

# Performance evaluation of an under-mattress sleep sensor versus polysomnography in $\geq 400$ nights with healthy and unhealthy sleep

Jack Manners<sup>1</sup>, Eva Kemps<sup>2</sup>, Bastien Lechat<sup>1</sup>, Peter Catcheside<sup>1</sup>, \*Danny Eckert<sup>1</sup>,  
and \*Hannah Scott<sup>1</sup>

<sup>1</sup>Flinders Health and Medical Research Institute: Sleep Health, Flinders University

<sup>2</sup>College of Education, Psychology, and Social Work, Flinders University

<sup>3</sup>College of Medicine and Public Health, Flinders University

July 1, 2024

\* Co-senior authors

## **Acknowledgements:**

The authors thank and acknowledge the support from our colleagues at Flinders University and FHMRI: Sleep Health. The authors also thank all participants involved in this study.

**NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.**

## Abstract

Consumer sleep trackers can provide useful insight into sleep and sleep patterns. However, large scale performance evaluation studies against direct sleep measures are needed to comprehensively understand sleep tracker accuracy. This study evaluated performance of an under-mattress sensor to estimate sleep and wake versus polysomnography, during multiple in-laboratory protocols in a large sample including individuals with and without sleep disorders and during day versus night sleep opportunities.

183 participants (51% male, mean[SD] age=45[18] years) attended the sleep laboratory for a research study that included simultaneous polysomnography and under-mattress sensor (Withings Sleep Analyzer [WSA]) recordings. Epoch-by-epoch analyses with confusion matrices were used to determine accuracy, sensitivity, and specificity of the WSA versus polysomnography. Bland-Altman plots examined bias in sleep duration, efficiency, onset-latency, and wake after sleep onset.

Overall WSA sleep-wake classification accuracy was 83%, sensitivity 95%, and specificity 37%. The WSA significantly overestimated total sleep time (48[81]minutes), Sleep efficiency (9[15]%), sleep onset latency (6[26]), and underestimated wake after sleep onset (54[78]),  $p < 0.05$ . Accuracy and specificity were higher for night versus daytime sleep opportunities in healthy individuals (89% and 47% versus 82% and 26% respectively,  $p < 0.05$ ). Accuracy and sensitivity were also higher for healthy individuals (89% and 97%) versus those with sleep disorders (81% and 91%,  $p < 0.05$ ).

WSA performance is comparable to other consumer sleep trackers, with high sensitivity but poor specificity compared to polysomnography. Poorer accuracy and specificity during daytime versus night-time sleep opportunities is likely due to increased wake time and reduced sleep efficiency. Contactless, under-mattress sleep sensors show promise for accurate sleep monitoring, noting the tendency to over-estimate sleep particularly where wake time is high.

## Introduction

Sleep is critical for optimal daytime function, performance, safety, and health. Thus, accurate and reliable estimates of sleep are important to help manage these key outcomes. However, sleep is difficult to objectively evaluate, particularly over extended multi-day periods. This is partly because gold-standard quantification of sleep, polysomnography, is expensive and complex to administer over multiple nights. Wearable and "nearable" (devices that are not in direct contact with an individual) sleep tracking devices are used to simplify sleep estimation in the home, such as wrist-based actigraphy, bedside radar, or mattress sensor devices [1, 2]. Deciding which devices are best suited for sleep estimation requires critical appraisal of device performance and practicality. Accordingly, the present study evaluated the performance of an under-mattress device, the Withings Sleep Analyzer (WSA), to evaluate sleep in a diverse cohort and various sleep opportunities (day and night).

To help circumvent cost and measurement complexity for multi-day assessments, wearable sleep trackers, such as actigraphy based devices, are often used to infer wake and sleep from body movements. This approach is sensitive to detect sleep but cannot reliably discriminate wake from sleep when people lie still awake [3]. Additional signals such as heart rate, heart rate variability, and breathing can improve device performance, but poor specificity remains a key limitation of this technology. Another problem is that wearable devices require the user to ensure the device is charged and properly worn. Consequently, data loss is common [4], variable [5], and performance can drastically drop when devices are not worn correctly [6]. "Nearable" sleep trackers that infer sleep from detected motion, including respiratory and cardiac motion, avoid the need to charge or wear any device. These devices are typically designed to be placed on or under the mattress, or at the bedside, with little subsequent intervention required once properly set up [1, 7]. However, some evidence suggests that these devices may be less accurate than wearable counterparts [8]. Thus, rigorous performance evaluation is essential to determine device reliability, practical benefits, and sleep tracking performance.

The WSA has been previously validated in smaller trials (N=18-118 [49-118 nights]) against polysomnography to estimate sleep and identify breathing disturbances [8-11]. The device uses a pneumatic sensor placed under the mattress to detect air-pressure changes, from which movement, sleep, respiration and heart rate are inferred. The device is attractive for both clinical

and research use, as once set up and connected to Wi-Fi and mains power, minimal ongoing user input is required. Indeed, recent large scale, long-term monitoring studies that have used this technology have been able to address key sleep health questions that were not previously possible with polysomnography or existing wearables [11–15]. Furthermore, this technology could markedly reduce data loss compared to wearable sleep trackers that require regular device charging and correct daily wear. The unobtrusive nature of this approach is also potentially amenable to patient/participant monitoring in environments less conducive to wearable devices such as hospitals, nursing homes, and in high-risk on-call workplace settings. Given poor specificity of sleep tracking devices, sleep classification may be less accurate in these environments as sleep is more likely to be impaired, compared to in the home or laboratory. Despite this, few devices have been rigorously evaluated across multi-night sleeps, and fewer devices have been assessed to any degree with non-standard sleep schedules. Device accuracy under these conditions is particularly important if they are to be used for readiness evaluation in shift workers, or to examine clinical outcomes in people with sleep disorders. Therefore, comprehensive performance evaluation of the WSA during sleep opportunities of varied timing and in people with and without sleep disordersd remains important to establish the potential utility of nearable sleep trackers for extended multi-day monitoring in more challenging sleep assessment settings.

This study used comprehensive objective measurements of sleep in several laboratory sleep research studies to evaluate the classification accuracy of the WSA compared to polysomnography. Secondary aims explored how classification accuracy 1) was affected by the timing of the sleep opportunity, 2) differed in people with versus without sleep disorders, 3) was explained by polysomnography sleep efficiency, and 4) compared to a validated consumer wearable, the Fitbit device.

## Methods

Data were utilized from 13 studies conducted at Flinders Health and Medical Research Institute: Sleep Health, from 2021-2023, where data from the WSA was collected during the sleep study. These comprised of ten studies with between 1-3 nights in the sleep laboratory for participants with suspected or diagnosed obstructive sleep apnea (OSA), insomnia, co-morbid insomnia and OSA (COMISA), cardiovascular disease, or general sleep complaint; two studies with single-night laboratory visits with healthy volunteers; and one study with two 8-day laboratory visits comprised of one nighttime sleep followed by five daytime sleep opportunities. Across all research studies, 416 sleep recordings (224 nighttime and 192 daytime recordings) with simultaneous polysomnography and WSA recordings were available for analysis. Data were time-matched based on the clock-times in the polysomnography and WSA recordings. Subsets of data were used to address secondary aims. Specifically, only data from healthy participants were available for the time of recording comparisons (given that only one study collected data from daytime sleep opportunities), only nighttime recordings were available for the healthy versus sleep disorder comparisons, and only one study of healthy participants had data available from both the WSA and the Fitbit to enable device comparisons. Detailed study information can be found in Table S1.

