Equivalence of activity outcomes derived from three research grade accelerometers worn simultaneously on each wrist

Buchan, Duncan S.

Published in:
Journal of Sports Sciences

DOI:
10.1080/02640414.2021.2019429

Published: 03/04/2022

Document Version
Publisher's PDF, also known as Version of record

Routledge
Taylor & Francis Group

Journal of Sports Sciences

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/rjsp20

Equivalence of activity outcomes derived from three research grade accelerometers worn simultaneously on each wrist

Duncan S. Buchan

To cite this article: Duncan S. Buchan (2022) Equivalence of activity outcomes derived from three research grade accelerometers worn simultaneously on each wrist, Journal of Sports Sciences, 40:7, 797-807, DOI: 10.1080/02640414.2021.2019429

To link to this article: https://doi.org/10.1080/02640414.2021.2019429

Published online: 28 Dec 2021.

JOURNAL OF SPORTS SCIENCES
2022, VOL. 40, NO. 7, 797–807
https://doi.org/10.1080/02640414.2021.2019429

Routledge
Taylor & Francis Group

PHYSICAL ACTIVITY, HEALTH AND EXERCISE

Equivalence of activity outcomes derived from three research grade accelerometers worn simultaneously on each wrist

Duncan S. Buchan

Division of Sport and Exercise, School of Health and Life Sciences, University of the West of Scotland, Lanarkshire, UK

ABSTRACT

This study evaluated the equivalence of activity outcomes from three accelerometer brands worn on both wrists during free living. Forty-four adults wore a GENEActiv, ActiGraph and Axivity accelerometer for 7 days. Outcomes were assessed between and within accelerometer brand and wrist location with average acceleration and the intensity gradient (IG) being of particular interest. Pairwise 95% equivalence tests and intra-class correlation coefficients (ICC) evaluated agreement. Average acceleration and the IG were largely equivalent between combinations of accelerometer device and wrists when applying a 10% equivalence zone. There was largely a lack of equivalence between pairings for time spent in acceleration values ≥100 m*g*. However, equivalence was largely achieved when applying an equivalence zone that encompassed values ranging from 0.3 to 0.45 SDs for IG and time spent above 100 m*g* and 150 m*g*. Agreement between pairings tended to be stronger between different brands on the non-dominant (ICCs ≥ 0.73–0.97) versus the dominant wrist (ICCs ≥ 0.57–0.97) and between wrists for the same accelerometer (ICCs ≥ 0.59–0.97) for average acceleration and the IG. These are important findings since device placement is not consistent in studies. Further work that applies an equivalence zone reflecting the variability of the outcome measure is encouraged.

ARTICLE HISTORY
Accepted 10 December 2021

KEYWORDS
GGIR; agreement; intensity gradient; physical activity; average acceleration

Introduction

Accelerometers are often used to measure habitual physical activity (PA) to understand the prevalence of this behaviour and to evaluate the effectiveness of public health initiatives (Troiano et al., 2014). Three research-grade accelerometer brands that are widely used by researchers include the ActiGraph (ActiGraph LLC, Pensacola, FL, USA), GENEActiv (ActivInsights Ltd., Cambridgeshire, UK) and the Axivity (Axivity Ltd., Newcastle, UK). Establishing whether these three accelerometer brands provide equivalent estimates of activity outcomes is important for researchers to have confidence in pooling data across studies using different accelerometer brands. For instance, different brands of wrist-worn accelerometers have been deployed in large-scale studies including the 2011–2012 and 2013–2014 cycles of the U.S. National Health and Nutrition Examination Survey (NHANES; NHANES, 2021), the UK Biobank study (Doherty et al., 2017) and the Pelotas Birth Cohort (Ricardo et al., 2020). In these studies, the ActiGraph GT3X+ was used in NHANES, the GENEActiv was used in the Pelotas Birth Cohort and the Axivity was used in the UK Biobank study. Furthermore, participants wore the ActiGraph and GENEActiv on their nondominant wrists in the NHANES and Pelotas Birth Cohort studies, respectively, whereas the Axivity was deployed on participants' dominant wrist in the UK Biobank study.

There is a paucity of published research that has examined the equivalence of outcomes from the ActiGraph, Axivity and GENEActiv when worn on the same and different wrist locations. The use of open-source resources such as the GGIR package (Migueles, Rowlands et al., 2019) in R [http:/cran.r-prouect.org] provides a means of processing data from the three accelerometer brands in an identical and transparent way that can facilitate data harmonisation between studies. When using the GGIR package, average acceleration is often compared between accelerometer brands since it reflects the volume of PA undertaken and can be used to classify time spent in different activity intensities (Rowlands et al., 2018). Previous work comparing output from the GENEActiv, Axivity (taped together) and ActiGraph (worn proximal to the GENEActiv and Axivity devices) on the non-dominant wrist found that average acceleration was approximately 10% lower from ActiGraph devices (Rowlands et al., 2017, 2016). These findings are in contrast to a recent study, which found equivalent average acceleration outputs between GENEActiv, Axivity and ActiGraph GT9X devices when worn on the non-dominant wrist but almost 10% lower average acceleration values from the ActiGraph in contrast to the GENEActiv and Axivity devices when worn on the dominant wrist (Rowlands, Plekhanova et al., 2019). The use of different participants, wrist locations, monitoring periods, device pairings and settings may explain the contrasting study findings.

For instance, in these earlier studies, subjects wore the GENEActiv, Axivity and ActiGraph GT9X devices on their non-dominant wrist with the Axivity taped to the GENEActiv and worn distal to the ActiGraph for 2 h in a stimulated living space (Rowlands et al., 2017). Elsewhere participants wore the

CONTACT Duncan S. Buchan duncan.buchan@uws.ac.uk Division of Sport and Exercise, School of Health and Life Sciences, University of the West of Scotland, Lanarkshire Campus G72 0LH, UK

GENEActiv and ActiGraph GT3X+ (worn proximal to the GENEActiv) for 2 days to ensure one complete nocturnal sleep was captured (Rowlands et al., 2016). However, in the more recent study, participants wore the GENEActiv, Axivity and ActiGraph GT9X devices on each wrist and were instructed to wear them for 24 hours a day for up to 7 days (Rowlands, Plekhanova et al., 2019). The Axivity was taped on top of the ActiGraph device with the positional order of monitors randomised, but identical on each wrist, between participants (Rowlands, Plekhanova et al., 2019). It has also been shown that on-board filtering can occur in ActiGraph devices, which can lead to accelerations being filtered out when captured during moderate-higher intensity activities (Fridolfsson et al., 2019). This, as well as the differences noted between the studies, could explain the lack of equivalence for average acceleration between the ActiGraph, Axivity and GENEActiv devices and the contrasting results between the studies.

Ensuring comparability between outcomes across accelerometer brands and wrist locations is crucial if data is to be pooled from different studies. However, the use of different models of ActiGraph accelerometers may also limit the pooling of data across studies. The most recent release of ActiGraph devices include the GT9X (2014), wGT3X-BT (2013) and the GT3X+ (2010). Despite the GT9X and wGT3X-BT appearing to have the same internal accelerometer, they differ in size (wGT3X-BT: 4.6 × 3.3 × 1.5 cm, 19 g; GT9X: 3.5 × 3.5 × 1.0 cm, 14 g), shape, method of attachment and in other internal components (e.g., the GT9X has an inertial measurement unit, whereas a lux sensor is used within the wGT3X-BT (ActiGraph, 2019). Conversely, the wGT3X-BT and GT3X+ (with the latter used in the NHANES 2011–2014) are the same size and shape but have different dynamic ranges (wGT3X-BT: ± 8 *g*; GT3X+: ± 6 *g*), which suggests these models contain different internal accelerometers (Actigraph, 2013; ActiGraph, 2019; Clevenger et al., 2020). It is plausible that these differences could affect acceleration output between ActiGraph models and therefore, impact on the pooling of data from studies deploying Axivity and GENEActiv accelerometers, with studies using previous models of ActiGraph devices.

Indeed, this was evident from recent studies that compared acceleration output between hip worn ActiGraph GT3X+ vs. GT9X accelerometers (Montoye et al., 2018) and wrist worn GT3X+ or wGT3X-BT versus GT9X accelerometers worn simultaneously on each wrist (Clevenger et al., 2020). Findings from both studies suggest that the GT9X device appears to record higher acceleration values than previous ActiGraph models. Given these findings, further work is needed to examine the equivalence of activity outcomes from Axivity, GENEActiv and previous generations of ActiGraph accelerometers. Examining the comparability of outcomes from previous generations of ActiGraph models with Axivity and GENEActiv devices may also facilitate comparisons and/or pooling of data from large scale studies that have used these three accelerometer devices (Doherty et al., 2017; NHANES, 2021; Ricardo et al., 2020). Therefore, this study aimed to establish the equivalence of activity outcomes from three research-grade accelerometers as follows: 1) between wrist, within brand; 2) within wrist, between brand and finally 3) between wrist, between brand. Since average acceleration and the IG have been proposed as metrics that can facilitate the pooling/comparison of PA data (Rowlands, Plekhanova et al., 2019), findings from these metrics were of particular interest.

