EMERGING TECHNOLOGIES

Sleep Parameter Assessment Accuracy of a Consumer Home Sleep Monitoring Ballistocardiograph Beddit Sleep Tracker: A Validation Study

Jarno Tuominen, MA[1]; Karoliina Peltola, MA[1]; Tarja Saaresranta, MD, PhD[2,3]; Katja Valli, PhD[1,4]

[1]Department of Psychology and Speech-Language Pathology, University of Turku, Finland; [2]Division of Medicine, Department of Pulmonary Diseases, Turku University Hospital, Turku, Finland; [3]Sleep Research Centre, Department of Pulmonary Diseases and Clinical Allergology, University of Turku, Finland; [4]Department of Cognitive Neuroscience and Philosophy, School of Bioscience, University of Skövde, Sweden

Study Objectives: Growing interest in monitoring sleep and well-being has created a market for consumer home sleep monitoring devices. Additionally, sleep disorder diagnostics, and sleep and dream research would benefit from reliable and valid home sleep monitoring devices. Yet, majority of currently available home sleep monitoring devices lack validation. In this study, the sleep parameter assessment accuracy of Beddit Sleep Tracker (BST), an unobtrusive and non-wearable sleep monitoring device based on ballistocardiography, was evaluated by comparing it with polysomnography (PSG) measures. We measured total sleep time (TST), sleep onset latency (SOL), wake after sleep onset (WASO), and sleep efficiency (SE). Additionally, we examined whether BST can differentiate sleep stages.

Methods: We performed sleep studies simultaneously with PSG and BST in ten healthy young adults (5 female/5 male) during two non-consecutive nights in a sleep laboratory.

Results: BST was able to distinguish SOL with some accuracy. However, it underestimated WASO and thus overestimated TST and SE. Also, it failed to discriminate between non-rapid eye movement sleep stages and did not detect the rapid eye movement sleep stage.

Conclusions: These findings indicate that BST is not a valid device to monitor sleep. Consumers should be careful in interpreting the conclusions on sleep quality and efficiency provided by the device.

Keywords: Beddit Sleep Tracker, validation, polysomnography, sleep parameter accuracy, sleep research methods

INTRODUCTION

Recent rise of interest in consumer health monitoring has spurred the development of mobile devices designed to measure objective sleep parameters.[1–5] Additionally, there is an urgent need in both clinical and research settings for a cost-effective, portable and reliable measuring device for studying sleep[6] in home settings. The current gold standard, polysomnography (PSG), is labor-intensive and costly in terms of time and resources, and as such often unfeasible for clinical or research settings. Thus far, however, only a few affordable home sleep monitoring devices have been validated against PSG.[1,5,7–9] Most of the sleep measures rely on either electroencephalography (EEG) or movement (cardiac, respiratory or body movements) data. Of the former, the most promising EEG methods (Zeo headband with 74 % PSG agreement[3] and the Nightcap with 93 % PSG agreement[10]) have unfortunately been discontinued and are no longer commercially available. The latter, movement measurement devices, include increasingly popular actigraphs, which are light wearable devices, usually connected to wrist, ankle or hip that provide information via an accelometer.[2,4,11,12] Actigraphy studies have provided mixed results, yet are currently used, for example, as an adjunct measure in sleep apnea monitoring.[12] Actigraphs have problems in detecting wakefulness, and as a result, overestimate both

total sleep time and sleep efficiency.[2,4,13] Recently, multidata approaches have gained ground and actigraphy devices have been used aside other data sources, such as the peripheral arterial blood flow tone.[7] One such novel device is the ŌURA ring, which in addition to actigraph data, collects heartbeat variation, heart rate and other variables from the finger to evaluate sleep parameters. The preliminary research findings have shown promise for this multidata approach,[14] but further studies are still required.

One of the recent commercially available home sleep monitoring devices is the Beddit Sleep Tracker (BST), a thin strip sensor placed under the mattress or mattress topper. BST relies on a 3-channel movement detection method, originally designed for the Static Charge Sensitive Bed.[15,16] BST transmits body, respiratory and heart (ballistocardiograph) movement data via a Bluetooth connection to a commercially designed app to calculate sleep parameters. An automated algorithm then transforms these aggregated sleep measures into an easy-to-read graph, compares it to previous nights, and provides information about sleep parameters in a cloud service. BST seems to be a promising tool, as it is based on a previously validated principle, and its demonstrated ability to measure heart rate[11] and respiration[17] has been found accurate, and it has undergone a single subject validation study.[18] It should be noted, however, that these validation studies have been performed as a part of

product development or by persons with corporate interests, and therefore an independent validation study is warranted.

