Comparison of Free-Living and Laboratory Activity Outcomes from ActiGraph Accelerometers Worn on the Dominant and Non-Dominant wrists.

Duncan S. Buchan[1]*, Lynne M. Boddy[2] and Gillian McLellan[1].

* Corresponding author

[1]University of the West of Scotland, School of Health and Life Sciences, Institute of Clinical Exercise and Health Science, UWS Lanarkshire Campus, Stephenson Place, Hamilton International Technology Park, Blantyre, G72 0LH, Scotland, UK.

[2]The Physical Activity Exchange, Research Institute for Sport and Exercise Sciences, Liverpool John Moores University, Liverpool, UK

Email: duncan.buchan@uws.ac.uk. Tel: +44 01698 283100 Ex. 825.

Email: gillian.mclennan@uws.ac.uk

Email: L.M.Boddy@ljmu.ac.uk

Word count: 4863

Running head: Comparing activity outcomes from wrist-worn accelerometers

Abstract

This study evaluated agreement in activity outcomes from ActiGraph accelerometers worn on both wrists in a laboratory and free-living setting. Part 1: Thirty-seven participants (25.5 ± 10.5 years) completed laboratory activities. Part 2: Thirty-nine participants (28.5 ± 9.8 years) wore accelerometers for 7 days. Outcomes included average acceleration and the intensity gradient (IG). Part 1: Average acceleration was equivalent at the group level between devices across all activity intensities. Wide limits of agreement ranging from 20.6% lower to 34.9% higher for the dominant wrist across all activities were observed. Part 2: The IG was equivalent between wrist locations, but average acceleration was approximately 8.5% higher when measured from the dominant wrist. Adjusting average acceleration values by -8.5% from the dominant wrist resulted in average acceleration falling within a strict 5% equivalence zone. Reducing average acceleration values from the dominant wrist by 8.5% results in equivalent outcomes between wrists during free-living.

Keywords: GGIR, agreement, equivalence, intensity gradient, physical activity, average acceleration.

Introduction:

Accurate measurement of physical activity (PA) is crucial for exploring relationships with numerous health indicators and for assessing the effectiveness of public health strategies to increase activity levels. Accelerometers are commonly used to provide an objective and more accurate estimate of activity levels in comparison to self-report methods (Troiano, McClain, Brychta, & Chen, 2014). There are several research grade accelerometers available (Axivity, GENEActiv) with the ActiGraph (Pensacola, FL, USA) devices being the most commonly used when measuring activity levels in adults (Wijndaele et al., 2015). The GT3X model was deployed on the non-dominant wrist in the 2011–2012 and 2013–2014 cycles of the NHANES with data collected from approximately 9000-10,000 participants (NHANES, 2018). Both the NHANES protocol and ActiGraph suggest that the accelerometer is worn on the individual's non-dominant wrist which is based on historical precedent for sleep research, and early wrist accelerometer calibration studies that did not appear to favour either wrist location for PA monitoring (Esliger et al., 2011; Phillips, Parfitt, & Rowlands, 2013; Zhang, Rowlands, Murray, & Hurst, 2012). Nonetheless, there is a lack of consensus regarding which wrist an accelerometer should be worn, since the NHANES used the non-dominant wrist whereas other large population surveys such as the UK BioBank deployed the Axivity accelerometer on the dominant wrist (Doherty et al., 2017).

An important technological advancement in accelerometers has seen raw acceleration signal data become available on various devices including the ActiGraph, Axivity and the GENEActiv. Processing raw data can be undertaken using open-source resources such as the GGIR package in R [http:/cran.r-prouect.org] affording greater transparency and consistency of methodologies, which may facilitate data harmonisation between studies deploying different brands of accelerometer. Before data sets from studies using these raw acceleration

devices are pooled, there is a need to demonstrate whether dominant and non-dominant wrist outputs are equivalent. Moreover, establishing how equivalent outcomes are between wrist locations could facilitate the use of either wrist site in future studies and enhance potential autonomy for research subjects. It is currently unclear whether ActiGraph devices when worn concurrently on the dominant and non-dominant wrist provide equivalent activity outcomes. One study recently evaluated the counts per minute (cpm) outcomes from ActiGraph devices worn simultaneously on the dominant and non-dominant wrist during 1 day (24h) of free-living and reported no significant differences in PA estimates over several axis (Dieu et al., 2017). Nonetheless, the findings are limited to overall activity measures derived from proprietary counts which limits comparability with studies using different devices.

Recently, Buchan et al., examined the comparability of data outcomes from ActiGraph accelerometers worn on both wrist locations with data processed using GGIR (Buchan, McSeveney, & McLellan, 2018). Findings revealed a high agreement between average acceleration, time spent in moderate-to-vigorous PA (MVPA), wear time and the distribution of time across acceleration levels from the two wrist locations. Yet, the use of an 8 h wear-time inclusion criteria across 1 day may not have been representative of habitual activity and limits the generalizability of their findings. In a free-living study of adults, Rowlands et al. examined whether several activity outcomes were equivalent between three research-grade accelerometer brands (Axivity, GENEActiv and ActiGraph) when worn concurrently on both wrists (Rowlands, Plekhanova, et al., 2019). Findings from the ActiGraph GT9X devices revealed that several activity outcomes could be considered equivalent whilst demonstrating excellent reliability between wrist locations.

An interesting finding reported by Rowlands et al., was that the average accelerations values measured from the dominant wrist were approximately 10% higher than values from the non-dominant wrists for both the Axivity and GENEActive devices, but this was not evident from the ActiGraph devices (Rowlands, Plekhanova, et al., 2019). The use of the most recent version of the ActiGraph, the GT9X Link, by Rowlands et al., (2019) which has a different design to the previous GT3X model, may be one explanation for the disparity in findings between accelerometer brands. Another reason suggested by Rowlands et al., could be due to the on-board filtering of higher intensity accelerations which may have suppressed the accelerations from the dominant wrist. Further work is needed nonetheless to confirm these observations. Moreover, as the majority of adults' waking time is spent being inactive during free-living studies (Colley et al., 2011; Matthews et al., 2008), it is also important to examine the equivalency of data outcomes across a range of activity intensities since any disparities at higher intensities may be obscured in free-living settings.

In an attempt to overcome the reliance upon population and device specific activity thresholds (Troiano et al., 2014), we were particularly interested in comparing how equivalent outcomes were between wrist locations for the following activity metrics; average acceleration, intensity gradient (IG) and the average acceleration value above which an individual's most active number of minutes are obtained (i.e. MX_{ACC}) (Rowlands, 2018; Rowlands et al., 2018; Rowlands, Sherar, et al., 2019). The average acceleration reflects the volume of activity whereas the IG reflects the intensity distribution throughout the entire monitoring period to provide an activity profile. Since these metrics are not population specific, they may be suitable for pooling and comparing activity outcomes across accelerometer datasets (Rowlands et al., 2018; Rowlands, Plekhanova, et al., 2019). Thus, the aims of this study were to examine the equivalence of activity outcomes from ActiGraph

GT3X+ accelerometers when worn on the dominant and non-dominant wrist during controlled laboratory activities (part 1) and free-living (part 2).