Materials and methods

A sample of 53 participants were recruited by email sent to the University of the West of Scotland's student body and word of mouth between October 2019 and January 2020. Written informed consent was provided by all participants, and the ethics committee of the University of the West of Scotland approved all study procedures.

The participants stature and mass were measured without shoes and in light clothing using a calibrated scale (Seca 880 and 770, Digital Scales, Seca Ltd, Birmingham, UK) and stadiometer (Seca Stadiometer, Seca Ltd, Birmingham, UK), respectively. After confirmation of their dominant wrist, the participants were fitted with two Axivity AX3 (herein: Axivity) (Axivity Ltd, Newcastle, UK), two ActiGraph wGT3X-BT (herein: ActiGraph) (ActiGraph LLC, Pensacola, FL, USA) and two GENEActiv (ActivInsights Ltd, Cambridgeshire, UK) devices and instructed to wear them 24 hr/day for 7 days. The Axivity was taped on top of the GENEActiv with the ActiGraph worn alongside this combination. Taping devices on top of another is common in studies to remove the need for additional straps and for comfort reasons (Rowlands et al., 2017; Rowlands, Plekhanova et al., 2019; Rowlands et al., 2016). The order of monitors on the participants' wrist was randomized (i.e., proximal vs. distal) but the order was consistent between wrists.

Accelerometers

All three devices are triaxial accelerometers. ActiGraph and GENEActiv devices have a dynamic range of ±8 *g*, whereas the Axivity allows users to select the dynamic range up to a maximum of ±16 *g*, where *g* is equal to the Earth's gravitational pull. Axivity devices were initialised with the dynamic range set to 8 *g* with data subsequently downloaded using the open-source software OmGui (OmGui, version 1.0.0.43, Open Movement, Newcastle, UK). ActiGraph devices were initialised and data downloaded using ActiLife v6.13.3 (Actigraph, Pensacola, FL, USA) with data files saved in raw format as ActiGraph.gt3x files, before being converted to .csv format for data processing. The "idle sleep mode" was not enabled in ActiLife v6.13.3. The GENEActiv devices were initialised with data downloaded using the GENEActiv PC software v3.2 (ActivInsights Ltd, Cambridgeshire, UK), which saved the raw data as .bin files. To ensure consistency, all devices were initialized using the same PC and configured to record data at 100 Hz. All accelerometers were set to commence data collection once fitted to the participants' wrists.

Data processing

All files were processed using the GGIR package version 2.3–0 in R statistical software (R Foundation for Statistical Computing, Vienna, Austria, https://cran.r-project.org/), which detected sustained and abnormally high values and autocalibrated the files (Van Hees et al., 2014). The GGIR package also computed Euclidean Norm Minus One (ENMO, herein average acceleration)

over a 5-s epoch, expressed in milli-gravitational units (m*g*) as the square root of the sum of the squared values of the raw acceleration signals, minus 1, with negative values rounded up to zero (Van Hees et al., 2013). Several outcomes were provided from each device. These included wear time (days), average acceleration (m*g*), intensity gradient (IG) and time accumulated above incremental acceleration thresholds including >0, >50, >100, >150, >200, >250, >300, >350, >400 m*g*. Briefly, average acceleration provides a proxy measure of the volume of activity, whereas the IG describes the intensity distribution of accelerations across the monitoring period (Rowlands, 2018). The IG describes the negative curvilinear relationship between the intensity of PA and the time accumulated at that intensity across the monitoring period and is always negative (Rowlands et al., 2018). A more negative (lower) IG value is reflective of little time accumulated at mid-higher intensities, whereas a less negative (higher) IG value is reflective of more time spent across the intensity range. These two metrics provide detail of the volume and intensity of activity undertaken across the monitoring period and have been proposed as useful metrics for comparing and/or pooling PA data (Rowlands, Plekhanova et al., 2019).

Files were excluded from subsequent analyses if post-calibration error was >0.01 *g* or participants has less than 1 day of valid wear data (defined as ≥16 h per day (Rowlands et al., 2018)) or wear data was not present for each 15 min period of the 24-h cycle. Confirmation of comparable classification of wear and non-wear between devices was confirmed numerically and by reviewing the accelerometer traces provided by GGIR. Outcomes were calculated from valid days only. We used the default non-wear setting, whereby invalid data were imputed by the average at similar time points on different days of the week (Van Hees et al., 2013). This ensured that outcomes were calculated based on the entire 24 h cycle.

Statistical analysis

Participants had to provide valid files from all six devices to be included within subsequent analyses. Descriptive statistics were calculated for all outcomes (mean (SD) or median (25^{th} – 75^{th} percentile) after testing for normality. Level of agreement between outcomes were examined using intraclass correlation coefficients (ICC, two-way mixed effects, single measures, absolute agreement) with 95% confidence intervals (95% CI) and limits of agreement (LoA; Bland & Altman, 1986). Based on the 95% CI of the ICC estimate, values <0.5, 0.5–0.75, 0.75–0.9 and >0.90 were indicative of poor, moderate, good and excellent agreement, respectively (Koo & Li, 2016). The equivalence of data outcome between devices were examined using pairwise 95% equivalence tests to establish whether the 95% CI for the mean of one accelerometer fell within the proposed equivalence zone of the alternate accelerometer (Dixon et al., 2018; Wellek, 2003). We defined our equivalence zone as ±10% of the chosen reference method as used previously (Buchan et al., 2020; Rowlands, Plekhanova et al., 2019). Equivalence analyses were performed using the log transformation of the original data when data was not normally distributed. Finally, equivalence tests were carried out twice with each device used as the reference monitor with comparisons only reported as equivalent if equivalence was achieved when both monitors were used as the reference device. To aid interpretations, findings from the equivalence analyses and ICC's were the main outcomes of interest with mean bias and LoA included in supplementary files. Statistical analyses were undertaken using IBM SPSS statistical software for Windows version 25 (IBM, Armonk, NY). Equivalence testings were undertaken in Minitab (v17). Alpha was set at 0.05.

Results

Upon completion of the study, 44 participants (18 female, mean (SD) age: 24.3 (7.9) y; stature 170.1 (10.1) cm; mass (69.4 (10.7) kg) provided valid data from the six accelerometer devices to be included within subsequent analysis. Across the six devices, the median (25th-75th percentile) for wear time was 24 h per day (21–24 h) with participants providing a median of 5.1 days (4.8–5.4 days) of monitoring days from the GENEActiv (both wrists) and 5 days (4.8–5.4 days) of monitoring days from both the ActiGraph and Axivity devices (both wrists). Data from nine participants were removed due to data not being available for each 1-hour period of the 24-hour cycle (n = 4), unknown technical issue (n = 2), post-calibration error (n = 2) and participant withdrawal (n = 1). Descriptive data from the six accelerometer devices can be found in Table 1.

Between wrists for the same accelerometer brand

Figure 1 provides the results for a) GENEActiv, b) Axivity and c) ActiGraph comparisons with additional analysis covering mean bias and LoA provided within Table S1. Across all three

Table 1. Activity outcomes from the GENEActiv, Axivity and ActiGraph accelerometers. N = 44.

