Comparison of Sleep and Physical Activity Metrics From Wrist-Worn ActiGraph wGT3X-BT and GT9X Accelerometers During Free-living in Adults

Duncan S Buchan

Division of Sport and Exercise, School of Health and Life Sciences, University of the West of Scotland, Lanarkshire Campus, G72 0LH, Scotland, UK.

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

Abstract

Background: ActiGraph accelerometers can monitor sleep and physical activity (PA) during free-living, but there is a need to confirm agreement in outcomes between different models. ***Methods:*** Sleep and PA metrics from two ActiGraphs were compared after participants (N = 30) wore a GT9X and wGT3X-BT on their non-dominant wrist for 7 days during free-living. PA metrics including total steps, counts, average acceleration (AA) - Euclidean Norm Minus One (ENMO) and Mean Amplitude Deviation (MAD), intensity gradient, the minimum acceleration value of the most active 10 and 30 mins (M10, M30), time spent in activity intensities from vector magnitude (VM) counts and ENMO cut-points and sleep metrics (sleep period time window, sleep duration, sleep onset and waking time) were compared. ***Results:*** Excellent agreement was evident for AA-MAD, counts, total steps, M10 and light PA (VM counts) with good agreement evident from the remaining PA metrics apart from moderate-vigorous PA (MVPA) (VM counts) which demonstrated moderate agreement. Mean bias for all PA metrics were low as were the limits of agreement (LoA) for the intensity gradient, AA-MAD and inactive time (ENMO and VM counts). The LoA for all other PA metrics were >10%. Excellent agreement, low mean bias and narrow LoA were evident for all sleep metrics. All sleep and PA metrics demonstrated equivalence (equivalence zone of ≤ 10%) apart from MVPA (ENMO) which needed an equivalence zone of 16%. ***Conclusions:*** Equivalent estimates of almost all PA and sleep metrics are provided from the GT9X and wGT3X-BT worn on the non-dominant wrist.

Keywords: GGIR, equivalence, steps, counts, ENMO

The use of device-based measures of physical activity (PA) are common when assessing habitual levels of this behaviour and evaluating intervention effectiveness (Troiano et al., 2014). The ActiGraph accelerometer is one such device that is often deployed by researchers given its ability to capture all components of PA including intensity, duration and frequency as well as sporadic activity throughout the day (Migueles et al., 2017). Since the release of the first ActiGraph model in 1993, newer models have since become available including the GT3X+ (2010), wGT3X-BT (2013), and the GT9X (2014). Although the two earlier models have identical dimensions (4.6 × 3.3 × 1.5cm, 19 g) they have different dynamic ranges (GT3X+: ± 6 *g*; wGT3X-BT: ± 8 *g*) which suggests that these two models have different internal accelerometers (ActiGraph, 2013). Although the GT9X has the same primary accelerometer as the wGT3X-BT (ADXL362 by Analog Devices, Norwood, MA) (Personal communication with ActiGraph), these models have different dimensions and unique components with the GT9X having a secondary accelerometer and an inertial measurement unit and the wGT3X-BT having a lux sensor (ActiGraph, 2019). Furthermore, technological advances have seen the GT9X become smaller and lighter compared to the wGT3X-BT (wGT3X-BT: 4.6 × 3.3 × 1.5 cm, 19g; GT9X: 3.5 × 3.5 × 1.0cm, 14g). With the GT9X being used in recent randomized controlled trials (Alley et al., 2022; McDonough et al., 2022; Nathan et al., 2021), it is important to establish whether similar outcomes are provided from accelerometer devices when newer models become available.

To the best of our knowledge, only two studies have examined the comparability of metrics from the GT9X and earlier ActiGraph models in adults during free-living (Clevenger et al., 2020a; Shah et al., 2023). In the Clevenger et al., (2020a) study participants wore the GT9X and wGT3X-BT at the hip and the GT9X and either the wGT3X-BT or GT3X+ on each wrist for 4 days during waking hours. In the main, Clevenger et al., (2020a) found good agreement between vector magnitude (VM) counts, Mean Amplitude Deviation (MAD) and time spent

in activity intensities at wear locations, regardless of device pairings. The authors did however observe a lack of equivalence for Euclidean Norm Minus One (ENMO) at both wear locations, regardless of the device pairings. Whereas in the study by Shah et al., (2023) participants wore a GT3X and an GT9X which were strapped together on their dominant wrist for 7 days during free-living, and found that total activity counts and time spent in activity intensities were comparable. Whilst these findings are useful, there are limitations in these studies. For instance, Clevenger et al., (2020a) had participants wear devices during waking hours only which precludes any comparisons between sleep metrics. Being able to establish the extent of agreement between the GT9X and earlier models for sleep metrics would be of particular interest to those reporting 24 h movement behaviours. Furthermore, whilst Shah et al., (2023) did employ a 24 h wear time protocol, comparisons were made between devices worn on the dominant wrist and their analysis was limited to outcomes provided by the proprietary software ActiLife only. As the non-dominant wrist seems to be a popular location to deploy ActiGraph devices (Migueles et al., 2017), further study is needed to address current evidence gaps.

When implementing a 24 h wear time protocol, researchers can use the open-source software GGIR to process accelerometer data (Migueles, Rowlands, et al., 2019). This is appealing for researchers as sleep and PA metrics can be captured from a single device. Moreover, processing acceleration data using GGIR removes the reliance upon proprietary software, can facilitate reproducible analysis and can incorporate researcher developed external functions into data processing (Migueles, Rowlands, et al., 2019). An example of external functions that can be applied within GGIR include the Verisense step-count algorithm (Patterson, 2021), and the actilifecounts R package (Migueles, 2022). Since stepping behaviour is intuitive and easily understandable by the general population (Bassett et al., 2017), being able to quantify steps within GGIR could enhance understanding of the associations between

stepping behaviour and health outcomes. Similarly, ActiGraph accelerometers have been widely used by researchers to quantity PA and energy expenditure, using the main output 'counts' which is generated within the proprietary ActiLife software. Until now, researchers using accelerometers other than ActiGraph were unable to generate count data (Neishabouri et al., 2022). Being able to generate count data within GGIR alongside a variety of activity and sleep metrics is appealing, since the same non-wear detection algorithms can be applied.

Recent findings have reported that equivalent estimates of sleep quality and timing are provided from a wrist worn GT9X when applying an automated sleep detection algorithm against the gold standard Polysomnography (PSG) (Plekhanova et al., 2022). It is unclear whether comparable estimates of sleep duration and timing are provided from the GT9X and earlier ActiGraph models however. Researchers can apply the automated Heuristic Algorithm looking at Change of Z-Angle (HDCZA) algorithm within the open-source software GGIR, which then combines sleep and activity data across the entire day (Migueles, Rowlands, et al., 2019; Van Hees et al., 2018). Clearly, understanding the extent of agreement for PA and sleep metrics between accelerometer devices is important for researchers using ActiGraph devices. Whether that is for pooling outcomes from studies using different ActiGraph models, or to maximize the pool of devices used in future studies if different models are available. Therefore, the aim of this study is to compare non-acceleration, acceleration and counts-based PA metrics as well as sleep metrics from the wGT3X-BT and GT9X worn concurrently on the non-dominant wrist during free-living in adults.

Methods

A total of 34 apparently healthy adults (14 females) aged 30.1 ± 5.4 years were recruited from South Lanarkshire, Scotland between October 2021, and February 2022. Participants

provided informed consent upon study approval from the ethics committee of the University of the West of Scotland (Approval number 16792-14091).

