postural and physical activity outcomes are comparable across different accelerometer brands worn on the thigh when processed using open-source methods. Twenty participants wore four accelerometers (Axivity, ActiGraph, activPAL, GENEActiv) for three free-living days. Postural and physical activity outputs (average acceleration, intensity gradient, intensity of the most active 30 min, 60 min, and 8 h) were generated.

Postural outputs: Mean absolute percent errors (MAPEs) were low, reliability excellent, and equivalency within the 5% zone across all monitor pairings for sitting/lying and upright times, but not specific lying postures. Physical activity outputs: MAPEs were higher and reliability lower than for sitting/lying and upright time. However, the majority of the outcomes were within the 10% equivalency zone for Axivity/GENEActiv and Axivity/ActiGraph pairings. Total sitting/lying and upright times show strong potential for harmonization across studies utilizing different thigh-worn accelerometers. The majority of acceleration outcomes compare well for Axivity, GENEActiv, and ActiGraph.

KEYWORDS

Activpal; geneactiv; axivity; actigraph; sedentary behavior; sitting

Introduction

Accelerometer-based devices worn on the thigh are increasingly being used in large research studies to assess physical behaviors (Stamatakis et al., 2020). The thigh has become a wear location of interest due to its accuracy for estimating the postural component of sedentary behavior (i.e., lying, sitting, and reclining postures) as well as physical activity (Edwardson et al., 2016; Montoye et al., 2016b; Lyden et al., 2017; Sellers et al., 2016). The activPAL device (PAL Technologies Ltd, Glasgow, UK) is specifically designed to be worn on the thigh but more recently other brands of accelerometer, such as ActiGraph, Axivity, and GENEActiv, are being used in this wear location (Daugaard et al., 2018; Hartley et al., 2018; Jørgensen et al., 2019; Ryan et al., 2019). These devices all measure raw acceleration across three axes so there is potential for data harmonization across devices. This is particularly pertinent given the recently established Prospective Physical Activity, Sitting, and Sleep Consortium (ProPASS), which aims to combine existing and future observational studies utilizing thigh worn accelerometry to “produce evidence on the associations of physical activity, sitting and sleep and health outcomes” (Stamatakis et al., 2020).

In a controlled laboratory setting, it has previously been demonstrated that accelerometers worn on the thigh produced similar accuracy for identifying lying, sitting, and upright postures irrespective of whether proprietary postural allocation algorithms or open-source algorithms were used to process the data (Edwardson et al., 2016). However, not all accelerometer brands were processed using a single method and the structured activities completed within the laboratory may not reflect free-living behaviors. More recently, by using the Acti4 software package to process all data, Crowley et al. (2019) compared time spent on physical behavior types (e.g., sitting, standing, walking, running) from three different accelerometer brands (activPAL, ActiGraph, and Axivity) worn on the thigh. They concluded that during a 7-day free living protocol, the difference in the average time spent on the physical behavior types was negligible between the accelerometer brands. When averaged over 24 h, the absolute differences were as little as 0.1 min for running to 3.5 min for moving between accelerometers. These results show
promise for the harmonization of data collected from different brands of thigh worn accelerometers. This study focuses on physical behavior types classified using Acti4 software; this is not open source, although it is freely available upon agreement with the National Research Center for the Working Environment, Copenhagen (Crowley et al., 2019). In addition to classification of behavior type, it is important to assess whether metrics that directly reflect measured acceleration from the accelerometers are comparable. These include the magnitude of the acceleration, the orientation of the monitor relative to gravity, and the pattern of the acceleration. Algorithms to classify behavior type will continue to evolve but will be based on direct measurements from the accelerometer (e.g., Montoye et al., 2016a; Tjurin et al., 2019). Further, direct measures may be used alongside behavior-type algorithms; for example, the intensity (acceleration magnitude) when someone is walking. Therefore, knowing which of the directly measured acceleration metrics compare well is crucial if considering harmonizing data across brands.

Direct measures of acceleration have been used to describe the physical activity profile and its association with health (e.g., Buchan et al., 2019; Fairclough et al., 2019; Rowlands et al., 2018a). For example, the average acceleration reflects the overall level of activity, the intensity gradient (Rowlands et al., 2018a) describes the distribution of acceleration intensity across the 24 h profile, and the MX metrics describe the intensity of the most active periods of the day (e.g., M30 for the most active 30 min) (Rowlands et al., 2019a, 2019c). We have previously shown the average acceleration and intensity gradient were equivalent between accelerometer brands worn on the non-dominant wrist, with only the intensity gradient being equivalent between brands on the dominant wrist (Rowlands et al., 2019b). Given comparability can differ between accelerometer brands for the dominant and non-dominant wrist (Rowlands et al., 2019b) and that different movement patterns occur from the wrist and thigh, we cannot assume results from studies comparing the metrics at the wrist will generalize to the thigh. Furthermore, the equivalency of these metrics has yet to be investigated between the activPAL (the most widely used thigh-worn accelerometer) and different accelerometer brands, and between all brands when worn on the thigh.

