JCSM | Journal *of* Clinical Sleep Medicine

SCIENTIFIC INVESTIGATIONS

Validation of a nonwearable device in healthy adults with normal and short sleep durations

David A. Hsiou, BS[1,2]; Chenlu Gao, PhD[3,4]; Robert C. Matlock, MD[5]; Michael K. Scullin, PhD[1]

[1]Department of Psychology and Neuroscience, Baylor University, Waco, Texas; [2]School of Medicine, Baylor College of Medicine, Houston, Texas; [3]Division of Sleep and Circadian Disorders, Departments of Medicine and Neurology, Brigham and Women's Hospital, Boston, Massachusetts; [4]Division of Sleep Medicine, Department of Medicine, Harvard Medical School, Boston, Massachusetts; [5]Ascension Medical Group Providence, Department of Sleep Medicine, Waco, Texas

Study Objectives: To determine the accuracy of early and newer versions of a nonwearable sleep tracking device relative to polysomnography and actigraphy, under conditions of normal and restricted sleep duration.

Methods: Participants were 35 healthy adults (mean age = 18.97; standard deviation = 0.95 years; 77.14% female; 42.86% White). In a controlled sleep laboratory environment, we randomly assigned participants to go to bed at 10:30 PM (normal sleep) or 1:30 AM (restricted sleep), setting lights-on at 7:00 AM. Sleep was measured using polysomnography, wristband actigraphy (the Philips Respironics Actiwatch Spectrum Plus), self-report, and an early or newer version of a nonwearable device that uses a sensor strip to measure movement, heart rate, and breathing (the Apple, Inc. Beddit). We tested accuracy against polysomnography for total sleep time, sleep efficiency, sleep onset latency, and wake after sleep onset.

Results: The early version of the nonwearable device (Beddit 3.0) displayed poor reliability (intraclass correlation coefficient [ICC] < 0.30). However, the newer nonwearable device (Beddit 3.5) yielded excellent reliability with polysomnography for total sleep time (ICC = 0.998) and sleep efficiency (ICC = 0.98) across normal and restricted sleep conditions. Agreement was also excellent for the notoriously difficult metrics of sleep onset latency (ICC = 0.92) and wake after sleep onset (ICC = 0.92). This nonwearable device significantly outperformed clinical-grade actigraphy (ICC between 0.44 and 0.96) and self-reported sleep measures (ICC < 0.75).

Conclusions: A nonwearable device showed better agreement than actigraphy with polysomnography outcome measures. Future work is needed to test the validity of this device in clinical populations.

Keywords: consumer sleep technology, nonwearable device, sleep tracking

Citation: Hsiou DA, Gao C, Matlock RC, Scullin MK. Validation of a nonwearable device in healthy adults with normal and short sleep durations. *J Clin Sleep Med.* 2022;18(3):751–757.

BRIEF SUMMARY

Current Knowledge/Study Rationale: Polysomnography is the gold standard for clinical sleep assessment and diagnosis, but this method is too cumbersome and costly for long-term monitoring of natural sleep patterns. This study compares the accuracy of sleep metrics collected from a nonwearable device, Apple Inc.'s Beddit Sleep Monitor (versions 3.0 and 3.5), against research/clinical standards such as polysomnography, actigraphy, and self-report.

Study Impact: Although many commercial sleep devices have displayed poor accuracy, the current findings show that a nonwearable device can have excellent accuracy in healthy adults. Reducing the burden on participants in sleep studies (via nonwearable devices) has broad implications for improving accessibility, monitoring duration, study costs, and representativeness of findings.

INTRODUCTION

Polysomnography (PSG) is the gold standard measurement of sleep. In research settings, PSG allows for the precise measurement of sleep stages and the detection of sleep microevents (utilizing electroencephalography monitoring). In clinical settings, PSG enables diagnoses and therapeutic testing (eg, positive airway pressure titration). Despite its strengths, PSG has some disadvantages such as being expensive and burdensome on patients and participants. In addition, sleeping in a laboratory is known to detrimentally affect sleep quality,[1] and key sleep measures can fluctuate from night to night, limiting the value of PSG if only a single night can be obtained.[2] One alternative (or complementary) approach is to use wristband actigraphy, in which the sleep/wake state is estimated via accelerometry in home settings.[3] Actigraphy has acceptable levels of accuracy for total sleep time (TST),[4] but the cost of clinical-grade actigraphy devices remains high and some individuals do not like to sleep while wearing devices (some populations, such as children with autism, do not tolerate actigraphy or PSG).[5] Therefore, there is a need for devices that can inexpensively and accurately measure sleep in the home environment with minimal or zero contact to the person (ie, nonwearables).

As public interest in sleep health has surged,[6,7] there has been a proliferation of commercial devices to track sleep.[8] However, when these commercial devices are tested relative to PSG and actigraphy, many of them show poor congruence. For example, Meltzer et al[9] found that the Fitbit Ultra either

overestimated or underestimated TST by more than 1 hour, depending on whether the normal or sensitive analysis mode was used. Similar inaccuracies in device measurement have been reported in other independent tests of commercial sleep tracking devices.[10–12] Nevertheless, the consumer sleep tracking market has continued a compound annual growth rate of 18.5%[13] and now has attracted leading technology companies such as Apple, Samsung, Bose, and others.

