Original Paper

Absolute Agreement Between a Consumer Wearable and Research Grade Accelerometry in Free-Living Conditions: Comparison Study

Karianne Hagerupsen[1], MSc; André Henriksen[2], PhD; Bente Morseth[3], PhD; Sameline Grimsgaard[1], MD, PhD; Jonas Johansson[1], PhD

[1]Department of Community Medicine, UiT The Arctic University of Norway, Tromsø, Norway
[2]Department of Computer Science, UiT The Arctic University of Norway, TROMSØ, Troms, Norway
[3]School of Sport Sciences, UiT The Arctic University of Norway, TROMSØ, Troms, Norway

Corresponding Author:

Karianne Hagerupsen, MSc
Department of Community Medicine
UiT The Arctic University of Norway
Hansine Hansens veg 78
Tromsø 9019
Norway
Phone: 47 77646088
Email: karianne.hagerupsen@uit.no

Abstract

Background: Research-grade accelerometers are widely used to monitor physical activity in research but are often expensive, have limited battery and storage, and impose a high participant burden. Consumer wearables are designed for long-term monitoring and are being extensively used in research. However, evidence regarding their accuracy remains inconsistent.

Objective: This study aimed to examine agreement between the Garmin Forerunner 55 and the ActiGraph wGT3X-BT for steps, moderate-to-vigorous physical activity (MVPA), and activity energy expenditure (AEE) in older adults under free-living conditions.

Methods: We included baseline data from 82 community-dwelling older adults (60-75 years) in the RESTART (Re-Inventing Strategies for Healthy Ageing; Recommendations and Tools) trial who wore the ActiGraph wGT3X-BT and the Garmin Forerunner 55 simultaneously for 3 to 7 days. Valid days were defined as ≥10 hours of wear time per day. We compared mean steps, MVPA, and AEE between devices. Agreement was examined using intraclass correlation coefficients (ICCs) and Bland-Altman plots.

Results: The Garmin Forerunner 55 recorded 990 (limits of agreement [LoA] –1123 to 3104) steps/day more than the ActiGraph wGT3X-BT with an ICC of 0.81, indicating good agreement but with substantial uncertainty due to the wide CI (95% CI 0.34-0.92). For AEE, the Garmin device reported 41.8 kcal/day lower values, with poor agreement (ICC=0.31; LoA –582.9 to 499.4 kcal/day). MVPA showed the largest discrepancy, with the Garmin device recording 23 minutes/day less and poor agreement (ICC=0.17; LoA –86.7 to 40.6 minutes/day). Bland-Altman plots indicated proportional bias across all metrics.

Conclusions: In older adults with low activity levels, the Garmin Forerunner 55 demonstrated good agreement for daily step count but poor agreement for AEE and MVPA compared with the ActiGraph wGT3X-BT. Consumer devices such as the Garmin Forerunner 55 may be suitable for the monitoring of step-based metrics at the group level, but intensity-based measures should be interpreted with caution.

Trial Registration: ClinicalTrials.gov NCT06122441; https://clinicaltrials.gov/study/NCT06122441

JMIR Form Res 2026;10:e105610; doi: 10.2196/105610

Keywords: physical activity; accelerometery; consumer wearables; older adults; absolute agreement

Introduction

Physical inactivity is a major global health challenge and is now the fourth leading risk factor of worldwide mortality. It contributes substantially to the burden of noncommunicable diseases, such as cardiovascular disease and diabetes [1]. The current global recommendations for physical activity advise adults to achieve at least 150 minutes of moderate-intensity aerobic exercise per week [2]. However, a large proportion fails to meet the guidelines and approximately one-third of the world's adult population is physically inactive [3]. Objective methods that accurately capture physical activity in free-living conditions and can be applied easily and cost-effectively are therefore essential for understanding population-level activity patterns and for evaluating interventions aimed at increasing physical activity [4].

Widely used research-grade accelerometers such as the ActiGraph have been applied in both clinical trials and large observational studies, such as the National Health and Nutrition Examination Survey (NHANES) [5] and the Tromsø Study [6,7]. These devices provide high-resolution movement data and are often considered the criterion standard for objective assessment of physical activity [8,9]. However, research-grade instruments are often expensive, have limited battery-life and storage capacity, and impose a relatively high participant burden [10]. These limitations may restrict their usage in large-scale or long-term studies. In addition, such devices are generally designed for research purposes and provide no immediate user feedback, such as step counts, limiting their value outside research settings [11].

In contrast, consumer wearables have grown in popularity and are increasingly used in both research and everyday life to monitor and motivate physical activity behavior and other health-related measures [12,13]. These devices are affordable, widely available, and therefore accessible to a wide range of people, in addition to being designed to promote continuous wear and thus allowing long-term collection of data [14]. Moreover, most activity trackers are user-friendly and can collect a spectrum of data, including steps, physical activity intensity, heart rate, and energy expenditure [15,16]. As a result, consumer wearables have the potential to complement, and, in some settings, replace research-grade devices. Larger population studies and long-term clinical studies are particularly relevant study designs in which consumer wearables come forward as high-potential monitoring tools [17,18]. Recent reviews have highlighted a rapid expansion in the use of consumer wearables across epidemiological and clinical studies, reflecting their growing methodological importance in health research [14,19].

However, such systematic large-scale application of consumer wearables in research is hampered by the limited and inconsistent evidence regarding their accuracy compared with research-grade devices [20,21]. New models and firmware updates are frequently released, and device algorithms are often proprietary and subject to change [8,22]. As the use of these devices continues to expand, the accuracy of wearable device metrics becomes essential. Consequently, this highlights the need for rigorous scientific evaluation of newly released devices and their agreement with research-grade devices under free-living conditions [20,23].

The aim of this study was therefore to investigate the agreement between the commercially available and previously nonevaluated wrist-worn Garmin Forerunner 55 and the widely used ActiGraph wGT3X-BT accelerometer in terms of steps per day, moderate-to-vigorous physical activity (MVPA), and activity energy expenditure (AEE) in older adults under free-living conditions.