Materials and Methods:

Part 1: Laboratory-based

A convenience sample of 37 adults (age: 25.5 ± 10.5 years, BMI: 24.9 ± 2.4 kg/m^2, 10 females) were recruited from South Lanarkshire, Scotland. All participants provided informed consent after study approval from the ethics committee of the University of the West of Scotland. Data collection took place between October 2017 and February 2018.

Procedures

Upon arrival at the laboratory, study procedures were explained to participants. Thereafter, participants height and weight 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. Each participant then wore two ActiGraph GT3X+ monitors (from herein: ActiGraph), one on each wrist, once they had identified their dominant wrist. Prior to distribution, both accelerometers were synchronised with Greenwich Mean time and initialized to capture data at 60Hz using ActiLife software V6.13.3 on the same computer. After participants had warmed up, they performed a set of 5 activities (1 light, 2 moderate and 2 vigorous) which were completed in a random order. Activities were performed on a treadmill at the following intensities:

- 5 min walk at 3 mph (Light intensity)
- 5 min walk at 3.5 mph (Moderate intensity)
- 5 min walk at 4 mph (Moderate intensity)

- 5 min running at 6 mph (Vigorous intensity)
- 5 min running at 6.5 mph (Vigorous intensity)

These activities were defined a-priori on the basis of predicted METs from the Compendium of Physical Activities (Ainsworth et al., 2011) whereby: 1.0 to ≤1.5 METs represented Sedentary behaviour, 1.5 to ≤3.0 METs for light intensity, 3.0 to ≤6.0 METs for moderate intensity and >6.0 METs for vigorous intensity. Each activity was performed for 5 min in a supervised laboratory setting with no rest period between activities. To eliminate the potential influence from participants moving from one treadmill intensity to another, data from the first and last minute of each activity were removed from the subsequent analysis.

Data reduction and processing

To calculate the average acceleration metric, the ActiGraph .gt3x files were converted to time-stamp free .csv files which were exported into R v3.5.3 (R Foundation for Statistical Computing, Vienna, Austria, https://cran.r-project.org/). These csv files were then processed using the GGIR package V1.9.1 which auto-calibrated the raw triaxial accelerometer signals and computed average acceleration expressed as Euclidean Norm Minus One (ENMO) in milli-gravitational units (m*g*) averaged over 5-s epochs, with negative values rounded up to zero and corrected for gravity (Van Hees et al., 2014). The package regenerated the time-stamps with files subsequently exported to Microsoft Excel 2016 (Microsoft, Redmond, WA) for analysis using a macro developed by the 1st author. The macro provided average acceleration, expressed in m*g* and calculated over 5-s epochs for minutes 2-4 of each activity from each device. Given the short duration of the laboratory protocol and the lack of sufficient non-movement periods available for auto-calibration, we used back-up calibration

coefficients derived from free-living data collected with the same accelerometer unit as reported elsewhere (Rowlands et al., 2017). All activity outcomes were calculated for both devices.

Part 2: Free-living

A convenience sample of 51 adults (students and non-university attending friends and family; age: 27.7 ± 9.2, 18 females) were recruited. All participants provided informed consent after study approval from the ethics committee of the University of the West of Scotland. Data collection took place between October 2018 and February 2019. After confirmation of their dominant wrist, participants were instructed to wear one ActiGraph accelerometer on each wrist for 24 h a day for 7 days. Prior to distribution, both accelerometers were synchronised with Greenwich Mean time and initialized to capture data at 90Hz. The "idle sleep mode" in ActiLife V6.13.3 was not enabled. Participants were instructed to remove the devices during water-based activities. Both accelerometers were set to commence data collection immediately after distribution.

Data reduction and processing

Upon the return of both devices, data were downloaded using ActiLife v6.13.3 (Actigraph, Pensacola, FL, USA) and saved in raw format as ActiGraph .gt3x files. To calculate the ENMO metric, the ActiGraph .gt3x files were converted to time-stamp free .csv files which were exported into R statistical software v3.5.3 (R Foundation for Statistical Computing, Vienna, Austria, https://cran.r-project.org/). These csv files were then processed using the GGIR package V1.9.1 which auto-calibrated the raw triaxial accelerometer signals

and computed the ENMO metric in m*g* averaged over 5-s epochs with negative values rounded up to zero and corrected for gravity (Van Hees et al., 2014). Files were excluded from subsequent analyses if post-calibration error was > 0.01 *g*, there were less than 4 days, which included 1 weekend day, of valid wear (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. 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 outcome variables were calculated based on the entire 24 h cycle.

Thereafter, the following metrics were generated in GGIR. Activity was expressed as average acceleration, time spent in MVPA, time spent inactive and the average acceleration value above which an individual's most active 2, 30 and 60 minutes of activity were obtained ($M2_{ACC}$, $M30_{ACC}$, $M60_{ACC}$; mg) (Rowlands, Sherar, et al., 2019). Furthermore, the distribution of time spent in intensity categories of 25 mg resolution (i.e. 0-25 mg, 25-50 mg, 50-75 mg….4000 mg, >4000 mg) was determined in GGIR using the argument 'iglevels = TRUE' to describe a participant's activity intensity distribution across the monitoring period. This represented the IG. Time spent in MVPA was defined as the time accumulated above an acceleration of 100 m*g* (Hildebrand, Van Hees, Hansen, & Ekelund, 2014) whereas time spent inactive was defined as time accumulated below 50 m*g* (Hildebrand, Hansen, van Hees, & Ekelund, 2017). All activity outcomes were calculated for both devices.

Statistical Analysis

Level of agreement across activities between wrist placements 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 during each laboratory activity and free-living 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 in previous studies (Buchan & McLellan, 2019; Rowlands, Plekhanova, et al., 2019). Where data was not normally distributed, equivalency analyses was performed using the log transformation of the original data. Since neither wrist is thought of as the criterion wear site, equivalence tests were carried out twice with each device used as the reference. We only classified outcome comparisons as equivalent if equivalency was achieved when both monitors were used as the reference device. Statistical analyses were undertaken using IBM SPSS statistical software for Windows version 25 (IBM, Armonk, NY). Equivalency testing were undertaken in Minitab (v17). The extent of information presented in this study to evaluate agreement in outcomes between the two wrist worn accelerometers follows recent recommendations (Looney, 2018).