Participants were instructed to wear two ActiGraph accelerometers (one wGT3X-BT and one GT9X) on their self-reported non-dominant wrist. Both devices were initialized in ActiLife v6.13.3 (ActiGraph, Pensacola, FL, USA) to collect data at 100 Hz, and set to commence data collection immediately after distribution. Both devices are able to capture a dynamic range of ± 8*g*. The "idle sleep mode" in ActiLife v6.13.3 was not enabled. The devices were placed next to each other on the wrist with the order of devices randomized (i.e., proximal vs. distal) between participants. The GT9X was worn using the ActiGraph GT9X Link adjustable wear sensor wristband whereas a woven nylon wristband was used to feed through the two slots at either end of the wGT3X-BT. Participants were instructed to wear the devices for 24 h a day, for 7 days, but were to remove the devices during water–based activities.

Upon the return of the devices, data were downloaded using ActiLife v6.13.3 and saved as 60-sec AGD files using 3-axis (VM) with the low-frequency extension disabled, and in raw format as .gt3x files. The .gt3x files were subsequently processed using the GGIR package version 2.10-1 in R statistical software (R Foundation for Statistical Computing, Vienna, Austria, http://cran.r-project.org/) (Migueles, Rowlands, et al., 2019). When using GGIR, signal processing includes the detection of abnormally high values which can be replaced if specified by the user, detection of non-wear (Migueles, Rowlands, et al., 2019) and the calibration of data files using local gravity as a reference (van Hees et al., 2014). Two acceleration metrics were also calculated including ENMO and MAD, averaged over 5 s epochs and expressed in milli-gravitational units (m*g*). These metrics reflect body movement by removing the gravitational component of acceleration with ENMO reflecting the average magnitude of dynamic acceleration, minus 1 *g*, with negative values rounded up to zero (van Hees et al., 2013). Whereas MAD reflects the variability in acceleration around the mean

(Euclidean norm of each raw acceleration datapoint minus the mean value for a given time period) (Edwardson et al., 2022). Files were excluded from subsequent analyses if post-calibration error was > 0.01 *g*, or if participants had less than 1 valid wear day (defined as ≥ 16 h per day) or wear data was not present for each 15 min period of the 24-h cycle (Buchan, 2022). GGIR imputes missing non-wear data with the average of the same 15 time period of other days if available, which helps to enhance participant retention and retain available data. As the imputation feature (do.imputation) is the default configuration of GGIR, this setting was retained to enhance the generalazability of the findings. Users have the option of removing the imputation feature if desired.

To identify valid days, the wear and non-wear time classifications was confirmed numerically and by comparing the accelerometer traces provided by GGIR with metrics only calculated when the same valid days were provided by both devices. The default non-wear setting was applied during processing within GGIR whereby invalid data was imputed by the average at similar time points on different days of the week (van Hees et al., 2013).

VM counts were also calculated in ActiLife v6.13.3 using the 60-s AGD files. Wear time was determined using the Choi et al., algorithm provided within ActiLife, modified to include the use of triaxial data (Choi et al., 2012). To identify valid days, the wear and non-wear time classifications were confirmed numerically in ActiLife with metrics only calculated when the same valid days were provided by both devices. Thereafter, time spent in activity intensities were calculated at the non-dominant wrist using cut-points for VM counts (Montoye et al., 2020). These cut-points were applied to determine inactive time (inclusive of sleep) (< 2860 counts/min), light-intensity PA (LPA) (2860 – 3940 counts/min) and moderate-to-vigorous PA (MVPA) (≥ 3941 counts/min).

Metrics provided from each device using GGIR included wear time (days), average acceleration (mg) which reflects the overall volume of PA, and the intensity gradient (IG) which describes the acceleration intensity distribution across the monitoring period (Rowlands et al., 2018). When calculating average acceleration (AA), both the ENMO and MAD metrics were calculated and are reported as AA-ENMO and AA-MAD henceforth. Alongside these metrics, ENMO cut-points were applied to determine time spent in inactive time (inclusive of sleep) (< 40 mg) (Hildebrand et al., 2017), LPA (between 40 mg and 100 mg), and MVPA (> 100 mg) (Hildebrand et al., 2014). Finally, the minimal acceleration value above which the most active 10 and 30 min/day (M10 and M30) were accumulated from the ENMO metric were also reported (Rowlands, Sherar, et al., 2019). MAD cut-points were not available for adults and the non-dominant wrist.

ActiGraph counts·min^{-1} were also provided in GGIR using the actilifecounts R package external function algorithm (Migueles, 2022). The actilifecounts algorithm replicates the algorithm used in the ActiLife software and allows users to obtain activity counts as recently described (Neishabouri et al., 2022). For the count data, only the VM counts are reported which were calculated as the square root of the sum of the squared counts in each of the three axes and reported as counts·min^{-1}. Steps were estimated using the Verisense step-count external function algorithm which provided total step counts per day (Patterson, 2021). This metric is reported as Total steps. Finally, sleep metrics including the sleep period time window (SPT), sleep duration, sleep onset and waking time were calculated using the HDCZA algorithm (Van Hees et al., 2018). Focusing on these metrics were informed from recent findings which demonstrated they were comparable to PSG measures when provided from the GT9X worn on either wrist, and in the absence of a sleep log (Plekhanova et al., 2022).

Statistics

Descriptive statistics were calculated for all measures (mean ± SD) or median (25th –75th percentile) following normality testing. Agreement in metrics between devices were explored using intraclass correlation coefficients (ICC, two-way mixed effects, single measures, absolute agreement) with 95% confidence intervals (CI), equivalence tests and limits of agreement (LoA) (Bland & Altman, 1986). Depending upon the lower bound 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 confidence interval approach was used to undertake the equivalence tests (95% equivalence test) (Wellek, 2003). In the absence of a criterion device, equivalence tests were carried out twice with each device used as the reference device. Comparisons were only reported as equivalent if equivalence was achieved when both devices were used as the reference, and the mean differences between devices fell within a minimum required equivalence zone. To establish equivalence, the CI of the mean from one accelerometer device had to fall within a defined equivalence zone of the mean of the alternate accelerometer. An equivalence zone of ≤ 10% of the mean was used to establish equivalence between pairings which is often used in similar studies (Buchan et al., 2020; Rowlands, Plekhanova, et al., 2019). Finally, Bland-Altman plots were used to visualize the extent of the differences in metrics between devices and to assess agreement (Bland & Altman, 1986). Statistical analyses were undertaken using IBM SPSS statistical software for Windows version 25 (IBM, Armonk, NY, USA) and GraphPad Prism 9 was used to estimate LoA. Finally, equivalence testing was undertaken in Minitab (v17) with alpha set at 0.05.

Results

From the 34 participants who provided baseline measures and wore the accelerometer devices, two participants experienced a device malfunction with their wGT3X-BT device. This left data from 32 participants to be processed in GGIR. Data from an additional two subjects were removed as they failed to provide at least 1 valid day of wear time from one or more of the devices. This left data from 30 participants (12 females) to be used in subsequent analysis. No data files were removed due to post-calibration error. Devices were worn on average for 5.7 ± 1.2 days with the average wear time being 22.8 ± 2.4h per day. Descriptive statistics of PA and sleep metrics are provided in Table 1. Findings from the mean bias, 95% LoA and ICC's are provided in Table 2. Equivalence for PA and sleep metrics are displayed in Figure 1a-b. Bland-Altman plots for PA metrics are displayed in Figure 2a-h and for sleep metrics in Figure 3a-d.

For the PA metrics, excellent agreement (>0.90) was observed from the ICC's for the intensity gradient, AA-MAD, total steps, M10 and LPA (VM counts) with good agreement (0.75–0.9) for AA-ENMO, counts, inactive time (ENMO), LPA (ENMO), MVPA (ENMO), M30 and inactive time (VM counts). From the ICC's, moderate agreement (0.5–0.75) was evident for MVPA (VM counts). Mean bias for all PA metrics were low. The LoA for the intensity gradient, AA-MAD and inactive time (both ENMO and VM counts) were all below 10%. The LoA for all other metrics were greater than 10% with MVPA (VM counts) demonstrating the widest LoA at 68%. For the sleep metrics, excellent agreement (>0.90) from the ICC's were found for sleep onset, waking time, SPT-window and sleep duration. Mean bias for the sleep metrics were low and displayed narrow LoA.