Methods

Study Sample

This study is a part of the RESTART (Re-Inventing Strategies for Healthy Ageing; Recommendations and Tools) randomized controlled trial (ClinicalTrials.gov ID NCT06122441), a complex intervention aiming to achieve lasting lifestyle changes in high-risk older adults. Participation eligibility and exclusion criteria for the RESTART trial have been described elsewhere [24]. A total of 544 individuals from the seventh survey of the Tromsø Study (Tromsø7, 2015-16) [25] were invited, of whom 75 responded. Another 115 individuals responded to a local newspaper advertisement, resulting in 190 individuals residing in Tromsø municipality being screened for eligibility. Based on the screening interview, 71 individuals were excluded. A total of 119 individuals attended baseline screening and provided informed consent. Of these, 8 were excluded, as 6 did not meet inclusion criteria, 1 withdrew after the baseline visit, and 1 was lost to follow-up before the baseline visit at the testing laboratory. The present study focused on evaluating baseline data for the 111 participants recruited for the RESTART trial, who were included if they were aged 60 to 75 years and had a BMI ≥28 kg/m^2, a self-reported sedentary lifestyle, and increased risk of cardiovascular disease. Baseline data were collected from January until June 2024.

Instruments

Garmin Forerunner 55

The Garmin Forerunner 55, released in 2021, is a GPS sport watch equipped with a wrist-based optical heart rate sensor and an integrated accelerometer for physical activity monitoring. The device measures 42×42×11.6 mm, weighs 37 g, and provides up to 14 days of battery time.

Participants paired the device with their smartphone and were instructed to synchronize data weekly through the Garmin Connect mobile app. We used the mSpider experimental data collection tool to collect and store activity data from Garmin Connect [26]. The Garmin Forerunner 55 exports data for steps, moderate- and vigorous-intensity minutes, and AEE. The precise methods behind these measures rely on proprietary algorithms, which are not publicly disclosed. It is also unclear when these algorithms are updated or modified and how this may affect device features [16,27]. At least one firmware update occurred during the baseline data collection.

ActiGraph wGT3X-BT

The ActiGraph wGT3X-BT (Ametris LLC) is a triaxial accelerometer weighing 19 g, with dimensions of 4.6×3.3×1.5 cm and a 30-100 Hz sampling rate, and it has up to 25 days of battery life.

We used the ActiLife software (version 6.13.6; Ametris LLC) to export ActiGraph accelerometer data and generate variables. Triaxial activity counts were used to generate physical activity and energy expenditure variables in addition to directly reported step counts. Minutes of MVPA were calculated by the widely used cut points defined by Sasaki et al [28]. The Freedson Adult VM3 equation was used to calculate AEE [29]. These methods are commonly used in studies involving both adults and community-dwelling older adults, including community-dwelling adults aged 60 to 70 years [30-32]. We therefore selected these cut points and equations to enhance comparability with the accelerometry literature. This approach is important because the Garmin device does not disclose its proprietary algorithms, making the use of established accelerometry methods relevant for meaningful comparison.

We defined valid days of measurement as containing at least 10 hours of wear time. We analyzed wear time for the ActiGraph device by using the ActiLife default setting for the wear time algorithm by Choi et al [33]. The wear time algorithm for Garmin Forerunner 55 is not known but we inspected minute-level heart rate data for all days included in the analysis. Days with at least 10 hours of heart rate data during daytime were considered valid days. No days with insufficient heart rate coverage were present in the dataset.

Procedure

For each participant, the Garmin Forerunner 55 was initialized using objectively measured weight and height, along with birth year and sex. Participants were instructed to wear the Garmin device on their nondominant wrist. The ActiGraph recording was set at 100 Hz and placed on the right hip. Participants received both written and verbal instructions on how to wear the devices and were asked to wear both devices simultaneously for 7 days. Participants were instructed to always wear the devices during daytime, except for temporarily removing the ActiGraph device during water-related activities and showering. Participants were not required to wear the devices while sleeping. All participants with ≥3 valid days of data on both devices were included in the analyses.

Statistical Analyses

All analyses were performed using Stata (version 18; StataCorp LLC). We considered the ActiGraph wGT3X-BT as the reference monitor for steps, MVPA, and AEE due to its wide use and validity in measuring physical activity in free-living settings [34,35]. The number of steps per day, AEE, and minutes of MVPA were averaged over valid days of recording to reduce the influence of day-to-day variation and ensure equal contributions from the participants regardless of valid days. The Shapiro-Wilk test with inspection of Q-Q plots was used to assess whether data were normally distributed. Outliers were identified using z scores, with observations exceeding a z score of 3 classified as extreme values. One Garmin MVPA value met this criterion (z=7.06) and was excluded from the MVPA analysis. No outliers were detected for steps or AEE.

Agreement between the Garmin and ActiGraph devices was evaluated using the intraclass correlation coefficient (ICC), with 95% CIs of the ICC estimate to indicate poor (<0.5), moderate (0.5-0.75), good (0.75-0.9), and excellent (>0.9) agreement [36]. For each outcome variable—daily step count, AEE, and MVPA—we applied a 2-way mixed-effects model with absolute agreement for single measures (ICC[3,1]). In this model, participants are treated as random effects and the two devices as fixed effects. The ICC(3,1) therefore quantifies the extent to which the two devices provide interchangeable values for each outcome at the individual measurement level. Because the ICC is estimated based on participant-level average values, it quantifies between device agreement and not day-to-day variation.

In addition to the ICC analyses, agreement between the Garmin and ActiGraph devices was further examined using Bland-Altman plots. For each outcome variable, we calculated the mean difference between devices and the 95% limits of agreement (LoA; mean difference ± 1.96 SD). This approach provides a visual and quantitative assessment of systematic bias and the range within which individual differences are expected to lie [37].

Ethical Considerations

This study was conducted according to the Declaration of Helsinki. The study was approved by the Norwegian Regional Committees for Medical and Health Research Ethics North (approval number 584841). All participants received written and verbal information and gave informed consent. The participants were also informed of their right to withdraw from the study at any time without giving any reason. Prior to analysis, we deidentified data and no identifiable information was used in the dataset.