Results

Part 1: Laboratory-based

The 37 participants completed all activities and three participants declared their left wrist as dominant. The extent of agreement for average acceleration between wrist locations can be found in Table 1. Agreement between wrist locations tended to be poor to good moderate across activities except for running at 6.5 mph which demonstrated moderate to good reliability (ICC 95%CI 0.61-0.88). Due to the presence of heteroscedasticity for all comparisons, Bland-Altman analysis was undertaken using logarithmic transformation as

recommended (Bland & Altman, 1999). Therefore, the mean bias of average acceleration from the dominant wrist relative to the non-dominant wrist at 3 mph was 0.05, with LoA between -0.06 and 0.15. Back-transformation of the data revealed that the dominant wrist LoA were 22.4% lower to 25.9% higher than the non-dominant wrist. At 3.5 mph, mean bias was 0.05 with LoA between -0.05 and 0.16. Back-transformation of the data revealed that the dominant wrist LoA were 20.6% lower to 28.9% higher than the non-dominant wrist. The mean bias of average acceleration at 4 mph was 0.05, with LoA between -0.07 and 0.16. Back-transformation of the data revealed that the dominant wrist LoA were 24.1% lower to 28.9% higher than the non-dominant wrist. For the running activities, the mean bias of average acceleration at 6 mph was 0.01, with LoA between -0.12 and 0.14. Back-transformation of the data revealed that the dominant wrist LoA were 25.9% lower to 34.9% higher than the non-dominant wrist. Finally, the mean bias of average acceleration at 6.5 mph was 0.01, with LoA between -0.09 and 0.12. Back-transformation of the data revealed that the dominant wrist 95% LoA were 20.6% lower to 28.8% higher than the non-dominant wrist.

Figure 1 shows the equivalence zones for average acceleration estimates between wrist sites across all activity intensities, with the dashed lines representing the 10% equivalence zone and a stricter 5% equivalence zone indicated by the dotted lines. Across all activities, average acceleration values measured at the dominant wrist were higher than those from the non-dominant wrist (ratio non-dominant/dominant < 1.0) (Figure 1A) but all outcomes were equivalent on average at the group level and fell within the stricter 5% equivalence zone. The equivalency analyses observations were consistent across activities regardless of the reference device used (Figure 1B).

Part 2: Free-living

Accelerometer files were available for 51 participants. The files of 12 participants were excluded from subsequent analysis (11 failed to wear the device for ≥ 16 $h.d^{-1}$ for at least 4 days including 1 weekend day and 1 had missing data for any 15-min window over the 24-h cycle (as indicated in GGIR by the variable '24-h cycle' < 1). This left a final sample of 39 participants (age: 28.5 ± 9.8 years; 17 females) that were included within the subsequent analysis. Average wear time per day for the non-dominant and dominant wrists were identical at 23.1 $h.d^{-1}$ with devices worn on average for 6.3 ± 0.7 days.

The extent of agreement for activity outcomes between wrist locations can be found in Table 2. Agreement between wrist locations tended to be poor to good for average acceleration, IG and the magnitude of accelerations for an individual's most active 30 minutes whereas agreement for an individual's most active 2 and 60 minutes was moderate to good. Moderate to excellent agreement was evident for inactive time whereas moderate to good agreement was evident for LPA. The mean bias of IG from the dominant wrist relative to the non-dominant wrist was 0.01, with LoA between -0.24 and 0.26 (or -10% and 10%). For average acceleration, mean bias was 2.4 mg, with LoA between -7 and 12 mg (or -22% and 40%). Mean bias for inactive time was -19.9 $min.d^{-1}$, with LoA between -96 and 56 $min.d^{-1}$ (or -8% and 5%). For LPA, mean bias was 5.2 $min.d^{-1}$, with LoA between -37 and 47 $min.d^{-1}$ (or -24% and 33%) whereas for MVPA, mean bias was 14.8 $min.d^{-1}$, with LoA between -42 and 72 $min.d^{-1}$ (or -40% and 72%). Due to the presence of heteroscedasticity for the $M2_{ACC}$ outcome, Bland-Altman analysis was undertaken using logarithmic transformation. Therefore, mean bias was 0.00 with LoA between -0.22 and 0.23. Back transformation of the data revealed that the dominant wrist LoA were 39.7% lower to 69.8% higher that the non-dominant wrist. Finally, for $M30_{ACC}$ the mean bias was 11 mg with LoA

between -37 and 58 mg (or -22% and 35%) whereas for $M60_{ACC}$, the mean bias was 8 mg with LoA between -27 and 43 mg (or -21% and 34%).

Figure 2 shows the equivalence zones for activity outcomes between wrists across all activity intensities with the dashed lines representing the 10% equivalence zone and a stricter 5% equivalence zone indicated by the dotted lines. The IG, inactive time, LPA and the average acceleration above which an individual's most active 2, 30 and 60 minutes were obtained were equivalent on average at the group level when the non-dominant wrist was used as the reference monitor (Figure 2A). Time spent in MVPA and average acceleration were not found to be statistically equivalent. These observations were identical when the dominant wrist monitor was used as the reference (Figure 2B). Further analysis revealed that time spent in moderate PA (MPA) and vigorous PA (VPA) determined from device specific thresholds (Hildebrand et al., 2014), were not found to be statistically equivalent regardless of the wrist monitor used as the reference (data not displayed). Further, time spent in MPA between wrists approached equivalence (ratio non-dominant/dominant = 1.11, 95%CI 0.995, 1.205) and demonstrated moderate to good reliability (ICC = 0.731, 95%CI 0.54, 0.85) whereas time spent in VPA between wrists was not found to be statistically equivalent on average (ratio non-dominant/dominant = 0.18, 95%CI -0.08, 1.0) and demonstrated poor to moderate reliability (ICC = 0.495, 95%CI 0.22, 0.70) between wrist locations.

Average acceleration was approximately 8.5% higher when measured at the dominant wrist compared to the non-dominant wrist. We therefore adjusted the average acceleration values from the dominant wrist by -8.5% for each individual and re-ran the equivalency and ICC analyses. Analysis revealed that average acceleration values were within the 5% equivalent zone regardless of the wrist used as the reference (Figure 3) and demonstrated

moderate to good reliability (ICC = 0.767 95%CI 0.60, 0.87) when the non-dominant wrist was used as the reference. Further, mean bias was 0.1 m*g*, with LoA between -9 and 9 m*g* (or -33% and 33%). We also adjusted the average acceleration values from the dominant wrist by -10% as has been suggested elsewhere (Rowlands, Plekhanova, et al., 2019). We found that average acceleration values were within the 10% equivalent zone, but not the 5% equivalent zone, regardless of the wrist used as the reference (data not shown). Moderate to good reliability (ICC = 0.763 95%CI 0.59, 0.87) was also evident.