## Equipment

### Polysomnography

Polysomnography was collected during all sleep opportunities. Polysomnography setups were conducted in accordance with the standard 10-20 electroencephalography electrode placement system, using Compumedics Grael 4K PSG:EEG devices (Compumedics, Victoria, Australia). Sleep studies were independently scored using Profusion Compumedics software (v 4.0) according to standardized American Academy of Sleep Medicine polysomnography scoring criteria [16]. Polysomnography sleep stages were extracted as wake, rapid eye movement (REM) sleep, stage 1 sleep (N1), stage 2 sleep (N2), and stage 3 sleep (N3).

## **Withings Sleep Analyzer**

The WSA is an under-mattress device that uses a pneumatic sensor to detect changes in pressure in an air-bladder relative to atmospheric pressure (i.e., ballistography). The device uses this information to infer movement, from which respiration, heart rate and sleep stages are estimated via proprietary algorithms. The WSA was placed under the mattress, level with the chest of the sleeping individual. WSA-derived total sleep time, sleep efficiency, and wake after sleep onset has been validated, compared to polysomnography [8, 9, 11, 17], but extensive performance evaluation of sleep characteristics and sleep staging accuracy is lacking.

## **Fitbit Charge 4**

The Fitbit Charge 4 is a wrist-worn device that contains a tri-axial accelerometer to track movement (i.e., actigraphy) and a photoplethysmography sensor to estimate heart rate. This information is used to infer sleep stages and wake. Fitbit devices were placed on each participant's non-dominant wrist, with appropriate band-sized chosen to ensure proper fit. The Charge 4 model has been validated against polysomnography [18], and earlier Fitbit models similarly show reasonable accuracy to detect sleep compared to polysomnography. Individual validation studies show such devices typically overestimate total sleep time by 30-60 minutes and sleep efficiency by 5-10%, and underestimate wake after sleep onset by 20-60 minutes [19–21]. Meta-analyses suggest that newer Fitbit devices may not significantly differ from polysomnography in total sleep time, sleep efficiency, and wake after sleep onset [22], yet the latest validated devices still show poor specificity (e.g., 52% of wake correctly identified [18]). Accordingly, Fitbit devices are generally at least on-par with highly validated research-grade actiwatch devices [22, 23].

## **Statistical Analysis**

WSA performance was evaluated based on recommended guidelines [24, 25]. Device data (WSA and Fitbit) were extracted via custom software at [www.snapi.space](http://www.snapi.space), developed in Python (v3.11). Polysomnography data were extracted as European Data Format files using Python, and time-matched to WSA and Fitbit data to provide concurrent epoch-by-epoch data. As the WSA device only provides sleep stage classification data 60 second intervals, Fitbit and polysomnography data

were converted from 30s to 60s epochs. Where combined epochs differed, wake was scored if present in either, otherwise the first value was used [26]. Lights on and off were derived from polysomnography sleep reports to determine sleep opportunities. Where device data started or ended within these limits, preceding and trailing epochs were designated as wake.

Total sleep time (TST) was calculated as the sum of sleep epochs within the sleep opportunity. Sleep efficiency (SE) was calculated as the sum of sleep epochs divided by the total number of epochs within the sleep opportunity. Sleep onset latency (SOL) was calculated as the sum of wake epochs before the first sleep epoch, within the sleep opportunity. Wake after sleep onset (WASO) was calculated as the sum of wake epochs between the first and last sleep epoch within the sleep opportunity. Wake and sleep classification performance was determined for each sleep recording as accuracy (proportion of correctly scored epochs), sensitivity (proportion of polysomnography-derived sleep epochs that the WSA correctly scored as sleep), and specificity (proportion of polysomnography-derived wake epochs that the WSA correctly scored as wake), compared to polysomnography. Additionally, four-stage ('wake', 'light', 'deep', 'REM') sleep classification from the WSA was compared to polysomnography sleep stages, where polysomnography N2 and N3 sleep epochs were combined as 'deep' sleep.

Linear mixed model (LMM) analyses examined 1) performance (WSA vs polysomnography) by time of recording (daytime vs nighttime) in healthy participants, 2) performance (WSA vs polysomnography) by sleep disorder status (healthy sleep vs sleep disorder), and 3) performance by consumer sleep tracker ([WSA vs polysomnography] vs [Fitbit vs polysomnography]). LMM analyses were performed using the lme4 package [27] (R v4.2.2) to examine how sleep estimates (TST, SOL, WASO, SE) and WSA performance (accuracy, sensitivity, and specificity) were affected by these three fixed effects. Participant ID was entered as random effects in all models. Marginal  $R^2$  was calculated to estimate variance explained by fixed effects. Secondary analysis were conducted using mixed models adjusted for age, BMI and polysomnography-derived sleep efficiency to examine potential confounders in the relationship between sleep opportunity timing, sleep disorder status, and WSA performance.

Bland-Altman plots were used to examine bias and limits of agreement (LOAs), calculated as  $\pm 1.96$  the standard deviation of mean differences and their 95% confidence limits, between polysomnography and device-derived estimates of TST, SE, WASO and SOL for each recording. Proportional bias was also calculated and tested, where significant bias indicated that the mean

difference between device and polysomnography increased or decreased as a function of the size of measurement [28].

For participants with more than one recording, the coefficient of variation (CV; SD divided by mean) for accuracy, sensitivity, and specificity were calculated for each individual. To examine differences in performance variability across sleep disorder statuses and sleep opportunity timings, group differences in mean CV and the distribution of CVs were compared using linear regression and Levene tests, respectively.

Finally, in the subset where both Fitbit and WSA devices were used, data-loss was quantified and compared using a paired-samples *t*-test. Post-hoc comparisons with Bonferroni corrections were conducted where main effects were significant. All data are reported as mean (SD) unless otherwise specified.  $p < .05$  was considered statistically significant.

## Results

Final data included 416 recordings from 183 participants, collected across 13 sleep research studies at Flinders Health and Medical Research Institute: Sleep Health. Participant demographics can be found in Table 1.

Table 1: Participant demographics.

	Healthy Sleep	Sleep Disorder	Total
Sample size	82	101	183
Nighttime recordings	96	128	224
Daytime recordings	192	0	192
Age (years), mean (SD)	30.9 (11.9)	56.5 (14.2)	45 (18.4)
Sex, n (%) male	35 (43)	58 (57)	94 (51)
n (%) female	47 (57)	43 (43)	89 (49)
BMI (kg/m <sup>2</sup> ), mean (SD)	24 (4)	31.7 (10.4)	28.3 (9)
AHI (events/hour), mean (SD)	2.9 (7.1)	25.9 (26.6)	16.2 (23.6)

Note: For a full breakdown of sleep disorders refer to Table S4.  
AHI=apnea hypopnea index; BMI=body mass index; SD=standard deviation

### Accuracy, Sensitivity and Specificity

The WSA had high overall accuracy and sensitivity but moderately poor specificity, compared to polysomnography, seen in Table 2. LMMs showed significantly lower mean (standard error [SEM]) accuracy of 7% ( $\pm 1.4\%$ ),  $F(272)=25.2$ ,  $p < .001$ , marginal  $R^2=0.08$ , and specificity of 19.7% ( $\pm 2.5\%$ ),  $F(277.2)=64.2$ ,  $p < .001$ , marginal  $R^2=0.18$ , during daytime sleep opportunities compared



to nighttime sleep opportunities. There was no significant difference in sensitivity ( $p > .05$ ). LMMs also showed significantly higher accuracy of 7.2% ( $\pm 1.7\%$ ),  $F(181)=18.5$ ,  $p < .001$ , marginal  $R^2=0.08$ , and sensitivity of 4.4% ( $\pm 1.5\%$ ),  $F(190.9)=8.6$ ,  $p = .004$ , marginal  $R^2=0.04$ , for healthy sleepers compared to those with a diagnosed or suspected sleep disorder.