Results

Participant characteristics are presented in Table 1. Out of 111 participants, 82 had valid recordings on both devices simultaneously and were included in the final analyses. Out of the 29 participants excluded from the analyses, 24 participants did not provide data from the Garmin Forerunner 55, two participants did not provide data from the ActiGraph wGT3X-BT, and four participants did not provide data from either device. Each participant had between 3 to 7 valid days of recording and the average number of valid days on both devices was 6.7 (SD 0.8) days. Table 2 displays the mean number of steps, AEE, and MVPA from the Garmin Forerunner 55 and ActiGraph wGT3X-BT. Table 3 presents ICCs and mean differences with LoA for steps, AEE, and MVPA measures for both devices. Mean differences are

further illustrated in Figure 1 for steps, Figure 2 for AEE, and Figure 3 for MVPA.

Table 1. Participant characteristics (N=82).

Variable	Value	Range
Age (years), mean (SD)	64.7 (4.0)	60-75
Female, n (%)	35.0 (42.7)	—[a]
Height (cm), mean (SD)	172.7 (9.5)	143.1-191
Weight (kg), mean (SD)	98.4 (15.2)	68.5-137
BMI (kg/m^2), mean (SD)	33.0 (4.1)	27.7-45.2

[a]Not applicable.

Table 2. Physical activity metrics from the Garmin Forerunner 55 and ActiGraph wGT3X-BT devices.

	Garmin Forerunner 55		ActiGraph wGT3X-BT	
	Mean (SD)	Range	Mean (SD)	Range
Steps (per day)	6285 (2414)	1599-12836	5295 (2131)	1448-11283
AEE[a] (kcal/day)	482.6 (260.1)	51.4-1100.5	524.4 (209.5)	139.7-1250.7
MVPA[b] (min/day)[c]	27.1 (28.6)	0-116.3	50.1 (23.7)	3.8-120.2

[a]AEE: activity energy expenditure.
[b]MVPA: moderate-to-vigorous physical activity.
[c]One individual was excluded from the MVPA analysis.

Table 3. Intraclass correlation coefficient (ICC) and mean difference between the Garmin Forerunner 55 and ActiGraph wGT3X-BT devices.

Variable	ICC[a] (95% CI)	Mean difference (SD)	95% limits of agreement
Steps (per day)	0.81 (0.34 to 0.92)	990.3 (1078.4)	–1123.3 to 3104.0
AEE[b] (kcal/day)	0.31 (0.11 to 0.50)	–41.8 (276.1)	–582.9 to 499.4
MVPA[c] (min/day)[d]	0.17 (–0.03 to 0.37)	–23.0 (47.0)	–86.7 to 40.6

[a]ICC: intraclass correlation coefficient.
[b]AEE: activity energy expenditure.
[c]MVPA: moderate-to-vigorous physical activity.
[d]One individual was excluded from the MVPA analysis.

Figure 1. Bland-Altman plot showing the difference in daily step counts between devices (Garmin–ActiGraph) plotted against the mean step count for each participant. Numbers are mean difference and upper and lower limits of agreement.

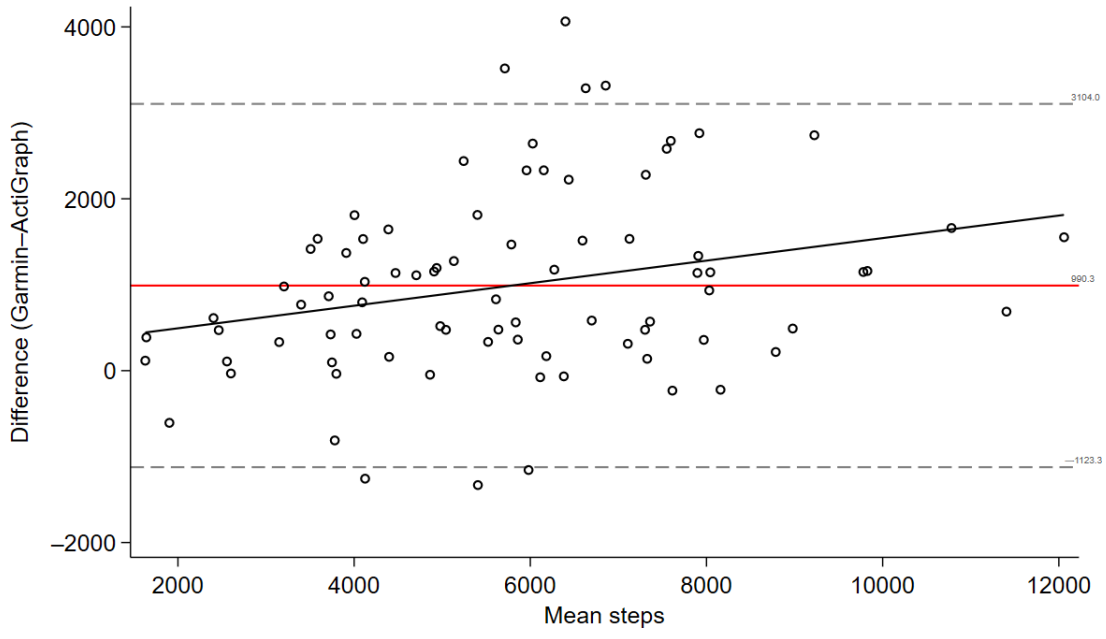


Figure 2. Bland-Altman plot showing the difference in daily activity energy expenditure (AEE) between devices (Garmin–ActiGraph) against the mean AEE for each participant. Numbers are mean difference and upper and lower limits of agreement.