	Non-dominant wrist			Dominant wrist		
Activity outcomes	GENEActiv	Axivity	ActiGraph	GENEActiv	Axivity	ActiGraph
Total wear time (days)	5.1 (4.8–5.4)	5.0 (4.8–5.2)	5.0 (4.8–5.2)	5.1 (4.8–5.4)	5.0 (4.8–5.2)	5.0 (4.8–5.2)
Average acceleration (m*g*)	26.9 (9.2)	27.9 (9.5)	26.8 (8.8)	28.4 (9.2)	29.7 (10.1)	27.5 (8.1)
Intensity gradient	−2.54 (0.20)	−2.56 (0.20)	−2.56 (0.21)	−2.52 (0.18)	−2.50 (0.18)	−2.50 (0.18)
Time below 50 m*g* (min)	1205.8 (90.4)	1199.3 (90.9)	1211.8 (89.0)	1198.8 (85.9)	1192.3 (85.4)	1211.6 (79.2)
Time above 50 m*g* (min)	234.2 (90.4)	240.7 (90.9)	228.2 (89.0)	241.3 (85.9)	247.6 (85.3)	228.4 (79.2)
Time above 100 m*g* (min)	90.7 (64.9–130.9)	93.4 (64.4–147.1)	87.5 (48.2–127.6)	95.3 (69.0–143.1)	110.4 (69.1–150.8)	102.9 (62.8–145.6)
Time above 150 m*g* (min)	41.5 (28.2–57.2)	39.9 (25.1–68.9)	36.0 (16.7–60.9)	42.6 (29.4–63.2)	47.6 (28.8–88.3)	39.1 (28.2–68.4)
Time above 200 m*g* (min)	15.8 (10.6–26.6)	15.3 (9.7–32.2)	16.4 (5.7–25.5)	18.0 (12.6–38.8)	20.0 (14.3–39.0)	17.3 (10.8–33.0)
Time above 250 m*g* (min)	7.1 (4.0–15.1)	6.1 (4.3–17.8)	7.3 (2.8–14.1)	9.4 (6.3–14.4)	10.8 (6.7–22.8)	7.9 (4.5–17.0)
Time above 300 m*g* (min)	3.9 (1.7–9.3)	3.6 (1.9–9.8)	3.6 (1.6–8.3)	5.5 (3.4–8.9)	5.4 (3.8–13.2)	4.8 (2.7–10.3)
Time above 350 m*g* (min)	2.2 (1.0–6.2)	2.2 (1.0–6.4)	2.2 (0.9–5.3)	3.5 (1.9–6.2)	3.5 (2.0–7.4)	3.0 (1.8–6.6)
Time above 400 m*g* (min)	1.5 (0.6–4.4)	1.6 (0.6–4.6)	1.7 (0.6–3.8)	2.2 (1.2–4.8)	2.4 (1.1–5.3)	1.9 (1.2–5.2)

Normally distributed data are presented as mean (SD), otherwise median (25th-75th percentile)

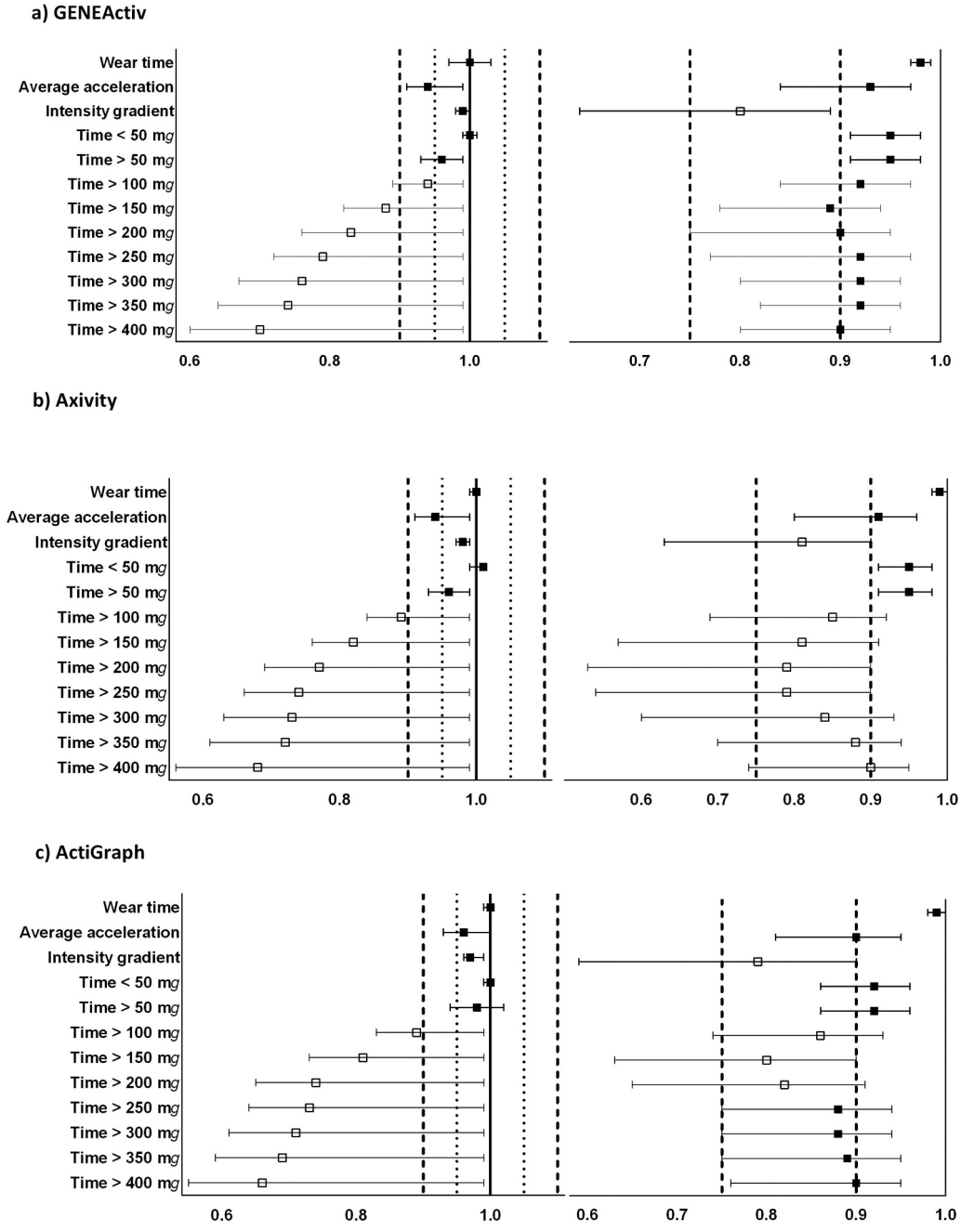

Figure 1. Equivalence and ICC between pairs of monitors between wrists for the same accelerometer brand: (a) GENEActiv, (b) Axivity and (c) ActiGraph. Equivalence analysis is presented in the left panel with dashed lines representing the 10% equivalence zone with a 5% equivalence zone indicated by the dotted lines. The ICC analysis is provided in the right panel with dashed lines representing ICCs >0.75 and >0.9. The horizontal lines within both panels represent the 95% CI. Solid squares in the left panel indicate that outcomes can be considered equivalent (within the 10% equivalence zone). Solid squares in the right panel indicate that the lower bound of the 95% CI of the ICC was >0.75. Hollow squares indicate that outcomes cannot be considered equivalent (within the 10% equivalence zone, left panel) or the lower bound of the 95% CI of the ICC was <0.75 (right panel).

accelerometer brands (Figure 1(a-c)), average acceleration and the IG were equivalent between wrists, with outcomes higher from the dominant wrist compared to the non-dominant wrist (ratio non-dominant/dominant <1.0). Similar findings were generally evident for time spent below 50 m*g* (except for Axivity) and above 50 m*g*, but there was a lack of equivalence evident across all three accelerometer brands for time spent in acceleration values ≥100 m*g*. Agreement (from the ICC's) between wrists for average acceleration was good to excellent and was moderate to good for the IG across all accelerometer brands. In the main, agreement in outcomes based on the time spent in acceleration thresholds were good to excellent for the GENEActiv and ActiGraph (Figure 1(a,c)) and moderate to excellent for the Axivity (Figure 1(b)).