In this study, we investigated the accuracy of the BST to measure objective parameters of sleep and to differentiate sleep stages. BST recordings were compared to PSG, and the parameters total sleep time (TST), sleep onset latency (SOL), wake after sleep onset (WASO) and sleep efficiency (SE) were investigated. Additionally, we analyzed whether BST could be used to detect sleep stages. If found to be sensitive and specific enough, BST would answer the need for a low cost and easy-to-use sleep monitoring method for consumer, clinical and scientific use.

METHODS

Ten right-handed non-smoking participants (5 female/5 male) aged 18–30 years (mean 24.5, SD 2.51, range 18–26), with a BMI < 30 kg/m^2 and no history of neurological or psychiatric diagnoses were selected based on an online screening with the Pittsburgh Sleep Quality Index (PSQI), excluding spousal questions.[19] A PSQI score 5 or less was considered a cutoff point. One participant reported a score of 6 due to daytime sleepiness and was individually interviewed before deemed applicable for the study.

Participants slept two non-consecutive nights at the Sleep Research Centre at the University of Turku, Finland, within the span of 1 week. Use of alcohol and medication was prohibited for 24 hours, and caffeine for 6 hours preceding the experiment. Simultaneous recordings of Embla (Medcare Flaga hf. Medical Devices, Reykjavik, Iceland) PSG (6 electrodes: C3–A2, C4–A1, O1–A2, O2–A1, F3–A2, F4–A1, Cz as Ref; EKG1&2; EOG1&2; EMG1&2; oximetry; thorax & abdomen belts), BST, and Interaxon Muse headband (data not available for the current study) were collected. BST sensor was placed under the mattress topper and connected via Bluetooth to a Nexus Android tablet device running a BST application 1.7.3 for the first ten observations, and 1.8.0. for the later ten observations, due to an automatic product update. The start of the recording was manually synchronized across devices. Additionally, before and after both nights, participants answered a subjective well-being questionnaire (results reported elsewhere).

PSG data for sleep parameters was analyzed with the Rem-Logic (Medcare Flaga hf. Medical Devices, Reykjavik, Iceland) program by an experienced sleep technician. Sleep data was analyzed in 30 second epochs following the American Academy of Sleep Medicine classification protocol.[20] BST score of "total amount of sleep" was compared to PSG TST, "time to fall asleep" to PSG SOL, and BST WASO and SE to their corresponding PSG categories. For the sleep stage classification BST data was categorized manually from the data graphs obtained from the BST cloud service. BST presents data in 2-minute segments, which were transformed to 30-second epochs, and then compared to corresponding PSG epochs. As the version of BST app used presents a graph on an axis of deep sleep–light sleep, the BST hypnogram was re-scored to four stages corresponding to PSG classification without REM sleep stage (wake = wake; light sleep = stage N1 and N2 sleep

with cutoff calculated from midpoint of the light sleep category in the BST hypnogram; deep sleep = stage N3 sleep). If the BST sleep stage score changed within the 2 minutes, 1 minute was scored as previous and 1 minute the next sleep stage. Automated BST scoring of REM sleep, available in the earlier version of the app, was no longer available in the version used in this study. Data from 1 night was omitted from the comparisons, due to BST losing wireless connection for an unknown reason. BST also did not provide total data for 2 nights: 1 night was missing SOL, and 1 night was missing SOL, WASO and SE. These, and their corresponding PSG data, were omitted from the comparisons.

Statistical analyses were performed using the IBM SPSS Statistics version 22, with the pre-set significance level of $P < .05$. Normality assumptions were tested using Kolmogorov-Smirnov test, and paired samples t tests were used for normally, and Mann-Whitney U test for non-normally, distributed variables. Sleep stage scoring was evaluated using Cohen kappa coefficient.

The study was approved by the Ethics Committee of the University of Turku, and a written informed consent was obtained prior to the study according to the declaration of Helsinki.