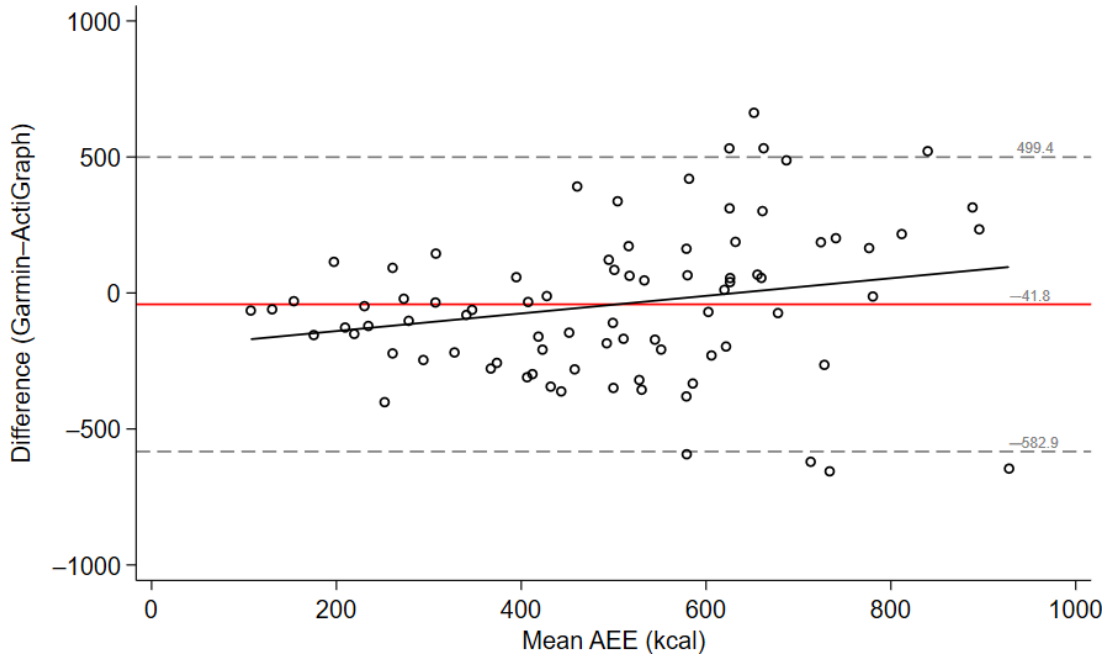
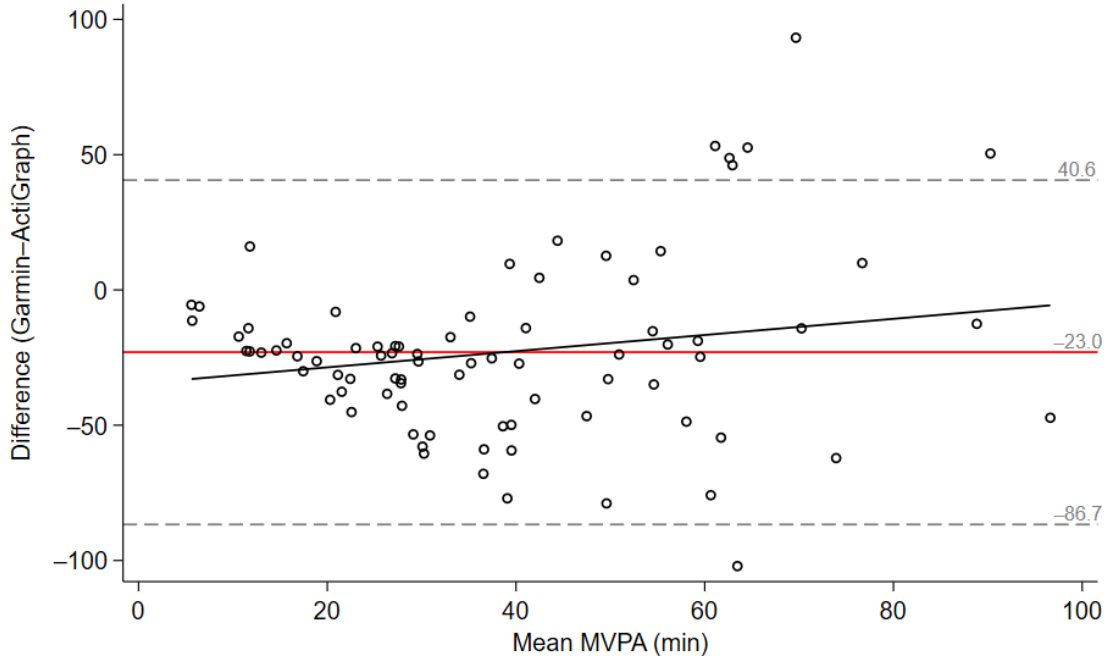


Figure 3. Bland-Altman plot showing the difference in daily moderate-to-vigorous physical activity (MVPA) between devices (Garmin–ActiGraph) against the mean MVPA for each participant. Numbers are mean difference and upper and lower limits of agreement.



There was a systematic difference between devices for daily step count (Table 3). The mean difference was 990.3 (SD 1078.4) steps/day, with the Garmin Forerunner 55 recording higher step counts on average compared with the ActiGraph wGT3X-BT. Analyses of absolute agreement between the two devices resulted in an individual ICC of 0.81, indicating good agreement. However, the wide CI included values ranging from poor to excellent agreement. The 95% LoA ranged from –1123 to 3104 steps/day (Table 3). Visual inspection of the Bland-Altman plot suggested an increase in the difference between Garmin and ActiGraph devices with increasing mean step count (Figure 1). The largest differences between devices occurred with moderate step volumes (approximately 5000-7000 steps/day), whereas both the least and most active participants showed smaller discrepancies. This pattern indicates that differences between devices vary across the activity range and the mean difference of 990.3 steps/day does not represent the discrepancy at either end of the distribution.

For AEE, the mean difference between devices was –41.8 (SD 276.1) kcal/day, indicating that the Garmin Forerunner 55 reported slightly lower values than the ActiGraph wGT3X-BT (Table 3). The individual ICC was 0.31, reflecting poor agreement between the two instruments for AEE. The 95%

LoA ranged from –582.9 to 499.4 kcal/day, and visual inspection of the Bland-Altman plot suggested that differences between devices tended to increase with higher mean AEE values (Figure 2). The largest discrepancies can be observed with moderate AEE levels, while low and high AEE values showed smaller differences. This indicates that discrepancy between devices varies across the AEE range.