In the current study, we evaluated the accuracy of a nonwearable device, the Beddit Sleep Monitor (hereafter, Beddit; Apple Inc., Cupertino, CA). Beddit is a flat strip sensor that is placed under the bedsheet. Apple acquired this startup company in 2017 (version 3.0) and released an updated version of Beddit in December 2018 (version 3.5). At the time of acquisition, Beddit 3.0 showed poor accuracy relative to PSG for TST, wake after sleep onset (WASO), and sleep efficiency (SE%).[12] The current study sought to replicate these findings for Beddit 3.0 and also test the accuracy of the subsequent Beddit version 3.5 against PSG (gold standard) along with actigraphy and self-report. Because the length and quality of sleep have been known to affect the accuracy of some sleep monitoring devices,[14] we randomly assigned participants to 8.5-hour or 5.5-hour time-in-bed opportunities.[15] Thus, the current work provides a comparison of research-grade, commercial, and self-reported measures of sleep across normal and restricted sleep schedules.

METHODS

Participants

Thirty-five adults (mean age = 18.97 years; standard deviation = 0.95 years; 77.14% female; 42.86% White, 14.29% Hispanic, 14.29% Asian, 5.71% African American, 20% other/multiple races) were recruited through advertisements at a university in the United States. All participants provided informed consent and completed the study between October 2019 and March 2020. Data collection ceased in response to the coronavirus pandemic. This study was approved by the Baylor University Institutional Review Board.

Sleep measurements

PSG

We used the Grass Comet XL Plus system (Astro-Nova, Inc., West Warwick, RI) to record sleep from electroencephalography (Fp1, Fp2, F3, F4, F7, F8, FZ, C3, C4, CZ, P3, P4, PZ, T3, T4, T5, T6, O1, and O2 sites; referenced to contralateral mastoids), bilateral electrooculography, and mentalis electromyography. Sleep stages were scored in 30-second epochs according to American Academy of Sleep Medicine guidelines by a registered polysomnography technician who was masked to participants' Beddit, actigraphy, and self-report data.[16]

Actigraphy

For actigraphy monitoring, we used Philips Respironics Actiwatch Spectrum Plus devices (Philips Electronics, Bend, OR). Relative to PSG, actigraphy tends to show modest specificity[14]

but strong sensitivity to sleep states.[17–19] Participants wore the actigraphy device on their nondominant wrist throughout the night. We set the devices to a medium-threshold sensitivity (40 activity counts per epoch) and recorded sleep in 30-second epoch lengths. Trained personnel then scored the data based on the scoring technique developed by Rijsketic and colleagues.[20]

Beddit

The Beddit sleep monitor device is a thin, flat sensor strip placed immediately under the bedsheet at chest level. Beddit records movement, heart rate, and breathing rate and provides automated estimates of time in bed, TST, sleep onset latency (SOL), WASO, time away from bed, SE%, average breathing rate, and average heart rate (along with maximum and minimum). Automated estimates are provided through Beddit's iphone-supported companion applications. Two successive Beddit versions were tested: Beddit 3.0 (released October 2016) and Beddit 3.5 (released December 2018). The Beddit device was connected to an Apple iPhone or iPad via Bluetooth technology. To avoid potential cross-device interference, only one Beddit version was used for each participant (n = 21 and 14 for versions 3.0 and 3.5, respectively).

Self-report

Upon awakening in the morning, participants estimated their sleep latency, TST, and WASO. Previous validation studies have shown moderate correlations ($r = .45$) between self-report and objective sleep measures.[21] In addition, participants completed the Pittsburgh Sleep Quality Index[22] to assess self-reported sleep quality during the past month.

Procedures

Participants arrived at a light- and sound-controlled sleep laboratory at approximately 8:00 PM. They completed a series of questionnaires, including the Pittsburgh Sleep Quality Index, and research personnel set up the Beddit devices, applied PSG electrodes, and configured the actigraphy devices. Participants were randomly assigned to go to bed at 10:30 PM (n = 19) or 1:30 AM (n = 16). After random assignment, participants were allowed to sit at a desk and complete homework, use their phones, and watch television until 15 minutes before their scheduled bedtime. Biocalibration testing was conducted at the scheduled bedtime and was immediately followed by lights-out. Lights-on occurred between 6:50 AM and 7:10 AM, depending on natural awakening or a switch to stage N1 sleep. Upon awakening, participants reported their TST, sleep latency, and WASO, before completing learning tasks that addressed a separate research question.