Table 2: Epoch-by-epoch performance of Withings Sleep Analyzer versus polysomnography.

	Accuracy %	Sensitivity %	Specificity %
<b>Healthy Sleeps</b>			
Nighttime	88.7 (7.8)	97.2 (3.3)	46.7 (19.9)
Daytime	81.8 (12)	96.5 (6.6)	26 (20.2)
<b>Sleep Disorders</b>			
Nighttime	80.6 (13.3)	91.4 (12.6)	46 (26.3)
<b>Overall</b>	83 (12)	95.1 (8.8)	36.9 (24.3)

Note: There were no data available for daytime sleep opportunities in patients with sleep disorders.

Fully adjusted models (Table 3) show that the group effect of sleep disorder status on accuracy was fully explained by age, BMI and polysomnography-derived sleep efficiency. The effect on specificity was partially explained by age. The group effect of nighttime versus daytime sleep opportunity on accuracy was partially explained by age and sleep efficiency. The effect on specificity was partially explained by sleep efficiency.

Table 3: Epoch-by-epoch performance of Withings Sleep Analyzer versus polysomnography.

Predictor	Accuracy %		Sensitivity %		Specificity %	
	Estimate	p	Estimate	p	Estimate	p
<b>Sleep Disorder Status</b>						
Sleep Disorder	1.95	.271	4.38	<b>.031</b>	-9.73	<b>.046</b>
Age	-0.10	<b>.042</b>	-0.01	.844	-0.41	<b>.002</b>
Body Mass Index	-0.18	<b>.018</b>	-0.12	.142	-0.12	.550
Sleep Efficiency	0.49	<b>&lt;.001</b>	-0.08	.0121	0.02	.896
<b>Sleep Opportunity Timing</b>						
Daytime	-5.50	<b>&lt;.001</b>	-0.86	.252	-20.95	<b>&lt;.001</b>
Age	-0.16	<b>.002</b>	-0.00	.941	-0.31	.071
Body mass index	0.12	.328	0.09	.385	0.06	.896
Sleep Efficiency	0.56	<b>&lt;.001</b>	0.02	.458	-0.20	<b>.020</b>

## Sleep Characteristics

Sleep characteristics as measured by polysomnography and estimated by the WSA are found in Table 4, including subsets of data for secondary analyses. Overall, the WSA significantly overestimated TST, SE, SOL, and significantly underestimated WASO, compared to polysomnography. This was also reflected in sleep stages, where light, deep and REM sleep are

typically underestimated, and wake is overestimated.

Participants had lower TST, SE, SOL and higher WASO in daytime sleep opportunities compared to nighttime sleep opportunities, and healthy participants had higher TST, SE, and lower WASO compared to those with a diagnosed or suspected sleep disorder. There were significant device by sleep opportunity timing interactions for all but SOL and REM sleep duration, where metrics were over or underestimated to a greater degree during daytime sleep opportunities. There were also significant device by sleep disorder status for light, deep, and REM sleep duration, where light sleep was overestimated and deep and REM sleep was underestimated to a greater degree in individuals with a sleep disorder, compared to healthy individuals. A detailed breakdown of interactions can be found in Supplementary Table [S4](#).

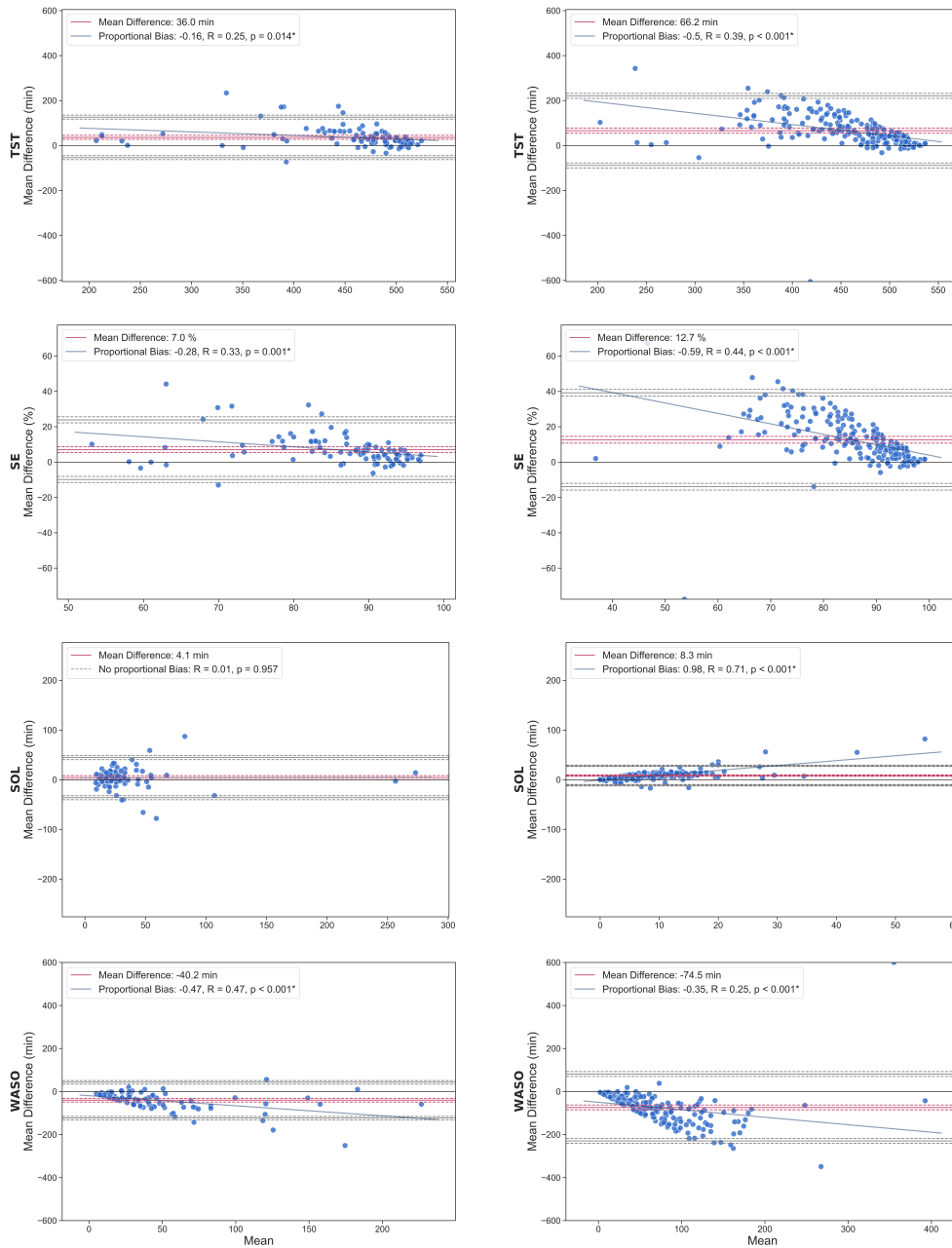
Bland-Altman plots in Figure [1](#) highlight the significant increase in mean bias for TST and SE estimates, and decrease for WASO estimates, during daytime versus nighttime sleep opportunities. Furthermore, these plots show the significant proportional bias that was evident in daytime and nighttime estimated TST, SOL, and WASO, as well as daytime SE. Bland-Altman plots of healthy sleepers compared to those with a diagnosed or suspected sleep disorder did not show notable differences, as found in supplementary Figure [.5](#).

