Title

Evaluation of the Agreement Between Research-Grade Actigraphy Sleep, Consumer-Grade Smartwatches and Self-reported Sleep Diaries in Masters Endurance Athletes

Running title

Sleep Tracking in Masters Athletes

Authors

Asli Devrim-Lanpir[1,2], Simon Devenney[1], Brendan Egan[1,3]

Affiliations

[1] School of Human Performance and Health, Dublin City University, Dublin, Ireland

[2] Department of Nutrition and Dietetics, Faculty of Health Sciences, Istanbul Medeniyet University, Istanbul, Turkey

[3] Florida Institute for Human and Machine Cognition, Pensacola, FL

Corresponding Author: Brendan Egan; brendan.egan@dcu.ie

ORCIDs:

Asli Devrim-Lanpir 0000-0002-4267-9950

Simon Deveney 0009-0007-3692-9312

Brendan Egan 0000-0001-8327-9016

CORRESPONDING AUTHOR

Brendan Egan, PhD

School of Health and Human Performance

Dublin City University
Glasnevin, Dublin D09 V209
Ireland
e: brendan.egan@dcu.ie
t: +353 1 7008803
f: +353 1 7008888

ABSTRACT WORD COUNT	**247**
MAIN TEXT WORD COUNT	**5530**
NUMBER OF REFERENCES	**53**
NUMBER OF TABLES	**4**
NUMBER OF FIGURES	**2**
SUPPLEMENTARY TABLES	**3**

Acknowledgements

The authors are grateful to all participants for their time and commitment to the study.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Funding Statement

This study was supported by an Irish Research Council Government of Ireland Postdoctoral Fellowship under grant number GOIPD/2022/312.

Conflict of Interest Disclosure

The authors declare no conflicts of interest.

Ethical Approval Statement

This study was conducted in accordance with the principles outlined in the Declaration of Helsinki, with the exception of being registered in a trials database. Ethical approval was granted by the Dublin City University Research Ethics Committee (permit: DCUREC/2023/042).

Patient Consent Statement

All participants were fully informed about the purpose, procedures, potential benefits, and possible risks of the study. Written informed consent was obtained from each athlete prior to their participation.

Permission to Reproduce Material from Other Sources

No previously published materials, images, or data from other sources were reproduced in this study.

Clinical Trial Registration

Not applicable.

Abstract

Sleep monitoring is a tool widely used to support recovery and performance in endurance athletes. This study aimed to assess agreement between research-grade actigraphy (ActiGraph GT9X), consumer-grade smartwatches (Garmin), and self-reported sleep diaries in masters endurance athletes. Seventy athletes (43 males, 46.3±7.3 y; 27 females, 49.3±8.3 y) wore ActiGraph and smartwatch devices on their non-dominant wrist while maintaining a self-reported sleep diary for seven consecutive nights. ActiGraph recorded the shortest total sleep time (332±87 min), whereas the diary and smartwatch recorded longer sleep durations by 109 and 126 min, respectively ($p<0.001$). Sleep efficiency (%) was also higher in the sleep diary and smartwatch compared to ActiGraph, with mean biases of -5.9% and -4.1%, respectively. Sleep diary values closely agreed with smartwatch values (ICC=0.880, 95% CI: 0.624 to 0.946), while poor agreement was found between ActiGraph and the sleep diary (ICC=0.190, 95% CI: -0.149 to 0.459). Proportional bias was evident in both the sleep diary and smartwatch, with greater differences in total sleep time and efficiency observed in athletes with shorter durations and lower sleep efficiency, respectively. Sex differences emerged, with stronger agreement between smartwatch and Actigraph in sleep efficiency in females (ICC=0.690, 95% CI: 0.336 to 0.857) than males (ICC=0.481, 95% CI: -0.020 to 0.723). Findings suggest that both consumer-grade devices and self-reported sleep diaries tend to report longer sleep durations and higher sleep efficiency relative to actigraphy. Sleep data from these methods should be interpreted with caution, particularly in athletes with shorter or more fragmented sleep.

Keywords: accuracy; ActiGraph GT9X; sleep measurement; exercise training; wearables

1. Introduction

Sleep is a crucial pillar of athletic performance and recovery, playing a restorative role in supporting the musculoskeletal, endocrine, immune, and nervous systems (Bonilla et al., 2021), affecting physical restoration, mental well-being, and overall training adaptation (Walsh et al., 2021). Disrupted sleep patterns not only affect physiology and performance but also contribute to mood disturbances, increased stress, and emotional dysregulation, which can further hinder training adaptation and competitive outcomes (Charest & Grandner, 2020; Hamlin et al., 2021). Polysomnography (PSG) remains the gold standard for sleep assessment as it captures multiple physiological signals—such as brain activity (mainly measured by electroencephalogram), eye movements, muscle activity, and cardiovascular parameters—enabling skilled technicians to identify distinct sleep stages in 30-second epochs (Rundo & Downey, 2019). However, the practicality of PSG is limited by the demanding schedules and frequent travel many athletes face (Doherty et al., 2021). As a result, an increasing number of athletes use consumer-grade sleep trackers, which provide moderate-to-good agreement with PSG (Sadeh, 2011), to obtain real-world data on sleep duration and quality. However, concerns persist about the validity and accuracy of these devices, particularly for populations with erratic sleep patterns, high training loads, or elevated stress levels (de Zambotti et al., 2019). These concerns underscore the need for research evaluating whether consumer devices can accurately assess sleep in active, high-performing individuals.

Although consumer-grade sleep trackers offer promising features for monitoring sleep patterns, studies indicate that inaccurate data—such as overstated awakenings or underestimated sleep duration—can cause unnecessary stress, anxiety, and healthcare costs (Baron et al., 2017; Kim et al., 2022). This phenomenon, known as orthosomnia, occurs when individuals become overly fixated on sleep data, leading to anxiety and disrupted sleep patterns (Baron et al., 2017).

This potential disruption is particularly relevant for endurance athletes, who rely on sleep for optimal recovery and performance (Bonilla et al., 2021). Misleading sleep data may cause athletes to believe that they are sleep-deprived, heightening pre-race anxiety and affecting training adaptations, while excessive focus on sleep metrics can contribute to orthosomnia, further disrupting sleep quality and recovery (Bonilla et al., 2021; Walsh et al., 2021). Therefore, it is essential to evaluate the accuracy of these devices in tracking sleep to ensure that they provide valuable insights rather than inducing unnecessary concern.

Examining between-sex differences in sleep patterns warrants further focus, as variations between male and female athletes have been observed in previous studies. A systematic review encompassing 81 studies (n = 1,830 athletes) revealed that male athletes, on average, obtained 7.2 ± 1.1 hours of sleep with a sleep onset latency (SOL) of 15.7 ± 17.0 minutes, whereas female athletes had a slightly longer sleep duration (7.5 ± 1.2 hours), shorter SOL (9.7 ± 11.2 minutes), and higher sleep efficiency (89.1% ± 4.8% vs. 85.2% ± 7.2% in males) (Vlahoyiannis et al., 2021). These findings highlight the necessity of integrating sex-specific considerations into sleep research, particularly as athletes generally demonstrate reduced sleep duration and efficiency relative to healthy non-athlete populations.

Given the growing use of consumer-grade wearables and subjective, self-reported sleep diaries in sports science and clinical settings, this study assesses their accuracy for measuring key sleep metrics such as Total Sleep Time (TST), Sleep Efficiency (SE%), and the number of nocturnal awakenings, compared to research-grade wrist-worn actigraphy (ActiGraph GT9X). This study focuses on a novel population—non-elite masters endurance athletes—whose sleep may follow different patterns compared to younger endurance athletes or indeed, athletes participating in other sports. Masters athletes are generally defined as individuals aged 35 years and older who continue to train and compete beyond the typical peak age for athletic performance, in alignment with international masters sport classifications (*HOME*, n.d.; Vlahoyiannis et al., 2021). Previous literature suggests that masters athletes do not necessarily experience poorer sleep quality but may display age-related changes in sleep architecture and duration (Vlahoyiannis et al., 2021; Wooten et al., 2021). These patterns are likely influenced by lifestyle demands, such as balancing training with work and family obligations, and may differ in sport-specific ways (Driller et al., 2023). Investigating this population is especially important as their sleep behaviours have received limited attention despite their increasing participation in endurance events (dos Santos et al., 2022). Moreover, examining whether between-sex differences in sleep patterns exist is important given previous findings such as female athletes having slightly longer TST and higher SE% as well as reporting more sleep disturbances around competition (Miles et al., 2022). We hypothesized that consumer-grade wearables in the form of wrist-worn smartwatches, and self-reported sleep diaries would report higher values for TST and SE% relative to actigraphy, and that self-reported sleep diaries would differ from actigraphy in measuring nocturnal awakenings.