The equivalence zone needed for PA metrics to be equivalent between the GT9X and wGT3X-BT were: AA-ENMO (7%), intensity gradient (2%), AA-MAD (5%), counts (5%),

total steps (4%), inactive time (ENMO) (2%), LPA (ENMO) (10%), MVPA (ENMO) (16%), M10 (7%), M30 (7%), inactive time (VM counts) (2%), LPA (VM counts) (4%) and MVPA (VM counts) (9%) (Figure 1 a). The equivalence zone needed for the sleep metrics to be considered equivalent between the GT9X and wGT3X-BT were: sleep onset (1%), waking time (2%), SPT-window (2%) and sleep duration (2%) (Figure 1 b).

Discussion

Findings from this study indicate strong agreement at the group level and equivalence for almost all, of the PA and sleep metrics compared in this study. The only metric that demonstrated a lack of equivalence between devices was MVPA (ENMO). Whilst the LoA were wide, mean bias was less than 8% for MVPA (ENMO). This was similar to the findings for the remaining metrics with mean bias for both the PA and sleep metrics being low (below 10%). Furthermore, the LoA tended to be wide for most PA metrics apart from the intensity gradient and inactive time regardless of the cut-point used. In contrast, the LoA were narrow for all sleep metrics. These findings suggest that group, but not individual, estimates of the PA metrics could be compared between the devices although caution is advised when comparing MVPA (ENMO) estimates between devices due to the lack of equivalence observed in this study. Whereas both group and individual estimates could be compared between devices for the sleep metrics.

Equivalent estimates for AA-ENMO from the non-dominant wrist is of particular interest given the availability of ActiGraph specific ENMO based cut-points to estimate time spent in activity intensities (Hildebrand et al., 2014, 2017). Our findings are in contrast to others who demonstrated poor agreement and a lack of equivalence for AA-ENMO between the ActiGraph GT9X and wGT3X-BT devices (Clevenger et al., 2020b, 2020a; Montoye et al., 2018). In these studies, differences between devices were a consequence of larger

acceleration values produced from the GT9X relative to previous models. When comparing AA-ENMO between the GT9X and wGT3X-BT worn on the non-dominant wrist, Clevenger et al., (2020a) reported a mean differences of 14.2 m*g*. In contrast, AA-ENMO was near identical between the devices (27.9 m*g* and 27.8 m*g*) in this study. Differences in AA-ENMO were also observed from the GT9X and wGT3X-BT when worn on the hip in youths during free living (mean difference of 13 m*g)* (Clevenger et al., 2020b) and in adults who wore both devices on the hip during an 80-min semi-structured protocol (Montoye et al., 2018). Findings from the latter study should be viewed with caution however, given the short duration of the protocol which meant it was not possible to calibrate the sensor data for gravity. The main difference in this study in comparison to those reporting differences in ENMO between the GT9X and wGT3X-BT devices (Clevenger et al., 2020a, 2020b), relates to the wear time requirements. Whereas participants were instructed to wear their devices whilst sleeping in this study, Clevenger et al., (2020a, 2020b) instructed their participants to remove devices whilst sleeping. It is plausible therefore that the different wear time requirements between studies explain the contrasting AA-ENMO findings.

Previous studies have demonstrated equivalent ENMO metrics between three different accelerometer brands (GENEActiv, Axivity and ActiGraph) when worn on the non-dominant wrist using either the GT9X (Rowlands, Plekhanova, et al., 2019) or wGT3X-BT models (Buchan, 2022). These are important findings which suggest that metrics collected from the non-dominant wrist may be pooled from different studies if either of these three accelerometer brands, and ActiGraph models GT9X and wGT3X-BT, have been used. Findings from this study found that AA-ENMO and ENMO based cut-points for estimating time spent inactive and in LPA are equivalent between the GT9X and wGT3X-BT devices. This is somewhat similar to previous findings who reported equivalent estimates for a combined sedentary/light intensity category using ENMO cut-points at the non-dominant

wrist when concurrently wearing the GT9X and wGT3X-BT (Clevenger et al., 2020a). These authors also observed a lack of equivalence for moderate intensity activity when using ENMO cut-points (Clevenger et al., 2020a) whereas in this study, a lack of equivalence was evident for MVPA. When applying VM counts cut-points, we found strong agreement and equivalent outcomes for inactive time, LPA and MVPA which is consistent with the findings of Clevenger et al., (2020a). Establishing equivalence in metrics across accelerometer models is important if data is to be pooled from different studies. It is encouraging therefore to demonstrate similar findings to those previously observed from the non-dominant wrist in adults when wearing the GT9X and wGT3X-BT concurrently, for both the ENMO and VM cut-points. Indeed, the findings observed from this study will likely support those interested in pooling metrics that rely upon ENMO and VM based cut-points from studies that utilized the wGT3X-BT and GT9X.

Good agreement and equivalence were also observed for the AA-MAD metric. This suggest that when adult ActiGraph AA-MAD cut-points are available for the non-dominant wrist, comparable estimates for time spent in activity intensities may be evident between the wGT3X-BT and GT9X. Whilst others have reported that the AA-MAD metric was not equivalent between the GT9X and wGT3X-BT when worn concurrently on the non-dominant wrist (Clevenger et al., 2020a), this is likely a consequence of the 5% equivalence zone used to establish equivalence. Although the equivalence zone needed to reach equivalence is not provided, from the available data it would appear that the AA-MAD metric was very close to reaching equivalence and may be similar to that reported in this study. As the intensity gradient describes the pattern of acceleration, the good agreement and equivalence evident here is consistent with the findings of previous studies, albeit when comparisons are made across different accelerometer brands (Buchan, 2022; Buchan & Maylor, 2023; Edwardson et al., 2022).

The limitations of cut-points to estimate time spent in activity intensities are well established (Migueles, Cadenas-Sanchez, et al., 2019). It was encouraging therefore to see good agreement and equivalence between devices for the non-acceleration based metrics including total steps and counts. Measuring step counts has been of interest to researchers for decades, but it wasn't until around 2011 where interest in step counting by the general population became widespread (Bassett et al., 2017). A recent study that calculated steps using the Verisene step-count algorithm for wrist-worn accelerometers, demonstrated that more daily steps were associated with a reduced risk of all-cause mortality and cancer mortality for up to 10,000 steps/day (del Pozo Cruz et al., 2022). Another important finding reported by the authors was that no minimal threshold for the beneficial association of increasing the number of daily steps with mortality and morbidity was evident. Being able therefore to generate step counts in GGIR from wrist worn accelerometers, may allow future research to disseminate recommendations in a manner that is more easily understandable by the general public.

Whilst these findings are promising, it is important to acknowledge that recording steps can vary widely depending on the device, attachment site and the algorithm used in adults (Toth et al., 2018; Tudor-Locke et al., 2015). For instance, adults who wore twelve wearable step counter devices placed throughout their body for 1 day during free-living, found that the number of recorded steps varied from 69% to 220% of actual recorded steps (Toth et al., 2018). Limited studies have assessed the performance of the Verisense Step Count algorithm during free-living. From these studies, findings revealed a modest positive mean bias of approx. 8% in steps/day from wrist-worn Axivity and GENEActive devices against the thigh worn activPAL during free-living (Maylor et al., 2022). Moreover, a mean bias of approx. 12% was noted for step cadence during outdoor walking and running activities after applying the Verisense algorithm to wrist worn ActiGraph devices compared to the reference device,

the hip worn ActiGraph devices (Rowlands et al., 2022). Basing future evidence-based PA recommendations using daily steps may be more intuitive for the general population, but further work is needed to refine the Verisense Step Count algorithm before such recommendations are provided.