Table 4: Sleep characteristics of Withings Sleep Analyzer versus polysomnography during nighttime and daytime recordings.

	PSG	WSA	Difference (WSA - PSG)
<b>Overall</b>			
Total Sleep Time	413.1 (84.4)	460.6 (75.3)	47.5 (81.2)*
Sleep Efficiency (%)	79.3 (14)	88.5 (11.5)	9.2 (14.8)*
Sleep Onset Latency	17.9 (30.6)	24.2 (29.7)	6.4 (26.1)*
Wake After Sleep Onset	90.1 (71.8)	36.2 (59.6)	-53.9 (77.9)*
Wake	108 (76.2)	60.4 (67.3)	-47.6 (81.2)*
Light Sleep	214.2 (63.3)	238 (67.2)	23.8 (86.4)*
Deep Sleep	113.7 (44.7)	119.4 (52.4)	5.7 (60)*
REM Sleep	84.7 (35)	102.7 (49.2)	18 (45.4)*
<b>Healthy Sleep - Nighttime</b>			
Total Sleep Time	436 (81)	472 (70.1)	36 (45.9)*
Sleep Efficiency (%)	82.3 (12)	89.2 (9.4)	7 (8.6)*
Sleep Onset Latency	30 (38.9)	34.1 (39)	4.1 (20.6)
Wake After Sleep Onset	64.3 (56)	24.1 (37.8)	-40.2 (42.5)*
Wake	94.3 (68.9)	58.3 (56.8)	-36 (45.9)*
Light Sleep	253.5 (57)	236.1 (67.2)	-17.4 (68.8)
Deep Sleep	98.1 (35.8)	129.8 (48)	31.8 (51.2)*
REM Sleep	84.2 (34)	105.9 (38.4)	21.7 (30.9)*
<b>Healthy Sleep - Daytime</b>			
Total Sleep Time	420.6 (84.5)	486.8 (58.4)	66.2 (79)*
Sleep Efficiency (%)	78.9 (14.5)	91.5 (9.4)	12.7 (13.6)*
Sleep Onset Latency	5.3 (5.1)	13.5 (11.2)	8.3 (10)*
Wake After Sleep Onset	107.4 (78.2)	32.9 (61.4)	-74.5 (79.2)*
Wake	112.7 (78.4)	46.3 (62.5)	-66.4 (78.9)*
Light Sleep	192.3 (50.1)	239.9 (60.2)	47.6 (77.1)*
Deep Sleep	132.1 (34.2)	125.2 (38.3)	-6.9 (46.3)
REM Sleep	95 (30.6)	120.7 (45.3)	25.7 (47.1)*
<b>Sleep Disorder - Nighttime</b>			
Total Sleep Time	384.6 (79.5)	412.7 (79.2)	28.1 (97.9)*
Sleep Efficiency (%)	77.9 (14.6)	83.5 (14.1)	5.7 (18.8)*
Sleep Onset Latency	27.6 (38)	32.9 (35.2)	5.2 (41.8)
Wake After Sleep Onset	83.5 (65.5)	50.2 (67.4)	-33.4 (88.4)*
Wake	111.2 (77.3)	83.1 (75.5)	-28.1 (97.9)*
Light Sleep	217.4 (70.7)	236.5 (77.1)	19.2 (98.6)
Deep Sleep	97.7 (53.6)	102.9 (68.1)	5.2 (76.4)
REM Sleep	69.5 (36.5)	73.2 (48.4)	3.8 (48.7)

Note: Data reflects mean (SD) and units are minutes unless otherwise indicated  
 $p < .05$  in mixed model analyses between Withings Sleep Analyzer and polysomnography  
 REM=rapid eye movement; SD=standard deviation

Figure 1: Bland-Altman plots of TST, SE, SOL, and WASO during nighttime (left) and daytime (right) sleep opportunities.

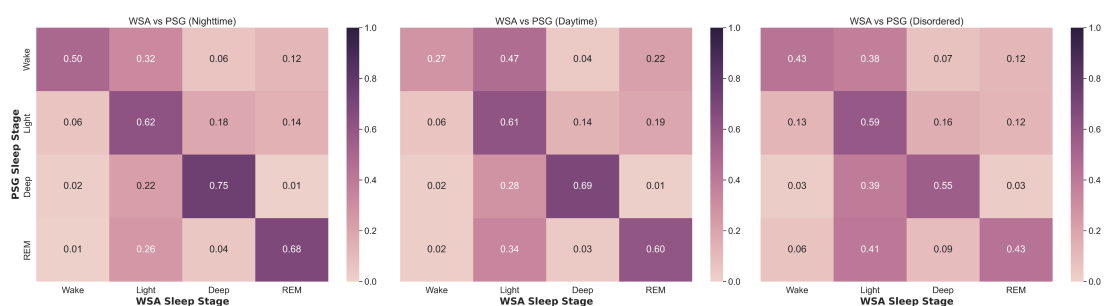


Note: Red solid lines indicate mean bias, with dashed red 95% CIs. Grey solid lines indicate LOAs (Mean bias  $\pm 1.96$  standard deviation), with dashed grey 95% CIs. TST=total sleep time; SE=sleep efficiency; SOL=sleep onset latency; WASO=wake after sleep onset

## Sleep Stages

Confusion matrices in Figure 2 demonstrate that during nighttime sleep opportunities with healthy sleepers, the WSA accurately classified 50% of wake and 91% of sleep epochs. For daytime sleep opportunities, the WSA accurately classified 27% of wake and 91% of sleep. Sensitivity for sleep stage classification ranged from 62-75% during nighttime sleep opportunities, and 60-69% during daytime sleep opportunities. For individuals with a diagnosed or suspected sleep disorder, the WSA accurately classified 43% of wake and 78% of sleep, with sleep stage accuracy ranging from 45-59%.

Figure 2: Confusion matrices showing Withings Sleep Analyzer versus polysomnography four-stage classification for healthy sleepers during nighttime and daytime recordings, and people with a sleep disorder during nighttime recordings.



## WSA versus Fitbit accuracy

Data from a subset of 22 healthy participants (nighttime recordings=35, daytime=160, mean[SD] age=31.3[12.4], 10 male, 12 female) were used to compare sleep estimation performance of the WSA and Fitbit versus polysomnography. Figure 3 shows the sleep-stage classification confusion matrices between devices for both nighttime and daytime sleep opportunities. There was a significant interaction effect of device by timing of sleep opportunity for accuracy,  $F(363.5)=6.5$ ,  $p=.01$ , marginal  $R^2=0.08$ , and specificity,  $F(365.3)=4.3$ ,  $p=.04$ , marginal  $R^2=0.17$ , whereby the Fitbit showed significantly higher accuracy (5%  $\pm$  1%) and specificity (19.1%  $\pm$  2.2%) than the WSA during daytime sleep opportunities, but did not significantly differ during nighttime sleep opportunities. In addition, the WSA had significantly poorer accuracy (8.7%  $\pm$  1.7%) and specificity (15.7%  $\pm$  3.6%) during daytime sleep opportunities, compared to night, while the Fitbit did not. This effect is evident in the confusion matrices, where daytime sleep opportunities saw a larger reduction in wake classification accuracy for the WSA (43% to 28%) compared to the

Fitbit (53% to 47%). The interaction effect of device by timing of sleep opportunity was also significant,  $p=.006$ , but post-hoc analyses revealed no significant comparisons.

Figure 3: Confusion matrices showing four-stage classification, compared to polysomnography, for Withings Sleep Analyzer versus Fitbit.