Within the non-dominant wrist between accelerometer brands

Figure 2 provides the results of the following comparisons: a) GENEActiv/Axivity, b) GENEActiv/ActiGraph and c) Axivity/ActiGraph with additional analysis covering mean bias and

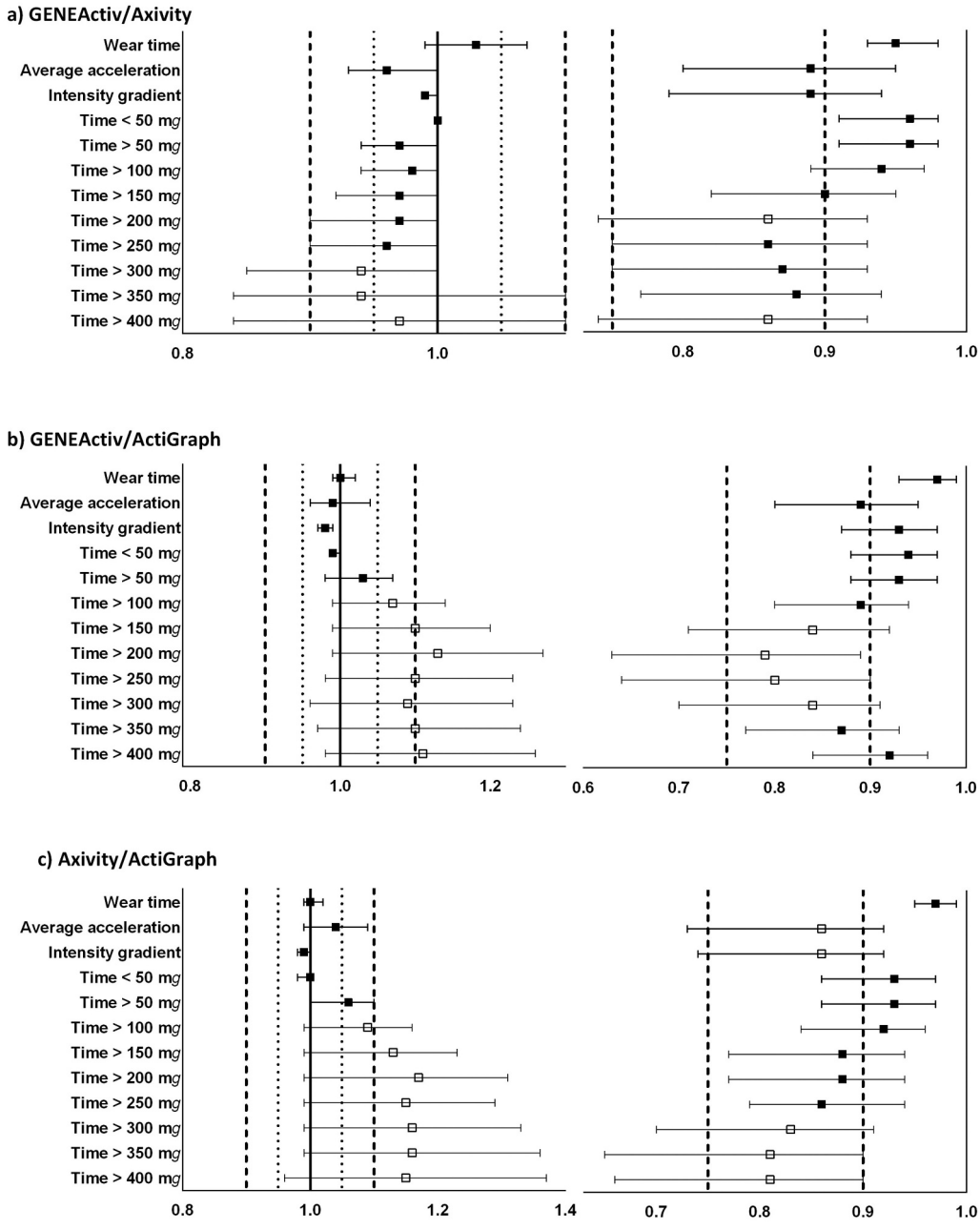

Figure 2. Equivalence and ICC between pairs of monitors on the non-dominant wrist and between accelerometer brands: (a) GENEActiv/Axivity, (b) GENEActiv/ActiGraph and (c) Axivity/ActiGraph. Equivalence analysis is presented in the left panel with dashed lines representing the 10% equivalence zone with a 5% equivalence zone indicated by the dotted lines. The ICC analysis is provided in the right panel with dashed lines representing ICCs >0.75 and >0.9. The horizontal lines within both panels represent the 95% CI. Solid squares in the left panel indicate that outcomes can be considered equivalent (within the 10% equivalence zone). Solid squares in the right panel indicate that the lower bound of the 95% CI of the ICC was >0.75. Hollow squares indicate that outcomes cannot be considered equivalent (within the 10% equivalence zone, left panel) or the lower bound of the 95% CI of the ICC was <0.75 (right panel).

LoA provided within Table S2. Across all three pairings (Figure 2 (a-c)), average acceleration, the IG and time spent below and above 50 m*g* were considered equivalent. Time spent in acceleration thresholds <250 m*g* were also found to be equivalent between GENEActiv/Axivity, but there was a lack of equivalence evident for time spent in acceleration thresholds >100 m*g* and above for the other pairings (Figure 2(b,c)). Outcomes recorded from the ActiGraph tended to be lower than GENEActiv and Axivity (ratio GENEActiv/ActiGraph >1.0, Figure 2(b), and Axivity/ActiGraph > 1.0, Figure 2(c), respectively), whereas outcomes tended to be higher from the Axivity compared to the GENEActiv (Figure 2(a)). Agreement (from the ICCs) between pairings for average acceleration and the IG was good to excellent between the GENEActiv and Axivity as well as the GENEActiv and ActiGraph (Figure 2(a,b)) but moderate to excellent between the Axivity and ActiGraph (Figure 2(c)). Finally, there tended to be a stronger agreement in time spent in acceleration thresholds between the GENEActiv and Axivity devices (Figure 2(a)) compared to the other two pairings (Figure 2(b,c)).

Within the dominant wrist between accelerometer brands

Figure 3 provides the results of the following comparisons: a) GENEActiv/Axivity, b) GENEActiv/ActiGraph and c) Axivity/ActiGraph with additional analysis covering mean bias and LoA provided within Table S3. Across all three pairings (Figure 3(a-c)), the IG and average acceleration could be considered equivalent apart from average acceleration values between the Axivity and ActiGraph (Figure 3(c)). The Axivity tended to record higher values of acceleration compared to the GENEActiv, resulting in a lack of equivalence in the time spent in acceleration values >100 m*g* and above (Figure 3(a)). The GENEActiv tended to record higher magnitudes of acceleration at the dominant wrist compared to ActiGraph, resulting in a lack of equivalence of outcomes for time spent in acceleration thresholds >200 m*g* and above (Figure 3(b)), whereas the Axivity recorded greater acceleration values compared to the ActiGraph, resulting in a lack of equivalence of outcomes for time spent in acceleration thresholds >50 m*g* and above (Figure 3(c)). Agreement (from the ICCs) between all three pairings for the IG were good to excellent (Figure 3(a-c)) but moderate to excellent for average acceleration between the GENEActiv and Axivity and the Axivity and ActiGraph pairings