Statistical analyses

First, we used a 2-way repeated-measures analysis of variance (device × sleep condition) to assess whether Beddit-measured bedtime, SOL, WASO, TST, and SE% differed from PSG-measured values across normal sleep and restricted sleep schedules. We focused on the main effect of device and interaction effects. We also assessed Pearson correlation coefficients to quantify the relationships between sleep measures. Notably, to

quantify the level of consistency between measures, we used the intraclass correlation coefficient (ICC), which is a standard measure of interrater reliability that is suitable for ratio variables and for studies with 2 or more raters (or, in the current study, 2 or more devices).[23] ICC values were computed through 2-way mixed-effect models that examined the consistency between measures and estimated the reliability of a single measure. We interpreted ICC values according to Koo and Li's guidelines,[23] in which 0.50 indicates moderate agreement, 0.75 indicates good agreement, and 0.90 indicates excellent agreement. We then computed and examined the 95% confidence intervals (CIs) of the ICCs to determine whether reliability differed significantly across devices. We plotted the consistency between measures using Bland-Altman plots.[24] On Bland-Altman plots, the mean difference represents systematic/bias error and the 95% CIs represent random/precision error. Because correlation measures (eg, ICC, Pearson correlation) can be influenced by restriction of range, we combined the normal and restricted sleep conditions in the ICC and correlational analyses when the initial analysis of variance showed no significant device by condition interactions.[25] Note that neither Beddit 3.0 nor Beddit 3.5 provided epoch-by-epoch sleep/wake data, and therefore we were unable to examine sensitivity and specificity outcome measures. All analyses were performed using SPSS software (version 27, IBM Corp, Armonk, NY). All statistical tests were 2-tailed, with alpha set to .05. Study data are publicly accessible at https://osf.io/vg345/.

RESULTS

Overview

Four actigraphy reports and one self-report sleep estimate were missing or incomplete. Therefore, the final dataset included PSG for 35 participants, Beddit data for 35 participants, actigraphy data for 31 participants, and self-report data for 34 participants.

Bedtime

We first investigated whether the actigraphy and Beddit devices would accurately capture participant bedtimes, which were experimentally assigned. For this initial measure, both actigraphy and Beddit 3.5 were successful (**Table 1**). **Figure 1A** shows that Beddit 3.5 had excellent reliability (ICC > 0.99), and the Bland-Altman plot in **Figure S1** in the supplemental material shows minimal bias error (< 5 minutes mean deviation; 95% CI, –11.79 to 2.00), relative to PSG. Bedtime was not measured by Beddit 3.0.

TST

Table 1 shows that TST was more difficult (than bedtime) to estimate for all non-PSG methods. Beddit 3.0 correlated weakly ($r = .38$; $P = .089$), and self-report correlated moderately ($r = .58$; $P < .001$) with PSG; however, the TST correlation with PSG was nearly perfect for actigraphy ($r = .96$; $P < .001$) and Beddit 3.5 ($r = .998$; $P < .001$; **Table S1** in the supplemental material). The ICCs showed a similar pattern. Beddit 3.5 and

actigraphy indicated excellent reliability relative to PSG in measuring TST (ICC > .95), significantly outperforming Beddit 3.0 and self-reported TST ($P < .05$). **Figure 2** shows that relative to PSG, the raw bias error for Beddit 3.5 was < 2 minutes (95% CI, –12.93 to 16.08), which was superior to the Beddit 3.0 deviation of 45 minutes. Similar outcomes for both Beddit versions were observed across the normal sleep and restricted sleep conditions (**Table 1**).

SE%

The SE% data mirrored the TST findings. Beddit 3.0 SE% correlated weakly with PSG SE% ($r = .34$; $P = .133$; **Table S2** in the supplemental material); by contrast, Beddit 3.5 correlated nearly perfectly with PSG for SE% ($r = .99$; $P < .001$) and showed only a 1.1% bias error (95% CI, –5.19 to 3.00; **Figure S2** in the supplemental material). Self-report ($r = .45$; $P = .008$) and actigraphy ($r = .81$; $P < .001$) estimates of SE% also correlated with PSG estimates, albeit not as highly as data from Beddit 3.5 ($r = .99$; $P < .001$). Relative to PSG, SE% ICC values for Beddit 3.5 (ICC = 0.98; **Figure 1B**) were significantly better than the levels of agreement with PSG for Beddit 3.0 or actigraphy ($P < .05$; **Table 1**). Self-reported SE% was more than 5% lower (standard deviation = 15.35%) than PSG-measured SE% and showed poor levels of agreement (ICC = 0.38).