Discussion

Few studies have compared activity outcomes based on acceleration values between ActiGraph accelerometers worn on the non-dominant and dominant wrists within laboratory and in free-living settings. Moreover, this is the first study to the best of our knowledge that has examined the equivalency of MX_{ACC} outcomes between wrist locations. In the present study, the dominant wrist produced higher accelerations than the non-dominant wrist across all walking and running laboratory activities, with the largest differences evident whilst walking. During free-living, most population-independent activity outcomes were equivalent at the group level. The dominant wrist produced higher average accelerations than the non-dominant wrist but when values from the dominant wrist were reduced by 8.5%, this metric became equivalent between wrists. Nonetheless, the wide LoA's still evident after this reduction suggests caution is advised when comparing outcomes between wrist locations at the individual level. These findings have important implications for researchers using raw acceleration ActiGraph wrist-worn accelerometers who look to pool findings from studies reporting population-independent activity outcomes from the dominant and non-dominant wrist.

The higher average acceleration values measured at the dominant wrist from ActiGraph accelerometers are in contrast to previous findings (Buchan et al., 2018; Rowlands, Plekhanova, et al., 2019). Recently Rowlands et al. reported from free-living adults that average acceleration was approximately 10% higher when measured at the dominant wrist for the GENEActiv and Axivity but not for the ActiGraph, when all devices were worn concurrently on both wrists (Rowlands, Plekhanova, et al., 2019). It is known that on-board filtering in ActiGraph devices occurs when generating counts whereby accelerations are filtered and rectified through a low-pass frequency filter, that can lead to accelerations being filtered out when captured during moderate-higher intensity activities (Fridolfsson et al., 2019; Rowlands, Plekhanova, et al., 2019). This could be a plausible explanation for the similar average acceleration values between the dominant and non-dominant wrists noted by Rowlands et al. (Rowlands, Plekhanova, et al., 2019). However, with these subjects spending approximately 20 h a day inactive (i.e. below 50 m*g*) and demonstrating equivalency between wrist locations for inactive time, the impact of the filter upon average acceleration levels is likely to be minimal.

Given the longer monitoring period of this study in comparison to others, the higher average acceleration values measured at the dominant wrist likely reflects the range of free-living behaviours including those that may involve more movements using the dominant wrist (e.g. mannerisms, writing, etc.). With Buchan et al. employing an 8hr monitoring period across 1 day it's reasonable to assume that activity outcomes may not be reflective of typical activity behaviours across the whole week (Buchan et al., 2018). Furthermore, despite Rowlands et al. requesting participants wear their devices for 24 hours a day for up to 7 days, the wear protocol was reduced to four days due to participants reporting discomfort from the

wrist straps needed for multiple devices (Rowlands, Plekhanova, et al., 2019). It's reasonable to assume therefore that this discomfort may have influenced habitual activity behaviours.

The narrower 95%CI for ICC and LoA for average acceleration after the most intense laboratory activity may be a consequence of filtering by ActiGraph during moderate-higher intensity activities. Equally, another plausible reason may be less decoupling during the most intense running activity where one would anticipate that both arms are used in a similar manner whereas during walking, the dominant arm may have more of an influence. Nonetheless, the lack of equivalence between wrists for time spent in MVPA during free-living is a consequence of the higher average acceleration values produced from the dominant wrist. The lack of equivalence in MVPA between wrists is contrary to recent findings from ActiGraph devices (Buchan et al., 2018; Rowlands, Plekhanova, et al., 2019) but is consistent with outcomes from the Axivity and GENEActiv devices when worn concurrently on both wrists during free-living (Rowlands, Plekhanova, et al., 2019).

It's important to note that different models of ActiGraph devices were used between studies with the previous version, the GT3X+ being used here and the most recent version, the GT9X Link being used by Rowlands et al. (Rowlands, Plekhanova, et al., 2019). Although it has been demonstrated that there is a high correlation between raw acceleration data produced from hip-worn GT9X and GT3X+ ActiGraph models when worn concurrently (Montoye et al., 2018), it is unclear whether similar findings are evident from the wrist. Recent findings have observed poor intermonitor agreement for step counts between wrist-worn wGT3X and GT9X devices despite the same sampling frequency and normal filter being used for both devices, whilst also controlling for placement effects (Hwang, Fernandez, & Lu, 2018).

As the same accelerometer sensor is used in both the GT3X+ and GT9X models, differences in the design of the devices and how they are worn has been offered as a possible reason which may affect acceleration output and subsequent MVPA estimates (Rowlands, Plekhanova, et al., 2019). The GT3X+ is larger than the GT9X device (46 × 33 × 15 mm, mass 19 g vs. 35 × 35 × 10 mm, mass 14 g) but our participants only wore one device on each wrist whereas Rowlands et al. had participants wear three devices concurrently using two wrist-straps, with the Axivity taped on top of the GT9X device (Rowlands, Plekhanova, et al., 2019). Given the discomfort noted by the participants, it's reasonable to assume that certain activities during free-living such as writing or using a computer mouse that would typically be undertaken using the dominant wrist may have been restricted and could explain the equivalent average acceleration output between wrist locations observed by Rowlands et al., (2019).

Previous studies comparing outputs from the GENEActiv, Axivity and ActiGraph devices when taped together at the wrist (Stiles, Griew, & Rowlands, 2013) or worn adjacent on the wrist (Rowlands et al., 2017; Rowlands, Yates, Davies, Khunti, & Edwardson, 2016) suggest that average acceleration output from ActiGraph devices deployed on the non-dominant wrist are approximately 10% lower than the GENEActiv and Axivity devices. In contrast, recent findings found average acceleration to be equivalent between ActiGraph, GENEActiv and Axivity devices worn on the non-dominant wrist but approximately 10% higher for the GENEActiv and Axivity devices in relation to the ActiGraph when measured from the dominant wrist (Rowlands, Plekhanova, et al., 2019). These studies suggest that equivalence in average acceleration outcomes between devices may differ depending on the wrist location, duration of monitoring period and setting (i.e. laboratory or free-living) which

our findings seem to support. Future work therefore may wish to examine the comparability of activity outcomes between the three-research grade raw accelerometer devices in both laboratory and in free-living settings to confirm our findings.

A strength of this study is the evaluation of data outcomes from the GT3X+ device worn on the dominant and non-dominant wrists in both a laboratory and free-living setting. This allowed for comparisons to be made across a range of defined intensities and to determine whether these differences affect daily free-living estimates. The high average wear time and number of days the devices were worn during free-living is a strength of this study, and provides confidence that data collected is representative of weekly activity behaviours. The lack of common free-living behaviour undertaken in the laboratory setting could be considered a limitation. However, we hypothesised that individuals would spend a large amount of their time inactive and in LPA during free-living and asking participants to undertake activities at these intensities in a laboratory setting over a short period of time felt unnecessary. Moreover, with the walking and running activities we wanted to ensure that participants were undertaking a range of activity intensities which could be controlled and monitored to compare data outcomes across activities of increasing intensities. Another limitation could be the choice of activities performed in the laboratory setting which required similar movements between the wrists lacking ecological validity. Finally, the self-selected homogenous cohort used in the laboratory and free-living settings may limit the generalizability of our findings.