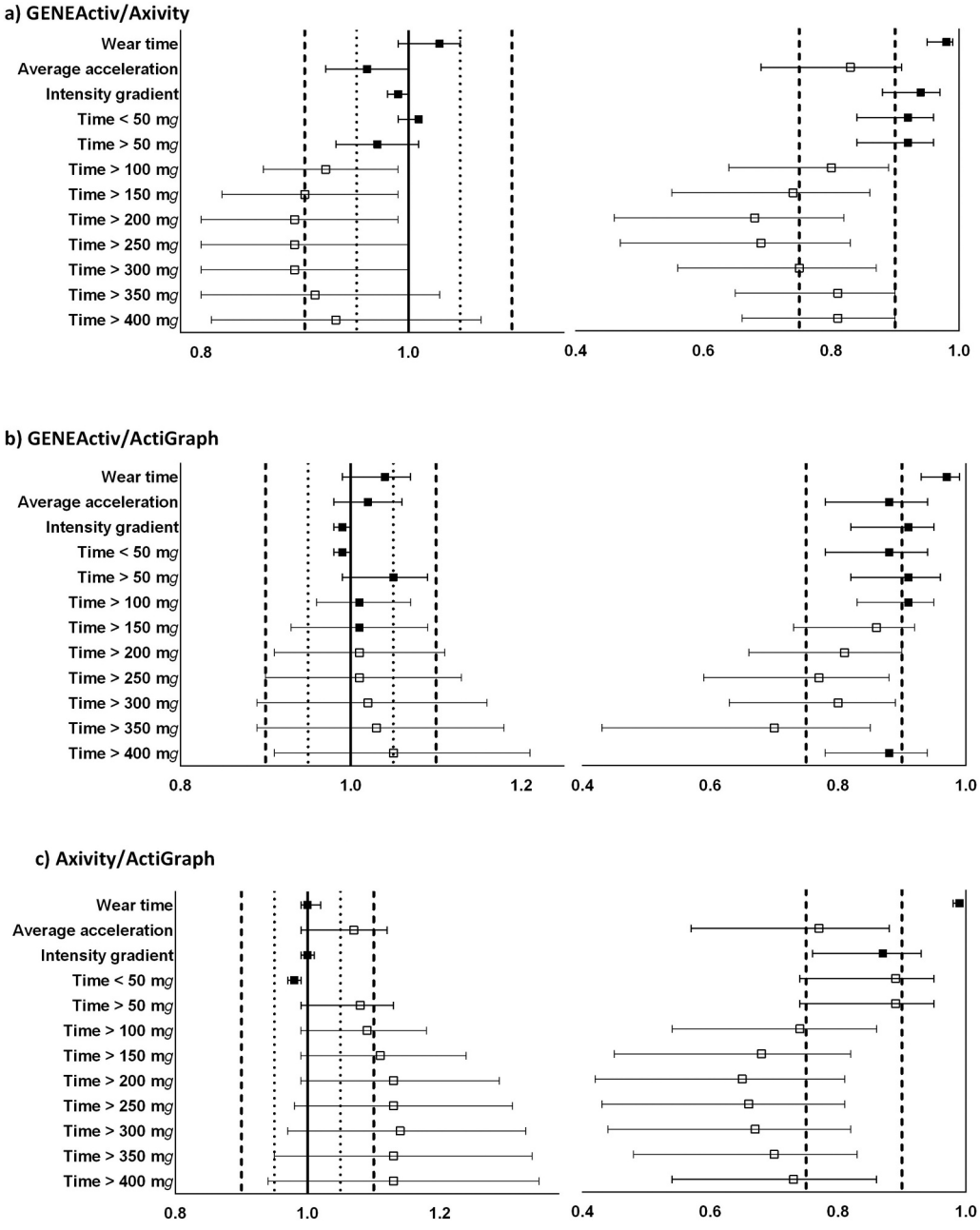

Figure 3. Equivalence and ICC between pairs of monitors on the dominant wrist and between accelerometer brands: (a) GENEActiv/Axivity, (b) GENEActiv/ActiGraph and (c) Axivity/ActiGraph. Equivalence analysis is presented in the left panel with dashed lines representing the 10% equivalence zone with a 5% equivalence zone indicated by the dotted lines. The ICC analysis is provided in the right panel with dashed lines representing ICCs >0.75 and >0.9. The horizontal lines within both panels represent the 95% CI. Solid squares in the left panel indicate that outcomes can be considered equivalent (within the 10% equivalence zone). Solid squares in the right panel indicate that the lower bound of the 95% CI of the ICC was >0.75. Hollow squares indicate that outcomes cannot be considered equivalent (within the 10% equivalence zone, left panel) or the lower bound of the 95% CI of the ICC was <0.75 (right panel).

(Figure 3(a,c)) and good to excellent between the GENEActiv and ActiGraph (Figure 3(b)). Agreement tended to be good to excellent across all three pairings for time spent below 50 m*g* and above 50 m*g* but tended to be moderate to good for time spent in acceleration thresholds >100 m*g* and above across several pairings (Figure 3(a-c)).

Between wrists and between accelerometer brands

Figure 4(a-f) displays the findings of the comparisons from two accelerometer brands worn on different wrists with additional analysis covering mean bias and LoA provided within Table S4. In general, accelerometers worn on the non-dominant wrist tended to record lower values of acceleration (apart from the Axivity(ND)/ActiGraph(D) pairing, Figure 4(d)), which resulted in lack of equivalence of outcomes for time spent in acceleration thresholds of >150 m*g* and above for all pairings (Figure 4(a-f)). Across all pairings, the IG, time spent below 50 m*g* and average acceleration outcomes (apart from the GENEActiv (ND)/Axivity(D) pairing, Figure 4(a)) could be considered equivalent. Agreement (from the ICCs) was in general excellent for wear time, moderate to good for average acceleration and the IG and good to excellent for time spent above and below 50 m*g*. Thereafter, the strength of agreement for time spent in acceleration values greater than 100 m*g* and above tended to vary across pairings and acceleration values.

Discussion

The aim of this study was to compare acceleration-based activity outcomes from three research-grade accelerometers worn concurrently on the dominant and non-dominant wrist during free living. Findings revealed that the IG was equivalent between any combination of device and wrist location when applying a 10% equivalence zone. Equivalent outcomes were evident for average acceleration between devices worn on the non-dominant and dominant wrist, apart from Axivity and ActiGraph devices worn on the dominant wrist. Equivalent outcomes for average acceleration were largely evident between most combinations of accelerometer devices between wrists, apart from the GENEActiv when worn on the non-dominant wrist and Axivity devices when worn on the dominant wrist.

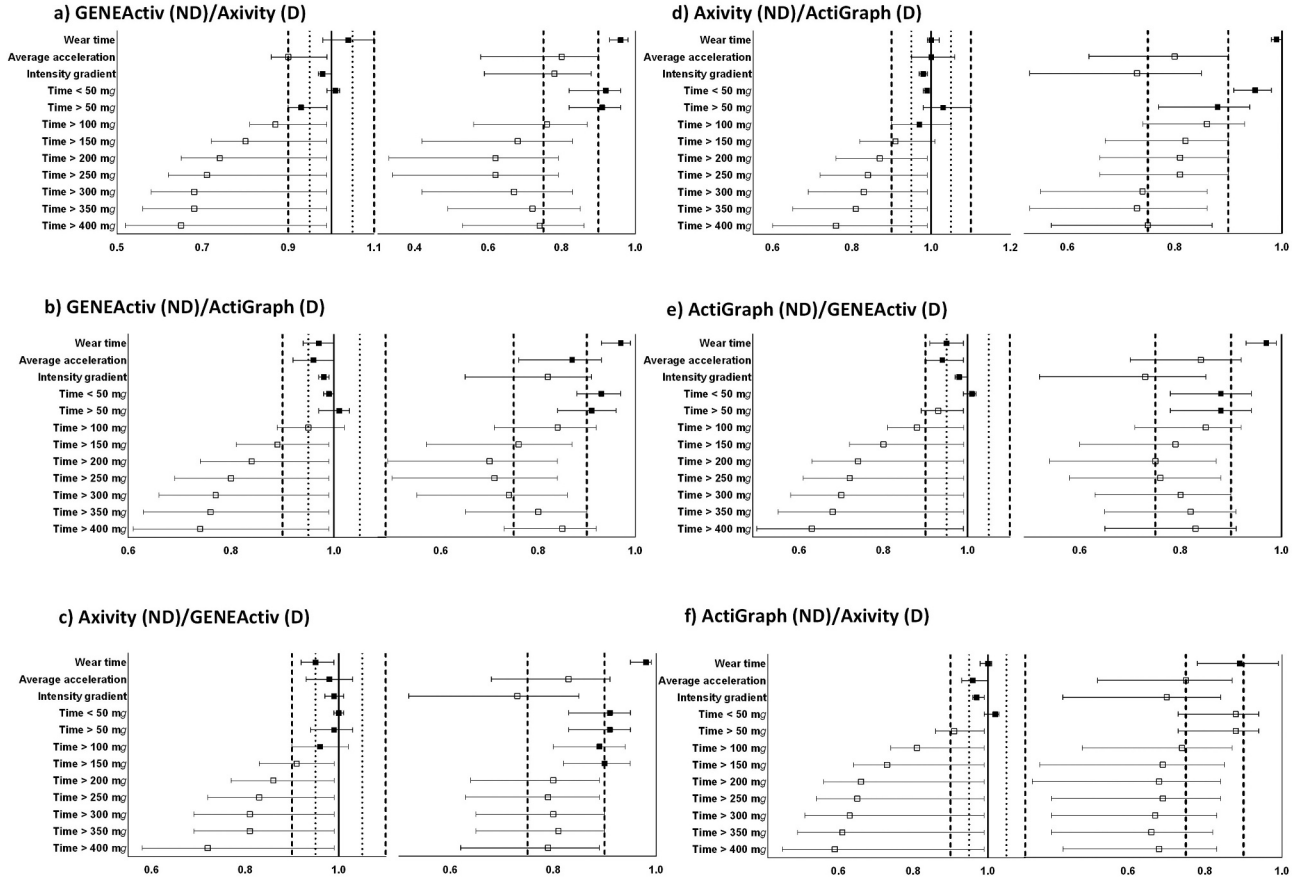

Figure 4. Equivalence and ICC between pairs of monitors between wrist and between accelerometer brands: (a) GENEActiv non-dominant (ND)/Axivity dominant (D), (b) GENEActiv (ND)/ActiGraph (D), (c) Axivity (ND)/GENEActiv (D), (d) Axivity (ND)/ActiGraph (D), (e) ActiGraph (ND)/GENEActiv (D) and (f) ActiGraph (ND)/Axivity (D). Equivalence analysis is presented in the left panel with dashed lines representing the 10% equivalence zone with a 5% equivalence zone indicated by the dotted lines. The ICC analysis is provided in the right panel with dashed lines representing ICCs >0.75 and >0.9. The horizontal lines within both panels represent the 95% CI. Solid squares in the left panel indicate that outcomes can be considered equivalent (within the 10% equivalence zone). Solid squares in the right panel indicate that the lower bound of the 95% CI of the ICC was >0.75. Hollow squares indicate that outcomes cannot be considered equivalent (within the 10% equivalence zone, left panel) or the lower bound of the 95% CI of the ICC was <0.75 (right panel).