RESULTS

Paired-samples t test revealed a statistically significant difference between BST and PSG in mean TST ($t = 44.17$, $P < .01$) (**Table 1A**). BST seems to incorrectly score some wake epochs as sleep and thus overestimate TST. SOL did not differ between PSG and BST ($Z = −1.14$, $P = .20$), but differences between PSG and BST were revealed in WASO ($Z = −3.72$, $P < .01$), and SE ($Z = −3.34$, $P < .01$). BST underestimated the amount of wake after sleep onset and thus overestimated sleep efficiency, but this does not explain the difference (**Table 1A**). **Figure 1** provides a graphical presentation of the results (for additional Bland Altman analyses, see **Figure S1** in the supplemental material).

For the cross tabulation of PSG and BST sleep stage comparison, see **Table 1B**. BST classified 6.1% of PSG scored REM sleep as wake, 30.4% as stage N1 sleep, 24.8% as stage N2 sleep, and 38.7% as stage N3 sleep (total 3,406 REM epochs). As PSG REM sleep did not correlate with any particular BST stage, further analyses were conducted without REM sleep data points based on the PSG data (total 14,900 epochs).

Agreement between devices for the remaining classifications (wake, stage N1, N2, and N3 sleep) was extremely poor (kappa = .095, $P < .001$). Mean inter-device Cohen kappa was .098 (standard error .015, range −.009 to .192, 95% CI .068, .129) (**Table 1B**). Due to this lack of consensus, further analyses were conducted with the original, coarser BST classification which includes stages wake, light sleep (PSG stages N1 and N2 sleep) and deep sleep (PSG stage N3 sleep) (**Table 1C**) with PSG data scored accordingly. Agreement between PSG and BST sleep stage scoring improved slightly, but was still very low (kappa = .101, $P < .001$). Further inter-device analyses resulted in mean kappa of .113 with mean standard error of .020 (range .008 to .237, 95% CI .074, .152).

**Table 1**—Sleep parameter and sleep stage agreement between PSG and BST.

A

	PSG			BST			P
	n	Mean (SD)	Range	n	Mean (SD)	Range	
TST (minutes)	19	412.8 (58.1)	333.0–524.4	19	456.3 (52.0)	363.0–531.6	< .001
SOL (minutes)	17	34.1 (18.0)	5.5–70.0	17	30.9 (20.4)	8.0–86.0	NS
WASO (minutes)	18	46.1 (33.0)	13.0–121.5	18	13.5 (17.3)	0.0–56.0	< .001
SE (%)	18	83.6 (7.9)	68.0–91.8	18	90.9 (6.7)	78.2–98.2	< .001

B

		PSG			
		Wake (%)	N1 (%)	N2 (%)	N3 (%)
BST	Wake (%)	42.1	9.9	40.2	7.8
	N1 = Light (%)	31.4	9.6	45.3	13.7
	N2 = Light (%)	15.4	6.7	49.6	28.3
	N3 = Deep (%)	5.8	4.8	51.9	37.5

C

		PSG		
		Wake (%)	Light Sleep (N1 & N2) (%)	Deep Sleep (N3) (%)
BST	Wake (%)	42.1	50.2	7.8
	Light Sleep (%)	22.8	55.6	21.5
	Deep Sleep (%)	5.8	56.7	37.5