Similarly, counts demonstrated excellent agreement and equivalence between devices. Activity counts provided by ActiGraph devices has been used extensively in previous studies to assess PA behaviour and estimate time spent in activity intensities (Migueles et al., 2017). These findings will be particularly relevant for those wishing to compare new data collected from the GT9X with data collected from the wGT3X-BT to estimate time spent in activity intensities, if processed in GGIR using the actilifecounts external function algorithm. Moreover, sleep metrics demonstrated excellent agreement and equivalence between devices. These observations are in line with recent findings which demonstrated equivalent estimates for sleep onset, waking time, SPT-window and sleep duration when applying the automated HDCZA algorithm to the GT9X or Axivity and comparing metrics against PSG (Plekhanova et al., 2022; Van Hees et al., 2018). The findings reported here demonstrate minimal differences for sleep onset, waking time, SPT-window and sleep duration, and shows that the automated HDCZA algorithm performs similarly when applied to the GT9X or wGT3X-BT. Overall, these findings provide support for the cross-generational comparability of the wGT3X-BT and GT9X devices but is dependent upon the outcome being compared.

There are some limitations to the present study. Agreement and equivalence of activity and sleep metrics were only compared between the ActiGraph GT9X and wGT3X-BT devices during free-living, and not against a criterion measure. Moreover, our findings are limited to the specific population used in the study which hinders the generalizability of the findings to other populations. A strength of this study includes the comparison of two devices from the most widely used accelerometer brand which were worn concurrently, with data processed

identically. A particular strength of the study relates to the findings of the non-acceleration-based metrics counts and steps which are reported for the first time. As is the comparison of sleep metrics between the ActiGraph GT9X and wGT3X-BT devices.

Conclusions

In summary, equivalent estimates of most PA and all sleep metrics are provided from the GT9X and wGT3X-BT when worn on the non-dominant wrist. Nonetheless, caution is advised comparing MVPA estimates between devices due to the lack of equivalence observed in this study when applying ENMO based cut-points.

Acknowledgements

We would like to thank the participants for their involvement in the study.

References

ActiGraph. (2013). *GT3X+ and wGT3X+ Device Manual*. https://s3.amazo naws.com/actigraphcorp.com/wp-content/uploads/2018/02/22094126/ GT3X-wGT3X-Device-Manual-110315.pdf

ActiGraph. (2019). *ActiGraph. User guide: ActiGraph GT9X link + ActiLife.* https://s3.amazonaws.com/actigraphcorp.com/ wp-content/uploads/2019/02/04090343/ActiGraph_Link_ UserGuide_02012019_Revision3.pdf

Alley, S. J., Uffelen, J. van, Schoeppe, S., Parkinson, L., Hunt, S., Power, D., Waterman, N., Waterman, C., To, Q. G., Duncan, M. J., Schneiders, A., & Vandelanotte, C. (2022). The Effectiveness of a Computer-Tailored Web-Based Physical Activity Intervention Using Fitbit Activity Trackers in Older Adults (Active for Life): Randomized Controlled Trial. *Journal of Medical Internet Research*, *24*(5), e31352. https://doi.org/10.2196/31352

Bassett, D. R., Toth, L. P., LaMunion, S. R., & Crouter, S. E. (2017). Step Counting: A Review of Measurement Considerations and Health-Related Applications. *Sports Medicine (Auckland, N.Z.)*, *47*(7), 1303–1315. https://doi.org/10.1007/s40279-016-0663-1

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.

Buchan, D. S. (2022). Equivalence of activity outcomes derived from three research grade accelerometers worn simultaneously on each wrist. *Journal of Sports Sciences*, *40*(7), 797–807. https://doi.org/10.1080/02640414.2021.2019429

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*, *24*(4), 247–257. https://doi.org/10.1080/1091367X.2020.1801441

Buchan, D. S., & Maylor, B. D. (2023). Comparison of physical activity metrics from two research-grade accelerometers worn on the non-dominant wrist and thigh in children. *Journal of Sports Sciences*, *41*(1), 80–88. https://doi.org/10.1080/02640414.2023.2197726

Choi, L., Ward, S. C., Schnelle, J. F., & Buchowski, M. S. (2012). Assessment of Wear/Nonwear Time Classification Algorithms for Triaxial Accelerometer. *Medicine and Science in Sports and Exercise*, *44*(10), 2009–2016. https://doi.org/10.1249/MSS.0b013e318258cb36

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

Clevenger, K. A., Pfeiffer, K. A., & Montoye, A. H. K. (2020b). Cross-Generational Comparability of Raw and Count-Based Metrics from ActiGraph GT9X and wGT3X-BT Accelerometers during Free-Living in Youth. *Measurement in Physical Education and Exercise Science*, *24*(3), 194–204. https://doi.org/10.1080/1091367X.2020.1773827

del Pozo Cruz, B., Ahmadi, M. N., Lee, I.-M., & Stamatakis, E. (2022). Prospective Associations of Daily Step Counts and Intensity With Cancer and Cardiovascular Disease Incidence and Mortality and All-Cause Mortality. *JAMA Internal Medicine*, *182*(11), 1139–1148. https://doi.org/10.1001/jamainternmed.2022.4000

Edwardson, C. L., Maylor, B. D., Dawkins, N. P., Plekhanova, T., & Rowlands, A. V. (2022). 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*, *26*(1), 39–50. https://doi.org/10.1080/1091367X.2021.1944154

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), Article 12. https://doi.org/10.1111/sms.12795

Hildebrand, M., van Hees, V. T., 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), Article 9. https://doi.org/10.1249/MSS.0000000000000289

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

Maylor, B. D., Edwardson, C. L., Dempsey, P. C., Patterson, M. R., Plekhanova, T., Yates, T., & Rowlands, A. V. (2022). Stepping towards More Intuitive Physical Activity Metrics with Wrist-Worn Accelerometry: Validity of an Open-Source Step-Count Algorithm. *Sensors*, *22*(24), Article 24. https://doi.org/10.3390/s22249984

McDonough, D. J., Helgeson, M. A., Liu, W., & Gao, Z. (2022). Effects of a remote, YouTube-delivered exercise intervention on young adults' physical activity, sedentary behavior, and sleep during the COVID-19 pandemic: Randomized controlled trial. *Journal of Sport and Health Science*, *11*(2), 145–156. https://doi.org/10.1016/j.jshs.2021.07.009

Migueles, J. H. (2022). *Actilifecounts*. https://github.com/jhmigueles/actilifecounts

Migueles, J. H., Cadenas-Sanchez, C., Ekelund, U., Delisle Nyström, C., Mora-Gonzalez, J., Löf, M., Labayen, I., Ruiz, J. R., & Ortega, F. B. (2017). Accelerometer Data

Collection and Processing Criteria to Assess Physical Activity and Other Outcomes: A Systematic Review and Practical Considerations. *Sports Medicine (Auckland, N.Z.)*, *47*(9), Article 9. https://doi.org/10.1007/s40279-017-0716-0

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), Article 4. https://doi.org/10.1111/sms.13356

Migueles, J. H., Rowlands, A. V., Huber, F., Sabia, S., & Hees, V. T. van. (2019). 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., Clevenger, K. A., Pfeiffer, K. A., Nelson, M. B., Bock, J. M., Imboden, M. T., & Kaminsky, L. A. (2020). Development of cut-points for determining activity intensity from a wrist-worn ActiGraph accelerometer in free-living adults. *Journal of Sports Sciences*, *38*(22), 2569–2578. https://doi.org/10.1080/02640414.2020.1794244

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 and Science in Sports and Exercise*, *50*(5), 1103–1112. https://doi.org/10.1249/MSS.0000000000001534