The mean difference between devices for MVPA was –23 (SD 47) minutes/day, in which the Garmin Forerunner 55 reported lower values compared with the ActiGraph wGT3X-BT. The individual ICC was 0.17, indicating poor agreement between the two devices for this parameter. The 95% LoA ranged from –86.7 to 40.6 minutes/day (Table 3). Visual inspection of the Bland-Altman plot suggests variability in the differences across the range of mean MVPA values, with a tendency toward larger discrepancies at higher MVPA values (Figure 3). However, the largest differences were not constantly observed at highest MVPA levels; discrepancies varied across the MVPA range, indicating that the mean difference of –23 minutes/day does not represent the discrepancy at either end of the distribution.

One participant had an extreme mean difference value (286 minutes) between the two devices for MVPA. This value was visually inspected and considered implausible and therefore excluded from the MVPA analysis. With the outlier included, the individual ICC was 0.17 (95% CI –0.03 to 0.35) and the mean difference was –19.3 (SD 47) minutes/day. The LoA widened (–111.4 to 72.8) and the regression line was strongly influenced by the single value. Multimedia Appendix 1 illustrates a Bland-Altman plot for MVPA with the outlier included.

Discussion

Main Findings

This study investigated the absolute agreement during free-living conditions between the Garmin Forerunner 55 and ActiGraph wGT3X-BT accelerometer across several activity metrics. The main findings indicate that the level of agreement varied substantially across the different metrics. The ICC point estimate for daily step counts indicated good absolute agreement. However, the CI was wide and indicates considerable uncertainty around the population ICC. Although the present study point estimate suggests good agreement, we cannot rule out lower agreement levels.

Agreement was substantially lower for AEE and MVPA. For MVPA, individual ICC was poor, also demonstrated by the wide 95% LoA and variability in the differences across the MVPA range. Our results suggest that while the Garmin Forerunner 55 showed good individual-level agreement for steps, the device recorded higher step counts than the ActiGraph wGT3X-BT, indicating a systematic group-level bias. In contrast, estimates of intensity-based metrics diverge considerably from those derived from the ActiGraph wGT3X-BT, with MVPA showing both large group-level differences and poor individual-level agreement. This pattern reflects broader methodological challenges when measuring physical activity using consumer wearables, particularly in older adults with lower activity levels. Additionally, the main findings support the use of simple step metrics when evaluating physical activity in study populations, the benefit of it being that the number of steps is relatively easy to communicate compared to intensity-based metrics.

Comparison With Prior Work

The findings in this study align with previous studies showing that consumer wearables tend to perform reasonably well for step counting but show substantial variability when estimating energy expenditure and time spent in different intensity categories [17,20]. Earlier studies have also demonstrated higher step counts from wrist-worn devices compared to hip-worn accelerometers, including ActiGraph devices [23,38]. This discrepancy may be attributed to the fact that wrist placement is more susceptible to nonambulatory arm movements which can be misclassified as steps [38]. This pattern is consistent with the slight overestimation of steps observed in the present study. However, the magnitude of overestimation in our sample appears smaller than what has been reported for other consumer wearables in agreement studies [17,21], suggesting that the Garmin Forerunner 55 may introduce less wrist-related inflation of step counts than some other devices.

Previous studies investigating consumer wearables have reported wide 95% LoA and inconsistent performance for AEE and MVPA across available devices, including Garmin models [17,27]. The tendency toward larger differences at higher activity levels observed in the present study is consistent with previous studies, demonstrating that consumer wearables may become less accurate at higher activity intensities [10,17,18,39]. The low ICCs for MVPA support existing literature, indicating that consumer devices often struggle to classify activity intensity in older adults, especially in those with slower gait speed or altered movement efficiency [18,39]. These findings align with previous discussions in the literature describing why research-grade and consumer-grade devices may disagree. The differences in activity intensity cut points, signal processing, and the use of proprietary algorithms have been identified as common sources of variability between devices [40,41].

Despite the poor agreement among isolated activity parameters and challenges such as nondisclosure of proprietary data generating algorithms by companies, the growing availability of consumer wearables presents new opportunities for physical activity research. Compared to traditional research-grade accelerometers, which are commonly used to capture activity over relatively short and predefined periods, consumer wearables support continuous, long-term monitoring of daily activity patterns [14]. This is particularly valuable in studies where activity patterns may vary over several weeks or months, since short observation windows may not be able to capture day-to-day variability in activity. Additionally, continuous monitoring allows researchers to explore seasonal variation and long-term behavioral adaptations that can be challenging to detect with conventional accelerometer measurements [13,17]. Our findings support using the Garmin

55 Forerunner in this context to provide valuable step-related activity behavior data.

Strengths and Limitations

The main strength of this study is our collection of activity data from participants in a free-living setting using both devices simultaneously over multiple days, which enhances ecological validity and aligns with recommended collection and processing of accelerometer data [34,41]. In contrast, data collected in controlled or laboratory settings may not capture the natural variability of daily movement patterns, as such environments may limit natural behavior and not fully reflect real-world activity [16,17]. The participants in this study contributed an average of 6.7 (SD 0.9) matched days of recording from both devices, providing a robust dataset for absolute agreement analyses. The sample consisted of older adults with low activity levels and a high BMI, a population underrepresented in comparison studies but highly relevant for public health and clinical applications [6,24]. Combining Bland-Altman plots and ICCs provides complementary perspectives on agreement and captures both systematic bias and individual-level variability.

There are also limitations in this study. Absolute agreement for the consumer wearable was compared against a research-grade accelerometer rather than a gold-standard measure of physical activity. However, there is currently no universally accepted gold standard for assessment of physical activity in free-living conditions [20,42]. The ActiGraph has been widely used as a reference instrument and evaluated against laboratory-based methods and is therefore considered an appropriate reference device for free-living comparison studies [9,28,29]. In addition, although we collected data at the day level, we analyzed agreement based on participant-level average daily values. The findings therefore reflect agreement in average daily estimates over the monitoring period rather than agreement on individual days. Thus, our results cannot be used to infer the extent to which the devices agree on a given day or the within-participant variation in device differences over time.