Note: Sleep stage classification compared four-stage estimation (Device 'wake' = PSG wake, device 'light' = PSG N1, device 'deep' = PSG N2 + N3, device 'REM' = PSG REM). N1=non-REM stage 1; N2=non-REM stage 2; N3=non-REM stage 3; PSG=polysomnography; REM=rapid eye movement; WSA=Withings Sleep Analyzer

Comparisons were conducted between WSA and Fitbit estimations of TST, SE and WASO, with mean (SD) values seen in Figure 5. Main effects of sleep opportunity timing and device were significant for all sleep outcomes ( $p<.05$ ), but there were no significant interactions between timing of sleep opportunity and device. Post-hoc analyses revealed that both the WSA and Fitbit overestimated TST and SE, and underestimated WASO, while only the WSA significantly overestimated SOL ( $p<.05$ ). Post-hoc analyses also showed that the Fitbit was significantly closer than the WSA to polysomnography in estimates of TST by 20 ( $\pm 8.1$ ) minutes,  $p=.04$ , SE by 3.5% ( $\pm 1.3\%$ ),  $p=.03$ , and WASO by 21.5 ( $\pm 7.5$ ) minutes,  $p=.01$ .

Table 5: Sleep characteristics of Withings Sleep Analyzer and Fitbit versus polysomnography during nighttime and daytime recordings.

	PSG	WSA	WSA-PSG	Fitbit	Fitbit-PSG	WSA-Fitbit
<b>Overall</b>						
TST	426.1 (82.9)	486.3 (57)	+60.2 (77.4)	467.1 (68.6)	+41 (57.5)	+9.6 (33.6)
SE	79.9 (14.2)	91.4 (9.2)	+11.5 (13.3)	88 (10.6)	+8.2 (11.1)	+1.7 (5.3)
SOL	8.8 (11.8)	17 (15.8)	+8.2 (12.6)	11.6 (23.8)	+2.8 (20.6)	+2.7 (10.6)
WASO	98.7 (77.8)	30.2 (60.1)	-68.5 (78.3)	52.3 (52.5)	-46.5 (57.1)	-11 (31)
<b>Night</b>						
TST	454.9 (57.5)	494.5 (30.9)	+39.6 (42.1)	473.4 (51.1)	+18.5 (52.7)	+10.6 (22.4)
SE	84.7 (10.8)	92 (5.8)	+7.4 (7.9)	88.3 (8.6)	+3.7 (9.3)	+1.9 (3.7)
SOL	23.1 (20.2)	29.1 (24.1)	+6 (19.7)	27.4 (41.9)	+4.3 (37.2)	+0.9 (19.2)
WASO	59.4 (54)	13.8 (18.2)	-45.6 (45.1)	34.5 (25.8)	-24.9 (49.7)	-10.4 (11.7)
<b>Day</b>						
TST	419.8 (86.3)	484.5 (61.2)	+64.8 (82.6)	465.7 (71.9)	+45.9 (57.5)	+9.4 (35.6)
SE	78.8 (14.6)	91.2 (9.8)	+12.4 (14)	87.9 (11)	+9.1 (11.3)	+1.6 (5.6)
SOL	5.7 (5.2)	14.4 (11.9)	+8.7 (10.5)	8.2 (15.9)	+2.5 (14.9)	+3.1 (7.6)
WASO	107.3 (79.6)	33.8 (65.3)	-73.5 (83.1)	56.1 (56)	-51.2 (57.7)	-11.1 (33.9)

PSG=polysomnography; SD=standard deviation; SE=sleep efficiency; SOL=sleep onset latency; TST=total sleep time; WASO=wake after sleep onset; WSA=Withings Sleep Analyzer

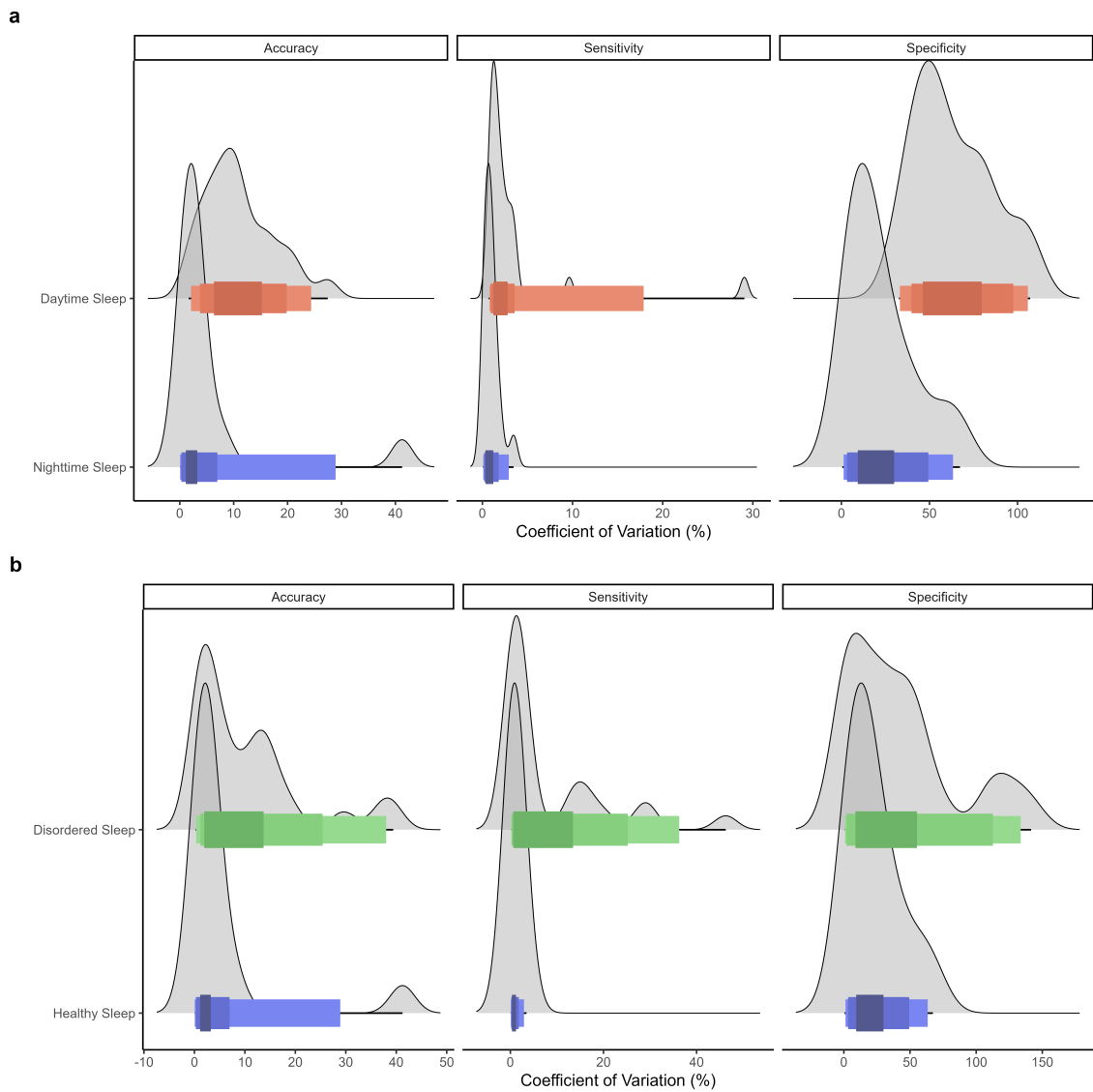
## Data Loss

Within the subset where both devices were used (n=25, 248 possible recordings), a total of 17 recordings (mean [SD] = 0.68 [0.9] per participant) were lost with the WSA compared to 53 (2.12 [3.38] per participant) with the Fitbit. The paired-samples *t*-test confirmed that this was a significant difference,  $t(24) = 2.42, p = .02, d = 0.58$ . WSA data loss occurred entirely due to mats being unintentionally left unpowered from prior sleep studies. Fitbit data loss occurred due to improper charging (both user and mechanical error), syncing errors, and improper wear.