Nathan, N., McCarthy, N., Hope, K., Sutherland, R., Lecathelinais, C., Hall, A., Lane, C., Trost, S., Yoong, S. L., & Wolfenden, L. (2021). The impact of school uniforms on

primary school student's physical activity at school: Outcomes of a cluster randomized controlled trial. *International Journal of Behavioral Nutrition and Physical Activity*, *18*(1), 17. https://doi.org/10.1186/s12966-021-01084-0

Neishabouri, A., Nguyen, J., Samuelsson, J., Guthrie, T., Biggs, M., Wyatt, J., Cross, D., Karas, M., Migueles, J. H., Khan, S., & Guo, C. C. (2022). Quantification of acceleration as activity counts in ActiGraph wearable. *Scientific Reports*, *12*(1), 11958. https://doi.org/10.1038/s41598-022-16003-x

Patterson, M. (2021). *Verisense-Toolbox/Verisense_step_algorithm at master · ShimmerEngineering/Verisense-Toolbox*. GitHub. https://github.com/ShimmerEngineering/Verisense-Toolbox

Plekhanova, T., Rowlands, A. V., Davies, M. J., Hall, A. P., Yates, T., & Edwardson, C. L. (2022). Validation of an automated sleep detection algorithm using data from multiple accelerometer brands. *Journal of Sleep Research*, *n/a*(n/a), e13760. https://doi.org/10.1111/jsr.13760

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., Maylor, B., Dawkins, N. P., Dempsey, P. C., Edwardson, C. L., Soczawa-Stronczyk, A. A., Bocian, M., Patterson, M. R., & Yates, T. (2022). Stepping up with GGIR: Validity of step cadence derived from wrist-worn research-grade accelerometers using the verisense step count algorithm. *Journal of Sports Sciences*, *40*(19), 2182–2190. https://doi.org/10.1080/02640414.2022.2147134

Rowlands, A. V., Plekhanova, T., Yates, T., Mirkes, E. M., Davies, M., Khunti, K., & Edwardson, C. L. (2019). 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., Sherar, L. B., Fairclough, S. J., Yates, T., Edwardson, C. L., Harrington, D. M., Davies, M. J., Munir, F., Khunti, K., & Stiles, V. H. (2019). A data-driven, meaningful, easy to interpret, standardised accelerometer outcome variable for global surveillance. *Journal of Science and Medicine in Sport*. https://doi.org/10.1016/j.jsams.2019.06.016

Shah, V. V., Brumbach, B. H., Pearson, S., Vasilyev, P., King, E., Carlson-Kuhta, P., Mancini, M., Horak, F. B., Sowalsky, K., McNames, J., & El-Gohary, M. (2023). Opal Actigraphy (Activity and Sleep) Measures Compared to ActiGraph: A Validation Study. *Sensors (Basel, Switzerland)*, *23*(4), 2296. https://doi.org/10.3390/s23042296

Toth, L. P., Park, S., Springer, C. M., Feyerabend, M. D., Steeves, J. A., & Bassett, D. R. (2018). Video-Recorded Validation of Wearable Step Counters under Free-living Conditions. *Medicine & Science in Sports & Exercise*, *50*(6), 1315. https://doi.org/10.1249/MSS.0000000000001569

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

Tudor-Locke, C., Barreira, T. V., & Schuna, J. M. (2015). Comparison of step outputs for waist and wrist accelerometer attachment sites. *Medicine and Science in Sports and Exercise*, *47*(4), 839–842. https://doi.org/10.1249/MSS.0000000000000476

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). Autocalibration of

accelerometer data for free-living physical activity assessment using local gravity and temperature: An evaluation on four continents. *Journal of Applied Physiology (Bethesda, Md.: 1985)*, *117*(7), 738–744. https://doi.org/10.1152/japplphysiol.00421.2014

Van Hees, V. T., Sabia, S., Jones, S. E., Wood, A. R., Anderson, K. N., Kivimäki, M., Frayling, T. M., Pack, A. I., Bucan, M., Trenell, M. I., Mazzotti, D. R., Gehrman, P. R., Singh-Manoux, B. A., & Weedon, M. N. (2018). Estimating sleep parameters using an accelerometer without sleep diary. *Scientific Reports*, *8*(1), 12975. https://doi.org/10.1038/s41598-018-31266-z

van Hees, V. T. van, Gorzelniak, L., León, E. C. D., 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), Article 4. https://doi.org/10.1371/journal.pone.0061691

Wellek, S. (2003). In *Testing Statistical Hypotheses of Equivalence* (pp. xv, 284 p.). Chapman & Hall/CRC. 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

Figure 1. Equivalence between GT9X minus GT3X- BT monitors worn at the non-dominant wrist for (a) Physical Activity metrics and (b) Sleep metrics. Dashed lines represent the 10% equivalence zone. Equivalence = 1 (solid line). The horizontal lines represent the 95% confidence interval of the ratio. Counts were estimated using the actilifecounts R package. VM cut-points were estimated using ActiLife v6.13.3.

Figure 2. Bland-Altman plot displaying mean bias and 95% limit of agreement level for PA metrics between GT9X minus GT3X- BT monitors worn at the non-dominant wrist. Counts were estimated using the actilifecounts R package. VM cut-points were estimated using ActiLife v6.13.3.

Figure 3. Bland-Altman plot displaying mean bias and 95% limit of agreement level for sleep metrics between GT9X minus GT3X- BT monitors worn at the non-dominant wrist.

Table 1. Descriptive statistics of physical activity and sleep metrics from ActiGraph devices (N = 30).

Physical Activity metric	**wGT3X-BT**	**GT9X**
AA-ENMO (mg/day)	27.9 ± 7.2	27.8 ± 8.1
Intensity Gradient (mg/day)	-2.58 ± 0.2	-2.56 ± 0.2
AA-MAD (mg/day)	40.7 ± 11.7	41.4 ± 11.4
Counts (counts·min^{-1})[†]	79.32 ± 45.7	81.9 ± 26.5
Total steps (steps)	54328 ± 30183	54930 ± 30715
M10 (mg)	256.7 ± 96.6	262.9 ± 92.3
M30 (mg)	171.2 ± 54.1	178.1 ± 55.7
ENMO cut-points		
Inactive time (min/day)	1141.5 ± 92.4	1137.2 ± 105.7
LPA (min/day)	203.3 ± 59.3	200.1 ± 71.1
MVPA (min/day)	96.6 ± 46.8	104.2 ± 49.3
VM counts cut-points[‡]		
Inactive time (min/day)	1053.2 ± 121.3	1055.4 ± 131.1
LPA (min/day)	197.6 ± 75.5	200.1 ± 76.1
MVPA (min/day)	104.9 ± 62.8	104.5 ± 60.1
Sleep metric		
Sleep onset, clock time (Hr ± min)	23:34 pm. ± 88.7	23.33 pm. ± 87.5
Waking time, clock time (Hr ± min)	08:06 am. ± 92.2	08:10 am. ± 95.2
SPT-window (Hr ± min)	8.5 ± 73	8.6 ± 73
Sleep duration (Hr ± min)	7.2 ± 58	7.2 ± 57

Data are presented as mean ± SD. Hr, hour. AA-ENMO, average acceleration Euclidean Norm Minus One; AA-MAD, average acceleration Mean Amplitude Deviation; M10, the minimum acceleration value of the most active 10 mins; M30 the minimum acceleration value of the most