2. Methods

2.1. Study Design & Participants

A total of 70 endurance athletes (n=43 males, 46.3 ± 7.3 years, 1.79 ± 0.05 m, 74.5 ± 8.0 kg, 10.1 ± 3.9% body fat, 20.9 ± 6.9 kg/m² fat-free mass index (FFMI)); and n=27 females, 49.3 ± 8.3 years, 1.65 ± 0.06 m, 58.8 ± 6.9 kg, 20.3 ± 4.2% body fat, 19.7 ± 1.9 kg/m² FFMI) participated in this study. The study included endurance athletes over the age of 35 years who trained for a minimum of 240 minutes across at least five sessions per week (verified by checking the training logs), and had participated in more than three endurance running events in the past 12 months, as verified through publicly-available race results. The ≥35 years age cutoff is consistent with classifications used in international Masters competitions and sport science literature (World Masters Athletics, 2025; Vlahoyiannis et al., 2021). Athletes with known sleep disorders, chronic illnesses, or medication use affecting sleep were excluded. All anthropometric measurements were performed by a certified ISAK Level 1 Anthropometrist following the standardized protocols of the International Society for the Advancement of Kinanthropometry (ISAK). Data collection took place during the base training phase of their periodization cycle, ensuring that sleep patterns were not influenced by active race participation. This cross-sectional study complied with the principles of the Declaration of Helsinki, with the exception of being registered in a trials database, and received approval from the Dublin City University Research Ethics Committee (permit: DCUREC/2023/042). Athletes expressing interest in participating were fully briefed on the study's objectives, procedures, potential benefits, and risks, and provided written informed consent prior to commencing.

2.2. Sleep Data Collection

Sleep data were continuously recorded over seven days using consumer-grade (Garmin, Olathe, Kansas, USA) and research-grade (ActiGraph GT9X, ActiGraph Inc., Florida, USA) wearable fitness devices, both worn on the non-dominant wrist. Participants were instructed to wear the devices at all times, removing it only for showering or swimming, and were encouraged to note any non-wear time. Data from Garmin smartwatches were synced and uploaded each morning to the respective device clouds via the Garmin Connect smartphone apps. After syncing, they logged this smartwatch-recorded information into an Excel spreadsheet. Garmin smartwatches were used for comparison as they are widely adopted by athletes for both training and sleep tracking. A total of 18 different Garmin models were included in the analysis, ranging from entry-level to high-end devices. Among male athletes, 30% used high-end, 58.1% mid-range, and 11.9% low-end Garmin devices. For female athletes, the distribution was 29.6% high-end, 48.1% mid-range, and 22.3% low-end (Supplementary Table 1).

The ActiGraph GTX9, an accelerometer-based sensor with a 60Hz sampling rate, collected triaxial movement data, which were downloaded via a custom docking station and processed into 1-minute epochs using the ActiLife Data Analysis Software, specifically designed to interpret ActiGraph data (Liu et al., 2023). The algorithm used to calculate ActiGraph activity counts has been previously documented (Neishabouri et al., 2022). Additionally, nights displaying unrealistically high or low movement levels were excluded, as they likely indicated device malfunction or removal. Similar screening procedures were applied to the self-reported and smartwatch data. Entries were excluded if participants left fields blank, reported implausible values (e.g., TST > 12 hours or SE% > 100%), or failed to sync device data for that night. In ActiGraph devices, the Sleep Fragmentation Index (SFI) was calculated as the proportion of one-minute sleep bouts relative to the total number of sleep bouts during the sleep period, and Wake After Sleep Onset (WASO) was determined as the total duration of wakefulness occurring between the time a person initially fell asleep and their final awakening. ActiGraph-derived nocturnal awakenings were operationalised based on the manufacturer's instructions as discrete periods of wakefulness detected based on increased movement intensity, with each awakening defined as an activity epoch exceeding a pre-defined threshold following sleep onset.

Participants simultaneously maintained a self-reported sleep diary, logging their subjective estimate of TST, SE%, and nocturnal awakenings. For self-reported nocturnal awakenings, participants were instructed to only report perceived conscious awakenings during the night. Self-reported sleep efficiency (SE%) was assessed using a visual analogue scale (VAS) ranging from 0 to 100. Prior to rating, participants were informed that sleep efficiency refers to the proportion of time spent asleep relative to time spent in bed. They were then asked to rate how efficiently they felt they had slept the previous night, with 0 indicating "very poor sleep efficiency" and 100 indicating "very good sleep efficiency. This approach is consistent with previous sleep research utilizing VAS scales to capture subjective perceptions of sleep quality and related domains (e.g., Leeds Sleep Evaluation Questionnaire, Verran & Snyder-Halpern Sleep Scale) (Parrott & Hindmarch, 1980; Snyder-Halpern & Verran, 1987). Although this method does not provide a calculated ratio of TST-to -time in bed, it allows for a simple and intuitive self-assessment of perceived sleep efficiency, which may better reflect the athlete's own interpretation of sleep performance. To minimize potential bias, participants were advised to record their self-reported sleep metrics before syncing their smartwatch with the Garmin Connect app.

For clarity, these methods will be referred to as **ACTI** (ActiGraph), **SMART** (smartwatch), and **SELF** (self-reported sleep diary) throughout.

2.3. Statistical Analysis

All statistical analyses were conducted using SPSS (version 25). Data normality was assessed using the Shapiro–Wilk test. For TST and SE%, a repeated-measures ANOVA was performed to compare ACTI, SMART, and SELF, with a Greenhouse–Geisser correction applied when Mauchly's test indicated a violation of sphericity. Post-hoc pairwise comparisons were conducted with Bonferroni adjustments to control for multiple comparisons, and partial eta-squared (η^2p) was calculated as a measure of effect size. Independent t-tests were used for comparisons of the number of awakenings, and differences between male and female participants, with Cohen's *d* calculated to determine effect sizes. Pairwise differences are reported as mean difference (95% confidence interval). Agreement between ACTI, SMART, and SELF was assessed using Bland–Altman plots, which provided mean bias and 95% limits of agreement. Proportional bias was calculated using linear regression methods for Bland–Altman plots, which indicates the reliability of the bias of one method versus another (y-axis value) and whether it changes in proportion to the mean of the those methods (x-axis value) in the Bland–Altman plots. Intraclass correlation coefficients (ICCs) were used to evaluate reliability and consistency, with values interpreted as poor (<0.50), moderate (0.50–0.75), good (0.75–0.90), and excellent (>0.90) (Koo & Li, 2016). Statistical significance was set at $p < 0.05$ for all tests.

3. Results

Table 1 presents the sleep characteristics of masters endurance athletes as assessed using actigraphy. On average, athletes had a TST of 334 ± 88 min (5.6 ± 1.5 hours), with a SE% of 88.6 ± 4.6%. There was no difference between female and males for TST (Females, 348 ± 94 min vs. Males, 324 ± 82 min; p = 0.228), but SFI was significantly higher in males (27.0 ± 7.9) compared to females (22.8 ± 7.0, p = 0.017, Cohen's *d* = 0.56).

Table 1. Characteristics of masters endurance athletes, according to age and sex

Characteristics	**Males**		**Females**		**M vs. F p value; ES**
	n=43	**Range**	**n=27**	**Range**	
Age (years)	46.3 ± 7.3	35-65	49.3 ± 8.3	35-72	0.099; 0.38
Height (m)	1.79 ± 0.05	1.69-1.89	1.65 ± 0.06	1.54-17.4	<0.001; 2.53
BM (kg)	74.5 ± 8.0	62.0-94.9	58.8 ± 6.9	44.5-73.3	<0.001; 2.10
BMI (kg/m^2)	23.3 ± 2.1	20.0-28.5	21.6 ± 2.2	18.7-27.3	0.002; 0.79

BF (%)	10.1 ± 3.9	3.0-19.5	20.3 ± 4.2	13.2-31.2	<0.001; 2.51
FFMI (kg FFM/m^2)	20.9 ± 6.9	18.3-25.0	19.7 ± 1.9	16.9-24.2	0.007; 0.23
TST (min) [a]	324 ± 82	147-519	348 ± 94	154-543	0.228; 0.27
SE (%) [a]	87.9 ± 5.1	70-96	89.5 ± 4.0	81-96	0.179; 0.35
Number of Awakenings [a]	17.7 ± 5.3	4.2-31.9	16.1 ± 4.9	6.8-24.1	0.184; 0.31
Average Awakening Length (min) [a]	2.26 ± 0.42	1.49-3.28	2.38 ± 0.50	1.35-3.24	0.258; 0.26
TIB (min) [a]	364 ± 81	209-559	387 ± 93	186-577	0.264; 0.26
WASO (min) [a]	39.5 ± 13.0	9.7-63.7	37.7 ± 11.5	18.1-55.6	0.570; 0.15
SFI [a]	27.0 ± 7.9	10-49.9	22.8 ± 7.0	9.7-41.4	0.017; 0.56

BF: body fat; BM: body mass; BMI: body mass index; ES, effect size calculated as Cohen's *d*; FFM, fat-free mass; FFMI: FFM index; TIB: time in bed; TST: total sleep time; WASO: wake after sleep onset; SE: sleep efficiency; SFI: sleep fragmentation index. [a]: from ActiGraph GTX9 data.