In summary, average acceleration measured from the dominant wrist was equivalent at the group level to the output from the non-dominant wrist across a range of defined intensities in a controlled laboratory setting. During free-living, most of the population

independent activity metrics were equivalent at the group level between wrist locations. Adjusting average acceleration values from the dominant wrist by -8.5% resulted in equivalent outputs at the group level for this metric between the wrists. At the individual level, wide LoA's across all comparisons suggest that caution is needed when comparing activity outcomes between wrist locations at the individual level. Nonetheless, further work involving different populations is needed to confirm these observations.

Acknowledgements

This study was partly funded by The University of the West of Scotland, VP Research Fund. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. We would like to thank Craig Napier, Grant Brownlie and Christopher Shields for their assistance with the data collection. We would also like to thank Dr Alex V. Rowlands for his help in processing the back-up calibration coefficients used in part 1 and Prof. Stuart Fairclough for his comments on an earlier draft of this manuscript.

Declaration of interest

The authors declare that they have no competing interests.

Availability of data and material

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request once ethical approval from The University of the West of Scotland Ethics committee has been obtained.

References

Ainsworth, B. E., Haskell, W. L., Herrmann, S. D., Meckes, N., Bassett, D. R., Tudor-Locke, C., … Leon, A. S. (2011). 2011 Compendium of Physical Activities: A second update of codes and MET values. *Medicine and Science in Sports and Exercise*, *43*(8), 1575–1581. https://doi.org/10.1249/MSS.0b013e31821ece12

Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet (London, England)*, *1*(8476), 307–310.

Bland, J. M., & Altman, D. G. (1999). Measuring agreement in method comparison studies. *Statistical Methods in Medical Research*, *8*(2), 135–160. https://doi.org/10.1177/096228029900800204

Buchan, D. S., & McLellan, G. (2019). Comparing physical activity estimates in children from hip-worn Actigraph GT3X+ accelerometers using raw and counts based processing methods. *Journal of Sports Sciences*, *37*(7), 779–787. https://doi.org/10.1080/02640414.2018.1527198

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

Colley, R. C., Garriguet, D., Janssen, I., Craig, C. L., Clarke, J., & Tremblay, M. S. (2011). Physical activity of Canadian adults: Accelerometer results from the 2007 to 2009 Canadian Health Measures Survey. *Health Reports*, *22*(1), 7–14.

Dieu, O., Mikulovic, J., Fardy, P. S., Bui-Xuan, G., Béghin, L., & Vanhelst, J. (2017). Physical activity using wrist-worn accelerometers: Comparison of dominant and non-dominant wrist. *Clinical Physiology and Functional Imaging*, *37*(5), 525–529. https://doi.org/10.1111/cpf.12337

Dixon, P. M., Saint-Maurice, P. F., Kim, Y., Hibbing, P., Bai, Y., & Welk, G. J. (2018). A Primer on the Use of Equivalence Testing for Evaluating Measurement Agreement. *Medicine and Science in Sports and 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., … 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

Esliger, D. W., Rowlands, A., Hurst, T. L., Catt, M., Murray, P., & Eston, R. G. (2011). Validation of the GENEA Accelerometer. *Medicine and Science in Sports and Exercise*, *43*(6), 1085–1093. https://doi.org/10.1249/MSS.0b013e31820513be

Fridolfsson, J., Börjesson, M., Buck, C., Ekblom, Ö., Ekblom-Bak, E., Hunsberger, M., … Arvidsson, D. (2019). 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., & 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, VT., Hansen, B. H., & Ekelund, U. (2014). Age group comparability of raw accelerometer output from wrist- and hip-worn monitors. *Medicine and Science in Sports and Exercise*, *46*(9), 1816–1824. https://doi.org/10.1249/MSS.0000000000000289

Hwang, J., Fernandez, A. M., & Lu, A. S. (2018). Application and Validation of Activity Monitors' Epoch Lengths and Placement Sites for Physical Activity Assessment in Exergaming. *Journal of Clinical Medicine*, *7*(9). https://doi.org/10.3390/jcm7090268

Koo, T. K., & Li, M. Y. (2016). 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

Looney, M. A. (2018). Assessment of interrater and intermethod agreement in the kinesiology literature. *Measurement in Physical Education and Exercise Science*, *22*(2), 116–128. https://doi.org/10.1080/1091367X.2017.1395742

Matthews, C. E., Chen, K. Y., Freedson, P. S., Buchowski, M. S., Beech, B. M., Pate, R. R., & Troiano, R. P. (2008). Amount of Time Spent in Sedentary Behaviors in the United States, 2003–2004. *American Journal of Epidemiology*, *167*(7), 875–881. https://doi.org/10.1093/aje/kwm390

Montoye, A. H. K., Nelson, M. B., Bock, J. M., Imboden, M. T., Kaminsky, L. A., Mackintosh, K. A., … Pfeiffer, K. A. (2018). Raw and Count Data Comparability of Hip-Worn ActiGraph GT3X+ and Link Accelerometers. *Medicine and Science in Sports and Exercise*, *50*(5), 1103–1112. https://doi.org/10.1249/MSS.0000000000001534

NHANES. (2018, February 16). NHANES - National Health and Nutrition Examination Survey Homepage. Retrieved 26 February 2018, from https://www.cdc.gov/nchs/nhanes/index.htm

Phillips, L. R. S., Parfitt, G., & Rowlands, A. V. (2013). Calibration of the GENEA accelerometer for assessment of physical activity intensity in children. *Journal of Science and Medicine in Sport*, *16*(2), 124–128. https://doi.org/10.1016/j.jsams.2012.05.013

Rowlands, A. V. (2018). 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. V., Edwardson, C. L., Davies, M. J., Khunti, K., Harrington, D. M., & Yates, T. (2018). Beyond Cut Points: Accelerometer Metrics that Capture the Physical Activity Profile. *Medicine and Science in Sports and 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., & L Edwardson, C. (2017). Accelerometer-assessed Physical Activity in Epidemiology: Are Monitors Equivalent? *Medicine & Science in Sports & Exercise*, *50*, 1. https://doi.org/10.1249/MSS.0000000000001435

Rowlands, A. V., Plekhanova, T., Yates, T., Mirkes, E., Davies, M., Khunti, K., & Edwardson, C. (2019). A basis for harmonisation of accelerometer physical activity outcomes in epidemiology. *Journal for the Measurement of Physical Behaviours*. Retrieved from https://lra.le.ac.uk/handle/2381/43838