Finally, most combinations of accelerometer devices and wrist location for time spent in accelerations ≥100 m*g* lacked equivalence and demonstrated poor agreement.

Previous findings from free-living adults found that average acceleration and the IG were equivalent between GENEActiv, Axivity and the ActiGraph GT9X devices when worn concurrently on the non-dominant wrist but average acceleration was approximately 10% lower for the ActiGraph compared to the GENEActiv and Axivity when worn on the dominant wrist (Rowlands, Plekhanova et al., 2019). We demonstrated identical findings from the non-dominant wrist in respect to average acceleration and the IG between pairings (Figure 2(a-c)). Findings from the dominant wrist were similar to the those non-dominant wrist in respect to average acceleration and the IG between pairings (Figure 3(a,b)), with a lack of equivalence in average acceleration only evident between the Axivity and ActiGraph pairing (Figure 3(c)). The Axivity recorded higher acceleration values than the ActiGraph (ratios Axivity/ActiGraph >1.0, Figures 2(c) and 3(c)) by ~1.1 m*g* (4%) from the non-dominant wrist and ~2.2 m*g* (7.4%) from the dominant wrist (Table 1), with the larger difference likely contributing to the lack of equivalence observed. Whilst these differences may seem minimal, Rowlands et al. (2021) recently proposed a minimum clinically important difference (MCID) in daily average acceleration measured at the wrist for health benefits in inactive people of approx. 0.8 to 1.0 m*g* (A Rowlands et al., 2021). While the authors contend that this estimate is preliminary, caution is advised if researchers intend to pool/compare average acceleration data captured from studies deploying the ActiGraph on the non-dominant wrist and the Axivity on the dominant wrist given the differences observed in average acceleration reported in this study between these devices.

It is known that accelerations captured from ActiGraph devices during moderate-vigorous intensity activities are filtered and rectified through a low-pass frequency filter, which may result in accelerations being filtered out (Fridolfsson et al., 2019; Rowlands, Plekhanova et al., 2019). Lower average acceleration values from the ActiGraph GT9X compared to the Axivity were also found from adults after being observed in a stimulated living space for 2 h while wearing GENEActiv, ActiGraph and Axivity devices on their non-dominant wrist (Rowlands et al., 2017). Here the authors reported a low mean bias of 3.6 m*g* from Bland-Altman plots (Axivity/ActiGraph pairing), which was greater than the mean bias reported in the present study, regardless of the device pairing within either wrist (range: 0.05 m*g* to 2.22 m*g*, Tables S1-S3). The mean bias reported in the present study is also comparable to recent findings from adults that wore the GENEActiv, Axivity and ActiGraph GT9X concurrently on both wrists during free living, regardless of the device pairing within either wrist (range: 0.50 m*g* to 4.58 m*g*; Rowlands, Plekhanova et al., 2019).

Reporting estimates of habitual PA is common from wrist-worn accelerometers with authors often using a threshold of 50 m*g* to report inactive time and a threshold of 100 m*g* for MVPA as proposed by Hildebrand et al. (Buchan et al., 2020, 2018; Hildebrand et al., 2017, 2014; Rowlands, Plekhanova et al., 2019). The lack of equivalence and poor agreement for time spent across acceleration ranges ≥100 m*g* reported here are consistent with the findings of others (Rowlands et al., 2017; Rowlands, Plekhanova et al., 2019; Rowlands et al., 2016). Higher acceleration values recorded from the dominant wrist in respect to the non-dominant wrist (Figure 1(a-c)) and from the GENEActiv and Axivity when paired with the ActiGraph, regardless of wrist, is likely to contribute to the lack of equivalence demonstrated between pairings (Figures 2(b,c) and 3(b, c)). The high variability and limited time spent in higher acceleration ranges will also have contributed to the lack of equivalence, as would the equivalence zone applied to these outcomes.

As highlighted recently, the use of a 10% equivalence zone may not be appropriate for all outcomes (Edwardson et al., 2021). The use of a 10% equivalence zone can be lax when values are very high (i.e., time spent <50 m*g*) and across a narrow range of values (i.e. the IG) but strict when values are smaller and highly variable (i.e., time spent in acceleration thresholds greater than 100 m*g*; Rowlands, Plekhanova et al., 2019; Table 1). When you consider the lack of equivalence evident for time spent in accelerations ≥100 m*g*, caution is advised in the interpretation of these findings since the equivalence zone covers a range of values that were a small fraction of the standard deviation (SD). Moreover, the 10% equivalence zone for the IG encompassed values that differed by an SD, which limits the physiological and clinical significance of the IG findings. Conversely, the 10% The use of equivalence zones based on the SD of the measure is encouraged since it considers the variability of the movement behaviour (O'Brien, 2021). Applying an equivalence zone of 0.3 SD for the IG revealed equivalent outcomes between brands when worn on the same wrist, for the GENEActiv between wrists, and between the Axivity worn on the non-dominant wrist and the GENEActiv worn on the dominant wrist (Figure 5(a)). Moreover, outcomes were largely equivalent when applying an equivalence zone that encompassed values ranging from 0.3–0.35 SDs, which strengthens the significance of these IG findings.

Despite using a different pairing configuration, Rowlands et al. (2019) also found a lack of equivalence for average acceleration between the Axivity device worn on the dominant wrist and the GENEActiv worn on the non-dominant wrist. Like this study, the Axivity device was placed on top of another accelerometer, which may have influenced the average acceleration outcome. The dominant arm tends to be used more often than the non-dominant arm for normal lifestyle activities, which our data seem to support (Table 1). Due to the pairing configuration, it is plausible that the GENEActiv device could have been restricted somewhat and unable to capture accelerations associated with certain activities unlike the Axivity device. While this is speculative, future work should consider randomizing the configuration of device pairings alongside device placement to confirm these findings.

As our findings suggest, small differences in acceleration values can adversely affect the equivalence and agreement findings when evaluating threshold-based outcomes such as time spent in acceleration thresholds greater than 100 m*g* (i.e. adult MVPA). The limitations of cut points to estimate time spent in activity intensities are well established since cut points are population and protocol specific, which can hinder the comparison of findings between studies (Migueles, Cadenas-Sanchez et al., 2019). The use of cut points can also lead to