Table 2 presents TST, SE%, and number of awakenings from ACTI, SMART, and SELF data for males and females. For TST, ACTI recorded significantly shorter sleep duration than both SELF and SMART ($p < 0.001$ for all comparisons, with large effect sizes ranging from partial $\eta^2 = 0.592$ to 0.705). In males, ACTI estimated an average TST of 324 ± 82 min, whereas SELF and SMART reported 436 ± 43 min and 451 ± 43 min , respectively. A similar finding was observed in females (350 ± 95, 477 ± 49, and 452 ± 43 min, respectively). For SE%, ACTI recorded lower SE% (82.6–82.8%), whereas SELF and SMART methods yielded higher SE% (86.2–87.9%). For number of awakenings, ACTI detected significantly more awakenings per night (17.3 ± 0.6 combined sample) than self-reports (0.9 ± 1.3) ($p < 0.001$).

Table 2. Total sleep time, sleep efficiency, and number of awakenings across ActiGraph (ACTI), smartwatch (SMART), and self-reported (SELF) data in male and female master endurance athletes

Measure	Group	ACTI Mean ± SD	SMART Mean ± SD	SELF Mean ± SD	F	*p* value	Effect Size
TST (min)	Males	324 ± 82	451 ± 43	436 ± 43	122.17	<0.001 *§^	0.705 [a]
	Females	350 ± 95	477 ± 49	452 ± 43	36.255	<0.001 *§^	0.592 [a]
	Combined	332 ± 87	459 ± 47	441± 44	151.980	<0.001 *§^	0.664 [a]
SE (%)	Males	82.2 ± 7.4	87.2 ± 4.3	87.9 ± 5.2	30.892	<0.001 *§	0.382 [a]
	Females	89.5 ± 4.0	85.6 ± 5.4	83.5 ± 8.4	9.292	<0.001 §^	0.263 [a]
	Combined	82.6 ± 7.76	86.6 ± 4.72	88.5 ± 4.8	34.397	<0.001 *§^	0.309 [a]
Number of awakenings*	Males	17.7 ± 5.3	-	0.79 ± 0.97	-	0.009 §	4.43 [b]
	Females	16.1 ± 4.9	-	0.89 ± 1.55	-	0.037 §	4.19 [b]
	Combined	17.2 ± 5.2	-	0.82 ± 1.19	-	0.001 §	4.34 [b]

Data are mean±SD. *paired-samples t test. Mauchly's test indicated a violation of the sphericity assumption for TST in males, $\chi^2(2) = 83.6$, $p < 0.001$; in females, $\chi^2(2) = 30.5$, $p < 0.001$; and in the combined sample, $\chi^2(2) = 109.3$, $p < 0.001$. Similarly, the assumption of sphericity was violated for Sleep Efficiency (SE%), with $\chi^2(2) = 11.5$, $p = 0.003$ in males; $\chi^2(2) = 9.4$, $p = 0.009$ in females; and $\chi^2(2) = 13.6$, $p = 0.001$ in the combined sample. Therefore, Greenhouse-Geisser corrections were applied. Effect sizes indicate by [a]: Partial η^2, and [b]: Cohen's *d*.
* p<0.05; ACTI vs SMART
§ p<0.05; ACTI vs SELF
^ p<0.05; SMART vs SELF

Table 3 presents the post-hoc pairwise comparisons of TST and SE% from ACTI, SMART, and SELF. Significant differences were observed for TST across all methods. Specifically, SMART TST exceeded ACTI TST values by 126 min in both males and females ($p < 0.001$). For SE%, significant differences were found between SELF and ACTI ($p < 0.001$ for males, $p = 0.002$ for females), with SELF showing consistently higher SE% values relative to ACTI. SMART SE% was also significantly higher than ACTI estimates in males [5.03 (3.20 to 6.87), p < 0.001] and in the combined sample [4.01 (2.49 to 5.52), p = 0.002], but this difference was not statistically significant in females (p = 0.123). SE% differences between SELF and SMART were not significantly different in males (p = 0.227) but were significantly different in females [3.92 (1.66-6.18), p = 0.001].

Table 3. Post-hoc pairwise comparison of sleep metrics across ActiGraph (ACTI), smartwatch (SMART), and self-reported (SELF) data in male and female master endurance athletes

Comparison	**Males MD (95% CI)**	***p***	**Females MD (95% CI)**	***p***	**Combined MD (95% CI)**	***p***
SELF vs. SMART (TST, min)	-14.8 (-20.5, -9.12)	<0.001	-25.3 (-27.9, -12.7)	0.001	-18.3 (-23.9, -12.7)	<0.001
SELF vs. ACTI (TST, min)	112 (90, 133)	<0.001	105 (66, 143)	<0.001	109 (90, 128)	<0.001
SMART vs. ACTI (TST, min)	126 (106, 148)	<0.001	126 (88, 165)	<0.001	126 (108, 145)	<0.001
SELF vs. SMART (%SE)	0.72 (-0.46, 1.90)	0.227	3.92 (1.66, 6.18)	0.001	1.81 (0.69, 2.93)	<0.001
SELF vs. ACTI (%SE)	5.84 (4.08, 7.58)	<0.001	5.99 (2.39, 9.59)	0.002	5.88 (4.24, 7.52)	<0.001
SMART vs. ACTI (%SE)	5.03 (3.20, 6.87)	<0.001	2.08 (-0.60, 4.75)	0.123	4.01 (2.49, 5.52)	0.002

CI: confidence interval; MD: Mean Difference; TST: total Sleep time; SE: sleep efficiency (%)

Table 4 presents the ICC values assessing the agreement between SELF, SMART, and ACTI measurements for TST, SE%, and number of awakenings across male, female, and combined samples. For TST, ICC values indicate *good* agreement between SELF and SMART (ICC = 0.880, 95% CI: 0.624 to 0.946), with slightly higher agreement in males (ICC = 0.914, 95% CI: 0.711 to 0.964; *excellent*) compared to females (ICC = 0.807, 95% CI: 0.324 to 0.929;

good). However, agreement was *poor* between SELF and ACTI (ICC = 0.190, 95% CI: -0.149 to 0.459), and SMART and ACTI (ICC = 0.212, 95% CI: -0.163 to 0.506) in combined samples. For SE%, SELF and SMART exhibited *moderate* agreement (ICC = 0.591, 95% CI: 0.354 to 0.740), with higher agreement in males (ICC = 0.744, 95% CI: 0.557 to 0.853; *moderate*) compared to females (ICC = 0.342, 95% CI: -0.220 to 0.468; *poor*). Agreement between SELF and ACTI was *poor* in the combined sample (ICC = 0.413, 95% CI: -0.029 to 0.656), with a particularly low value in females (ICC = 0.064). Notably, SMART and ACTI had *moderate* agreement overall (ICC = 0.552, 95% CI: 0.209 to 0.736), but agreement was stronger in females (ICC = 0.690, 95% CI: 0.336 to 0.857; *moderate*) compared to males (ICC = 0.481, 95% CI: -0.020 to 0.673; *poor*). For number of awakenings, ICC values were close to zero across all comparisons.