## Multi-night Performance

For participants with more than one recording (n=49, 282 nights), variability in accuracy, sensitivity, and specificity is demonstrated in Figure 4. Mixed models showed significantly greater variability (represented as the mean coefficient of variation) in accuracy,  $2.33 \pm 2.59\%$ ,  $p = .03$ , and specificity,  $6.03 \pm 6.85\%$ ,  $p < .001$ , for daytime compared to nighttime sleep opportunities. Sensitivity was more variable in individuals with a sleep disorder,  $2.42 \pm 3.02\%$ ,  $p = .02$ . The Levene's test also indicated that the distribution of sensitivity variability differed,  $6.12, p = .02$ , seen in Figure 4 as the wider distribution of variability for individuals with disordered sleep. No other significant differences in distribution were found.

Figure 4: Withings Sleep Analyzer versus polysomnography variability in accuracy for a) daytime versus nighttime sleep opportunities, and b) individuals with healthy versus disordered sleep.



Note: Larger mean coefficients of variation reflect greater multi-night variability in Withings Sleep Analyzer performance. Wider plots reflect a greater range of variability in performance across individuals in each category.



## Discussion

This study comprehensively evaluated the performance of the WSA to estimate sleep and wake compared to polysomnography. Appropriate use of such consumer sleep technology requires adequate performance evaluation [24]. Here, we completed one of the largest independent performance evaluations of a consumer sleep tracker to date. Results indicate that the WSA was accurate compared to polysomnography to classify sleep and wake states, with moderate-to-poor accuracy for sleep staging, that was comparable to existing wearable devices [8, 29]. The WSA systematically overestimated TST, SE and SOL, and underestimated WASO, compared to polysomnography. Accuracy and specificity was comparable to existing devices during nighttime sleep opportunities with healthy sleepers, but poorer during daytime sleep opportunities and poorer to estimate sleep in people with sleep disorders. These effects were partially explained by lower polysomnography-derived sleep efficiency during these instances. Overall, the WSA demonstrates relatively minimal bias in TST and SE estimates that is comparable to other validated consumer sleep tracking devices [8, 22], but performance was worse for sleep periods with more wakefulness.

These data are consistent with previous studies in which movement-based sleep trackers often misclassify motionless wake as sleep [30]. However, few studies have investigated circumstances where poorer than normal sleep is likely to impact reliable sleep estimation. Compared to one study with actigraphy devices [31], the WSA performs relatively poorly for reliably detecting wake during daytime recordings, with 72% of wake misclassified as sleep. However, in the subset of participants with a diagnosed or suspected sleep disorder, the lower accuracy was driven by a reduction in sensitivity (i.e., relatively more sleep being misclassified as wake). This may be due to the especially high misclassification of light sleep by the WSA as wake or deeper sleep, and the relatively high proportion of such sleep in people with a sleep disorder (OSA in particular) [32–34]. This suggests that the WSA, and likely other sleep tracking devices [18, 35–37], are less accurate to classify sleep and wake in individuals with a sleep disorder. Whether these devices are accurate enough for a given purpose is difficult to determine from the accuracy data alone. Performance evaluation, together with investigation into the potential utility of such sleep estimations, is required to make this judgement. We have shown that the WSA has utility to estimate cognitive performance under simulated shift work conditions

[38], and other studies have found associations between wearable device estimations of sleep and cognitive fatigue or associated outcomes[39, 40]. Thus, further work is needed to test how the proportion of sleep misclassification observed in the current study may impact the device utility in clinical, occupational, and related contexts.

This study also showed that that the Fitbit had more reliable sleep classification accuracy than the WSA. Specifically, the Fitbit showed higher accuracy and specificity (more accurate wake classification) during daytime sleep opportunities. This was further evidenced in sleep characteristics, where the Fitbit overestimated TST by around 20 minutes less, overestimated SE by about 4% less, and underestimated WASO by around 20 minutes less than the WSA. This may be due to limitations in the modality of the two devices, where relatively motionless wake could be interpreted more accurately by a wrist-worn device than an under-mattress sensor. It may be possible that, given perceived restriction to movement imposed by polysomnography equipment, participants were less inclined to change position or move during wake. This effect would likely be more prominent in chest movement, as the WSA primarily estimates, than in wrist movements measured by the Fitbit. Other differences between device algorithms may also account for the differences between WSA and Fitbit accuracy, particularly in the incorporation of heart rate and breathing signals. Overall, the Fitbit device was more accurate at classifying sleep than the WSA, particularly during daytime sleep opportunities, but further comparisons are warranted to examine whether this is maintained in a naturalistic environment.

There are additional considerations that should be noted when critically evaluating sleep tracking devices. The WSA, unlike wrist-worn devices, does not need charging or manual synchronization of data. As such, in the study used for the sub-sample device comparison, we found a data-loss rate of 7% (compared to 21% data loss with the Fitbit). As opposed to the laboratory environment with technicians ensuring suitable device setup and suitable use to protect data fidelity, data loss would be expected to be higher in a naturalistic environment. This greater data loss may be higher for the Fitbit device than the WSA, given that, once set up, the WSA does not require user input or effort to continue recording whereas wrist-worn devices must be charged, worn correctly, and routinely synchronized for lossless data capture. Additionally, the WSA has been found to be highly accurate at bed occupancy timing (i.e., in/out of bed times) and duration compared to polysomnography with video [10], which wearable devices are typically less accurate at detecting in home environments [22, 41, 42]. Thus, the WSA has practical

advantages over other forms of sleep tracking devices that should be taken into consideration when selecting a device for use, alongside its accuracy for sleep/wake detection.

Although there are strengths to the large sample size, there are still limitations to this study. Some of the included research studies with individuals with a diagnosed or suspected sleep disorder had pre-sleep events that may have impacted sleep, such as drug vs. placebo interventions and respiratory testing. Given that sleep efficiency partially explained WSA performance compared to polysomnography (by about 30-35%) the direct effects of such covariates on sleep quality likely account for some variability in WSA classification performance. However, the extensive and heterogeneous sample provides opportunity to evaluate performance with the expected variability in real-world use (e.g., variable continuous positive airway pressure use, acute sedative use, noise disruptions, etc.). It should also be considered that this study was conducted using data from entirely in-laboratory sleep research protocols, and performance in a naturalistic setting may differ. Finally, there were insufficient numbers to compare device accuracy between individual sleep disorders, and this should be elucidated in further work.

## Conclusions

This study extensively compared the WSA to polysomnography in a large sample of healthy individuals and people with a diagnosed or suspected sleep disorder during nighttime and daytime sleep opportunities to provide novel insights into performance characteristics. Overall, the WSA was accurate at sleep and wake detection during nighttime recordings compared to other consumer sleep trackers but was less accurate at wake classification during daytime recordings. The WSA was also less accurate at sleep classification in people with a suspected or diagnosed sleep disorder. The WSA overestimated TST, SE, SOL, and underestimated WASO to comparable levels seen with other consumer sleep trackers [8, 29]. Lower polysomnography-derived sleep efficiency was associated with worse WSA wake classifications, but this only partially explained the effect of wake on classification accuracy. In a subset of participants, the Fitbit was highly comparable to the WSA during nighttime recordings, but consistent with better wake classification performance, was more accurate during daytime recordings. Ultimately, the WSA may be suitable to estimate sleep and wake in a variety of

naturalistic environments, and the final choice should depend on the performance, logistics, and pragmatic needs of the clinical practice or research study. The current study findings provide important novel data in which to inform these key decisions.