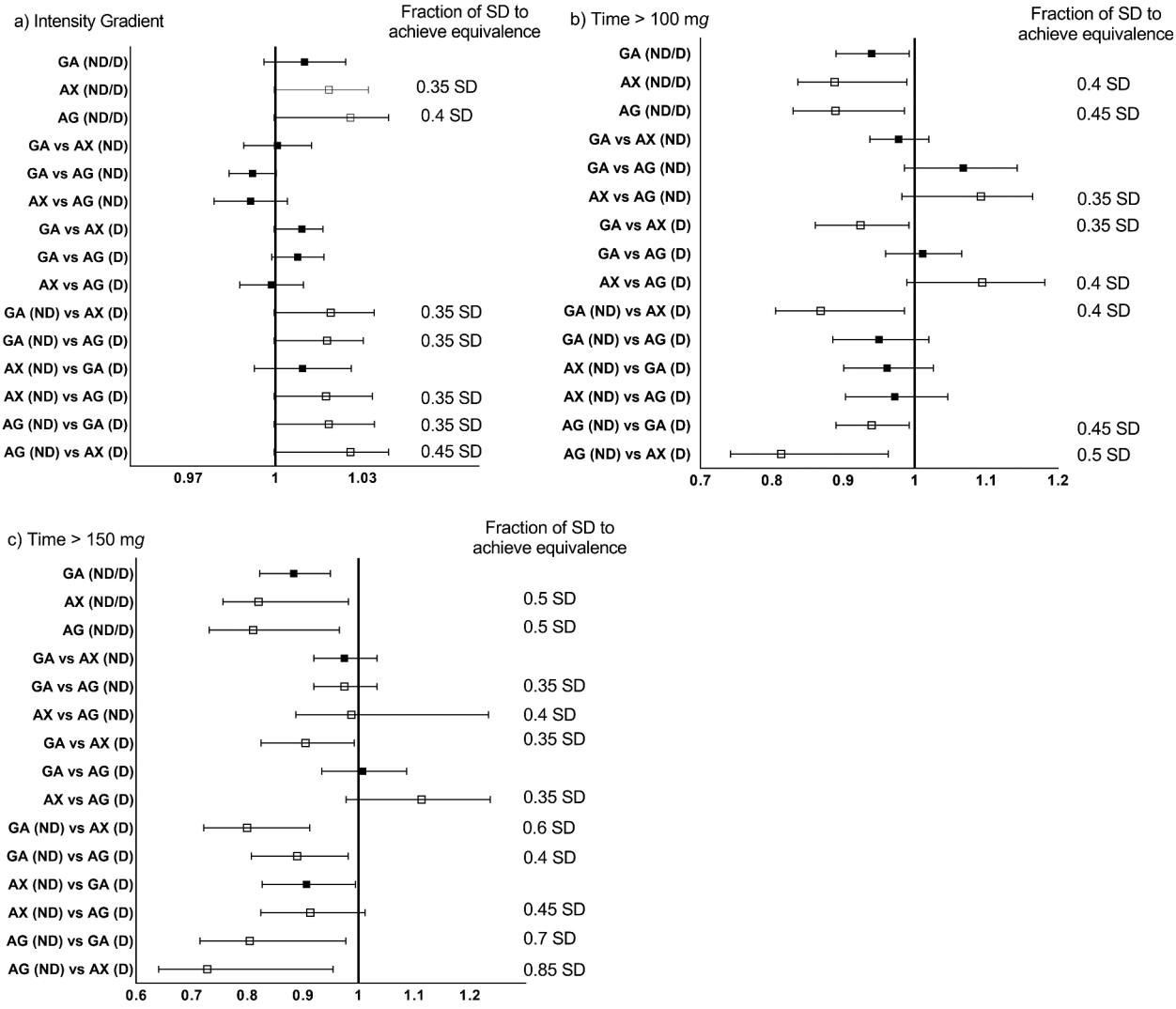

Figure 5. Equivalence between pairs of monitors between wrist and between accelerometer brands for (a) intensity gradient, (b) time > 100 m*g* and (c) time > 150 m*g*. An equivalence zone of 0.3 SD was applied with solid squares indicating that outcomes could be considered equivalent. Hollow squares indicate that outcomes could not be considered equivalent when applying an equivalence zone of 0.3 SD. The required equivalence zone to achieve equivalence was subsequently investigated and is provided where appropriate. GA: GENEActiv; AXE: Axivity; AG: ActiGraph; ND: non-dominant; D: dominant.

erroneous activity estimates between participants where despite demonstrating similar levels of activity, one's activity may fall just below an intensity cut point, resulting in a recording of zero minutes. An alternative approach that overcomes the limitations of using cut points is reporting the minimum acceleration value achieved across a period of time (Rowlands, Dawkins et al., 2019).

Using this approach, we recently reported equivalence in the minimal acceleration value (MX_{ACC} with X indicating the duration) for participants most active 2, 30 and 60 minutes from ActiGraph devices worn concurrently on the dominant and non-dominant wrist during free living (Buchan et al., 2020). Findings revealed equivalent outcomes for the $M2_{ACC}$, $M30_{ACC}$, $M60_{ACC}$ across wrist location when using a 10% equivalence zone that encompassed values ranging from 0.3 to 0.45 SDs. This in contrast to the findings of this study where there was a lack of equivalence in outcomes for time spent in acceleration values >100 m*g* and which cover the acceleration ranges of the MX metrics. When applying a similar equivalence zone of 0.3 SD as done for the IG (Figure 5(a)) for time spent above 100 m*g* (i.e, adult MVPA; Figure 5(b)) and above 150 m*g* (which covers the $M30_{ACC}$ range reported elsewhere (Buchan et al., 2020); Figure 5(c)), equivalence was largely achieved when applying an equivalence zone that encompassed values ranging from 0.3 to 0.45 SDs. These findings demonstrate the importance of applying appropriate equivalence zones and the effect these can have upon equivalence findings.

Findings from the ICCs were mixed depending on the pairing. The strength of agreement tended to be poorer when comparing outcomes between both wrists and brands (Figure 4(a-f)) and is in contrast with the agreement between accelerometer brands placed on either the non-dominant and dominant wrist (Figures 2(a-c) and 3(a-c)). Of note is the strength of agreement and equivalence in outcomes from the GENEActiv and Axivity pairing on the non-dominant wrist (Figure 2(a)), which may be a consequence of how the monitors were worn; the Axivity was taped on top of the GENEActiv. Nonetheless, the extent of

agreement between outcomes was not evident from the GENEActiv and Axivity pairing on the dominant wrist (Figure 3(a)).

Comparing findings between others is challenging. To the best of our knowledge, only one study has examined the equivalence of outcomes from ActiGraph (GT9X), Axivity and GENEActive devices when worn concurrently on both wrists (Rowlands, Plekhanova et al., 2019). Notwithstanding the different samples used between studies and differences in accelerometer devices, how the devices were worn can also hinder comparisons. For instance, average acceleration values provided from the wGT3X-BT, used in this study, and the GT9X, used by Rowlands et al. (Rowlands, Plekhanova et al., 2019), may not provide equivalent outcomes when worn concurrently on both the dominant and non-dominant wrists during free living (Clevenger et al., 2020). Nonetheless, the authors do warrant caution when interpreting these findings due to a small sample size and its impact upon power and accuracy of findings.

Further, a necessary requirement when comparing accelerometers brands worn concurrently is the need for devices to be worn proximal from the manufacturer's recommended location. In the study by Rowlands et al. (2019), the Axivity was taped on top of the ActiGraph GT9X monitor and placed alongside the GENEActiv to reduce the number of wrist straps worn by participants for comfort reasons. In this study, the Axivity was taped on top of the GENEActiv with the ActiGraph worn alongside this combination for the same reasons. Despite this, some participants did experience a degree of discomfort, which may have hindered their range of movement from the wrists. Future work may wish to assess the extent of discomfort experienced by participants when alternative device pairing combinations are used and explore whether this impacts study findings. In addition, further work assessing the cross-generational comparability of accelerometer devices is advised and should be considered when interpreting similar study findings or pooling/comparing outcomes from other accelerometer devices.

A strength of this study is the comparison of activity outcomes from three widely used research-grade accelerometers worn concurrently on both wrists. As is the matching of wear time between devices and the compliance demonstrated by participants. Moreover, the order of device placements on each wrist was randomized, which helped minimize the effects of device placement upon study findings. Since only healthy adult participants were recruited to the study, our findings may not be applicable for older adults or those with variable gait characteristics. Finally, the unique combination of monitors taped together on each wrist could be considered a limitation as it affects how the monitors were worn.

In summary, our findings suggest that average acceleration was equivalent between most combinations of accelerometer device and wrist locations apart from Axivity and ActiGraph when worn on the dominant wrist and between GENEActiv worn on the non-dominant wrist and Axivity worn on the dominant wrist. When applying a 10% equivalence zone, the IG was equivalent between any combination of accelerometer device and wrist location. Nonetheless, the use of a 10% equivalence zone limits the physiological and clinical significance of the IG findings. When applying an equivalence zone that reflects the narrower variability of the IG, equivalent outcomes were evident between brands when worn on the same wrist, for the GENEActiv between wrists and only between the Axivity worn on the non-dominant wrist and GENEActiv worn on the dominant wrist. Moreover, outcomes were largely equivalent when applying an equivalence zone that encompassed values ranging from 0.3 to 0.35 SDs. When applying a more appropriate equivalence zone to time spent in accelerations ≥100 m*g* and ≥150 m*g*, equivalence was largely achieved when applying an equivalence zone that encompassed values ranging from 0.3 to 0.45 SDs. These are important findings since device placement is not consistent in large cohort studies such as NHANES, which deployed the ActiGraph on the non-dominant wrist, and the UK Biobank, which deployed the Axivity on the dominant wrist. Nonetheless, further work that applies an equivalence zone that considers the variability of the outcome measure(s) from wrist-worn accelerometer devices is needed to confirm these findings. Finally, further work may also wish to consider randomizing device pairings between participants to minimize its potential effect upon study findings.

Acknowledgments

I would like to thank Euan Dalgleish for his assistance with the data collection and the participants for agreeing to participate in this study.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The author(s) reported there is no funding associated with the work featured in this article.