The aim of this study is to compare postural and physical activity outcomes across the GENEActiv, activPAL, ActiGraph, and Axivity accelerometers worn on the thigh when processed using the same open-source methods. We hypothesize that outcomes created using information on the orientation of the device (i.e., postural outcomes) will not differ between accelerometer brands, but there will be differences in acceleration metrics between accelerometer brands when processed using the open-source accelerometer processing and analyzing software GGIR (Migueles, Rowlands et al., 2019), particularly for the activPAL due to its lower dynamic range (± 2 g) and sampling frequency (20 Hz) relative to the other accelerometer brands. The results will inform the potential for data harmonization across studies for different postural and physical activity outcomes across different accelerometer brands when worn on the thigh.

**Methods**

**Participants and procedure**

A convenience sample of 23 adults (≥18 years of age) was recruited via word of mouth from the University of Leicester (staff and students). Ethical approval was requested in May 2019 and received from the University of Leicester’s College of Life Sciences ethical representatives in July 2019.

Following the provision of informed consent, participants’ height, weight, and body fat percentage (Tanita BC-418 MA, Tanita Europe BV) were measured and basic demographic variables (age, sex, and ethnicity) were collected. Participants were fitted with four accelerometers: activPAL micro, GENEActiv, Axivity, ActiGraph Link) on the upper mid-line of the right thigh. The devices were attached to the skin with medical dressing (Hypafix; BSN Medical, UK). The Axivity was taped on top of the ActiGraph and the activPAL taped on top of the GENEActiv with the order in which these were worn on the thigh randomized for each participant. The two randomization options were as follows: 1) GENEActiv/activPAL closest to the knee and the ActiGraph/Axivity closest to the hip; 2) GENEActiv/activPAL closest to the hip and the ActiGraph/Axivity closest to the knee. There was approximately a 1.5 cm gap between the pairs of devices when placed on the thigh. Participants were instructed to wear the monitors continuously for three nights and two complete waking days and complete a log of the time they went to bed, fell to sleep, woke up, and got out of bed as well as any monitor removal times.

**Accelerometers**

The activPAL3™ micro (PAL Technologies Ltd, Glasgow, UK) is a triaxial accelerometer with a dynamic range of ± 2 g, where g is equal to the Earth’s gravity. Default settings (20 Hz, 10 second
minimum sitting, and upright period) were used during initialization. activPAL monitors were initialized and data downloaded using PAL Connect version 8.10.6.63 and data saved in raw format as .csv through PAL Batch.

The ActiGraph GT9X Link (ActiGraph LLC, Pensacola, FL, USA), GENEActiv Original (ActivInsights Ltd., Cambridgeshire, UK), and Axivity AX3 (Axivity Ltd, Newcastle, UK) are triaxial accelerometers with a dynamic range of ± 8 g (ActiGraph GT9X Link and GENEActiv Original) and ± 16 g (Axivity AX3). These monitors can be worn on various body locations including wrist, waist, ankle, upper arm, and thigh. All accelerometers were configured on the same PC in the dynamic range ± 8 g and to record at a frequency of 100 Hz to reflect common practice (e.g., Doherty et al., 2017; Fairclough et al., 2019; Rowlands et al., 2018b). The Actigraph Link GT9X monitors were initialized and data downloaded using ActiLife version 6.13.4. Data were saved in raw format as .gt3x, and subsequently converted to .csv format for data processing. GENEActiv monitors were initialized and data downloaded using GENEActiv PC software version 3.2 and data were saved in raw format as .bin files. Axivity devices were set up and data downloaded with OmGui open-source software (OmGui Version 1.0.0.37, Open Movement, Newcastle, UK) and saved in .cwa format.

**Data processing**

**Postural outcomes**

GENEActiv: The raw .bin files were converted into 15 second epoch csv files using GENEActiv PC software v3.2. The 15-s epoch files were imported into a custom-built spreadsheet template in Excel originally designed for GENEActiv 15 s epoch data (ActivInsights postural allocation algorithm (Rowlands et al., 2014) available at: https://www.researchgate.net/project/AMBer-Assessment-of-Movement-Behaviors) that computes the most likely posture based on the relative values of the x, y, z vectors measured at the thigh.

ActiGraph, Axivity and activPAL: The raw .csv files were converted with a custom-built program (GT9X-to-SedSphere, Liverpool John Moores University, UK) written in MATLAB (R2017b, The MathWorks Inc., Natick, MA, USA). It calculates the vector magnitude with gravity subtracted, summed over 15 s epochs (as in GENEActiv PC software), and matches the orientation of each axis to those of the GENEActiv, as required for the ActivInsights postural allocation algorithm (Hurter et al., 2019). Data were then imported into a custom-built template in Excel processing as described for the GENEActiv. The following outcomes were generated and averaged across the 2 days: sitting/lying on the back, lying on the side, lying on the front, total time sitting/lying (summary of the previous three variables) and total upright time.

**Physical activity outcomes**

All accelerometer files were processed and analyzed with the R-package GGIR version 1.9–4 (http://cran.r-project.org) (Migueles, Rowlands et al., 2019) using default settings to maximize generalizability. Signal processing in GGIR included auto calibration using local gravity as a reference (Van Hees et al., 2014); calculation of two acceleration signal aggregation metrics (ENMO (Euclidean Norm minus 1 g) and MAD (Mean Amplitude Deviation)) averaged over 5 s epochs; detection of non-wear; and detection of abnormally sustained high values. Both metrics aim to remove the gravitational component of acceleration, thus reflecting body movement (Van Hees et al., 2013). ENMO is the average magnitude of dynamic acceleration with one g subtracted and resulting values rounded to zero (Van Hees et al., 2014, 2013). MAD is the variability in acceleration around the mean (Euclidean norm of each raw acceleration datapoint minus the mean of its corresponding 5 s epoch (Aittasalo et al., 2015; Migueles, Cadenas-Sanchez et al., 2019; Vähä-Yyppä et al., 2015a, 2015b)). Both were expressed in milli-gravitational units (mg). Non-wear was imputed using the default setting in GGIR i.e., invalid data were imputed by the average at similar time-points on different days of the week.