SOL

The early version of the nonwearable device (Beddit 3.0) underestimated SOL by 13 minutes (95% CI, –58.99 to 32.13) and showed essentially zero reliability (ICC = 0.02) relative to PSG (**Table 1** and **Figure S3** in the supplemental material). By contrast, **Figure 1C** shows that the newer version of the nonwearable device (Beddit 3.5) had excellent reliability with PSG (ICC = 0.92) with only 6 minutes of bias error (95% CI, –5.49 to 18.28; **Figure S3**). Actigraphy and self-report performed less well, showing moderate-to-good reliability (ICC = 0.74) and poor reliability (ICC = 0.37), respectively, relative to PSG.
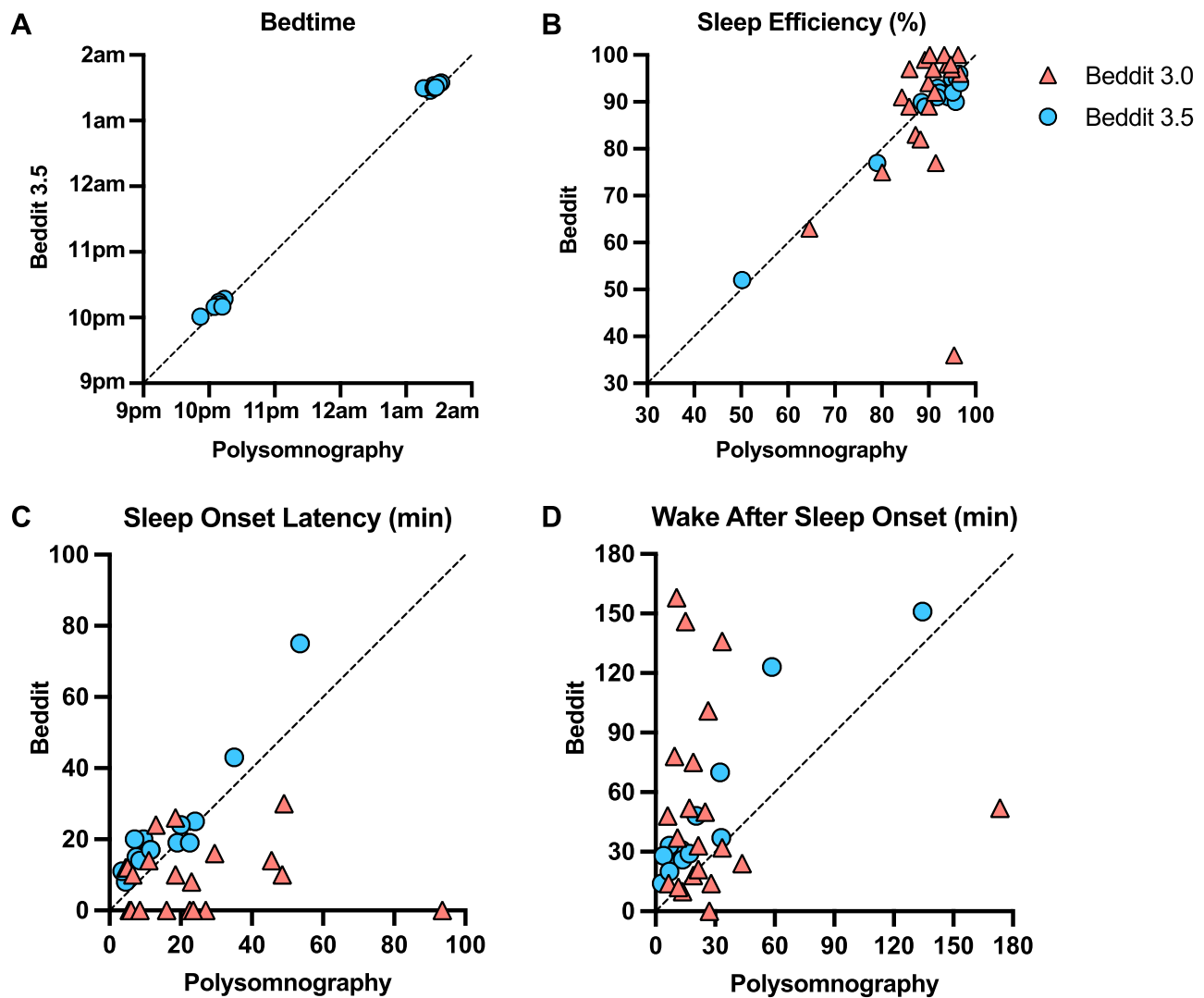
WASO

WASO often shows the weakest levels of agreement between sleep tracking devices and PSG. Indeed, Beddit 3.0 displayed very poor reliability for WASO (ICC = –0.02), overestimating WASO duration by 25 minutes relative to PSG (95% CI, –89.90 to 141.23; **Table 1** and **Figure S4** in the supplemental material). Interestingly, **Figure 1D** illustrates that the newer Beddit 3.5 device provided excellent reliability for WASO relative to PSG (ICC = 0.92), although the bias error remained high at 22 minutes (95% CI, –7.47 to 50.90; **Figure S4**). In other words, Beddit 3.5 overestimated WASO for all participants, but its overestimations were very consistent in relation to PSG-measured WASO variability. Actigraphy and self-report showed a different pattern: They were just as likely to overestimate or underestimate WASO for a given individual, which led to relatively low average bias error (actigraphy: mean = –3.19; 95% CI, –60.52 to 54.14; self-report: mean = 5.47; 95% CI, –52.57 to 63.51), but the ICC values revealed these measures'

Table 1—PSG in comparison to Beddit, actigraphy, and self-report measures of bedtime, TST, SE%, SOL, and WASO.