Table 4. Intraclass Correlation Coefficient (ICC) analysis between ActiGraph (ACTI), smartwatch (SMART), and self-reported (SELF) data in male and female masters endurance athletes

Comparison	**Males ICC (95% CI)**	**Females ICC (95% CI)**	**Combined ICC (95% CI)**
SELF vs. SMART (TST)	0.914	0.807	0.880
	(0.711, 0.964)	(0.324, 0.929)	(0.624, 0.946)
SELF vs. ACTI (TST)	0.207	0.119	0.190
	(-0.170, 0.508)	(-0.258, 0.468)	(-0.149, 0.459)
SMART vs. ACTI (TST)	0.212	0.169	0.212
	(-0.167, 0.526)	(-0.203, 0.511)	(-0.163, 0.506)
SELF vs. SMART (SE)	0.744	0.342	0.591
	(0.557, 0.853)	(-0.220, 0.673)	(0.354, 0.740)
SELF vs. ACTI (SE)	0.548	0.064	0.413
	(-0.055, 0.787)	(-0.550, 0.500)	(-0.029, 0.656)
SMART vs. ACTI (SE)	0.481	0.690	0.552
	(-0.020, 0.723)	(0.336, 0.857)	(0.209, 0.736)
SELF vs. ACTI	0.023	0.047	0.030
(Numbers of awakening)	(-0.019, 0.062)	(-0.054, 0.221)	(-0.04, 0.127)

ICC: Intraclass Correlation Coefficient; TST: total Sleep time; SE: sleep efficiency (%)

When examining whether the model tier (high-end, mid-range, or low-end) of Garmin smartwatch influenced agreement between ACTI and SELF, there were no significant differences in TST or SE% values across tiers when compared to ACTI (Supplementary Table 2). While TST values recorded by all Garmin tiers were consistently higher than those recorded by ACTI, agreement remained low (ICCs < 0.16), regardless of tier (Supplementary Table 2). In contrast, agreement of SMART with self-reported TST was good across all tiers (ICCs between 0.754 and 0.825) (Supplementary Table 2). For SE%, the mid-range tier showed the highest agreement with ACTI (ICC = 0.529, 95% CI: 0.269 to 0.717; *moderate*), while the low-end tier showed weaker agreement with both ACTI and SELF. Overall, device tier did not substantially affect the interpretation of Garmin-derived sleep metrics (Supplementary Table 2). Additionally, when examining whether age influenced agreement between methods, no significant differences in bias or agreement were observed across age tertiles (Supplementary Table 3).

Figure 1 presents Bland-Altman plots comparing TST measured by ACTI vs. SELF, SELF vs. SMART, and ACTI vs. SMART, and, stratified by male, female, and the combined sample. The SELF vs. SMART comparison shows a mean bias of -17.5 min (95% LoA: -67.9 to 32.8 min) in the combined sample, with similar trends in males (-14.8 min, LoA: -54.8 to 25.2 min) and females (-22.7 min, LoA: -88.4 to 42.9 min). In contrast, the ACTI vs. SMART comparison demonstrates that SMART reported longer TST values relative to ACTI, with a mean bias of 127 min (LoA: -286 to 32 min) in the combined sample. The longer TST values in SMART is consistent across males (126 min, LoA: -274 to 21 min) and females (126 min, LoA: -311 to 55 min). Similarly, the ACTI vs. SELF comparison reveals that SELF TST values were longer than ACTI, with a mean bias of 109 min (LoA: -275 to 56 min) in the combined sample. Linear regression analyses were conducted to assess proportional bias in the Bland-Altman plots. A significant proportional bias was observed in the comparisons between ACTI and SELF TST ($R^2 = 0.41$ for females, $R^2 = 0.34$ for males, $R^2 = 0.36$ combined; all $p < 0.05$) and between ACTI and SMART TST ($R^2 = 0.35$ for females, $R^2 = 0.37$ for males, $R^2 = 0.34$ combined; all $p < 0.05$), indicating that the difference between methods increased as average sleep duration decreased. In contrast, no proportional bias was observed in comparisons between SELF and SMART TST ($R^2 < 0.01$ in all groups).

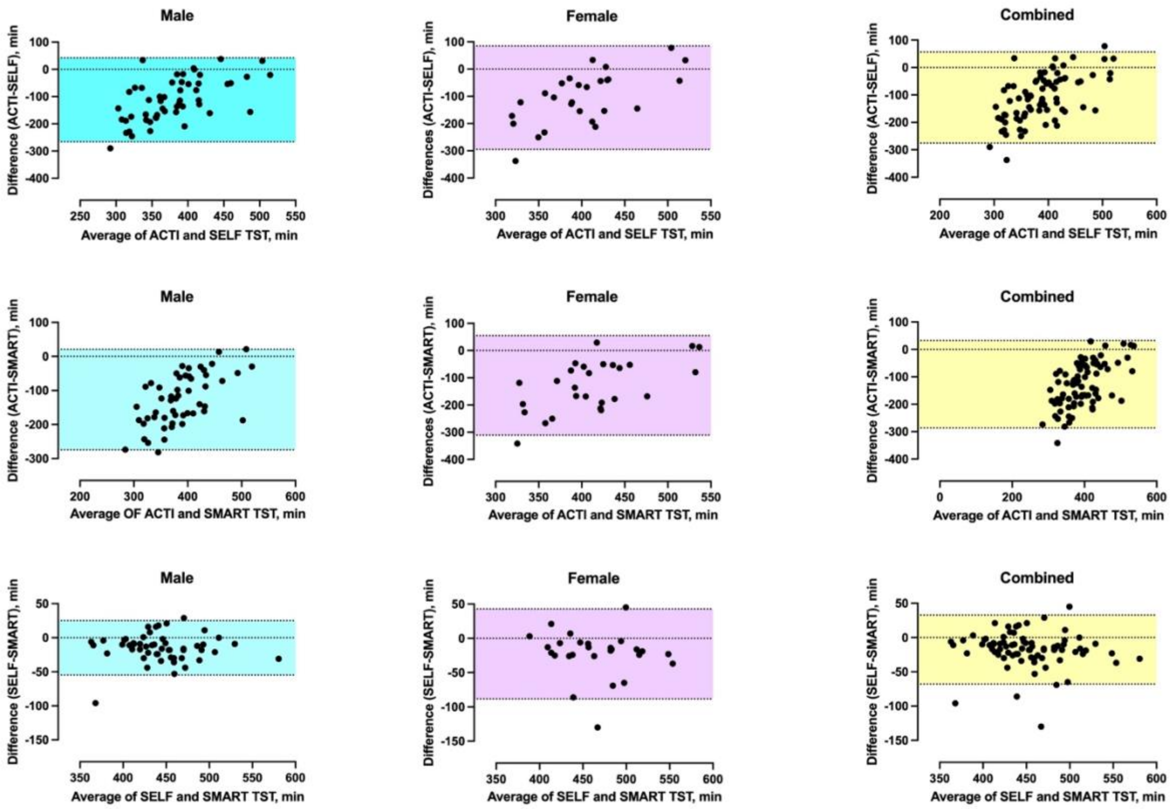

Figure 1. Method comparison of total sleep time evaluated using Bland-Altman plots for data from ActiGraph (ACTI), smartwatch (SMART), and self-reported (SELF) data in male and female masters endurance athletes. *$p < 0.05$ for the linear regression analysis indicating the presence of proportional bias.

Figure 2 shows the Bland-Altman plots comparing SE% between SMART vs. ACTI, SELF vs. SMART, and ACTI vs. SELF in masters endurance athletes. The ACTI vs. SMART comparison shows a mean bias of -4.1% (95% LoA: -17.17% to 9.04%) in the combined sample, indicating that SMART reported higher SE% relative to ACTI. The SELF vs. SMART comparison demonstrates a mean bias of 1.81% (LoA: -7.98% to 11.61%) in the combined sample, indicating that SELF recorded higher SE% compared to SMART. Differences were largest in females (3.91%, LoA: -7.28% to 15.11%), while males showed a smaller bias (0.72%, LoA: -7.60% to 9.04%). The ACTI vs. SELF comparison exhibited the largest discrepancy, with a mean bias of -5.87% (LoA: -20.05% to 8.30%) in the combined sample, indicating that SELF reported higher SE% relative to ACTI. The differences were similar between males (-5.81%, LoA: -17.86% to 6.23%) and females (-5.99%, LoA: -23.83% to 11.85%). Bland–

Altman regression analyses for sleep efficiency (%) revealed evidence of proportional bias in comparisons involving ACTI. Specifically, ACTI vs SELF comparisons yielded R^2 values of 0.17 (males), 0.40 (females), and 0.22 (combined) (all $p < 0.05$), while ACTI vs SMART comparisons showed R^2 values of 0.30, 0.25, and 0.26, respectively (all $p < 0.05$). These findings indicate that discrepancies between ACTI compared to SELF and SMART increased with lower SE%. In contrast, SELF vs. SMART comparisons showed minimal proportional bias, with R^2 values below 0.10 across all groups.