Rowlands, A. V., Sherar, L. B., Fairclough, S. J., Yates, T., Edwardson, C. L., Harrington, D. M., … Stiles, V. H. (2019). A data-driven, meaningful, easy to interpret, population-independent accelerometer outcome variable for global surveillance. *BioRxiv*, 604694. https://doi.org/10.1101/604694

Rowlands, A. V., Yates, T., Davies, M., Khunti, K., & Edwardson, C. L. (2016). Raw Accelerometer Data Analysis with GGIR R-package: Does Accelerometer Brand Matter? *Medicine and Science in Sports and Exercise*, *48*(10), 1935–1941. https://doi.org/10.1249/MSS.0000000000000978

Stiles, V. H., Griew, P. J., & Rowlands, A. V. (2013). Use of accelerometry to classify activity beneficial to bone in premenopausal women. *Medicine and Science in Sports and Exercise*, *45*(12), 2353–2361. https://doi.org/10.1249/MSS.0b013e31829ba765

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

Van Hees, V. T., Fang, Z., Langford, J., Assah, F., Mohammad, A., Silva, I. C. M. da, … Brage, S. (2014). 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., … 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* (pp. xv, 284 p.). Retrieved 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

Wijndaele, K., Westgate, K. L., Stephens, S. K., Blair, S. N., Bull, F. C., Chastin, S. F. M., … Healy, G. N. (2015). Utilization and Harmonization of Adult Accelerometry Data: Review and Expert Consensus. *Medicine and Science in Sports and Exercise*, *47*(10), 2129–2139. https://doi.org/10.1249/MSS.0000000000000661

Zhang, S., Rowlands, A. V., Murray, P., & Hurst, T. L. (2012). Physical activity classification using the GENEA wrist-worn accelerometer. *Medicine and Science in Sports and Exercise*, *44*(4), 742–748. https://doi.org/10.1249/MSS.0b013e31823bf95c

Table 1. Agreement between average acceleration (measured in milligravity units, m*g*) values per 5 s for walking (3-4 mph) and running (6-6.5 mph) from ActiGraph accelerometers worn on the non-dominant and dominant wrists (n = 37).

Activity	**Average acceleration (non-dominant wrist) Median (25^{th}-75^{th} percentile)**	**Average acceleration (dominant wrist) Median (25^{th}-75^{th} percentile)**	**ICC (95% CI)**	**Limits of agreement ǂ**	**Equivalency†**
3 mph	145.9 (120.4-158.1)	161.5 (137.5-182.1)	0.735 (0.26, 0.89)	22.4%, 25.9%	✓
3.5 mph	178.4 (157.5-189.1)	195.6 (176.5-224.3)	0.587 (0.01, 0.83)	20.6%, 28.9%	✓
4 mph	234.0 (208.6-260.3)	261.2 (214.2-272.5)	0.730 (0.30, 0.88)	24.1%, 28.9%	✓
6 mph	889.4 (775.1-975.7)	908.3 (825.5-1014.1)	0.601 (0.35, 0.77)	25.9%, 34.9%	✓
6.5 mph	978.9 (829.4-1074.8)	1006.8 (881.2-1074.5)	0.777 (0.61, 0.88)	20.6%, 28.8%	✓

Comparisons always made between the dominant versus non-dominant wrist. ICC = Intraclass Correlation Coefficients. ǂLog-transformed data back-transformed (antilog) and reported as percentages. †95% confidence interval (CI) for mean of the monitor worn on the dominant wrist falls within the proposed equivalence zone (i.e. ±10%) of the mean of the monitor worn on the non-dominant wrist. ✓ = outcomes considered equivalent.

Table 2. Summary of activity outcomes during free-living from ActiGraph accelerometers worn on the dominant and non-dominant wrists (n = 39).

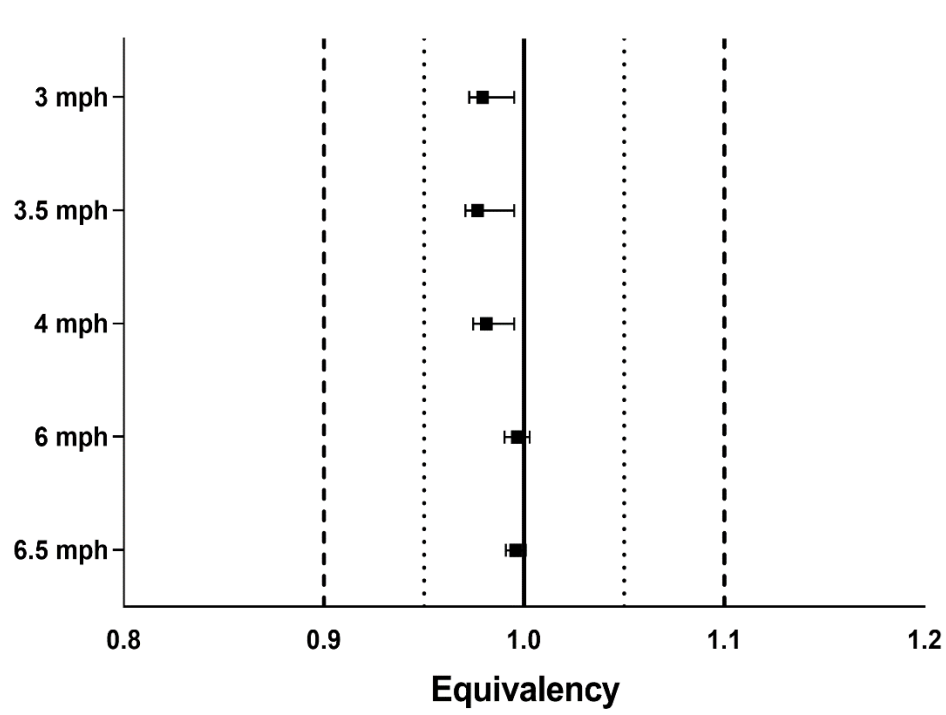
Activity outcome	**(non-dominant wrist) Mean or median (± SD or 25th-75th percentile)**	**(dominant wrist) Mean or median (± SD or 25th-75th percentile)**	**ICC (95% CI)**	**Limits of agreement**	**Equivalency†**
Intensity Gradient	-2.48 ± 0.17	-2.47 ± 0.14	0.671 (0.45, 0.81)	-0.24, 0.26	✓
Average acceleration (m*g*)	27.4 ± 6.9	29.8 ± 6.9	0.727 (0.48, 0.86)	-7, 12 m*g*	X
Inactive Time (min.d^{-1})	1212.9 ± 68.8	1192.9 ± 69.8	0.813 (0.61, 0.91)	-96, 56 mins	✓
LPA (min.d^{-1})	133.3 ± 34.3	138.4 ± 33.9	0.802 (0.65, 0.89)	-37, 47 mins	✓
MVPA (min.d^{-1})	94.5 ± 40.4	109.3 ± 43.8	0.724 (0.47, 0.86)	-42, 72 mins	X
$M2_{ACC}$ (m*g*)	451.1 (411.7-609.8)	479.4 (423.6-607.5)	0.746 (0.57, 0.86)	40%, 70% ǂ	✓
$M30_{ACC}$ (m*g*)	172.8 ± 44.1	178.7 ± 40.2	0.598 (0.35, 0.77)	-37, 58 m*g*	✓
$M60_{ACC}$ (m*g*)	126.1 ± 30.9	131.4 ± 30.2	0.702 (0.50, 0.83)	-27, 43 m*g*	✓