Participants were excluded if their accelerometer files showed: post-calibration errors greater than 0.01 g - (10 mg), or wear data were not present for each 15 min period of the 24 h cycle. Detection of non-wear has been described in detail previously (See “Procedure for non-wear detection” in supplementary document to Van Hees et al., 2013). Briefly, non-wear is estimated based on the standard deviation and value range of each axis, calculated for 60 min windows with a 15-min sliding window. The window is classified as non-wear if, for at least two out of the three axes, the SD (standard deviation) is less than 3 mg or the value range is less than 50 mg.

The following outcomes were generated and averaged across the 2 days: average acceleration (mg); intensity gradient; acceleration above which a person’s most active 8 hours, 60 and 30 minutes (M8h, M60, and M30, mg) are accumulated (within the GGIR package, these metrics are obtained in part 2 using: qlevelsqlevels M8h, M60, and M30). The M8h, M60, and M30 rank the accelerations for each epoch during the day in descending order to obtain the acceleration above which the person’s most active
The average acceleration reflects the overall level of physical activity. The intensity gradient reflects the distribution of acceleration intensity across the 24 h day and has been described elsewhere (Rowlands et al., 2018a); in brief, it describes the negative curvilinear relationship between physical activity intensity and the time accumulated at that intensity during the 24 h day. The intensity gradient is always negative, reflecting the drop in time accumulated as intensity increases; a more negative (lower) gradient reflects a steeper drop with little time accumulated at mid-range and higher intensities, while a less negative (higher) gradient reflects a shallower drop with more time spread across the intensity range. It was calculated as previously described (Rowlands et al., 2018a) and generated in GGIR (argument ig levels = TRUE).

### Statistical analysis

Only those who provided a full set of data (at least 20 hours of non-imputed data on each day across all four accelerometers: amount and timing of non-wear time the same between accelerometers, confirmed numerically and by visual inspection of the accelerometer traces) were included in analyses. This was determined for the physical activity outcomes and matched for the postural outcomes. Sleep data were not excluded as the purpose was to compare the posture and acceleration metrics as generated from the open-source algorithms. Descriptive statistics (mean ± SD) or median (interquartile range) where data were not normally distributed were calculated for all outputs. Pairwise 95% equivalence tests were used to determine whether the 95% CI for the mean of one accelerometer fell within a 10% equivalence zone of the second accelerometer (Wellek, 2003). The 10% equivalence zone might not be strict enough when the values are very high or the range is low, so a 5% equivalence zone was also used. Log transformation was applied to data that were not normally distributed. No accelerometer is considered a gold standard so equivalency tests were conducted with each accelerometer as the reference monitor for each accelerometer pairing. Postural and physical activity outcomes from accelerometer pairings were considered equivalent if equivalency was achieved irrespective of which accelerometer was the reference monitor. Equivalence results are shown with the ActiGraph as the reference when paired with the GENEActiv, Axivity, and ActiPAL, and Axivity as a reference when paired with the activPAL and GENEActiv, and the GENEActiv as a reference when paired with activPAL. Our previous work (Rowlands et al., 2018b), suggests that the ratio between the mean average acceleration outputs from the two accelerometers would be within 1 ± 0.05 and the standard deviation of the differences in the log transformed outputs would be less than 0.08. A sample size of 20 gives 80% power (alpha 0.05) to detect statistical equivalence of the anticipated effect size (Minitab v20.1; Minitab LLC, State College, PA).

The level of agreement between outcome variables was examined using mean absolute percent error (MAPE), pairwise Intra-class Correlation Coefficients (ICC) with 95% confidence intervals (CI) and mean bias and limits of agreement (LoA) (Bland & Altman, 1986). The level of reliability from ICCs was based on the lower bound of the 95% CI of the ICC values. If lower bound of the 95% CI was <0.5, reliability was classified as “poor,” 0.5–0.75 as “moderate,” >0.75–0.9 as “good” and >0.9 as “excellent” (Koo & Li, 2016). Percent agreement and kappa were used to determine epoch by epoch classification accuracy for upright vs sitting/lying (sitting/lying on the back, lying on the side or on the front).

Descriptive statistics such as MAPE, ICCs, LoA, and classification accuracy were calculated using IBM SPSS 25 Statistics, and equivalency tests were conducted in Minitab (v19). Alpha was set at 0.05.

### Results

Valid data across all accelerometers was provided by 20 of the 23 participants, consisting of seven males and 13 females (mean age: 28.7 ± 6.3 years [range 20–45 years]; mean BMI: 24.0 ± 3.8 kg/m² [range 19.8–31.3 kg/m²]). The data were invalid for two participants due to malfunction of the activPAL accelerometer and for one participant due to <20 hours wear-time on each day. Descriptive data (mean (SD), or median (25th–75th percentile) were not normally distributed) of postural and physical activities outcomes are presented in Table 1. Monitors were worn for 23.9 ± 0.4 h per 24 h day, with 4.9 ± 1.9 minutes of data imputed per 24 h day. Participants took 9825 ± 3948 steps per day (generated from the activPAL (Winkler et al., 2016)) and spent 10.7 ± 2.1 hours per day sedentary (excluding time reported sleeping).