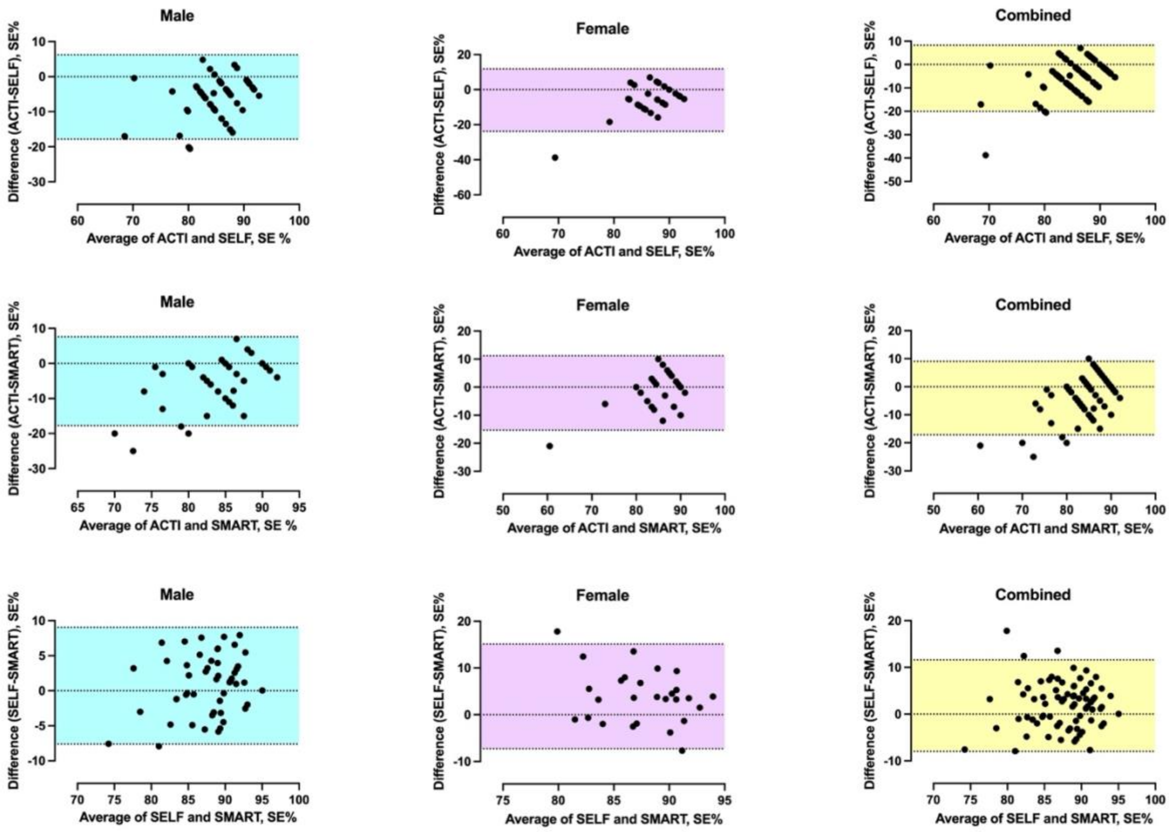

Figure 2. Agreement analysis of sleep efficiency (SE%) evaluated using Bland-Altman plots for data from ActiGraph (ACTI), smartwatch (SMART), and self-reported (SELF) data in male and female master endurance athletes. $*p < 0.05$ for the linear regression analysis indicating the presence of proportional bias.

4. Discussion

In this study, we compared the sleep metrics recorded by smartwatch (SMART) and self-reported (SELF) methods to those recorded by research-grade actigraphy (ActiGraph GTX9; ACTI) in masters endurance athletes, examining TST, SE%, and nocturnal awakenings. Across both sexes, ACTI consistently yielded lower values for TST and SE%, while SELF and SMART tended to yield higher estimates. Furthermore, nocturnal awakenings in SELF were notably fewer than those recorded by ACTI. Although SMART demonstrated slightly better agreement with ACTI than SELF in predicting TST and SE%, both SELF and SMART methods displayed notable biases in their lack of agreement. In exploratory analyses, neither age nor model of SMART influenced any of these main findings.

ACTI recorded an average sleep duration (TST) of ~5.5 hours, while both SELF and SMART indicated ~7 hours. This discrepancy is consistent with prior research on elite athletes across multiple sports (Lastella et al., 2015; Leeder et al., 2012; Sargent et al., 2014), where actigraphy often detects less than 7 hours of sleep, but self-reported methods—such as diaries (used in elite football (H. Fullagar et al., 2016; H. H. K. Fullagar et al., 2016) and rugby (Fowler et al., 2016)) or questionnaires (used in national team sports (Swinbourne et al., 2016) and elite multi-sport samples (Bleyer et al., 2015))—tend to report over 7 hours. This finding also aligns with research showing that individual-sport athletes sleep less—around 6.5 hours—compared to their team-sport counterparts, who average about 7.0 hours (Lastella et al., 2015). While the National Sleep Foundation recommends 7 to 9 hours of sleep for both young adults (18–25 years) and adults (26–64 years (Hirshkowitz et al., 2015)), athletes may require additional sleep to fully recover from both the physical and mental demands of their sport (Bird, 2013), and meeting these guidelines can be challenging for athletes due to demanding schedules, intensive training and competition (Erlacher et al., 2011; Juliff et al., 2015).

The present study simultaneously employed research-grade actigraphy and a consumer-grade wrist-worn device to assess sleep in athletes. This dual approach allowed us to directly compare the agreement of established research methodologies with the convenience of consumer devices. We observed an average discrepancy of 126 minutes in TST between research-grade and consumer-grade devices, a difference that we contend is too substantial for accurate monitoring. This discrepancy likely arises because consumer-grade devices tend to misclassify quiescent wakefulness as sleep, particularly when athletes remain still before falling asleep. Actigraphy devices are classified as research-grade wearable devices because they undergo rigorous validation against PSG, the gold standard for sleep assessment, and provide access to raw data for detailed analysis (Rundo & Downey, 2019). Actigraphy can partially classify sleep stages when compared to PSG (Yuan et al., 2024). In contrast, consumer-grade smartwatches are primarily designed for fitness and wellness tracking, relying on proprietary algorithms that are not always transparent or scientifically-validated (Shei et al., 2022). Their accuracy in detecting sleep stages varies, as they often estimate rather than directly measure sleep patterns (de Zambotti et al., 2019). As de Zambotti et al. noted (de Zambotti et al., 2024), despite some improvements, these algorithms rely on a smoothed moving average of movement data to determine sleep and wake states. Because quiescent wakefulness—when an individual is awake but not moving much—is physiologically-similar to non-rapid eye movement sleep, devices may misclassify quiet wake as sleep (de Zambotti et al., 2024).

Building on the example by Zeitzer (Zeitzer, 2024), who demonstrated that actigraphy overestimates sleep in individuals with fragmented sleep, below we describe a similar comparison to the athletic context to highlight how misclassification would differ between well-rested and poorly sleeping athletes. In well-rested athletes, the discrepancy may be minimal. For example, 432 minutes of PSG-confirmed sleep would compare favourably with 420 minutes recorded by a wrist-worn device over an 8-hour period (90% sleep efficiency). However, in athletes with disturbed sleep, the error would become more pronounced; such an athlete with 264 minutes of PSG-confirmed sleep (55% efficiency) could have their sleep time estimated at 350 minutes by the device, and thereby falsely elevating sleep efficiency to approximately 73%. These potential differences highlight the challenges in accurately detecting wakefulness in athletes and underscore the need for further refinement of sleep detection algorithms in this population.

In addition, we also detected a discrepancy of 109 minutes in TST between ACTI and SELF data. A similar finding emerged from a previous study involving female football players, which compared self-reported sleep data with actigraphy and revealed a discrepancy exceeding two hours (95% limits of agreement ranging from −1.49 to +0.40 hours (Gooderick et al., 2025)). One possible explanation for this discrepancy is social desirability bias, where individuals may overreport positive behaviours, such as sleep duration, to align with perceived societal expectations (Sultanov, 2024). This tendency has been observed in self-reports of physical activity, where social desirability influences participants to overestimate their activity levels (Adams et al., 2005). Similarly, athletes might overreport their sleep duration to present themselves favourably, leading to discrepancies between self-reported and objective sleep measures.

The presence of proportional bias in comparisons involving ACTI suggests that the degree of disagreement between ACTI and both SELF and SMART measurements varied with the duration of sleep recorded. Specifically, SELF and SMART recorded longer TST values than ACTI, particularly at shorter sleep durations. This finding is supported by the moderate R^2 values (up to 0.41), indicating that up to 41% of the variability in measurement difference was explained by the average TST. Similar findings have been observed in previous validation studies. For instance, proportional bias was reported for various commercial devices when estimating TST, with discrepancies increasing at lower sleep durations, especially for wearable-derived metrics (Kainec et al., 2024). Additionally, several popular consumer wearables—including Apple Watch, Garmin, and WHOOP—have shown greater deviations from comparison methods in individuals with lower SE%, highlighting device limitations in detecting wakefulness accurately (Miller et al., 2022). These results highlight the importance of considering sleep duration when interpreting TST from athletes or monitoring that involves SELF and SMART, particularly in individuals with short sleep durations. In practical terms, those with short sleep durations are more likely to self-report longer TST than actual TST achieved.