	PSG (Mean, SD)	Comparison Device (Mean, SD)	Difference With PSG (Mean, SD)	Agreement With PSG (ICC [95% CI])	Main Effect of Device	Device by Sleep Condition Interaction
Bedtime						
Beddit 3.5 (n = 14)	12:05 AM (102.93 min)	12:00 AM (102.14 min)	−4.90 (3.52) min	0.999 [0.998–0.999]	$F(1,12) = 24.88$, $P < .001$, $\eta_p^2 = 0.68$***	$F(1,12) = 0.82$, $P = .383$, $\eta_p^2 = 0.06$
Actigraphy (n = 31)	11:30 PM (96.83 min)	11:30 PM (95.66 min)	0.68 (6.38) min	0.998 [0.995–0.999]	$F(1,29) = 0.10$, $P = .755$, $\eta_p^2 = 0.003$	$F(1,29) = 1.36$, $P = .254$, $\eta_p^2 = 0.05$
TST (min)						
Beddit 3.5 (n = 14)	375.07 (115.63)	376.64 (115.47)	1.57 (7.40)	0.998 [0.994–0.999]	$F(1,12) = 0.48$, $P = .502$, $\eta_p^2 = 0.04$	$F(1,12) = 0.24$, $P = .633$, $\eta_p^2 = 0.02$
Beddit 3.0 (n = 21)	399.40 (85.42)	444.81 (487.65)	45.40 (462.02)	0.13 [−0.31 to 0.52]	$F(1,19) = 0.05$, $P = .831$, $\eta_p^2 = 0.002$	$F(1,19) = 0.85$, $P = .367$, $\eta_p^2 = 0.04$
Actigraphy (n = 31)	400.34 (98.76)	398.16 (99.82)	−2.18 (28.32)	0.96 [0.92–0.98]	$F(1,29) = 0.29$, $P = .598$, $\eta_p^2 = 0.01$	$F(1,29) = 0.29$, $P = .594$, $\eta_p^2 = 0.01$
Self-report (n = 34)	384.65 (94.50)	357.35 (134.41)	−27.29 (110.24)	0.55 [0.27–0.75]	$F(1,32) = 1.93$, $P = .174$, $\eta_p^2 = 0.06$	$F(1,32) = 0.45$, $P = .508$, $\eta_p^2 = 0.01$
SE%						
Beddit 3.5 (n = 14)	89.45 (12.21)	88.36 (11.43)	−1.10 (2.09)	0.98 [0.95–0.995]	$F(1,12) = 3.24$, $P = .097$, $\eta_p^2 = 0.21$	$F(1,12) = 0.88$, $P = .366$, $\eta_p^2 = 0.07$
Beddit 3.0 (n = 21)	89.27 (7.08)	88.24 (15.49)	−1.04 (14.69)	0.26 [−0.19 to 0.61]	$F(1,19) = 0.34$, $P = .564$, $\eta_p^2 = 0.02$	$F(1,19) = 1.29$, $P = .270$, $\eta_p^2 = 0.06$
Actigraphy (n = 31)	88.91 (9.64)	87.94 (8.80)	−0.97 (5.71)	0.81 [0.64–0.90]	$F(1,29) = 1.20$, $P = .282$, $\eta_p^2 = 0.04$	$F(1,29) = 0.65$, $P = .428$, $\eta_p^2 = 0.02$
Self-report (n = 34)	89.35 (9.44)	83.83 (17.05)	−5.52 (15.35)	0.38 [0.05–0.63]	$F(1,32) = 4.24$, $P = .048$, $\eta_p^2 = 0.12$*	$F(1,32) = 0.002$, $P = .960$, $\eta_p^2 < 0.001$
SOL (min)						
Beddit 3.5 (n = 14)	16.96 (13.77)	23.36 (16.97)	6.39 (6.06)	0.92 [0.78–0.98]	$F(1,12) = 14.24$, $P = .003$, $\eta_p^2 = 0.54$**	$F(1,12) = 0.02$, $P = .890$, $\eta_p^2 = 0.002$
Beddit 3.0 (n = 21)	22.86 (21.52)	9.43 (9.30)	−13.43 (23.24)	0.02 [−0.41 to 0.44]	$F(1,19) = 5.21$, $P = .034$, $\eta_p^2 = 0.22$*	$F(1,19) = 3.75$, $P = .068$, $\eta_p^2 = 0.17$
Actigraphy (n = 31)	21.94 (19.45)	17.52 (23.41)	−4.42 (15.62)	0.74 [0.52–0.86]	$F(1,29) = 1.65$, $P = .209$, $\eta_p^2 = 0.05$	$F(1,29) = 1.30$, $P = .264$, $\eta_p^2 = 0.04$
Self-report (n = 34)	20.07 (18.91)	27.21 (14.96)	7.13 (19.19)	0.37 [0.04–0.62]	$F(1,32) = 4.87$, $P = .035$, $\eta_p^2 = 0.13$*	$F(1,32) = 0.76$, $P = .390$, $\eta_p^2 = 0.02$
WASO (min)						
Beddit 3.5 (n = 14)	25.86 (34.71)	47.57 (40.51)	21.71 (14.89)	0.92 [0.78–0.97]	$F(1,12) = 32.90$, $P < .001$, $\eta_p^2 = 0.73$***	$F(1,12) = 1.72$, $P = .214$, $\eta_p^2 = 0.13$
Beddit 3.0 (n = 21)	27.24 (34.91)	52.90 (46.68)	25.67 (58.96)	−0.02 [−0.44 to 0.40]	$F(1,19) = 6.25$, $P = .022$, $\eta_p^2 = 0.25$*	$F(1,19) = 3.49$, $P = .077$, $\eta_p^2 = 0.16$
Actigraphy (n = 31)	28.24 (36.01)	25.05 (15.61)	−3.19 (29.25)	0.44 [0.11–0.69]	$F(1,29) = 0.49$, $P = .489$, $\eta_p^2 = 0.02$	$F(1,29) = 0.26$, $P = .612$, $\eta_p^2 = 0.01$
Self-report (n = 34)	26.51 (34.82)	31.99 (44.00)	5.47 (29.61)	0.72 [0.51–0.85]	$F(1,32) = 1.17$, $P = .288$, $\eta_p^2 = 0.04$	$F(1,32) = 0.12$, $P = .734$, $\eta_p^2 = 0.004$