Tables 2–5 present the MAPE (Table 2), ICCs (Table 3), mean bias, and LoA (Table 4) and a summary of the
equivalency (Table 5) between monitor pairing. Figures 1 and 2 displays the equivalency results.

**Postural Outputs**

Descriptive statistics were similar for sitting/lying and upright time across all monitors (Table 1). However, mean values suggested the activPAL tended to give higher estimates of sitting/lying on the back and lower estimates of lying on the side. MAPEs were lowest (0.3–0.8% for sitting/lying time and 1.5–2.5% for upright time) and reliability excellent (lower bound ICCs ≥0.99) for all monitor pairings for sitting/lying time and upright time. MAPEs were higher for the specific lying positions (back: 1.1–8.5%; side: 3.1–21.0%), with the monitor pairings which included the activPAL having consistently higher MAPEs. Reliability was good for sitting/lying on the back (lower bound ICCs 0.87–0.99), lying on the side (lower bound ICCs 0.85–0.98) and lying on the front (lower bound ICCs 0.81–0.95). All monitor pairings were within the 5% equivalency zone for the average time spent sitting/lying and upright. Results for the specific lying positions were less consistent, but for sitting/lying on the back and lying on the side all monitor pairings that did not include the activPAL were within the 10% equivalency zone. Epoch-by-epoch classification accuracy for upright vs sitting/lying was generally high irrespective of monitor pairing (mean percent accuracy 96.7% (6.3)), Table 2.

**Physical activity outputs**

Mean values for the activPAL were lower across all physical activity outputs in comparison to other monitors as were values from the ActiGraph compared to the GENEActiv and Axivity across the majority of outputs. The MAPEs were less than 8% (3.6–8.0%) for all monitor pairings for intensity gradients generated from ENMO and MAD. For all other outputs, there was greater variation in MAPE across monitor pairings, with the ActiGraph/Axivity pairing generally having lower MAPEs (ENMO: 3.6–11.3%; MAD: 3.0–15.5%) than other monitor pairings across most physical activity outputs. The M8h followed a different pattern with monitor pairings including the Axivity having the highest MAPE for ENMO outputs (38.2–44.2%), but monitor pairings including the activPAL having the highest MAPE for MAD outputs (63.2–177.0%). The monitor pairings which included the activPAL consistently had higher MAPEs (ENMO: 6.9–29.3%; MAD: 6.0–35.5% for GENEActiv/activPAL pairing; ENMO: 6.5–25.5%; MAD: 6.8–25.2 for ActiGraph/activPAL pairing; ENMO: 7.9–32.8%; MAD: 5.5–32.5 for Axivity/activPAL pairing) across all other physical activity outputs compared to the other monitors.

| Postural Outputs | activPAL | Axivity | GENEActiv | ActiGraph |
|------------------|----------|---------|-----------|-----------|
| Sitting/Lying on Back (mins/day) | 763.0 ± 141.6 | 721.6 ± 142.8 | 712.7 ± 139.2 | 725.6 ± 144.2 |
| Lying on Side (mins/day) | 225.5 ± 70.4 | 266.9 ± 70.2 | 257.8 ± 76.5 | 264.9 ± 70.5 |
| Lying on Front (mins/day) | 87.8 ± 66.6 | 71.9 (27.3–134.7) | 98.9 ± 71.8 | 81.1 ± 65.3 |
| Sitting/Lying (min/day) | 1075.6 | 1073.4 | (1014.8–1108.5) | (1019.3–1107.4) |
| Upright (min/day) | 364.4 | 366.56 | 368.1 | 363.6 |
| | (327.9–411.3) | (331.5–425.2) | (332.6–420.8) | (330.9–421.6) |

**Physical activity outputs**

| Physical Activity Outputs | activPAL | Axivity | GENEActiv | ActiGraph |
|---------------------------|----------|---------|-----------|-----------|
| **ENMO** | | | | |
| Average acceleration (mg) | 21.1 (18.1–28.8) | 27.4 (21.1–36.0) | 26.2 (20.7–32.8) | 24.2 (19.1–31.3) |
| Intensity Gradient | −1.99 ± 0.36 | −1.89 ± 0.29 | −1.89 ± 0.31 | −1.93 ± 0.31 |
| M30 (mg) | 240.2 ± 79.3 | 312.4 ± 126.4 | 311.6 ± 119.6 | 291.6 ± 110.1 |
| M60 (mg) | 162.0 ± 71.6 | 199.3 ± 108.2 | 194.9 ± 95.4 | 188.6 ± 98.4 |
| M8h (mg) | 9.5 ± 3.3 | 9.0 ± 5.0 | 7.7 ± 3.8 | 8.4 ± 4.4 |
| **MAD** | | | | |
| Average acceleration (mg/day) | 30.19 (21.04–33.55) | 37.36 (25.99–41.22) | 40.76 (29.67–44.92) | 35.08 (24.64–38.08) |
| Intensity Gradient (mg/day) | −1.68 ± 0.27 | −1.71 ± 0.28 | −1.77 ± 0.27 | −1.74 ± 0.32 |
| M30 (mg) | 344.5 ± 100.8 | 420.0 ± 142.2 | 425.6 ± 139.8 | 394.0 ± 129.0 |
| M60 (mg) | 242.7 ± 107.3 | 279.1 ± 129.3 | 283.3 ± 129.3 | 265.0 ± 123.2 |
| M8h (mg) | 5.8 ± 5.5 | 8.5 ± 5.4 | 10.4 ± 6.7 | 7.1 ± 5.7 |
MAD (ENMO: ICC 0.95–0.99). For M8h ENMO, reliability was moderate for the GENEActiv/activPAL, ActiGraph/GENEActiv, ActiGaph/Axivity (ICC: 0.52–0.65), but otherwise poor. All the monitored pairings were within the 10% equivalency zone for the intensity gradient for ENMO and MAD. The M30 and M60 were within 10% equivalence zone for ENMO and MAD for the Axivity and the GENEActiv and the ActiGraph and Axivity. The average acceleration was within 10% for the Axivity and GENEActiv (ENMO) and for the ActiGraph and Axivity (MAD). An inconsistent pattern of results were observed for the other outcomes and monitor pairings.