ORCID

Duncan S. Buchan http://orcid.org/0000-0002-4320-4615

References

Actigraph. (2013). *GT3X+ and wGT3X+ device manual [Internet]*. Retrieved February 26, 2020, from https://s3.amazonaws.com/actigraphcorp.com/wp-content/uploads/2018/02/22094126/GT3X-wGT3X-device-manual-110315.pdf

ActiGraph. (2019). *ActiGraph. User guide: ActiGraph GT9X link + ActiLife*. [Internet]. Retrieved February 26, 2020, from https://s3.amazonaws.com/actigraphcorp.com/wp-content/uploads/2019/02/04090343/ActiGraph_Link_UserGuide_02012019_Revision3.pdf

Bland, J. M., & Altman, D. G. (1986, February). 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., Boddy, L. M., & McLellan, G. (2020). Comparison of free-living and laboratory activity outcomes from actigraph accelerometers worn on the dominant and non-dominant wrists: Measurement in physical education and exercise science: Vol 24, No 4. *Measurement in Physical Education and Exercise Science*, *24*(4), 247–257. https://doi.org/10.1080/1091367X.2020.1801441

Buchan, D. S., McSeveney, F., & McLellan, G. (2018, July). A comparison of physical activity from Actigraph GT3X+ accelerometers worn on the dominant and non-dominant wrist. *Clinical Physiology and Functional Imaging*, *39*(1), 51–56. https://doi.org/10.1111/cpf.12538

Clevenger, K. A., Pfeiffer, K. A., & Montoye, A. H. K. (2020, December). Cross-generational comparability of hip- and wrist-worn ActiGraph GT3X+, wGT3X-BT, and GT9X accelerometers during free-living in adults. *Journal of Sports Sciences*, *38*(24), 2794–2802. https://doi.org/10.1080/02640414.2020.1801320

Dixon, P. M., Saint-Maurice, P. F., Kim, Y., Hibbing, P., Bai, Y., & Welk, G. J. (2018, April). A primer on the use of equivalence testing for evaluating measurement agreement. *Medicine & Science in Sports & Exercise*, *50*(4), 837–845. https://doi.org/10.1249/MSS.0000000000001481

Doherty, A., Jackson, D., Hammerla, N., Plötz, T., Olivier, P., Granat, M. H., White, T., van Hees, V. T., Trenell, M. I., Owen, C. G., Preece, S. J., Gillions, R., Sheard, S., Peakman, T., Brage, S., & Wareham, N. J. (2017, February). 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., Maylor, B. D., Dawkins, N. P., Plekhanova, T., & Rowlands, A. V. (2021, June 20). Comparability of postural and physical activity metrics from different accelerometer brands worn on the thigh: Data harmonization possibilities. *Measurement in Physical Education and Exercise Science*, 1–12. https://doi.org/10.1080/1091367X.2021.1944154

Fridolfsson, J., Börjesson, M., Buck, C., Ekblom, Ö., Ekblom-Bak, E., Hunsberger, M., Ler, L., & Arvidsson, D. (2019, January). Effects of frequency filtering on intensity and noise in accelerometer-based physical activity measurements. *Sensors*, *19*(9), 2186. https://doi.org/10.3390/s19092186

Hildebrand, M., Hansen, B. H., van Hees, V. T., & Van, Ekelund, U. (2017). Evaluation of raw acceleration sedentary thresholds in children and adults. *Scandinavian Journal of Medicine & Science in Sports*, *27*(12), 1814–1823. https://doi.org/10.1111/sms.12795

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

Koo, T. K., & Li, M. Y. (2016, June). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Migueles, J. H., Cadenas-Sanchez, C., Tudor-Locke, C., Löf, M., Esteban-Cornejo, I., Molina-Garcia, P., Mora-Gonzalez, J., Rodriguez-Ayllon, M., Garcia-Marmol, E., Ekelund, U., & Ortega, F. B. (2019). Comparability of published cut-points for the assessment of physical activity: Implications for data harmonization. *Scandinavian Journal of Medicine & Science in Sports*, *29*(4), 566–574. https://doi.org/10.1111/sms.13356

Migueles, J. H., Rowlands, A. V., Huber, F., Sabia, S., & van Hees, V. T. (2019, September). GGIR: A research community–driven open source R package for generating physical activity and sleep outcomes from multi-day raw accelerometer data. *Journal for the Measurement of Physical Behaviour*, *2*(3), 188–196. https://doi.org/10.1123/jmpb.2018-0063

Montoye, A. H. K., 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 & Science in Sports & Exercise*, *50*(5), 1103–1112. https://doi.org/10.1249/MSS.0000000000001534

NHANES. (2021). *NHANES - National Health and Nutrition Examination Survey: Physical Activity Monitor (PAM) procedures manual* [Internet]. Retrieved February 23, 2021, from https://wwwn.cdc.gov/nchs/data/nhanes/2011-2012/manuals/Physical_Activity_Monitor_Manual.pdf

O'Brien, M. W. (2021). Implications and Recommendations for Equivalence Testing in Measures of Movement Behaviors: A Scoping Review. Journal for the Measurement of Physical Behaviour, 4(4), 353–362. https://doi.org/10.1123/jmpb.2021-0021

Ricardo, L. I. C., Wendt, A., Galliano, L. M., de Muller, W. A., Cruz, G. I. N., Wehrmeister, F., Brage, S., Ekelund, U., & Crochemore M. Silva, I. (2020, January). Number of days required to estimate physical activity constructs objectively measured in different age groups: Findings from three Brazilian (Pelotas) population-based birth cohorts. *PLOS ONE*, *15*(1), e0216017. https://doi.org/10.1371/journal.pone.0216017

Rowlands, A. V., Dawkins, N. P., Maylor, B., Edwardson, C. L., Fairclough, S. J., Davies, M. J., Harrington, D. M., Khunti, K., & Yates, T. (2019, December). Enhancing the value of accelerometer-assessed physical activity: Meaningful visual comparisons of data-driven translational accelerometer metrics. *Sports Medicine - Open*, *5*(1), 47. https://doi.org/10.1186/s40798-019-0225-9

Rowlands, A. V., Edwardson, C. L., Davies, M. J., Khunti, K., Harrington, D. M., & Yates, T. (2018, June). Beyond cut points: Accelerometer metrics that capture the physical activity profile. *Medicine & Science in Sports & Exercise*, *50*(6), 1323–1332. https://doi.org/10.1249/MSS.0000000000001561

Rowlands, A. V., Mirkes, E., Yates, T., Clemes, S., Davies, M., Khunti, K., et al. (2017, October 1). Accelerometer-assessed physical activity in epidemiology: Are monitors equivalent? *Medicine and Science in Sports and Exercise*, *50*(2), 257-265.

Rowlands, A. V., Plekhanova, T., Yates, T., Mirkes, E. M., Davies, M., Khunti, K., & Edwardson, C. L. (2019, September). Providing a basis for harmonization of accelerometer-assessed physical activity outcomes across epidemiological Datasets. *Journal for the Measurement of Physical Behaviour*, *2*(3), 131–142. https://doi.org/10.1123/jmpb.2018-0073

Rowlands, A. V., Yates, T., Davies, M., Khunti, K., & Edwardson, C. L. (2016, May). Raw accelerometer data analysis with GGIR R-package: Does accelerometer brand matter? *Medicine & Science in Sports & Exercise*, *48*(10), 1935–1941. https://doi.org/10.1249/MSS.0000000000000978

Rowlands, A. V. (2018, November). Moving forward with accelerometer-assessed physical activity: Two strategies to ensure meaningful, interpretable, and comparable measures. *Pediatric Exercise Science*, *30*(4), 450–456. https://doi.org/10.1123/pes.2018-0201

Rowlands, A., Davies, M., Dempsey, P., Edwardson, C., Razieh, C., & Yates, T. (2021, July). Wrist-worn accelerometers: Recommending ~1.0 mg as the minimum clinically important difference (MCID) in daily average acceleration for inactive adults. *British Journal of Sports Medicine*, *55*(14), 814–815. https://doi.org/10.1136/bjsports-2020-102293

Troiano, R. P., McClain, J. J., Brychta, R. J., & Chen, K. Y. (2014, July). 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 Hees, V. T., Fang, Z., Langford, J., Assah, F., Mohammad, A., Da Silva, I. C. M., Trenell, M. I., White, T., Wareham, N. J., & Brage, S. (2014, October). Autocalibration 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 Leon, 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). *In: Testing statistical hypotheses of equivalence* [Internet]. Chapman & Hall/CRC (pp. xv, 284). Retrieved July 19, 2019, from https://books.google.co.uk/books?id=1nXLBQAAQBAJ&printsec=frontcover&dq=Testing+Statistical+Hypotheses+of+Equivalence.&hl=en&sa=X&ved=0ahUKEwjxyKnPw8DjAhWDSRUIHde5DOcQ6AEIMDAB#v=onepage&q=Testing%20Statistical%20Hypotheses%20of%20Equivalence.&f=false