*$P < .05$, **$P < .01$, ***$P < .001$. CI = confidence interval, ICC = intraclass correlation coefficient, η_p^2 = partial eta squared, PSG = polysomnography, SD = standard deviation, SE% = sleep efficiency, SOL = sleep onset latency, TST = total sleep time, WASO = wake after sleep onset.

Figure 1—Comparisons of Beddit and PSG estimates.

Comparisons of Beddit and PSG estimates of bedtime (ICC > 0.99 for Beddit 3.5) **(A)**, SE% (ICC = 0.98 for Beddit 3.5, ICC = 0.26 for Beddit 3.0) **(B)**, SOL (ICC = 0.92 for Beddit 3.5, ICC = 0.02 for Beddit 3.0) **(C)**, and WASO (ICC = 0.92 for Beddit 3.5, ICC = –0.02 for Beddit 3.0) **(D)**. The diagonal lines indicate perfect agreement between Beddit and PSG. ICC = intraclass correlation coefficient, PSG = polysomnography, SE% = sleep efficiency, SOL = sleep onset latency, WASO = wake after sleep onset.

weak to modest reliability (actigraphy: ICC = 0.44; self-report: ICC = 0.72).

DISCUSSION

Previous work found that an early nonwearable device, Beddit 3.0, yielded poor accuracy across measures of sleep (SOL, WASO, TST, and SE%).[12] We replicated these findings. The newer version of Beddit (3.5), however, distinguished itself from its earlier version with strong agreement with PSG. Compared to PSG, the mean bias (and 95% CIs) of Beddit 3.5 was within the field's standards for TST (threshold: 40 minutes), SE% (threshold: 5%), SOL (threshold: 30 minutes), and WASO

(threshold: 30 minutes).[26] Notably, the agreement levels between Beddit 3.5 and PSG were stronger than the agreement levels between actigraphy and PSG for measures of SE%, SOL and WASO. Agreement levels were consistent across longer and shorter time-in-bed opportunities (ie, the two sleep conditions). Therefore, at least in healthy adults, Beddit 3.5 is a valid measure of TST, SE%, SOL, and WASO.

The current findings with Beddit 3.5 add to the developing literature on measuring sleep using minimal or no-contact sensors.[27] Such nonwearable devices estimate sleep by detecting movement, heart rate, and/or respiration using radiofrequency waves (eg, ResMed S+),[28] impulse radio ultra-wideband technology (Somnofy),[29] cardioballistic sensors (Beddit),[12] and piezoelectric sensors placed under the mattress (EarlySense).[30]

Figure 2—Bland-Altman plots.
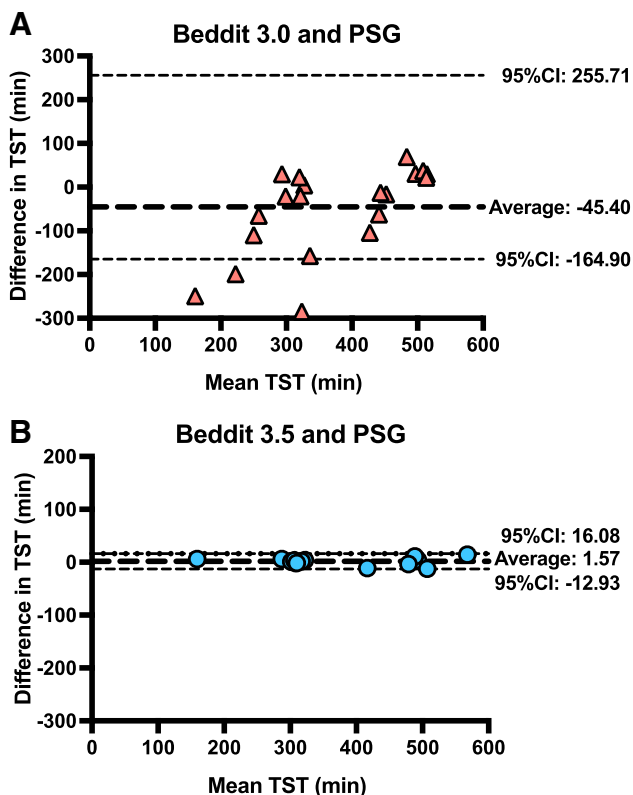
Bland-Altman plots for TST between Beddit 3.0 and PSG **(A)** and Beddit 3.5 and PSG **(B)**. Note that 1 outlier with TST > 40 hours was excluded in **(A)**. CI = confidence interval, PSG = polysomnography, TST = total sleep time.