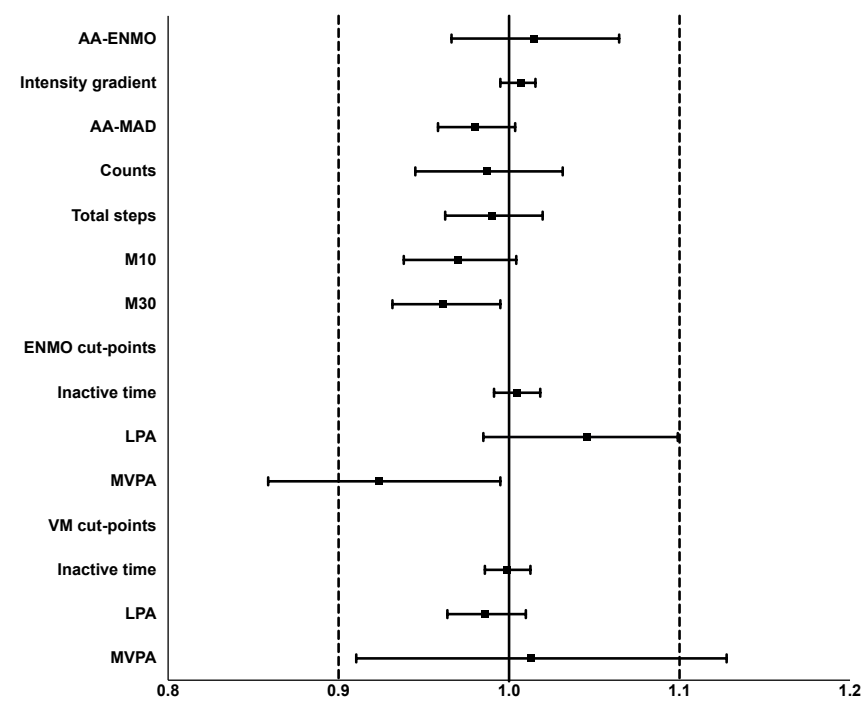
active 30 mins; LPA, light physical activity; MVPA, moderate-vigorous physical activity; SPT-window, sleep period time window. † Estimated using the actilifecounts R package. ‡ Estimated using ActiLife v6.13.3.

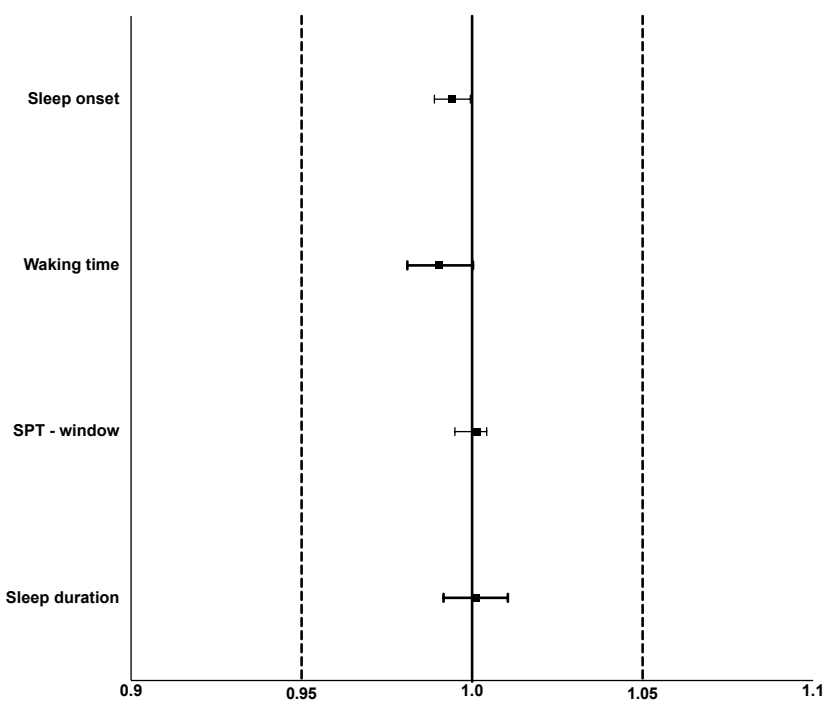
Table 2. Agreement, intraclass correlations and equivalence zones between the GT9X and wGT3X-BT for physical activity and sleep metrics

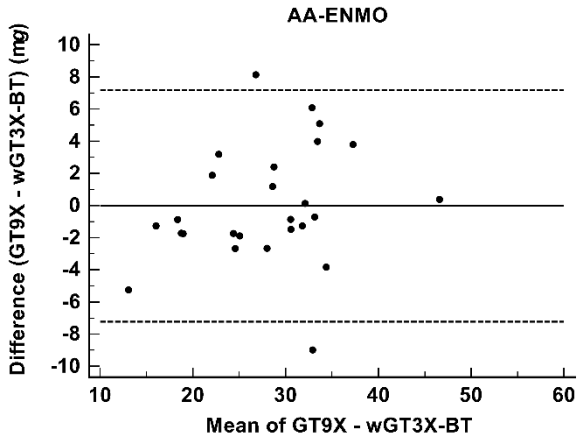
Physical Activity metric	Mean bias (95% LoA)	ICC (95%CI)	Equivalence zone, %
AA-ENMO (mg/day)	0 (7.2)	0.89 (0.77 to 0.95)	7
Intensity Gradient (mg/day)	0 (0.16)	0.96 (0.91 to 0.98)	2
AA-MAD (mg/day)	0.7 (6.7)	0.97 (0.93 to 0.98)	5
Counts (counts·min^{-1})[†]	2.6 (21.6)	0.92 (0.83 to 0.96)	5
Total steps (steps)	937 (7580)	0.99 (0.98 to 0.99)	4
M10 (mg)	6.3 (54.4)	0.97 (0.93 to 0.98)	7
M30 (mg)	6.9 (38.9)	0.95 (0.88 to 0.98)	7
ENMO cut-points			
Inactive time (min/day)	-4.3 (89.4)	0.91 (0.80 to 0.96)	2
LPA (min/day)	-3.3 (60.2)	0.91 (0.80 to 0.97)	10
MVPA (min/day)	7.6 (51.2)	0.86 (0.76 to 0.95)	16
VM counts cut-points[‡]			
Inactive time (min/day)	2.2 (81.7)	0.95 (0.89 to 0.98)	2
LPA (min/day)	2.4 (31.7)	0.98 (0.96 to 0.99)	4
MVPA (min/day)	-0.4 (71.3)	0.83 (0.66 to 0.92)	9
Sleep metric			
Sleep onset (min)	-1.8 (30)	0.99 (0.98 to 0.99)	1
Waking time (min)	6 (18)	0.99 (0.98 to 0.99)	2
SPT-window (min)	6 (36)	0.98 (0.95 to 0.99)	2
Sleep duration (min)	-0.6 (24)	0.98 (0.95 to 0.99)	2

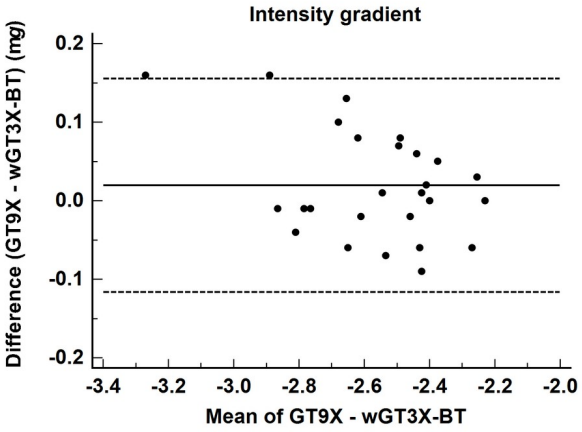
LoA, Limits of Agreement; ICC, intraclass Correlation Coefficients; CI, confidence intervals; Hr, hour GT9X vs wGT3X-BT. AA-ENMO, average acceleration Euclidean Norm Minus One; AA-MAD, average acceleration Mean Amplitude Deviation; M10, the minimum acceleration value of the most active 10 mins; M30 the minimum acceleration value of the most active 30 mins; LPA, light physical activity; MVPA,

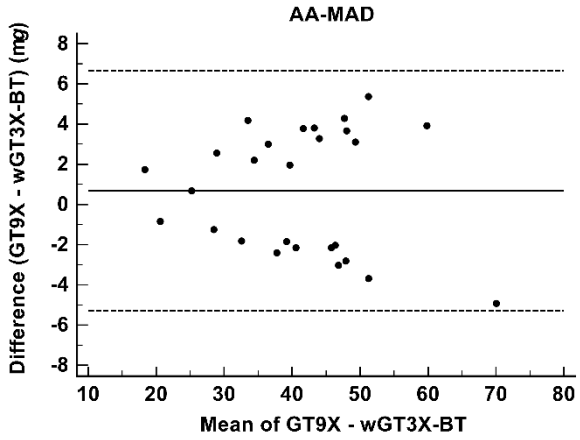
moderate-vigorous physical activity; SPT-window, sleep period time window. † Estimated using the actilifecounts R package. ‡ Estimated using ActiLife v6.13.3.