Comparisons always made between the dominant versus non-dominant wrist. ICC = Intraclass Correlation Coefficients. ǂ Log-transformed data back-transformed (antilog) and reported as percentages due to the presence of heteroscedasticity. †95% confidence interval (CI) for mean of the monitor worn on the dominant wrist falls within the proposed equivalence zone (i.e. ±10%) of the mean of the monitor worn on the non-dominant wrist. ✓ = outcomes considered equivalent. X = outcomes not considered equivalent. LPA = light physical activity, MPA = moderate physical activity, VPA = vigorous physical activity, MVPA = moderate-to-vigorous physical activity. MX_{ACC} = average acceleration above which the most active X mins was obtained.

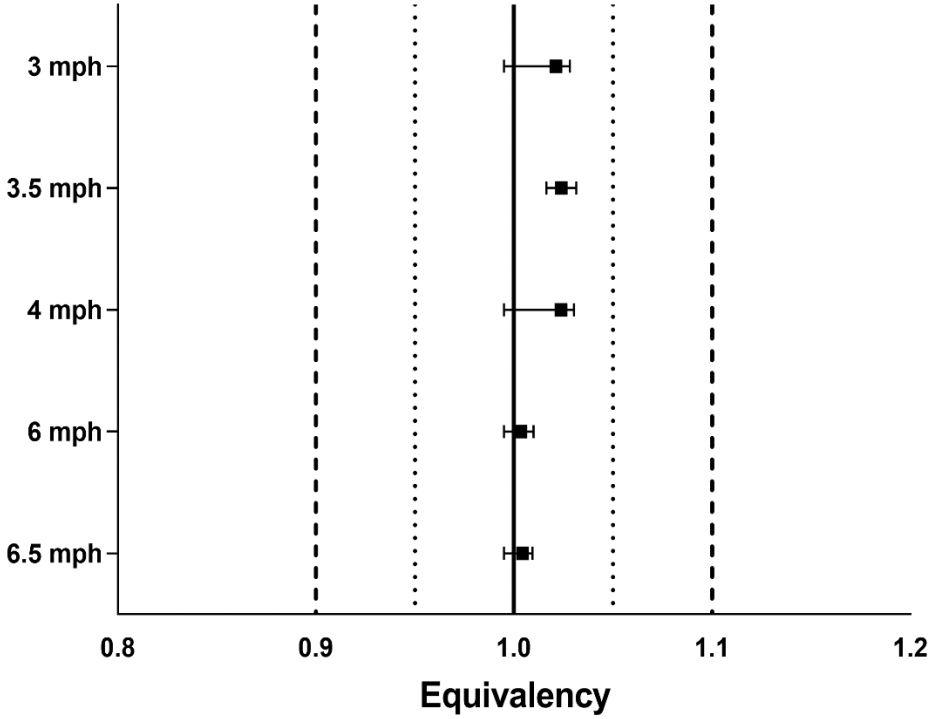
Figure 1. Equivalence of average acceleration for walking (3-4 mph) and running (6-6.5 mph) per 5 s between ActiGraph accelerometers worn on the non-dominant and dominant wrist. A) non-dominant wrist used as reference; B) dominant wrist used as the reference. The 10% zone of equivalence is indicated by the dashed lines and dotted lines indicate a stricter 5% equivalence zone.

Figure 2. Equivalence between ActiGraph accelerometers worn on the non-dominant and dominant wrist. A) non-dominant wrist used as reference; B) dominant wrist used as the reference. The 10% zone of equivalence is indicated by the dashed lines and dotted lines indicate a stricter 5% equivalence zone. LPA = light physical activity, MVPA = moderate-to-vigorous physical activity, MX_{ACC} = average acceleration above which the most active X mins was obtained.

Figure 3. Equivalence between ActiGraph accelerometers worn on the non-dominant and dominant wrist for average acceleration, adjusted by -8.5% for the dominant wrist A) non-dominant wrist used as reference; B) dominant wrist used as the reference. The 10% zone of equivalence is indicated by the dashed lines and dotted lines indicate a stricter 5% equivalence zone. Note: no adjustment was made to the intensity gradient values as variable is included only to demonstrate the equivalence of these activity metrics.