Discussion

Accelerometer-based devices worn on the thigh are now being used in large observational studies (Stamatakis et al., 2020). To aid comparisons across studies and to combine data from studies that have employed different accelerometer brands, it is important to understand which postural and physical activity metrics are comparable at the individual level. When information on the angle of the thigh is used for classification of posture, total sitting/lying time and upright time compare well across all monitor pairings. For directly measured acceleration, the majority of outcomes compared well for the Axivity, GENEActiv, and ActiGraph. Overall, the upright and sitting/lying postural outputs show the highest potential for data pooling across all four brands.

Recently, Crowley et al. (2019) compared time spent in types of physical behaviors (i.e., sitting/lying, standing, moving, walking, stair climbing, running, and cycling) across three different accelerometer brands (activPAL, Axivity and ActiGraph) worn simultaneously on the thigh during a semi-structured protocol of activities and a 7-day free-living measurement period. All data were processed using a custom developed software package called Acti4 and results showed only small absolute standard deviation values for lying/sitting (1.2 minutes/day), standing (3.4 minutes/day), moving (3.5 minutes/day), walking (1.9 minutes/day), running (0.1 minutes/day), stair climbing (1.2 minutes/day), and cycling (1.9 minutes/day). Considering these results alongside the present study, data harmonization using data collected from different accelerometer brands worn on the thigh appears to be most appropriate when outcomes of interest are broad postures (sit/lie or upright) or physical behavior types rather than acceleration magnitude-based metrics, even when the same processing methods are used. This is consistent with previous research that has demonstrated greater consistency between monitor brands for the orientation of the monitor (Edwardson et al., 2016; Rowlands et al., 2016) and acceleration features from the frequency domain (John et al., 2013; Rowlands et al., 2015) than for the magnitude of acceleration. For acceleration magnitude, the Axivity/GENEActiv (ENMO) and the Axivity/ActiGraph (MAD) pairings show the greatest potential for data harmonization as they could be considered

| Monitor 1 | GE | AP | GE | AP |
|---|---|---|---|---|
| Postural Outputs | | | | |
| Sitting/Lying on Back | 7.4 (5.3) | 8.0 (8.4) | 8.5 (8.1) | 4.2 (5.6) |
| Lying on Side | 16.2 (17.0) | 20.1 (21.2) | 21.0 (20.6) | 9.2 (8.9) |
| Lying on Front | N/A* | N/A* | N/A* | N/A* |
| All Sitting/Lying | 0.8 (0.8) | 0.6 (0.8) | 0.8 (0.8) | 0.5 (0.7) |
| Upright | 2.5 (3.1) | 1.7 (2.1) | 2.2 (2.5) | 1.5 (2.5) |
| Epoch-by-epoch percent classification accuracy (SD) | | | | |
| Upright vs All Sitting/Lying | 94.4 (8.4) | 97.0 (3.1) | 96.7 (3.3) | 96.5 (8.5) |
| Physical Activity Outputs | | | | |
| ENMO | | | | |
| Average acceleration | 26.4 (6.8) | 22.3 (18.2) | 29.4 (20.4) | 12.5 (9.8) |
| Intensity Gradient | 6.5 (4.2) | 6.6 (4.4) | 8.0 (5.1) | 5.2 (5.2) |
| M30 | 29.8 (10.1) | 25.5 (18.9) | 32.8 (18.8) | 10.5 (9.7) |
| M60 | 24.8 (12.5) | 24.8 (12.5) | 27.6 (20.2) | 10.4 (9.3) |
| M8h | 23.7 (14.5) | 32.9 (25.7) | 38.2 (33.5) | 25.6 (30.0) |
| MAD | | | | |
| Average acceleration | 34.6 (9.0) | 16.6 (8.5) | 24.4 (10.0) | 13.0 (6.9) |
| Intensity Gradient | 6.3 (6.9) | 7.2 (7.1) | 5.3 (5.8) | 4.8 (5.5) |
| M30 | 21.1 (9.8) | 13.1 (9.4) | 20.2 (11.0) | 8.4 (5.8) |
| M60 | 14.7 (6.6) | 9.7 (8.5) | 14.9 (10.0) | 8.4 (6.4) |
| M8h | 177.0 (276.3) | 63.2 (123.8) | 123.4 (218.0) | 33.7 (10.7) |

GE: GENEActiv; AP: activPAL; AG: ActiGraph; AX: Axivity. * MAPEs could not be calculated due to zeros values recorded by an accelerometer in this pairing for some participants.
equivalent for all ENMO outputs except for the intensity of the least active 8 h.