In general, these nonwearables have shown high sensitivity (detecting sleep state), modest specificity (detecting wake state), and poor sleep staging relative to PSG.[27] In some cases, however, commercial sleep tracking devices have shown equal or better accuracy than wristband actigraphy for TST.[31] The current findings add to this literature that the nonwearable Beddit 3.5 outperforms wristband actigraphy under multiple sleep schedules (8.5-hour and 5.5-hour time-in-bed opportunities). Like other wearable and nonwearable devices, Beddit 3.5 has difficulty precisely measuring WASO (overestimates by 22 minutes), but at least for Beddit 3.5, its levels of overestimation were consistent across participants. Therefore, the Beddit 3.5 WASO estimate would be appropriate for research studies focused on interindividual variability in WASO but inappropriate for studies in which absolute threshold values must be identified.

This study has limitations. First, the single-center design and limited study population restricts data generalization. All participants were healthy, young adults who presented with normal sleep patterns and behaviors. Thus, these findings may fail to be replicated in other populations, such as those with increased variability in sleep or frequent disturbances, such as older adults or critically ill patients. Second, the Beddit 3.5 device does not provide epoch-by-epoch data, which are necessary to calculate sensitivity and specificity. Third, because Apple has not shared what was changed across the Beddit devices, it remains unclear

whether changes in firmware, device sensitivity, algorithms, or some other modification was responsible for the improved accuracy. Fourth, this validation study was conducted in a controlled sleep laboratory, and it remains to be tested whether Beddit 3.5 is equally accurate when there are bed partners or pets who sleep in bed, as commonly occurs in home settings.

In conclusion, some commercial technology, including the nonwearable Beddit 3.5, has reached acceptable levels of accuracy for sleep research in healthy adults. This finding reveals a clear advancement over the state of the commercial sleep tracking field just 6–8 years ago.[9–12] Future studies on Beddit 3.5 and other devices will need to determine whether the high levels of accuracy are retained in diverse groups and clinical populations. Furthermore, future work will need to determine whether these devices can be used accurately outside of a sleep laboratory (eg, home settings, in-patient settings). If the validity of commercial sleep tracking devices is upheld in such contexts, then there will be broad implications for sleep research, along with clinical applications. For example, in clinical settings, nonwearable devices could be utilized in conjunction with traditional sleep testing to inform circadian rhythm and insomnia disorders that are dominant in home settings. Such long-term, low-cost monitoring would provide depth to the clinical picture while avoiding first-night effects and other disturbances to sleep quality that are commonly observed with PSG and wearable sleep monitors.

ABBREVIATIONS

CI, confidence interval
ICC, intraclass correlation coefficient
PSG, polysomnography
SE%, sleep efficiency
SOL, sleep onset latency
TST, total sleep time
WASO, wake after sleep onset

REFERENCES

1. Agnew HW Jr, Webb WB, Williams RL. The first night effect: an EEG study of sleep. *Psychophysiology.* 1966;2(3):263–266.

2. Alshaer H, Ryan C, Fernie GR, Bradley TD. Reproducibility and predictors of the apnea hypopnea index across multiple nights. *Sleep Sci.* 2018;11(1):28–33.

3. Stone KL, Ancoli-Israel S. Actigraphy. In: Kryger MH, Roth T, Dement WC, eds. *Principles and Practice of Sleep Medicine.* 6th ed. Philadelphia, PA: Elsevier; 2016: 1671–1678.

4. Conley S, Knies A, Batten J, et al. Agreement between actigraphic and polysomnographic measures of sleep in adults with and without chronic conditions: a systematic review and meta-analysis. *Sleep Med Rev.* 2019;46: 151–160.

5. Adkins KW, Goldman SE, Fawkes D, et al. A pilot study of shoulder placement for actigraphy in children. *Behav Sleep Med.* 2012;10(2):138–147.

6. Perry GS, Patil SP, Presley-Cantrell LR. Raising awareness of sleep as a healthy behavior. *Prev Chronic Dis.* 2013;10:130081.

7. Buysse DJ. Sleep health: can we define it? Does it matter? *Sleep.* 2014; 37(1):9–17.

8. Shelgikar AV, Anderson PF, Stephens MR. Sleep tracking, wearable technology, and opportunities for research and clinical care. *Chest.* 2016;150(3):732–743.