TST and SE% did not differ significantly between sexes; however, differences between the sexes in the agreement between methods were observed. SELF and SMART exhibited strong agreement overall, with slightly higher agreement in males for TST. In contrast, SMART and ACTI demonstrated moderate agreement in females, with SE% values from SMART being approximately 2% higher than those from ACTI, but showed poor agreement in males. Similarly, for SE%, SELF and ACTI displayed moderate agreement in males but poor agreement in females, suggesting that self-reported sleep data may be less reliable for assessing sleep patterns in female endurance athletes. Additionally, SFI was higher in males, indicating greater variability in sleep distribution among men. The results suggest potential discrepancies between males and females in the accuracy of sleep measurement across different methods, consistent with previous literature (Roberts et al., 2021; Vlahoyiannis et al., 2021). These discrepancies may be influenced by biological differences in sleep patterns, differences in sleep perception, or greater variability in sleep-tracking algorithm performance across sexes (Miles et al., 2022). These findings suggest that between-sex differences in wearable-derived and self-reported sleep metrics should be considered in future research, particularly regarding algorithm calibration and self-report accuracy.

The high level of agreement observed between self-reported TST and smartwatch-measured TST could reflect contamination of the measure caused by athletes looking at their device before completing the diary, rather than reflecting true agreement between the methods. Despite this methodological consideration, the practical relevance is that athletes' sleep estimates closely match the data provided by their devices, indicating that athletes likely trust and rely on their smartwatch readings, reinforcing their confidence in the device's accuracy. As a result, athletes may not critically assess that accuracy, as their self-reported sleep largely reflects what the smartwatch records, despite our data and others observing that both consumer-grade devices and self-monitored data are not always accurate (Driller et al., 2023). While self-monitoring can raise awareness of sleep-related issues, it also poses risks of false positives and heightened anxiety (Van den Bulck, 2015). A cross-sectional study on consumer-grade device users revealed that 66% found sleep tracking useful, and 24% reported improved sleep patterns, highlighting the behavioural influence of these technologies (Maher et al., 2017). A systematic review on consumer sleep tracking revealed that among 25 studies, 52% showed a positive link between self-reported TST and next-day mood, and almost all found a positive association between SE and mood across all ages (Hickman et al., 2024). These findings underscore the importance of ensuring the accuracy of sleep wearables, as misinterpretations of sleep data could misguide exercise training and recovery decisions.

Cognitive performance is also influenced by sleep perception; individuals receiving negative sham feedback about their sleep exhibit poorer cognitive function, while those receiving positive feedback report improved mood and alertness (Gavriloff et al., 2018). Given that not all wearables achieve acceptable agreement with PSG, particularly in detecting different sleep stages, it is crucial to educate athletes on the limitations of these devices to prevent overreliance on potentially inaccurate sleep metrics. consumer-grade devices, despite advancements, remain less accurate than clinical tools due to opaque "black-box" algorithms, leading to inconsistent data and misleading recommendations (Jahrami & Pandi-Perumal, 2023). Similarly, De Zambotti et al. (2024) emphasize the need for standardized protocols, rigorous performance evaluations, and open-source analytical pipelines to validate sleep metrics across diverse populations and real-world settings (de Zambotti et al., 2024). Therefore, ensuring greater transparency, standardisation, and validation in wearable sleep technology might be practical for improving its reliability and applicability in both research and practical settings, particularly for athlete monitoring and sleep health assessment.

Key strengths of this study include a large sample size and multi-night sleep monitoring, enhancing the accuracy and representativeness of sleep measurements and providing a more comprehensive reflection of typical sleep patterns in this unique athlete population. Yet, while this study provides valuable insights into sleep tracking in masters endurance athletes, certain limitations should be noted. We did not use a gold standard method such as PSG to track sleep characteristics. However, instead of PSG, we use actigraphy as research-grade device, as the FDA has approved the use of actigraphy for assessing sleep in healthy individuals**,** and it is also recommended by the American Academy of Sleep Medicine (AASM) (Smith et al., 2018)**.** We only included the Garmin brand, as it is the most commonly used among athletes. Future studies should incorporate a variety of smartwatch brands to evaluate measurement accuracy and reliability across different devices. Although neither method is a gold standard, our study aims to highlight these discrepancies and raise awareness among sport scientists and professionals about the limitations of relying solely on sleep tracking devices. Another limitation of this study is the potential for self-report biases, such as recall errors and the tendency to overestimate sleep duration, which may have affected the accuracy of self-reported sleep measures. A related limitation is the inability to ensure that participants did not consult their Garmin smartwatch before recording their self-reported sleep data. Participants were instructed to record their subjective diary first, and then sync their devices with the Garmin Connect app, but we cannot verify that there was full compliance with this instruction. Any deviation from our suggested instructions may have introduced contamination into the SELF data, and could partially explain the high agreement observed between TST in SMART and

SELF. Therefore, we cannot rule out that some participants viewed their smartwatch data before entering sleep diary values, and thereby inflating the sleep diary-to-smartwatch agreement. In addition, ACTI may detect micro-arousals or brief movement episodes that do not necessarily correspond to consciously-perceived awakenings. As such, ACTI will often detect a higher number of nocturnal awakenings compared to self-reported sleep diaries or gold-standard PSG, particularly in individuals with fragmented or light sleep (Marino et al., 2013; Sadeh & Acebo, 2002). This distinction is important when interpreting our findings, as the relatively high number of awakenings recorded by ACTI may reflect the device's sensitivity to movement rather than clinically meaningful sleep disruptions. Although we applied basic plausibility checks across all three data sources, differences in preprocessing requirements between ACTI, SELF, and SMART methods may still introduce minor systematic bias. This issue highlights the challenge of ensuring methodological consistency when comparing devices with differing output formats and user dependencies. An additional limitation is that participants wore two devices on the same (non-dominant) wrist, which may have introduced minor discomfort or adherence variability. Although this approach was chosen to minimize interference with daily activities and maintain consistent wear location across participants, we recognize that the combined bulk or fit of both devices may have influenced device accuracy or participant compliance. Lastly, although our sleep data were collected continuously over a 7-day period, a narrative review on sleep tracking in athletes recommends longer monitoring durations (14, 21, or even 28 nights) to better capture habitual sleep patterns (Driller et al., 2023).

Considering the practical relevance of these findings, in this cohort of masters endurance athletes, both SLEEP and SMART recorded total sleep times approximately 2 h longer per night and SE% values about ~4 to 6 percentage points higher than the ACTI. Masters athletes typically have higher training loads (e.g.; 10–15 h per week) while managing full-time work and family responsibilities (Wooten et al., 2021), and either they or their coaches may rely on consumer wearables for quick feedback on recovery (Procida et al., 2024). When a device or a diary influenced by that device reports longer sleep or higher efficiency than what objective monitoring shows, athletes and coaches may assume recovery is adequate, and postpone adjustments to training load, nutrition, or sleep hygiene strategies. Short or fragmented sleep in endurance populations is linked to higher perceived stress, reduced training quality, and a greater incidence of illness during demanding training blocks (Hausswirth et al., 2014). Given that age-related changes can already slow physiological recovery (Reaburn et al., 2019), feedback that paints an overly favourable picture of sleep may hasten maladaptation, dampen performance, and threaten long-term participation. Our findings therefore highlight the need to validate consumer devices in this population and to educate athletes and practitioners about their limitations before using them to guide exercise training decisions.

5. Conclusions

In this study of masters endurance athletes, consumer-grade smartwatch (Garmin brand) and self-reported sleep measurements systematically recorded longer TST and higher SE% values compared to research-grade actigraphy (ActiGraph GT9X), with both methods exhibiting systematic biases that limit their accuracy in sleep tracking, especially in athletes with shorter sleep durations. Agreement between methods varied between male and female athletes, which underscores the importance of evaluating between-sex differences in the accuracy of sleep tracking in athletes. Because consumer wearables and self-reported sleep diaries recorded longer sleep and higher sleep efficiency than research-grade actigraphy in masters endurance athletes, the former methods may give a misleading impression of recovery and readiness to train; validation of these tools and user education around these concepts are therefore required before relying on them to guide exercise training. Overall, these findings highlight the need for caution when using consumer-grade devices or self-reported sleep diaries to assess sleep in athletic populations, given their tendency to report longer sleep durations and higher efficiency values while failing to capture wake episodes accurately.