The intensity gradient (calculated from either ENMO or MAD) showed comparability across all monitor pairings, likely because this metric describes the distribution (or pattern) of acceleration intensity across the 24 h profile metric, rather than the magnitude of acceleration. This is in line with previous research where the intensity gradient was equivalent between all accelerometer brands worn on either wrist (Rowlands et al., 2019b). However, the intensity gradient has a narrow range of values, and thus the 10% equivalence zone encompasses values that differ by half a standard deviation, which limits its physiological and clinical relevance. The range of average acceleration encompassed by the 10% zone is 0.2 standard deviations, which equates to a 3% equivalency zone for the intensity gradient. Our results suggest that the intensity gradient falls into this 3% equivalency zone for the Axivity/GENEActiv and ActiGraph/Axivity pairings, which is consistent with equivalency results for the other acceleration outcomes. The results for the intensity of the most active 8 h suggest that the equivalence between monitors is lowest and MAPE highest for very low acceleration magnitudes, particularly for the MAD metric. This is important as it may have implications for low-active populations where accelerations throughout the day are lower; however, reliability was moderate

Table 3. Intra-class correlations between pairs of monitors.

| Monitor 1 | GE | AP | AG | AX | GA | GE | AX | GA | AX | GE | AG | AX | GA |
|-----------|----|----|----|----|----|----|----|----|----|----|----|----|----|
| GE        | 0.98 (0.94,0.99) | 0.95 (0.97,0.98) | 0.95 (0.87,0.98) | 0.98 (0.94,0.99) | 0.98 (0.95,0.99) | 0.99 (0.99,0.99) |
| AP        | 0.96 (0.89,0.98) | 0.94 (0.85,0.98) | 0.95 (0.88,0.98) | 0.95 (0.87,0.98) | 0.96 (0.90,0.99) | 0.99 (0.98,0.99) |
| AG        | 0.95 (0.88,0.98) | 0.84 (0.60,0.94) | 0.81 (0.53,0.93) | 0.90 (0.76,0.96) | 0.95 (0.88,0.98) | 0.94 (0.84,0.98) |
| AX        | 0.99 (0.99,0.99) | 0.99 (0.99,0.99) | 0.99 (0.99,0.99) | 0.99 (0.99,0.99) | 0.99 (0.99,0.99) | 0.99 (0.99,0.99) |
| Intra-class correlation (ICC) (95% CI) |

Table 4. Agreement between pairs of monitors.

| Monitor 1 | GE | AP | AG | AX | GA | GE | AG | AX | GA | GE | AG | AX | GA |
|-----------|----|----|----|----|----|----|----|----|----|----|----|----|----|
| GE        | 50.3 | −37.5 | −41.4 | 12.9 | 8.9 | 4.0 | 85.4 | 126.3 | 125.0 | 85.3 | 79.1 | 24.1 |    |
| AP        | 32.4 | 39.5 | 41.4 | 7.1 | 9.0 | 1.9 | 91.9 | 65.1 | 58.8 | 63.0 | 54.9 | 25.1 |    |
| AG        | 11.1 | −6.7 | −7.2 | −17.8 | −18.3 | 0.4 | 58.3 | 95.2 | 102.1 | 57.5 | 65.0 | 10.3 |    |
| AX        | −6.9 | −4.7 | −7.2 | 2.2 | −0.3 | 2.4 | 20.3 | 17.9 | 21.2 | 17.4 | 20.0 | 10.4 |    |
| Physical Activity Outputs |

Table 3. Intra-class correlations between pairs of monitors.

| Monitor 1 | GE | AP | AG | AX | GA | GE | AG | AX | GA | GE | AG | AX | GA |
|-----------|----|----|----|----|----|----|----|----|----|----|----|----|----|
| GE        | 0.95 (0.87,0.98) | 0.93 (0.83,0.97) | 0.91 (0.78,0.97) | 0.96 (0.90,0.98) | 0.96 (0.91,0.99) | 0.96 (0.90,0.98) |
| AP        | 0.96 (0.89,0.98) | 0.94 (0.84,0.98) | 0.91 (0.77,0.96) | 0.96 (0.89,0.98) | 0.93 (0.82,0.97) | 0.96 (0.90,0.98) |
| AG        | 0.94 (0.84,0.98) | 0.91 (0.76,0.96) | 0.89 (0.73,0.96) | 0.94 (0.85,0.98) | 0.95 (0.88,0.98) | 0.99 (0.96,0.99) |
| AX        | 0.96 (0.89,0.98) | 0.93 (0.82,0.97) | 0.91 (0.78,0.97) | 0.97 (0.92,0.99) | 0.97 (0.92,0.99) | 0.99 (0.98,0.99) |
| ENMO      | 0.86 (0.65,0.95) | 0.64 (0.09,0.86) | 0.55 (−0.14,0.82) | 0.86 (0.64,0.94) | 0.77 (0.42,0.91) | 0.81 (0.52,0.92) |
| MAD       | 0.98 (0.96,0.99) | 0.98 (0.97,0.99) | 0.98 (0.96,0.99) | 0.98 (0.96,0.99) | 0.98 (0.96,0.99) | 0.98 (0.96,0.99) |