## References

1. Yoon, H. & Choi, S. H. Technologies for Sleep Monitoring at Home: Wearables and Nearables. *Biomedical Engineering Letters* **13**, 313–327. doi:[10.1007/s13534-023-00305-8](https://doi.org/10.1007/s13534-023-00305-8) (Aug. 2023).
2. Tobin, S. Y., Williams, P. G., Baron, K. G., Halliday, T. M. & Depner, C. M. Challenges and Opportunities for Applying Wearable Technology to Sleep. *Sleep Medicine Clinics* **16**, 607–618. doi:[10.1016/j.jsmc.2021.07.002](https://doi.org/10.1016/j.jsmc.2021.07.002) (Dec. 2021).
3. Paquet, J., Kawinska, A. & Carrier, J. Wake Detection Capacity of Actigraphy During Sleep. *Sleep* **30**, 1362–1369 (Oct. 2007).
4. Fleur, R. G. S. *et al.* Use of Fitbit Devices in Physical Activity Intervention Studies Across the Life Course: Narrative Review. *JMIR mHealth and uHealth* **9**, e23411. doi:[10.2196/23411](https://doi.org/10.2196/23411) (May 2021).
5. Huang, Y., Upadhyay, U., Dhar, E., Kuo, L.-J. & Syed-Abdul, S. A Scoping Review to Assess Adherence to and Clinical Outcomes of Wearable Devices in the Cancer Population. *Cancers* **14**, 4437. doi:[10.3390/cancers14184437](https://doi.org/10.3390/cancers14184437) (Jan. 2022).
6. Sun, X., Qiu, L., Wu, Y., Tang, Y. & Cao, G. SleepMonitor: Monitoring Respiratory Rate and Body Position During Sleep Using Smartwatch. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* **1**, 1–22. doi:[10.1145/3130969](https://doi.org/10.1145/3130969) (Sept. 2017).
7. Schutte, N. S. *et al.* Evaluating Consumer and Clinical Sleep Technologies: An American Academy of Sleep Medicine Update. *Journal of Clinical Sleep Medicine* **17**, 2275–2282. doi:[10.5664/jcsm.9580](https://doi.org/10.5664/jcsm.9580).
8. Lee, T. *et al.* Accuracy of 11 Wearable, Nearable, and Airable Consumer Sleep Trackers: Prospective Multicenter Validation Study. *JMIR mHealth and uHealth* **11**, e50983. doi:[10.2196/50983](https://doi.org/10.2196/50983) (Nov. 2023).
9. Edouard, P. *et al.* Validation of the Withings Sleep Analyzer, an under-the-Mattress Device for the Detection of Moderate-Severe Sleep Apnea Syndrome. *Journal of Clinical Sleep Medicine* **17**, 1217–1227. doi:[10.5664/jcsm.9168](https://doi.org/10.5664/jcsm.9168).

10. Ravindran, K. K. G. *et al.* Validation of Technology to Monitor Sleep and Bed Occupancy in Older Men and Women. *Alzheimer's & Dementia* **17**, e056018. doi:[10.1002/alz.056018](https://doi.org/10.1002/alz.056018) (2021).
11. Scott, H. *et al.* Sleep Irregularity Is Associated With Hypertension: Findings From Over 2 Million Nights With a Large Global Population Sample. *Hypertension* **80**, 1117–1126. doi:[10.1161/HYPERTENSIONAHA.122.20513](https://doi.org/10.1161/HYPERTENSIONAHA.122.20513) (May 2023).
12. Scott, H. *et al.* Are We Getting Enough Sleep? Frequent Irregular Sleep Found in an Analysis of over 11 Million Nights of Objective in-Home Sleep Data. *Sleep Health* **10**, 91–97. doi:[10.1016/j.sleh.2023.10.016](https://doi.org/10.1016/j.sleh.2023.10.016) (Feb. 2024).
13. Lechat, B. *et al.* Multi-Night Measurement for Diagnosis and Simplified Monitoring of Obstructive Sleep Apnoea. *Sleep Medicine Reviews* **72**, 101843. doi:[10.1016/j.smrv.2023.101843](https://doi.org/10.1016/j.smrv.2023.101843) (Dec. 2023).
14. Lechat, B. *et al.* Multinight Prevalence, Variability, and Diagnostic Misclassification of Obstructive Sleep Apnea. *American Journal of Respiratory and Critical Care Medicine*. doi:[10.1164/rccm.202107-17610C](https://doi.org/10.1164/rccm.202107-17610C) (Mar. 2022).
15. Lechat, B. *et al.* Regular Snoring Is Associated with Uncontrolled Hypertension. *npj Digital Medicine* **7**, 1–8. doi:[10.1038/s41746-024-01026-7](https://doi.org/10.1038/s41746-024-01026-7) (Feb. 2024).
16. Iber, C. The AASM Manual for the Scoring of Sleep and Associated Events: Rules. *Terminology and Technical Specification* (2007).
17. Kainec, K. A. *et al.* Evaluating Accuracy in Five Commercial Sleep-Tracking Devices Compared to Research-Grade Actigraphy and Polysomnography. *Sensors* **24**, 635. doi:[10.3390/s24020635](https://doi.org/10.3390/s24020635) (Jan. 2024).
18. Dong, X. *et al.* Validation of Fitbit Charge 4 for Assessing Sleep in Chinese Patients with Chronic Insomnia: A Comparison against Polysomnography and Actigraphy. *PLOS ONE* **17**, e0275287. doi:[10.1371/journal.pone.0275287](https://doi.org/10.1371/journal.pone.0275287) (Oct. 2022).
19. de Zambotti, M., Goldstone, A., Claudatos, S., Colrain, I. M. & Baker, F. C. A Validation Study of Fitbit Charge 2™ Compared with Polysomnography in Adults. *Chronobiology International* **35**, 465–476. doi:[10.1080/07420528.2017.1413578](https://doi.org/10.1080/07420528.2017.1413578) (Apr. 2018).