Author Contributions

ADL: Conceptualization, Methodology, Investigation, Project Administration, Formal Analysis, Writing – Original Draft; **SD**: Methodology, Project Administration **BE**: Conceptualization, Methodology, Writing – Review & Editing, Supervision. All authors read and approved the final version of the manuscript.

References

Adams, S. A., Matthews, C. E., Ebbeling, C. B., Moore, C. G., Cunningham, J. E., Fulton, J., & Hebert, J. R. (2005). The Effect of Social Desirability and Social Approval on Self-Reports of Physical Activity. *American Journal of Epidemiology*, *161*(4), 389–398. https://doi.org/10.1093/aje/kwi054

Baron, K. G., Abbott, S., Jao, N., Manalo, N., & Mullen, R. (2017). Orthosomnia: Are Some Patients Taking the Quantified Self Too Far? *Journal of Clinical Sleep Medicine : JCSM : Official Publication of the American Academy of Sleep Medicine*, *13*(2), 351–354. https://doi.org/10.5664/jcsm.6472

Bird, S. P. (2013). Sleep, Recovery, and Athletic Performance: A Brief Review and Recommendations. *Strength & Conditioning Journal*, *35*(5), 43. https://doi.org/10.1519/SSC.0b013e3182a62e2f

Bleyer, F. T. de S., Barbosa, D. G., Andrade, R. D., Teixeira, C. S., & Felden, É. P. G. (2015). Sleep and musculoskeletal complaints among elite athletes of Santa Catarina. *Revista Dor*, *16*, 102–108. https://doi.org/10.5935/1806-0013.20150020

Bonilla, D. A., Pérez-Idárraga, A., Odriozola-Martínez, A., & Kreider, R. B. (2021). The 4R's Framework of Nutritional Strategies for Post-Exercise Recovery: A Review with Emphasis on New Generation of Carbohydrates. *International Journal of Environmental Research and Public Health*, *18*(1), Article 1. https://doi.org/10.3390/ijerph18010103

Charest, J., & Grandner, M. A. (2020). Sleep and Athletic Performance: Impacts on Physical Performance, Mental Performance, Injury Risk and Recovery, and Mental Health. *Sleep Medicine Clinics*, *15*(1), 41. https://doi.org/10.1016/j.jsmc.2019.11.005

de Zambotti, M., Cellini, N., Goldstone, A., Colrain, I. M., & Baker, F. C. (2019). Wearable Sleep Technology in Clinical and Research Settings. *Medicine and Science in Sports and Exercise*, *51*(7), 1538–1557. https://doi.org/10.1249/MSS.0000000000001947

de Zambotti, M., Goldstein, C., Cook, J., Menghini, L., Altini, M., Cheng, P., & Robillard, R. (2024). State of the science and recommendations for using wearable technology in sleep and circadian research. *Sleep*, *47*(4), zsad325. https://doi.org/10.1093/sleep/zsad325

Doherty, R., Madigan, S. M., Nevill, A., Warrington, G., & Ellis, J. G. (2021). The Sleep and Recovery Practices of Athletes. *Nutrients*, *13*(4), 1330. https://doi.org/10.3390/nu13041330

dos Santos, N. E., Daniel, N. V. S., Franco, B., Bastos, A. M., Belli, T., & Esteves, A. M. (2022). Sleep and nutritional profile of endurance and ultra-endurance running athletes. *Sleep Science*, *15*(4), 441–447. https://doi.org/10.5935/1984-0063.20220076

Driller, M. W., Dunican, I. C., Omond, S. E. T., Boukhris, O., Stevenson, S., Lambing, K., & Bender, A. M. (2023). Pyjamas, Polysomnography and Professional Athletes: The Role of Sleep Tracking Technology in Sport. *Sports*, *11*(1), Article 1. https://doi.org/10.3390/sports11010014

Erlacher, D., Ehrlenspiel, F., Adegbesan, O. A., & El-Din, H. G. (2011). Sleep habits in German athletes before important competitions or games. *Journal of Sports Sciences*. https://www.tandfonline.com/doi/abs/10.1080/02640414.2011.565782

Fowler, P. M., Duffield, R., Lu, D., Hickmans, J. A., & Scott, T. J. (2016). Effects of Long-Haul Transmeridian Travel on Subjective Jet-Lag and Self-Reported Sleep and Upper Respiratory Symptoms in Professional Rugby League Players. *International Journal of Sports Physiology and Performance*, *11*(7), 876–884. https://doi.org/10.1123/ijspp.2015-0542

Fullagar, H., Duffield, R., Skorski, S., White, D., Bloomfield, J., Kölling, S., & Meyer, T. (2016). Sleep, Travel, and Recovery Responses of National Footballers During and After Long-Haul International Air Travel. *International Journal of Sports Physiology and Performance*, *11*(1), 86–95. https://doi.org/10.1123/ijspp.2015-0012

Fullagar, H. H. K., Skorski, S., Duffield, R., Julian, R., Bartlett, J., & Meyer, T. (2016). Impaired sleep and recovery after night matches in elite football players. *Journal of Sports Sciences*. https://www.tandfonline.com/doi/abs/10.1080/02640414.2015.1135249

Gavriloff, D., Sheaves, B., Juss, A., Espie, C. A., Miller, C. B., & Kyle, S. D. (2018). Sham sleep feedback delivered via actigraphy biases daytime symptom reports in people with insomnia: Implications for insomnia disorder and wearable devices. *Journal of Sleep Research*, *27*(6), e12726. https://doi.org/10.1111/jsr.12726

Gooderick, J., Wood, T., Abbott, W., Hayes, M., & Maxwell, N. (2025). Does a self-reported sleep duration reflect actigraphy reported sleep duration in female football players? *Science and Medicine in Football*. https://www.tandfonline.com/doi/abs/10.1080/24733938.2023.2297903

Hamlin, M. J., Deuchrass, R. W., Olsen, P. D., Choukri, M. A., Marshall, H. C., Lizamore, C. A., Leong, C., & Elliot, C. A. (2021). The Effect of Sleep Quality and Quantity on Athlete's Health and Perceived Training Quality. *Frontiers in Sports and Active Living*, *3*. https://doi.org/10.3389/fspor.2021.705650

Hausswirth, C., Louis, J., Aubry, A., Bonnet, G., Duffield, R., & LE Meur, Y. (2014). Evidence of disturbed sleep and increased illness in overreached endurance athletes. *Medicine and Science in Sports and Exercise*, *46*(5), 1036–1045. https://doi.org/10.1249/MSS.0000000000000177

Hickman, R., D'Oliveira, T. C., Davies, A., & Shergill, S. (2024). Monitoring Daily Sleep, Mood, and Affect Using Digital Technologies and Wearables: A Systematic Review. *Sensors*, *24*(14), Article 14. https://doi.org/10.3390/s24144701

Hirshkowitz, M., Whiton, K., Albert, S. M., Alessi, C., Bruni, O., DonCarlos, L., Hazen, N., Herman, J., Katz, E. S., Kheirandish-Gozal, L., Neubauer, D. N., O'Donnell, A. E., Ohayon, M., Peever, J., Rawding, R., Sachdeva, R. C., Setters, B., Vitiello, M. V., Ware, J. C., & Hillard, P. J. A. (2015). National Sleep Foundation's sleep time duration recommendations: Methodology and results summary. *Sleep Health: Journal of the National Sleep Foundation*, *1*(1), 40–43. https://doi.org/10.1016/j.sleh.2014.12.010

HOME. (n.d.). World Masters Athletics. Retrieved 6 June 2025, from https://world-masters-athletics.org/

Jahrami, H., & Pandi-Perumal, S. R. (2023). Smart, Personalized Sleep: The Benefits and Risks of Artificial Intelligence and Wearables in the Consumer Sleep Technology Market. *Sleep and Vigilance*, *7*(1), 105–107. https://doi.org/10.1007/s41782-023-00230-x

Juliff, L. E., Halson, S. L., & Peiffer, J. J. (2015). Understanding sleep disturbance in athletes prior to important competitions. *Journal of Science and Medicine in Sport*, *18*(1), 13–18. https://doi.org/10.1016/j.jsams.2014.02.007

Kainec, K. A., Caccavaro, J., Barnes, M., Hoff, C., Berlin, A., & Spencer, R. M. C. (2024). Evaluating Accuracy in Five Commercial Sleep-Tracking Devices Compared to Research-Grade Actigraphy and Polysomnography. *Sensors*, *24*(2), Article 2. https://doi.org/10.3390/s24020635