Furthermore, the ActiGraph wGT3X-BT estimates of AEE and MVPA were derived using the Freedson VM3 combination equation and Sasaki cut points, respectively. Both equations were developed in younger adults and may be less accurate in older individuals with slower gait speed or higher BMI [28,29]. This may influence the accuracy of the reference estimates and could partly explain discrepancies between devices. However, both equations have been used previously in studies with samples that also included older adults, making them comparable for population-based research [5-7,31]. Another limitation is that Garmin proprietary algorithms are not publicly disclosed and may change during firmware updates, limiting transparency and reproducibility, a well-known challenge in consumer wearable research [8,16,22]. Since the Garmin Forerunner 55 does not provide information on wear time, we inspected minute-level heart rate data for all days included, confirming all days contained ≥10 hours of heart rate data during daytime, indicating substantial wear time. This reduces the likelihood that days with ActiGraph wear but Garmin nonwear time were included in the analysis.

Device placement represents another important challenge when comparing devices, as differences in placement may introduce discrepancies regarding step counts, intensity estimates, and energy expenditure [34]. Placement-related differences, together with other methodological factors, likely contribute to the group-level differences observed in the present study, even when individual-level agreement was acceptable for some metrics. We evaluated absolute agreement between a wrist-worn consumer wearable and a research-grade hip-worn accelerometer, the latter serving as a criterion measure of habitual physical activity.

Hip-worn accelerometers are extensively validated for capturing ambulatory movements, such as walking [9,29], which is likely the predominant activity in this population of older adults. Although differences in sensor placement may influence signal characteristics, the comparison remains appropriate since wrist wear is typical real-world placement for consumer devices [20,21]. By combining a validated reference method with a pragmatic wear location for the index device, the study design maintains ecological relevance while focusing on absolute agreement. Finally, although the sample size is adequate for analyses of agreement, it may limit the generalizability to other age groups and activity profiles.

Conclusions

In summary, the Garmin Forerunner 55 consumer wearable demonstrated good absolute agreement with the ActiGraph wGT3X-BT for daily step count but showed poor agreement for AEE and intensity-based physical activity metrics in older adults with low activity levels and elevated BMI. Our findings suggest that the Garmin Forerunner 55 may be suitable for monitoring step-based metrics at the group level in free-living conditions. However, its limited agreement for AEE and MVPA suggests that these metrics should be interpreted with caution. Future studies are needed to determine whether the device can detect changes in physical activity over time or in response to lifestyle interventions.

Acknowledgments

The authors would like to thank all the participants in the RESTART (Re-Inventing Strategies for Healthy Ageing; Recommendations and Tools) trial for their contribution.

Generative AI Use

The authors used AI (ChatGPT-5 mini; OpenAI) to assist with language editing only. The authors remain fully responsible for the final manuscript.

Funding

The RESTART trial was funded by Research Council Norway (grant number 336341) and the Simon Fougner Hartmann Family Fund. No further funding was required for this study. Open access funding is provided by UiT - The Arctic University of Norway. Open access publication charges for this article were covered by a publishing agreement provided by UiT - The Arctic University of Norway.

Data Availability

This study uses data from the RESTART randomized controlled trial (ClinicalTrials.gov ID NCT06122441). Data can be made available upon reasonable request to the RESTART trial principal investigator. Access is subject to approval under Norwegian data protection regulations and the study's ethical approval. Deidentified data can be shared for research purposes only.

Authors' Contributions

Conceptualization: KH, JJ, AH
Data curation: KH, AH, BM
Investigation: JJ, AH, KH, SG, BM
Methodology: KH, JJ, AH
Project administration: JJ
Supervision: JJ, AH, SG
Visualization: KH
Writing – original draft: KH
Writing – review & editing: KH, JJ, AH, SG, BM

Conflicts of Interest

None declared.

Multimedia Appendix 1

Bland-Altman plot with outlier included. The plot illustrates the difference in daily moderate-to-vigorous physical activity (MVPA) between devices against the mean MVPA for each participant. Numbers are mean difference and upper and lower limits of agreement.
[PNG File (Portable Network Graphics File), 48 KB-Multimedia Appendix 1]

References

1. Santos AC, Willumsen J, Meheus F, Ilbawi A, Bull FC. The cost of inaction on physical inactivity to public health-care systems: a population-attributable fraction analysis. Lancet Glob Health. Jan 2023;11(1):e32-e39. [doi: 10.1016/S2214-109X(22)00464-8] [Medline: 36480931]
2. Bull FC, Al-Ansari SS, Biddle S, et al. World Health Organization 2020 guidelines on physical activity and sedentary behaviour. Br J Sports Med. Dec 2020;54(24):1451-1462. [doi: 10.1136/bjsports-2020-102955] [Medline: 33239350]
3. Strain T, Flaxman S, Guthold R, et al. National, regional, and global trends in insufficient physical activity among adults from 2000 to 2022: a pooled analysis of 507 population-based surveys with 5·7 million participants. Lancet Glob Health. Aug 2024;12(8):e1232-e1243. [doi: 10.1016/S2214-109X(24)00150-5] [Medline: 38942042]
4. Silfee VJ, Haughton CF, Jake-Schoffman DE, et al. Objective measurement of physical activity outcomes in lifestyle interventions among adults: a systematic review. Prev Med Rep. Sep 2018;11:74-80. [doi: 10.1016/j.pmedr.2018.05.003] [Medline: 29984142]
5. Füzéki E, Engeroff T, Banzer W. Health benefits of light-intensity physical activity: a systematic review of accelerometer data of the National Health and Nutrition Examination Survey (NHANES). Sports Med. Sep 2017;47(9):1769-1793. [doi: 10.1007/s40279-017-0724-0] [Medline: 28393328]
6. Morseth B, Hopstock LA. Time trends in physical activity in the Tromsø study: an update. PLoS ONE. Apr 14, 2020;15(4):e0231581. [doi: 10.1371/journal.pone.0231581] [Medline: 32287319]
7. Sagelv EH, Ekelund U, Pedersen S, et al. Physical activity levels in adults and elderly from triaxial and uniaxial accelerometry. The Tromsø Study. PLoS ONE. Dec 3, 2019;14(12):e0225670. [doi: 10.1371/journal.pone.0225670] [Medline: 31794552]
8. Henriksen A, Haugen Mikalsen M, Woldaregay AZ, et al. Using fitness trackers and smartwatches to measure physical activity in research: analysis of consumer wrist-worn wearables. J Med Internet Res. Mar 22, 2018;20(3):e110. [doi: 10.2196/jmir.9157] [Medline: 29567635]
9. Troiano RP, McClain JJ, Brychta RJ, Chen KY. Evolution of accelerometer methods for physical activity research. Br J Sports Med. Jul 2014;48(13):1019-1023. [doi: 10.1136/bjsports-2014-093546] [Medline: 24782483]