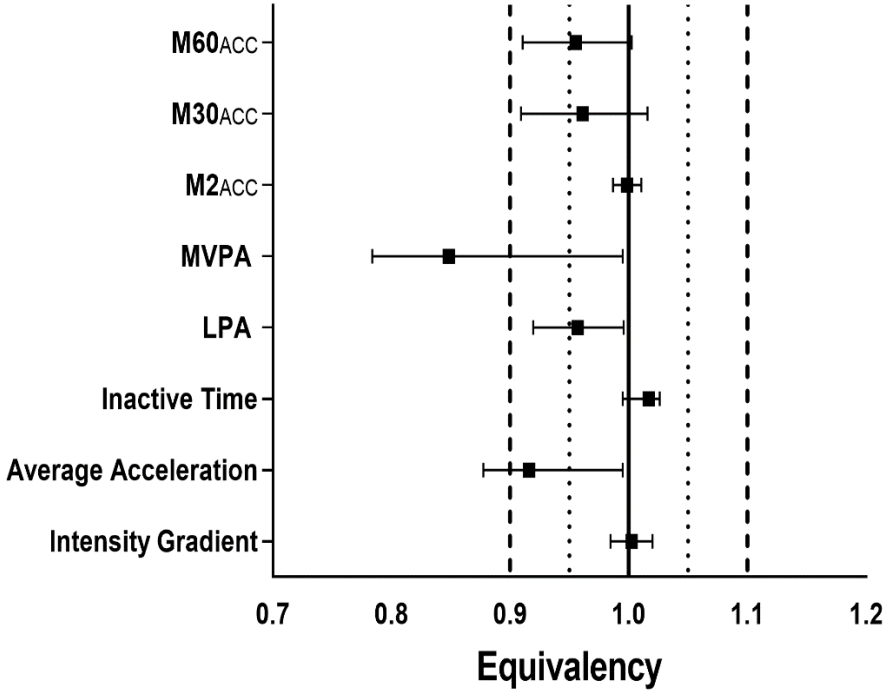

3 mph
3.5 mph
4 mph
6 mph
6.5 mph
0.8
0.9
1.0
1.1
1.2
Equivalency

A

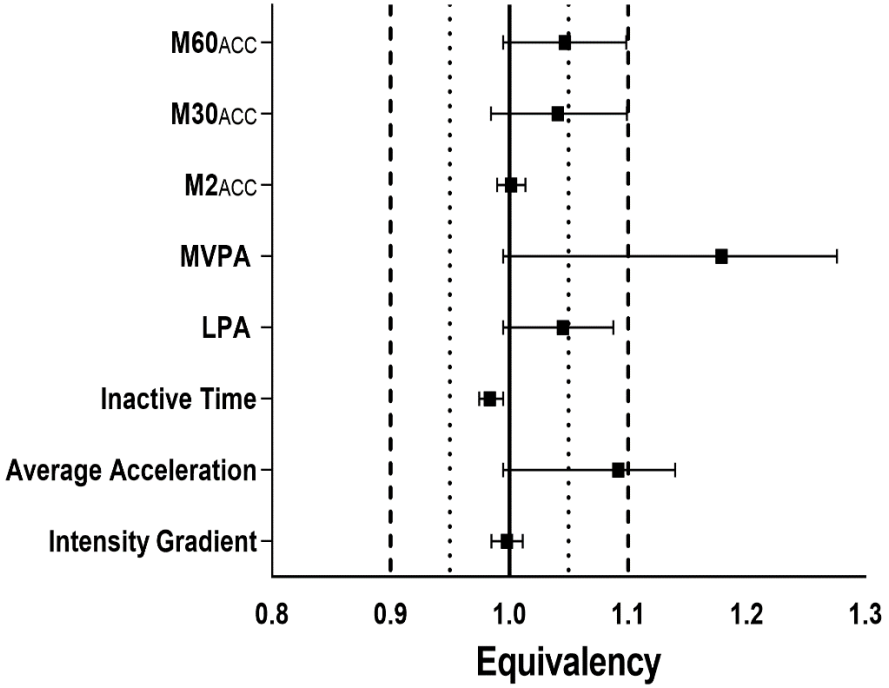

3 mph
3.5 mph
4 mph
6 mph
6.5 mph
0.8
0.9
1.0
1.1
1.2
Equivalency

B

Figure 1.

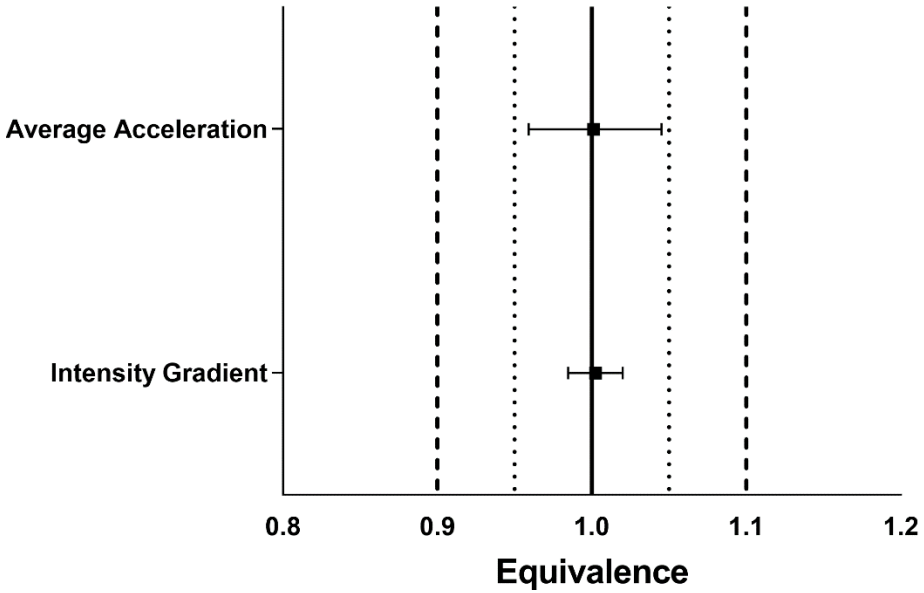

M60ACC
M30ACC
M2ACC
MVPA
LPA
Inactive Time
Average Acceleration
Intensity Gradient
0.7
0.8
0.9
1.0
1.1
1.2
Equivalency

A

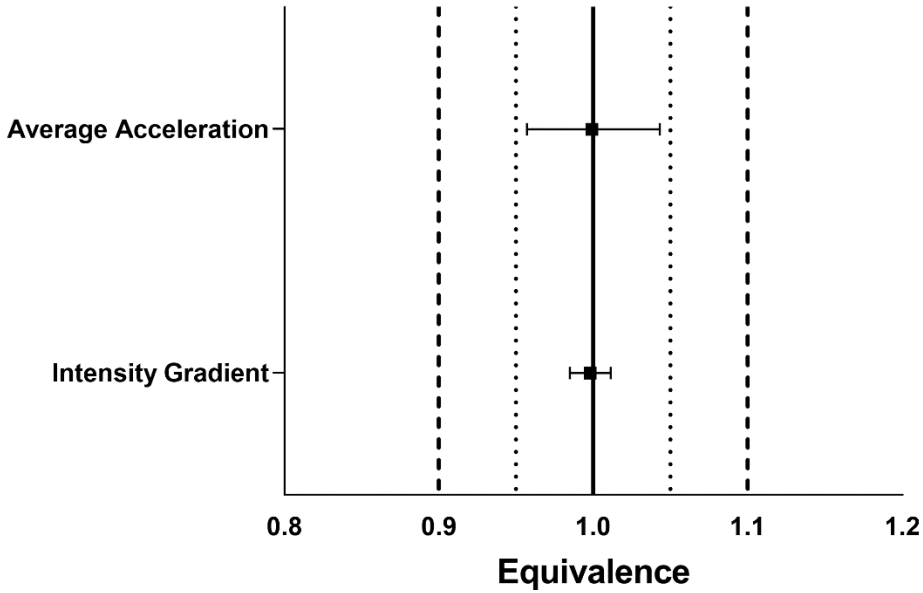

M60ACC
M30ACC
M2ACC
MVPA
LPA
Inactive Time
Average Acceleration
Intensity Gradient
0.8
0.9
1.0
1.1
1.2
1.3
Equivalency

B

Figure 2.


Average Acceleration
Intensity Gradient
0.8
0.9
1.0
1.1
1.2
Equivalence

A


Average Acceleration
Intensity Gradient
0.8
0.9
1.0
1.1
1.2
Equivalence

B

Figure 3.