9. Meltzer LJ, Hiruma LS, Avis K, Montgomery-Downs H, Valentin J. Comparison of a commercial accelerometer with polysomnography and actigraphy in children and adolescents. *Sleep.* 2015;38(8):1323–1330.

10. Liu W, Ploderer B, Hoang T. In Bed with Technology: Challenges and Opportunities for Sleep Tracking. Presented at: *Proceedings of the Annual Meeting of the Australian Special Interest Group for Computer Human Interaction*; December 2015; Parkville, Victoria, Australia.

11. Lee J, Finkelstein J. Consumer sleep tracking devices: a critical review. *Stud Health Technol Inform.* 2015;210:458–460.

12. Tuominen J, Peltola K, Saaresranta T, Valli K. Sleep parameter assessment accuracy of a consumer home sleep monitoring ballistocardiograph Beddit Sleep Tracker: a validation study. *J Clin Sleep Med.* 2019;15(3):483–487.

13. Research Nester. Smart sleep tracking device market: global demand analysis and opportunity outlook 2024. https://www.researchnester.com/reports/smart-sleep-tracking-device-market/373. Accessed October 12, 2021.

14. Sadeh A. The role and validity of actigraphy in sleep medicine: an update. *Sleep Med Rev.* 2011;15(4):259–267.

15. Chee NIYN, Ghorbani S, Golkashani HA, Leong RLF, Ong JL, Chee MWL. Multi-night validation of a sleep tracking ring in adolescents compared with a research actigraph and polysomnography. *Nat Sci Sleep.* 2021;13:177–190.

16. Iber C, Ancoli-Israel S, Chesson AL, Quan SF. The new sleep scoring manual—the evidence behind the rules. *J Clin Sleep Med.* 2007;3(2):107.

17. Kushida CA, Chang A, Gadkary C, Guilleminault C, Carrillo O, Dement WC. Comparison of actigraphic, polysomnographic, and subjective assessment of sleep parameters in sleep-disordered patients. *Sleep Med.* 2001;2(5):389–396.

18. Lichstein KL, Stone KC, Donaldson J, et al. Actigraphy validation with insomnia. *Sleep.* 2006;29(2):232–239.

19. Morgenthaler T, Alessi C, Friedman L, et al; American Academy of Sleep Medicine. Practice parameters for the use of actigraphy in the assessment of sleep and sleep disorders: an update for 2007. *Sleep.* 2007;30(4):519–529.

20. Rijsketic JM, Dietch JR, Wardle-Pinkston S, Taylor DJ. Actigraphy (Actiware) Scoring Hierarchy Manual. https://insomnia.arizona.edu/actigraphy. Accessed October 12, 2021.

21. Lauderdale DS, Knutson KL, Rathouz PJ, Yan LL, Hulley SB, Liu K. Cross-sectional and longitudinal associations between objectively measured sleep duration and body mass index: the CARDIA Sleep Study. *Am J Epidemiol.* 2009;170(7):805–813.

22. Buysse DJ, Reynolds CF III, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. *Psychiatry Res.* 1989;28(2):193–213.

23. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med.* 2016;15(2):155–163.

24. Altman DG, Bland JM. Measurement in medicine: the analysis of method comparison studies. *The Statistician.* 1983;32(3):307–317.

25. Sackett PR, Laczo RM, Arvey RD. The effects of range restriction on estimates of criterion interrater reliability: implications for validation research. *Person Psychol.* 2002;55(4):807–825.

26. Smith MT, McCrae CS, Cheung J, et al. Use of actigraphy for the evaluation of sleep disorders and circadian rhythm sleep-wake disorders: an American Academy of Sleep Medicine systematic review, meta-analysis, and GRADE assessment. *J Clin Sleep Med.* 2018;14(7):1209–1230.

27. Lujan MR, Perez-Pozuelo I, Grandner MA. Past, present, and future of multisensory wearable technology to monitor sleep and circadian rhythms. *Front Digit Health.* 2021;3:721919.

28. Schade MM, Bauer CE, Murray BR, et al. Sleep validity of a non-contact bedside movement and respiration-sensing device. *J Clin Sleep Med.* 2019; 15(7):1051–1061.

29. Toften S, Pallesen S, Hrozanova M, Moen F, Grønli J. Validation of sleep stage classification using non-contact radar technology and machine learning (Somnofy). *Sleep Med.* 2020;75:54–61.

30. Tal A, Shinar Z, Shaki D, Codish S, Goldbart A. Validation of contact-free sleep monitoring device with comparison to polysomnography. *J Clin Sleep Med.* 2017; 13(3):517–522.

31. Chinoy ED, Cuellar JA, Huwa KE, et al. Performance of seven consumer sleep-tracking devices compared with polysomnography. *Sleep.* 2021;44(5): zsaa291.

SUBMISSION & CORRESPONDENCE INFORMATION

DISCLOSURE STATEMENT