10. Henriksen A, Grimsgaard S, Horsch A, Hartvigsen G, Hopstock L. Validity of the Polar M430 activity monitor in free-living conditions: validation study. JMIR Form Res. Aug 16, 2019;3(3):e14438. [doi: 10.2196/14438] [Medline: 31420958]
11. Migueles JH, Molina-Garcia P, Torres-Lopez LV, et al. Equivalency of four research-grade movement sensors to assess movement behaviors and its implications for population surveillance. Sci Rep. Apr 1, 2022;12(1):5525. [doi: 10.1038/s41598-022-09469-2] [Medline: 35365696]
12. Brickwood KJ, Watson G, O'Brien J, Williams AD. Consumer-based wearable activity trackers increase physical activity participation: systematic review and meta-analysis. JMIR mHealth uHealth. Apr 12, 2019;7(4):e11819. [doi: 10.2196/11819] [Medline: 30977740]
13. Ferguson T, Olds T, Curtis R, et al. Effectiveness of wearable activity trackers to increase physical activity and improve health: a systematic review of systematic reviews and meta-analyses. Lancet Digit Health. Aug 2022;4(8):e615-e626. [doi: 10.1016/S2589-7500(22)00111-X] [Medline: 35868813]
14. Huhn S, Axt M, Gunga HC, et al. The impact of wearable technologies in health research: scoping review. JMIR mHealth uHealth. Jan 25, 2022;10(1):e34384. [doi: 10.2196/34384] [Medline: 35076409]
15. Kaewkannate K, Kim S. A comparison of wearable fitness devices. BMC Public Health. May 24, 2016;16:433. [doi: 10.1186/s12889-016-3059-0] [Medline: 27220855]
16. Shei RJ, Holder IG, Oumsang AS, Paris BA, Paris HL. Wearable activity trackers-advanced technology or advanced marketing? Eur J Appl Physiol. Sep 2022;122(9):1975-1990. [doi: 10.1007/s00421-022-04951-1] [Medline: 35445837]
17. Fuller D, Colwell E, Low J, et al. Reliability and validity of commercially available wearable devices for measuring steps, energy expenditure, and heart rate: systematic review. JMIR mHealth uHealth. Sep 8, 2020;8(9):e18694. [doi: 10.2196/18694] [Medline: 32897239]
18. Tedesco S, Barton J, O'Flynn B. A review of activity trackers for senior citizens: research perspectives, commercial landscape and the role of the insurance industry. Sensors (Basel). Jun 3, 2017;17(6):1277. [doi: 10.3390/s17061277] [Medline: 28587188]
19. Li R, Li Y, Wang L, et al. Wearable activity tracker-based interventions for physical activity, body composition, and physical function among community-dwelling older adults: systematic review and meta-analysis of randomized controlled trials. J Med Internet Res. Apr 3, 2025;27:e59507. [doi: 10.2196/59507] [Medline: 40179387]
20. Evenson KR, Goto MM, Furberg RD. Systematic review of the validity and reliability of consumer-wearable activity trackers. Int J Behav Nutr Phys Act. Dec 18, 2015;12:159. [doi: 10.1186/s12966-015-0314-1] [Medline: 26684758]
21. Feehan LM, Geldman J, Sayre EC, et al. Accuracy of Fitbit devices: systematic review and narrative syntheses of quantitative data. JMIR mHealth uHealth. Aug 9, 2018;6(8):e10527. [doi: 10.2196/10527] [Medline: 30093371]
22. Wright SP, Hall Brown TS, Collier SR, Sandberg K. How consumer physical activity monitors could transform human physiology research. Am J Physiol Regul Integr Comp Physiol. Mar 1, 2017;312(3):R358-R367. [doi: 10.1152/ajpregu.00349.2016] [Medline: 28052867]
23. Degroote L, De Bourdeaudhuij I, Verloigne M, Poppe L, Crombez G. The accuracy of smart devices for measuring physical activity in daily life: validation study. JMIR mHealth uHealth. Dec 13, 2018;6(12):e10972. [doi: 10.2196/10972] [Medline: 30545810]
24. Johansson J, Deraas TS, Hopstock LA, Henriksen A, Grimsgaard S. Improving and preserving cardiorespiratory fitness, muscle strength and adiposity through a complex lifestyle intervention in community-dwelling older adults with elevated cardiometabolic risk: study protocol for the RESTART randomised controlled trial. BMJ Open. Apr 19, 2025;15(4):e095810. [doi: 10.1136/bmjopen-2024-095810] [Medline: 40254301]
25. Hopstock LA, Grimsgaard S, Johansen H, Kanstad K, Wilsgaard T, Eggen AE. The seventh survey of the Tromsø Study (Tromsø7) 2015-2016: study design, data collection, attendance, and prevalence of risk factors and disease in a multipurpose population-based health survey. Scand J Public Health. Nov 2022;50(7):919-929. [doi: 10.1177/14034948221092294] [Medline: 35509230]
26. Henriksen A, Johannessen E, Hartvigsen G, Grimsgaard S, Hopstock LA. Consumer-based activity trackers as a tool for physical activity monitoring in epidemiological studies during the COVID-19 pandemic: development and usability study. JMIR Public Health Surveill. Apr 23, 2021;7(4):e23806. [doi: 10.2196/23806] [Medline: 33843598]
27. Evenson KR, Spade CL. Review of validity and reliability of Garmin activity trackers. J Meas Phys Behav. Jun 2020;3(2):170-185. [doi: 10.1123/jmpb.2019-0035] [Medline: 32601613]
28. Sasaki JE, John D, Freedson PS. Validation and comparison of ActiGraph activity monitors. J Sci Med Sport. Sep 2011;14(5):411-416. [doi: 10.1016/j.jsams.2011.04.003] [Medline: 21616714]
29. Freedson PS, Melanson E, Sirard J. Calibration of the Computer Science and Applications, Inc. accelerometer. Med Sci Sports Exerc. May 1998;30(5):777-781. [doi: 10.1097/00005768-199805000-00021] [Medline: 9588623]