GE: GENEActiv; AP: activPAL; AG: ActiGraph; AX: Axivity;
Table 5. Equivalency between pairs of monitors.

| Monitor 1 | GE | AG | AX | GE | AG | AX | GE | AG | AX | AG | AX | AG |
|-----------|----|----|----|----|----|----|----|----|----|----|----|----|
| Monitor 2 | AP | AP | AP | GE | GE | GE | AP | AP | AP | GE | GE | GE |

**Postural Outputs**

| Position | 10% | 5% |
|----------|-----|----|
| Sitting/Lying on | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Back | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Lying on Side | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Lying on Front* | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| All Sitting/Lying | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Upright | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |

**Physical Activity Outputs**

**ENMO**

| Measure | 10% | 5% |
|---------|-----|----|
| Average acceleration | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Intensity Gradient | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| M30 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| M60 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| M8h | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |

**MAD**

| Measure | 10% | 5% |
|---------|-----|----|
| Average acceleration | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Intensity Gradient | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| M30 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| M60 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| M8h | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |

Equivalence (can be considered equivalent)*

GE: GENEActiv; AP: activPAL; AG: ActiGraph; AX: Axivity; * = 95% confidence interval (CI) for mean of second monitor falls into the proposed equivalency zone of the mean of the first monitor and vice versa, thus can be considered equivalent. * Equivalence could not be calculated due to zeros values recorded by an accelerometer in this pairing for some participants.

Figure 1. Equivalence between pairs of monitors for postural outputs. Error bars represent the 95% confidence interval of the ratio. Equivalence = 1 (solid line). Dashed lines indicate the 10% equivalence zone and dotted lines indicate a 5% equivalence zone. GE: GENEActiv; AP: activPAL; AG: ActiGraph; AX: Axivity.
Figure 2. Equivalence between pairs of monitors for physical activity outputs. Error bars represent the 95% confidence interval of the ratio. Equivalence = 1 (solid line). Dashed lines indicate the 10% equivalence zone and dotted lines indicate a 5% equivalence zone. GE: GENEActiv; AP: activPAL; AG: ActiGraph; AX: Axivity.

(ENMO) to excellent (MAD), and mean bias was relatively low.

The activPAL tended to record lower acceleration values for the physical activity outputs calculated from direct acceleration relative to the other devices. The lower acceleration values observed with the activPAL device were anticipated due to the lower dynamic range (±2 g vs ±8 g) and frequency (20 Hz vs 100 Hz) of the device settings for data collection in comparison to other brands (John et al., 2019). The ActiGraph device also recorded lower accelerations for the physical activity outputs relative to the GENEActiv and Axivity accelerometers. This has been reported previously when accelerometers brands were worn concurrently on the erometers. This has been reported previously when
tivity outputs relative to the GENEActiv and Axivity accelerometers. The main strengths of this study were the use of open software methods to process the data from all monitors, the free-living assessment, and simultaneous wearing of four different accelerometer brands. However, a number of limitations should be noted. Firstly, in order for the monitors to be worn on a similar part of the thigh, we made the decision to tape monitors together. The activPAL was taped on top of the GENEActiv and the Axivity on top of the ActiGraph which may have affected the results, specifically this could have impacted on the different lying positions. Second, the epoch duration differed for postures and physical activity analyses; to ensure the results were relevant to people accessing the open-source methods the default epoch settings of 15 s for the posture spreadsheet and of 5 s for GGISR outcomes were used. This may have impacted on the findings. Finally, our sample size was small and did not include any participants over the age of 45 years. Our sample did however include a range of activity levels; the average acceleration ranged from 10 to 70 mg and the steps per day ranged from 3475 to 20,640. In view of the poorer results for the lowest acceleration outcomes (intensity of the most active
8 h), comparability of acceleration outcomes in a lower active population should be investigated.
In conclusion, the results suggest that average time spent in total sitting/lying, upright time derived from any of the monitors tested have a strong potential for harmonization. Overall, most acceleration outcomes compared well for the Axivity, GENEActiv and ActiGraph for both the ENMO and MAD acceleration metrics, but care is recommended when comparing acceleration magnitude outcomes between brands. For the widely used ENMO acceleration metric (Fairclough et al., 2019; Migueles, Cadenas-Sanchez et al., 2019; Rowlands et al., 2019a), the Axivity and the GENEActiv appear most suited to harmonization, however, the magnitude of acceleration from the activPAL is lower than from the other monitor brands. These findings will inform studies looking to harmonize individual-level data from studies deploying accelerometer data, e.g. ProPASS

Acknowledgments

Dr. Lynne Boddy, Liezel Hurter, and Dr. Karl Gibbons from Liverpool John Moores University for generously sharing their custom-built program (GT9X-to-SedSphere) written in MATLAB (R2017b, The MathWorks Inc., Natick, MA, USA).

ORCID

Benjamin D Maylor http://orcid.org/0000-0002-4503-0479
Alex V Rowlands http://orcid.org/0000-0002-1463-697X

Author contributions

CLE had the original idea for the study and all authors contributed to the study design. CLE wrote the first draft of the manuscript. AVR and BM contributed to data processing. CLE analyzed the data. All authors contributed to the interpretation of the data, reviewed/edited the manuscript and approved the final manuscript.