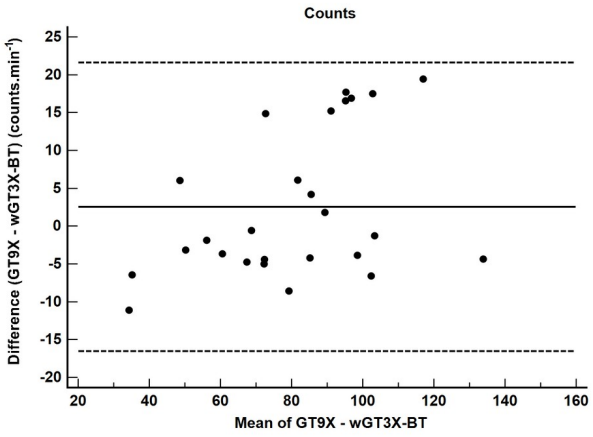

AA-ENMO
Intensity gradient
AA-MAD
Counts
Total steps
M10
M30
ENMO cut-points
Inactive time
LPA
MVPA
VM cut-points
Inactive time
LPA
MVPA
0.8
0.9
1.0
1.1
1.2

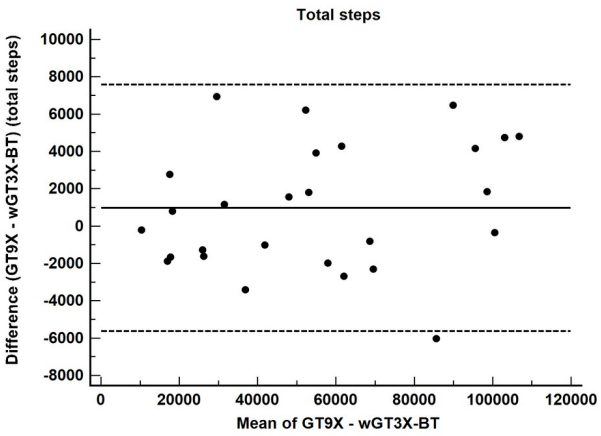

Sleep onset
Waking time
SPT - window
Sleep duration
0.9
0.95
1.0
1.05
1.1

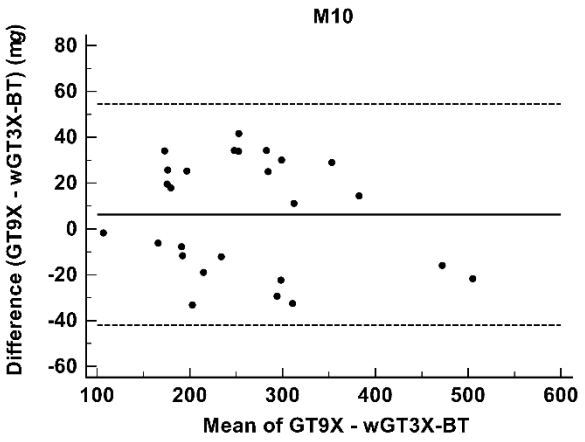

AA-ENMO
Difference (GT9X - wGT3X-BT) (mg)
Mean of GT9X - wGT3X-BT

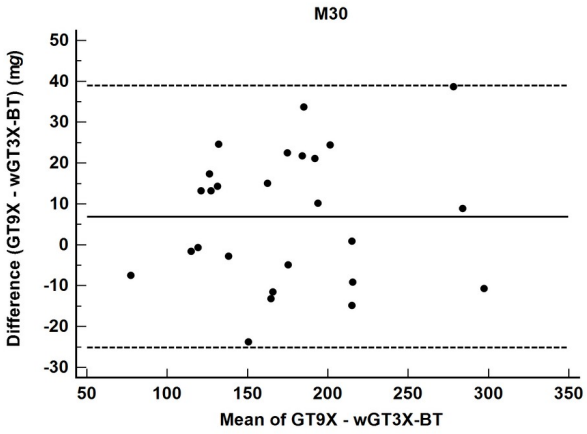

Intensity gradient
Difference (GT9X - wGT3X-BT) (mg)
Mean of GT9X - wGT3X-BT

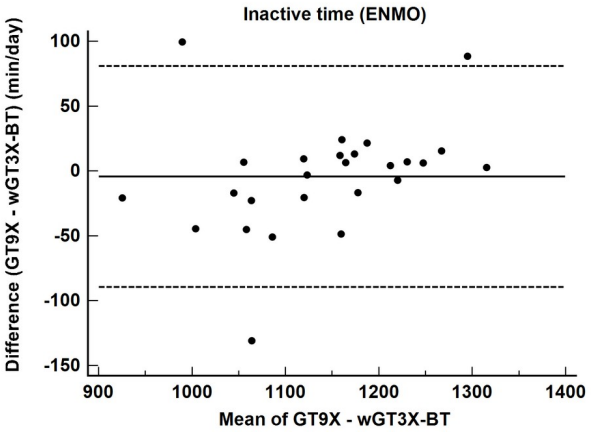

AA-MAD
Difference (GT9X - wGT3X-BT) (mg)
Mean of GT9X - wGT3X-BT

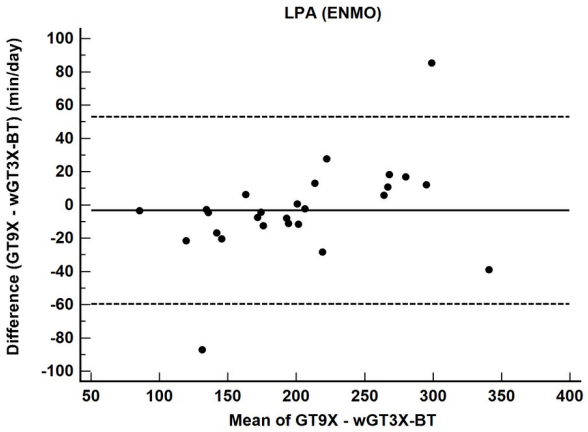

Counts
Difference (GT9X - wGT3X-BT) (counts.min-1)
Mean of GT9X - wGT3X-BT

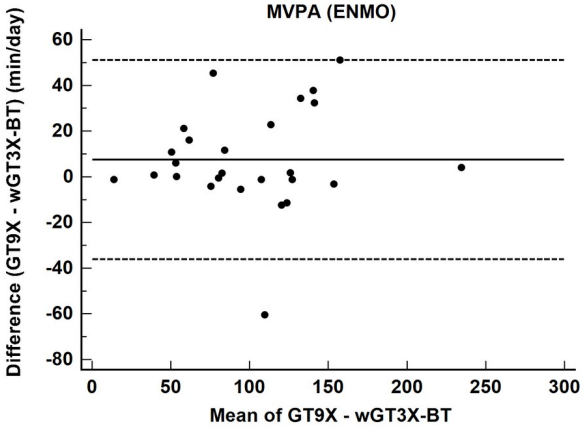

Total steps
Difference (GT9X - wGT3X-BT) (total steps)
Mean of GT9X - wGT3X-BT

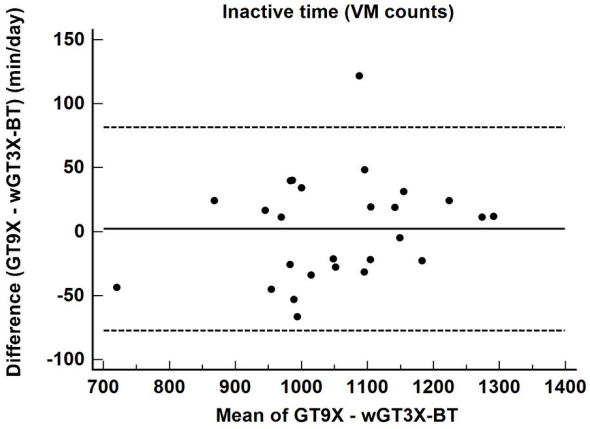

M10
Difference (GT9X - wGT3X-BT) (mg)
Mean of GT9X - wGT3X-BT

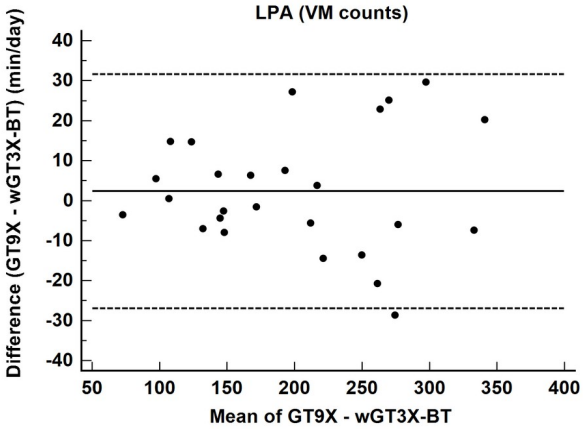

M30
Difference (GT9X - wGT3X-BT) (mg)
Mean of GT9X - wGT3X-BT

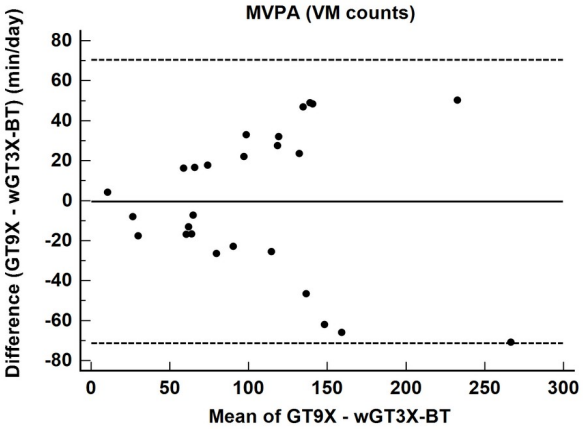

Inactive time (ENMO)
Difference (GT9X - wGT3X-BT) (min/day)