20. Stucky, B. *et al.* Validation of Fitbit Charge 2 Sleep and Heart Rate Estimates Against Polysomnographic Measures in Shift Workers: Naturalistic Study. *Journal of Medical Internet Research* **23**, e26476. doi:[10.2196/26476](https://doi.org/10.2196/26476) (Oct. 2021).
21. Moreno, F., Porras, A., L, ó.-E. P., Art, é. A. & Baca, E. Validation of Fitbit Charge 2 and Fitbit Alta HR Against Polysomnography for Assessing Sleep in Adults With Obstructive Sleep Apnea. *Journal of Clinical Sleep Medicine* **15**, 1645–1653. doi:[10.5664/jcsm.8032](https://doi.org/10.5664/jcsm.8032).
22. Haghayegh, S., Khoshnevis, S., Smolensky, M. H., Diller, K. R. & Castriotta, R. J. Accuracy of Wristband Fitbit Models in Assessing Sleep: Systematic Review and Meta-Analysis. *Journal of Medical Internet Research* **21**. doi:[10.2196/16273](https://doi.org/10.2196/16273) (Nov. 2019).
23. Conley, S. *et al.* Agreement between Actigraphic and Polysomnographic Measures of Sleep in Adults with and without Chronic Conditions: A Systematic Review and Meta-Analysis. *Sleep medicine reviews* **46**, 151–160. doi:[10.1016/j.smrv.2019.05.001](https://doi.org/10.1016/j.smrv.2019.05.001) (Aug. 2019).
24. de Zambotti, M., Cellini, N., Goldstone, A., Colrain, I. M. & Baker, F. C. Wearable Sleep Technology in Clinical and Research Settings. *Medicine and Science in Sports and Exercise* **51**, 1538–1557. doi:[10.1249/MSS.0000000000001947](https://doi.org/10.1249/MSS.0000000000001947) (July 2019).
25. Menghini, L., Cellini, N., Goldstone, A., Baker, F. C. & de Zambotti, M. A Standardized Framework for Testing the Performance of Sleep-Tracking Technology: Step-by-step Guidelines and Open-Source Code. *Sleep* **44**. doi:[10.1093/sleep/zsaa170](https://doi.org/10.1093/sleep/zsaa170) (Feb. 2021).
26. Sadeh, A., Sharkey, M. & Carskadon, M. A. Activity-Based Sleep-Wake Identification: An Empirical Test of Methodological Issues. *Sleep* **17**, 201–207. doi:[10.1093/sleep/17.3.201](https://doi.org/10.1093/sleep/17.3.201) (May 1994).
27. Bates, D. *et al.* Package ‘Lme4’. *convergence* **12**, 2 (2015).
28. Euser, A. M., Dekker, F. W. & Le Cessie, S. A Practical Approach to Bland-Altman Plots and Variation Coefficients for Log Transformed Variables. *Journal of Clinical Epidemiology* **61**, 978–982. doi:[10.1016/j.jclinepi.2007.11.003](https://doi.org/10.1016/j.jclinepi.2007.11.003) (Oct. 2008).
29. Chinoy, E. D. *et al.* Performance of Seven Consumer Sleep-Tracking Devices Compared with Polysomnography. *Sleep*. doi:[10.1093/sleep/zsaa291](https://doi.org/10.1093/sleep/zsaa291) (Dec. 2020).

30. Miller, D. J., Sargent, C. & Roach, G. D. A Validation of Six Wearable Devices for Estimating Sleep, Heart Rate and Heart Rate Variability in Healthy Adults. *Sensors* **22**, 6317. doi:[10.3390/s22166317](https://doi.org/10.3390/s22166317) (Jan. 2022).
31. Gao, C. et al. Actigraphy-Based Sleep Detection: Validation with Polysomnography and Comparison of Performance for Nighttime and Daytime Sleep During Simulated Shift Work. *Nature and Science of Sleep* **14**, 1801–1816. doi:[10.2147/NSS.S373107](https://doi.org/10.2147/NSS.S373107) (Dec. 2022).
32. Eckert, D. J. & Younes, M. K. Arousal from Sleep: Implications for Obstructive Sleep Apnea Pathogenesis and Treatment. *Journal of Applied Physiology* **116**, 302–313. doi:[10.1152/japplphysiol.00649.2013](https://doi.org/10.1152/japplphysiol.00649.2013) (Feb. 2014).
33. Ratnavadivel, R. et al. Marked Reduction in Obstructive Sleep Apnea Severity in Slow Wave Sleep. *Journal of Clinical Sleep Medicine* **05**, 519–524. doi:[10.5664/jcsm.27651](https://doi.org/10.5664/jcsm.27651) (Dec. 2009).
34. Shahveisi, K. et al. Sleep Architecture in Patients With Primary Snoring and Obstructive Sleep Apnea. *Basic and Clinical Neuroscience* **9**, 147–156. doi:[10.29252/NIRP.BCN.9.2.147](https://doi.org/10.29252/NIRP.BCN.9.2.147) (2018).
35. Kang, S.-G. et al. Validity of a Commercial Wearable Sleep Tracker in Adult Insomnia Disorder Patients and Good Sleepers. *Journal of Psychosomatic Research* **97**, 38–44. doi:[10.1016/j.jpsychores.2017.03.009](https://doi.org/10.1016/j.jpsychores.2017.03.009) (June 2017).
36. Te Lindert, B. H. W. et al. Optimizing Actigraphic Estimates of Polysomnographic Sleep Features in Insomnia Disorder. *Sleep* **43**, zsa090. doi:[10.1093/sleep/zsaa090](https://doi.org/10.1093/sleep/zsaa090) (Nov. 2020).
37. Cook, J. D., Eftekari, S. C., Leavitt, L. A., Prairie, M. L. & Plante, D. T. Optimizing Actigraphic Estimation of Sleep Duration in Suspected Idiopathic Hypersomnia. *Journal of Clinical Sleep Medicine* **15**, 597–602. doi:[10.5664/jcsm.7722](https://doi.org/10.5664/jcsm.7722).
38. Manners, J. et al. Estimating Vigilance from the Pre-Work Shift Sleep Using an under-Mattress Sleep Sensor. *Journal of Sleep Research* **n/a**, e14138. doi:[10.1111/jsr.14138](https://doi.org/10.1111/jsr.14138).
39. Adão Martins, N. R., Annaheim, S., Spengler, C. M. & Rossi, R. M. Fatigue Monitoring Through Wearables: A State-of-the-Art Review. *Frontiers in Physiology* **12**. doi:[10.3389/fphys.2021.790292](https://doi.org/10.3389/fphys.2021.790292) (Dec. 2021).



40. Zhu, Y., Jankay, R. R., Pieratt, L. C. & Mehta, R. K. Wearable Sensors and Their Metrics for Measuring Comprehensive Occupational Fatigue: A Scoping Review. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* **61**, 1041–1045. doi:[10.1177/1541931213601744](https://doi.org/10.1177/1541931213601744) (Sept. 2017).
41. Feehan, L. M. et al. Accuracy of Fitbit Devices: Systematic Review and Narrative Syntheses of Quantitative Data. *JMIR mHealth and uHealth* **6**, e10527. doi:[10.2196/10527](https://doi.org/10.2196/10527) (Aug. 2018).
42. Hamill, K. et al. Validity, Potential Clinical Utility and Comparison of a Consumer Activity Tracker and a Research-Grade Activity Tracker in Insomnia Disorder II: Outside the Laboratory. *Journal of Sleep Research* **29**. doi:[10.1111/jsr.12944](https://doi.org/10.1111/jsr.12944) (Feb. 2020).

## List of Figures

1	Bland-Altman plots of TST, SE, SOL, and WASO during nighttime (left) and daytime (right) sleep opportunities. . . . .	12
2	Confusion matrices showing Withings Sleep Analyzer versus polysomnography four-stage classification for healthy sleepers during nighttime and daytime recordings, and people with a sleep disorder during nighttime recordings. . . . .	13
3	Confusion matrices showing four-stage classification, compared to polysomnography, for Withings Sleep Analyzer versus Fitbit. . . . .	14
4	Withings Sleep Analyzer versus polysomnography variability in accuracy for a) daytime versus nighttime sleep opportunities, and b) individuals with healthy versus disordered sleep. . . . .	16
5	Bland-Altman plots of TST, SE, SOL, and WASO for healthy participants (left) and participants with suspected or diagnosed sleep disorders (right). . . . .	30

## List of Tables

1	Participant demographics. . . . .	8
2	Epoch-by-epoch performance of Withings Sleep Analyzer versus polysomnography. . . . .	9
3	Epoch-by-epoch performance of Withings Sleep Analyzer versus polysomnography. . . . .	9

4	Sleep characteristics of Withings Sleep Analyzer versus polysomnography during nighttime and daytime recordings. . . . .	11
5	Sleep characteristics of Withings Sleep Analyzer and Fitbit versus polysomnography during nighttime and daytime recordings. . . . .	15
S4	Detailed study breakdown. . . . .	28
S4	Model outcomes of Withings Sleep Analyzer versus polysomnography estimates between sleep disorder statuses and sleep opportunity timings. . . . .	29