Conflicts of Interest and Source of Funding

The analysis was supported by the National Institute for Health Research (NIHR) Leicester Biomedical Research Centre which is a partnership between University Hospitals of Leicester NHS Trust, Loughborough University, and the University of Leicester.

References

Aittasalo, M., Vähä-Yppä, 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. https://doi.org/10.1186/s13102-015-0010-0.

Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. The Lancet, 1(8476), 307–310. https://doi.org/10.1016/S0140-6736(86)90837-8

Buchan, D. S., McLeLanG, G., Donnelly, S., & Arthur, R. (2019). The use of the intensity gradient and average acceleration metrics to explore associations with BMI z-score in children. Journal of Sports Sciences, 37(23), 2751–2758. https://doi.org/10.1080/02640414.2019.1664536.

Crowley, P., Skotte, J., Stamatakis, E., Hamer, M., Aadhalm, M., Stevens, M. L., Rangul, V., Mork, P. J., & Hollermann, A. (2019). Comparison of physical behavior estimates from three different thigh-worn accelerometers brands: A proof-of-concept for the Prospective Physical Activity, Sitting, and Sleep consortium (ProPASS). The International Journal of Behavioral Nutrition and Physical Activity, 16(1), 65. https://doi.org/10.1186/s12966-019-0835-0.

Daugaard, R., Tjur, M., Sliepen, M., Lipperts, M., Grimm, B., & Mechlenburg, I. (2018). Are patients with knee osteoarthritis and patients with knee joint replacement as physically active as healthy persons? Journal of Orthopaedic Translation, 14, 8–15. https://doi.org/10.1016/j.jot.2018.03.001.

Doherty, A., Jackson, D., Hammerla, N., Plotz, T., Olivier, P., Granat, M. H., …. Wareham, N. J. (2017). Large Scale Population Assessment of Physical Activity Using Wrist Worn Accelerometers: The UK Biobank Study. PLoS ONE, 12(2), e0169649. https://doi.org/10.1371/journal.pone.0169649.

Edwardson, C. L., Rowlands, A. V., Bunnewell, S., Sanders, J., Eslinger, D., Gorely, T., O’Connell, S., Davies, M., Khunti, K., & Yates, T. (2016). Accuracy of posture allocation algorithms for thigh and waist-worn accelerometers. Medicine and Science in Sports and Exercise, 48(6), 1085–1090. https://doi.org/10.1249/MSS.0000000000000865.

Fairclough, S. J., Taylor, S., Rowlands, A. V., Boddy, L., & Noonan, R. (2019). Average acceleration and intensity gradient of primary school children and associations with indicators of health and wellbeing. Journal of Sports Sciences, 37(18), 2159–2167. https://doi.org/10.1080/02640414.2019.1624313.

Hartley, P., Keevil, V. L., Westgate, K., White, T., Brage, S., Romero-Ortuno, R., & Deaton, C. (2018). Using Accelerometers to Measure Physical Activity in Older Patients Admitted to Hospital. Current Gerontology and Geriatrics Research, 2018, Article ID: 3280240. https://doi.org/10.1155/2018/3280240.

Hurter, L., Rowlands, A. V., Fairclough, S. J., Knowles, Z. R., Porcellato, I. A., Cooper, A. M., & Boddy, L. M. (2019). Validating the Sedentary Sphere method in children: Does wrist or accelerometer brand matter. Journal of Sport Sciences, 37(16), 1910–1918. https://doi.org/10.1080/02640414.2019.1605647.

John, D., Sasaki, J., Staudemayer, J., Mavilia, M., & Freedson, P. S. (2013). Comparison of raw acceleration from the GENEA and ActiGraphTM GT3X+ activity monitors. Sensors, 13(11), 14754–14763. https://doi.org/10.3390/s131114754.

John, D., Tang, Q., Albinali, F., & Intille, S. (2019). An open source monitor-independent movement summary for accelerometer data processing. Journal for the Measurement of Physical Behaviour, 2(4), 268–281. https://doi.org/10.1123/jmpb.2018-0068.
in Physical Education and Exercise Science, 23(3), 249–257. https://doi.org/doi:10.1080/1091367X.2019.1613997

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

Vähä-Ypyä, H., Vasankari, T., Husu, P., Suni, J., & Sievänen, H. (2015b). A universal, accurate intensity-based classification of different physical activities using raw data of accelerometer. Clinical Physiology and Functional Imaging, 35(1), 64–70. https://doi.org/10.1111/cpf.12127.

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. Journal of Applied Physiology, 117(7), 738–744. https://doi.org/10.1152/japplphysiol.00421.2014.

van Hees, V. T., Gorzelniak, L., Dean León, E. C., Eder, M., Pias, M., Taherian, S., Ekelund, U., Renström, F., Franks, P. W., Horsch, A., & Brage, S. (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. https://doi.org/10.1371/journal.pone.0061691.

Wellek, S. (2003). Testing Statistical Hypotheses of Equivalence. Chapman & Hall/CRC.

Winkler, E. A., Bodicoat, D. H., Healy, G. N., Bakrania, K., Yates, T., Owen, N., Dunstan, D. W., & Edwardson, C. L. (2016). Identifying adults’ valid waking wear time by automated estimation in activPAL data collected with a 24 h wear protocol. Physiological Measurement, 37(10), 1653–1668. https://doi.org/10.1088/0967-3334/37/10/1653.