Kim, K., Park, D.-Y., Song, Y. J., Han, S., & Kim, H. J. (2022). Consumer-grade sleep trackers are still not up to par compared to polysomnography. *Sleep and Breathing*, *26*(4), 1573–1582. https://doi.org/10.1007/s11325-021-02493-y

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

Lastella, M., Roach, G. D., Halson, S. L., & Sargent, C. (2015). Sleep/wake behaviours of elite athletes from individual and team sports. *European Journal of Sport Science*, *15*(2), 94–100. https://doi.org/10.1080/17461391.2014.932016

Leeder, J., Glaister ,Mark, Pizzoferro ,Kathleen, Dawson ,Jean, & and Pedlar, C. (2012). Sleep duration and quality in elite athletes measured using wristwatch actigraphy. *Journal of Sports Sciences*, *30*(6), 541–545. https://doi.org/10.1080/02640414.2012.660188

Liu, F., Schrack, J., Wanigatunga, S. K., Rabinowitz, J. A., He, L., Wanigatunga, A. A., Zipunnikov, V., Simonsick, E. M., Ferrucci, L., & Spira, A. P. (2023). Comparison of sleep parameters from wrist-worn ActiGraph and Actiwatch devices. *Sleep*, *47*(2), zsad155. https://doi.org/10.1093/sleep/zsad155

Maher, C., Ryan, J., Ambrosi, C., & Edney, S. (2017). Users' experiences of wearable activity trackers: A cross-sectional study. *BMC Public Health*, *17*(1), 880. https://doi.org/10.1186/s12889-017-4888-1

Marino, M., Li, Y., Rueschman, M. N., Winkelman, J. W., Ellenbogen, J. M., Solet, J. M., Dulin, H., Berkman, L. F., & Buxton, O. M. (2013). Measuring Sleep: Accuracy, Sensitivity, and Specificity of Wrist Actigraphy Compared to Polysomnography. *Sleep*, *36*(11), 1747–1755. https://doi.org/10.5665/sleep.3142

Miles, K. H., Clark, B., Fowler, P. M., Miller, J., & Pumpa, K. L. (2022). What are the sleep characteristics of elite female athletes? A systematic review with meta-analysis. *Biology of Sport*, *39*(3), 751–763. https://doi.org/10.5114/biolsport.2022.108705

Miller, D. J., Sargent, C., & Roach, G. D. (2022). A Validation of Six Wearable Devices for Estimating Sleep, Heart Rate and Heart Rate Variability in Healthy Adults. *Sensors*, *22*(16), Article 16. https://doi.org/10.3390/s22166317

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

Parrott, A. C., & Hindmarch, I. (1980). The Leeds Sleep Evaluation Questionnaire in psychopharmacological investigations—A review. *Psychopharmacology*, *71*(2), 173–179. https://doi.org/10.1007/BF00434408

Procida, D., Mara, J., Mitchell, L., & Etxebarria, N. (2024). How Do Age-Group Triathlon Coaches Manage Training Load? A Pilot Study. *Sports*, *12*(9), Article 9. https://doi.org/10.3390/sports12090261

Reaburn, P., Doering, T., & Borges, N. (2019). Masters athletes take longer to recover from high intensity exercise than training- matched younger athletes. Does increased protein intake enhance recovery? *Journal of Science and Medicine in Sport*, *22*, S32–S33. https://doi.org/10.1016/j.jsams.2019.08.198

Roberts, S. S. H., Falkenberg, E., Stevens, A., Aisbett, B., Lastella, M., & Condo, D. (2021). The Sleep of Elite Australian Rules Footballers During Preseason: A Comparison of Men and Women. *International Journal of Sports Physiology and Performance*, *16*(5), 641–646. https://doi.org/10.1123/ijspp.2020-0340

Rundo, J. V., & Downey, R. (2019). Chapter 25—Polysomnography. In K. H. Levin & P. Chauvel (Eds.), *Handbook of Clinical Neurology* (Vol. 160, pp. 381–392). Elsevier. https://doi.org/10.1016/B978-0-444-64032-1.00025-4

Sadeh, A. (2011). The role and validity of actigraphy in sleep medicine: An update. *Sleep Medicine Reviews*, *15*(4), 259–267. https://doi.org/10.1016/j.smrv.2010.10.001

Sadeh, A., & Acebo, C. (2002). The role of actigraphy in sleep medicine. *Sleep Medicine Reviews*, *6*(2), 113–124. https://doi.org/10.1053/smrv.2001.0182

Sargent, C., Lastella ,Michele, Halson ,Shona L., & and Roach, G. D. (2014). The impact of training schedules on the sleep and fatigue of elite athletes. *Chronobiology International*, *31*(10), 1160–1168. https://doi.org/10.3109/07420528.2014.957306

Shei, R.-J., Holder, I. G., Oumsang, A. S., Paris, B. A., & Paris, H. L. (2022). Wearable activity trackers–advanced technology or advanced marketing? *European Journal of Applied Physiology*, *122*(9), 1975–1990. https://doi.org/10.1007/s00421-022-04951-1

Smith, M. T., McCrae, C. S., Cheung, J., Martin, J. L., Harrod, C. G., Heald, J. L., & Carden, K. A. (2018). Use of Actigraphy for the Evaluation of Sleep Disorders and Circadian Rhythm Sleep-Wake Disorders: An American Academy of Sleep Medicine Clinical Practice Guideline. *Journal of Clinical Sleep Medicine: JCSM: Official Publication of the American Academy of Sleep Medicine*, *14*(7), 1231–1237. https://doi.org/10.5664/jcsm.7230

Snyder-Halpern, R., & Verran, J. A. (1987). Instrumentation to describe subjective sleep characteristics in healthy subjects. *Research in Nursing & Health*, *10*(3), 155–163. https://doi.org/10.1002/nur.4770100307

Sultanov, M. B. (2024). The Association Between Social Desirability and Competitive Anxiety in Young Football Players with Different Qualifications. *Psychology in Russia: State of the Art*, *17*(3), 114–124. https://doi.org/10.11621/pir.2024.0308

Swinbourne, R., Gill, N., Vaile, J., & Smart, D. (2016). Prevalence of poor sleep quality, sleepiness and obstructive sleep apnoea risk factors in athletes. *European Journal of Sport Science*, *16*(7), 850–858. https://doi.org/10.1080/17461391.2015.1120781

Van den Bulck, J. (2015). Sleep apps and the quantified self: Blessing or curse? *Journal of Sleep Research*, *24*(2), 121–123. https://doi.org/10.1111/jsr.12270

Vlahoyiannis, A., Aphamis, G., Bogdanis, G. C., Sakkas, G. K., Andreou, E., & Giannaki, C. D. (2021). Deconstructing athletes' sleep: A systematic review of the influence of age, sex, athletic expertise, sport type, and season on sleep characteristics. *Journal of Sport and Health Science*, *10*(4), 387–402. https://doi.org/10.1016/j.jshs.2020.03.006

Walsh, N. P., Halson, S. L., Sargent, C., Roach, G. D., Nédélec, M., Gupta, L., Leeder, J., Fullagar, H. H., Coutts, A. J., Edwards, B. J., Pullinger, S. A., Robertson, C. M., Burniston, J. G., Lastella, M., Meur, Y. L., Hausswirth, C., Bender, A. M., Grandner, M. A., & Samuels, C. H. (2021). Sleep and the athlete: Narrative review and 2021 expert consensus recommendations. *British Journal of Sports Medicine*, *55*(7), 356–368. https://doi.org/10.1136/bjsports-2020-102025

Wooten, S. V., Mittag, U., Alvero Cruz, J. R., Stray-Gundersen, S., Hoffmann, F., Michély, S., Möstl, S., Sies, W., Mulder, E., Rauschendorfer, P., Chilibeck, P. D., Rittweger, J., & Tanaka, H. (2021). Life Satisfaction, Positive Affect, and Sleep Impairment in Masters Athletes: Modulation by Age, Sex, and Exercise Type. *Frontiers in Physiology*, *12*. https://doi.org/10.3389/fphys.2021.634433

Yuan, H., Hill, E. A., Kyle, S. D., & Doherty, A. (2024). A systematic review of the performance of actigraphy in measuring sleep stages. *Journal of Sleep Research*, *33*(5), e14143. https://doi.org/10.1111/jsr.14143

Zeitzer, J. M. (2024). The utility of sleep wearables—Well, that depends. *Sleep*, *47*(4), zsae019. https://doi.org/10.1093/sleep/zsae019