30. Fridolfsson J, Raustorp A, Börjesson M, Ekblom-Bak E, Ekblom Ö, Arvidsson D. Simple step counting captures comparable health information to complex accelerometer measurements. J Intern Med. May 2025;297(5):492-504. [doi: 10.1111/joim.20081] [Medline: 40165032]
31. Heitmann KA, Løchen ML, Hopstock LA, et al. Cross-sectional associations between accelerometry-measured physical activity, left atrial size, and indices of left ventricular diastolic dysfunction: the Tromsø Study. Prev Med Rep. Mar 2021;21:101290. [doi: 10.1016/j.pmedr.2020.101290] [Medline: 33425668]
32. Johansson J, Morseth B, Scott D, Strand BH, Hopstock LA, Grimsgaard S. Moderate-to-vigorous physical activity modifies the relationship between sedentary time and sarcopenia: the Tromsø Study 2015-2016. J Cachexia Sarcopenia Muscle. Aug 2021;12(4):955-963. [doi: 10.1002/jcsm.12718] [Medline: 34060236]
33. Choi L, Liu Z, Matthews CE, Buchowski MS. Validation of accelerometer wear and nonwear time classification algorithm. Med Sci Sports Exerc. Feb 2011;43(2):357-364. [doi: 10.1249/MSS.0b013e3181ed61a3] [Medline: 20581716]
34. Migueles JH, Cadenas-Sanchez C, Ekelund U, et al. Accelerometer data collection and processing criteria to assess physical activity and other outcomes: a systematic review and practical considerations. Sports Med. Sep 2017;47(9):1821-1845. [doi: 10.1007/s40279-017-0716-0] [Medline: 28303543]
35. Ngueleu AM, Barthod C, Best KL, Routhier F, Otis M, Batcho CS. Criterion validity of ActiGraph monitoring devices for step counting and distance measurement in adults and older adults: a systematic review. J Neuroeng Rehabil. Oct 17, 2022;19(1):112. [doi: 10.1186/s12984-022-01085-5] [Medline: 36253787]
36. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med. Jun 2016;15(2):155-163. [doi: 10.1016/j.jcm.2016.02.012] [Medline: 27330520]
37. Bland JM, Altman DG. Measuring agreement in method comparison studies. Stat Methods Med Res. Jun 1999;8(2):135-160. [doi: 10.1177/096228029900800204] [Medline: 10501650]
38. Tudor-Locke C, Barreira TV, Schuna JM Jr. Comparison of step outputs for waist and wrist accelerometer attachment sites. Med Sci Sports Exerc. Apr 2015;47(4):839-842. [doi: 10.1249/MSS.0000000000000476] [Medline: 25121517]
39. Tedesco S, Sica M, Ancillao A, Timmons S, Barton J, O'Flynn B. Validity evaluation of the Fitbit Charge2 and the Garmin vivosmart HR+ in free-living environments in an older adult cohort. JMIR mHealth uHealth. Jun 19, 2019;7(6):e13084. [doi: 10.2196/13084] [Medline: 31219048]
40. Evenson KR, Scherer E, Peter KM, Cuthbertson CC, Eckman S. Historical development of accelerometry measures and methods for physical activity and sedentary behavior research worldwide: a scoping review of observational studies of adults. PLoS ONE. Nov 21, 2022;17(11):e0276890. [doi: 10.1371/journal.pone.0276890] [Medline: 36409738]
41. Welk GJ, Bai Y, Lee JM, Godino J, Saint-Maurice PF, Carr L. Standardizing analytic methods and reporting in activity monitor validation studies. Med Sci Sports Exerc. Aug 2019;51(8):1767-1780. [doi: 10.1249/MSS.0000000000001966] [Medline: 30913159]
42. Keadle SK, Lyden KA, Strath SJ, Staudenmayer JW, Freedson PS. A framework to evaluate devices that assess physical behavior. Exerc Sport Sci Rev. Oct 2019;47(4):206-214. [doi: 10.1249/JES.0000000000000206] [Medline: 31524786]

Abbreviations

AEE: activity energy expenditure
ICC: intraclass correlation coefficient
LoA: limits of agreement
MVPA: moderate to vigorous physical activity
NHANES: National Health and Nutrition Examination Survey
RESTART: Re-Inventing Strategies for Healthy Ageing; Recommendations and Tools

Edited by Javad Sarvestan; peer-reviewed by Jonah J C Thomas; submitted 26.Jun.2026; initial comments to author 24.Aug.2026; final revised version received 20.Sep.2026; accepted 02.Nov.2026; published 24.Nov.2026

Please cite as:
Hagerupsen K, Henriksen A, Morseth B, Grimsgaard S, Johansson J
Absolute Agreement Between a Consumer Wearable and Research Grade Accelerometry in Free-Living Conditions: Comparison Study
JMIR Form Res 2026;10:e105610
URL: https://formative.jmir.org/2026/1/e105610
doi: 10.2196/105610

© Karianne Hagerupsen, André Henriksen, Bente Morseth, Sameline Grimsgaard, Jonas Johansson. Originally published in JMIR Formative Research (https://formative.jmir.org), 24.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.