Mean of GT9X - wGT3X-BT

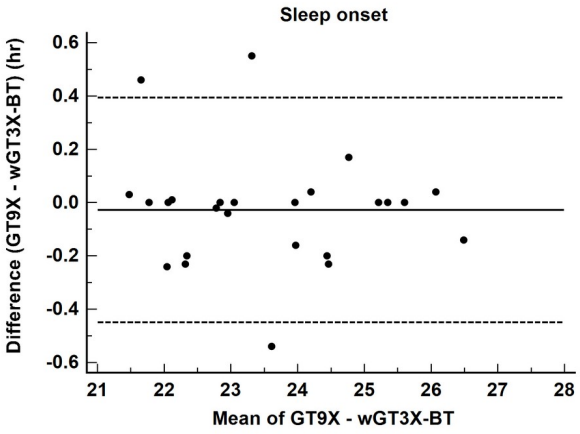

LPA (ENMO)
Difference (GT9X - wGT3X-BT) (min/day)
Mean of GT9X - wGT3X-BT

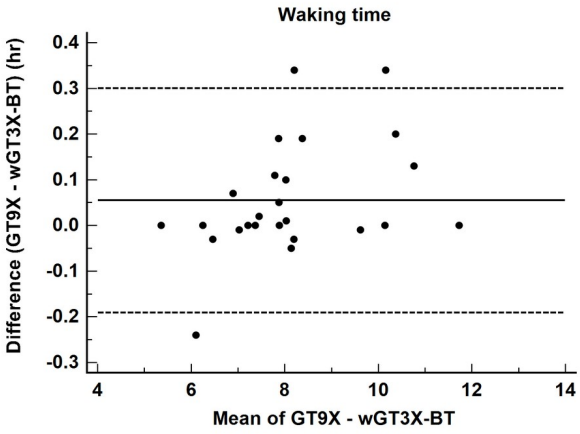

MVPA (ENMO)
Difference (GT9X - wGT3X-BT) (min/day)
Mean of GT9X - wGT3X-BT

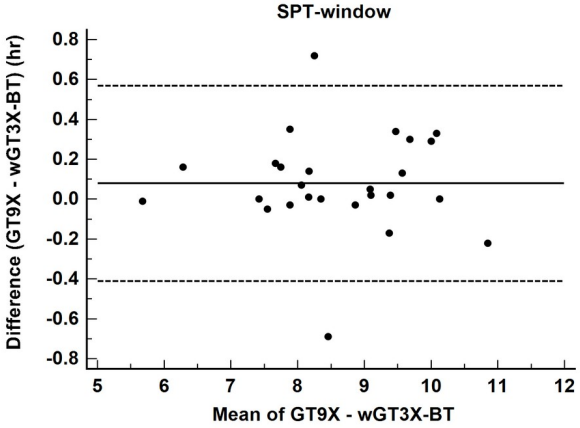

Inactive time (VM counts)
Difference (GT9X - wGT3X-BT) (min/day)
Mean of GT9X - wGT3X-BT

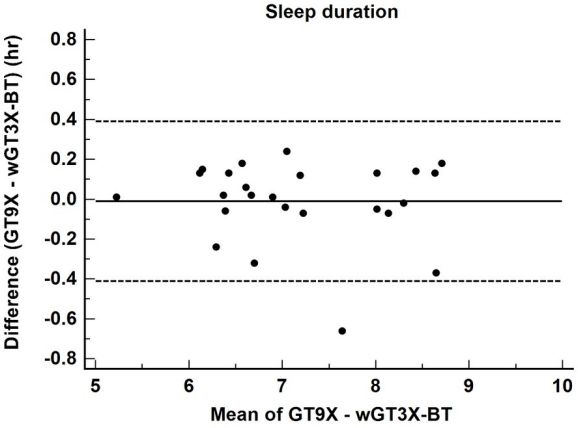

LPA (VM counts)
Difference (GT9X - wGT3X-BT) (min/day)
Mean of GT9X - wGT3X-BT


MVPA (VM counts)
Difference (GT9X - wGT3X-BT) (min/day)
Mean of GT9X - wGT3X-BT


Sleep onset
Difference (GT9X - wGT3X-BT) (hr)
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
21
22
23
24
25
26
27
28
Mean of GT9X - wGT3X-BT


Waking time
Difference (GT9X - wGT3X-BT) (hr)
0.4
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
4
6
8
10
12
14
Mean of GT9X - wGT3X-BT


SPT-window
Difference (GT9X - wGT3X-BT) (hr)
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
5
6
7
8
9
10
11
12
Mean of GT9X - wGT3X-BT


Sleep duration
Difference (GT9X - wGT3X-BT) (hr)
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
5
6
7
8
9
10
Mean of GT9X - wGT3X-BT