n varies due to loss of BST data. **(A)** Sleep parameter agreement: TST, SOL, WASO and SE. **(B)** Sleep stage agreement: wake, stage N1, N2 and N3 sleep. **(C)** Sleep stage agreement: wake, light and deep sleep. BST = Beddit Sleep Tracker, 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 BST and PSG sleep parameter estimations.
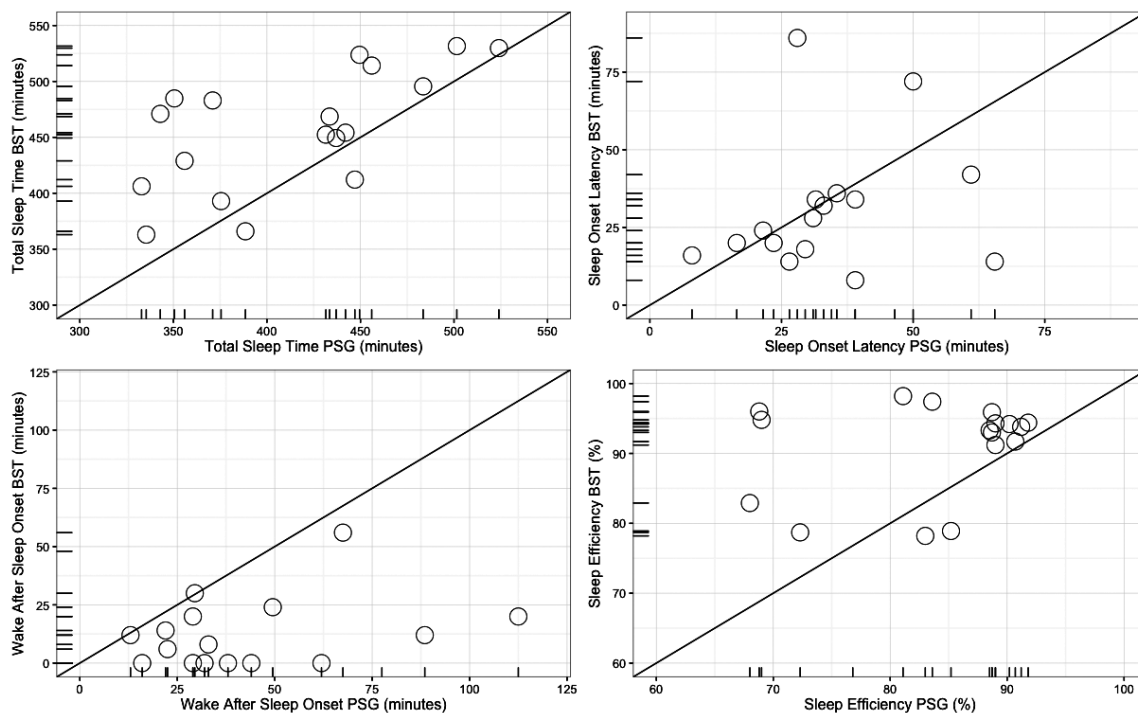
The diagonal line represents optimal accuracy. BST = Beddit Sleep Tracker, PSG = polysomnography.

DISCUSSION

This study investigated whether BST accurately measures sleep parameters (TST, SOL, WASO, SE), and whether it could be used to distinguish between sleep stages.

A consensus between BST and PSG was found only in SOL, and even there a large variance between the two measures was observed (**Figure 1**). BST underestimated WASO resulting in overestimation of TST and SE. However, this finding does not exhaustively explain the discrepancy between BST and PSG.

This, in turn, indicates that BST is not a reliable method for achieving comprehensive information about sleep in consumer use, and it misses important information about nocturnal awakenings informative both in clinical assessment and research. This finding should be kept in mind while evaluating previous research comparing, for example, FitBit commercial wristband accuracy to that of BST[21] without a PSG validation. There are several factors that make comparison between commercial devices difficult; frequent updates, variability in the definition of sleep parameters and especially lack of access to the underlying algorithms that quantify sleep parameters.[14,21]

Consensus between BST and PSG sleep stage classification was very poor when using a four-stage classification (wake, stage N1, N2 and N3 sleep), and only slightly better with a coarser classification (wake, light sleep, and deep sleep). Importantly, the currently available BST app lacks REM sleep classification, and PSG REM sleep did not predictably correlate with any BST stage. The problems of classifying stages of sleep under the rather arbitrary labels of "light" and "deep," or "light" and "sound" have been noted in previous studies of home sleep monitoring devices, and found to vary and not correspond with any specific sleep stage.[14,22,23] While interpreting our results, however, one should keep in mind that the necessary transformations of the sleep stage scoring may have provided a source of error, even after REM sleep was removed from the analysis and a more liberal classification was attempted.

To conclude, for sleep and dream research, or purposes of sleep disorder diagnostics,[20] BST is currently inapplicable in differentiating between sleep stages, calculating TST, WASO and SE. It can, however, estimate SOL to a reasonable degree. Despite good face validity and positive preliminary results,[18] BST at its current state provides consumers with inadequate and misleading information about their sleep parameters. Therefore, users should not make decisions regarding sleep quality and wellbeing based on BST data. It should be noted that this finding does not generalize to other home sleep measurements nor the future developments of the itself promising three channel method, but should be read as a call for further improvement and more large-scale validation studies. While the current surge of sleep monitoring devices is a welcome trend, further development work is still needed.

ABBREVIATIONS

BST, Beddit Sleep Tracker
PSG, polysomnography
PSQI, Pittsburgh Sleep Quality Index
SE, sleep efficiency
SOL, sleep onset latency
TST, total sleep time
WASO, wake after sleep onset

REFERENCES

1. Kelly JM, Strecker RE, Bianchi MT. Recent developments in home sleep-monitoring devices. *ISRN Neurol.* 2012;2012:768794.

2. Cellini N, Buman MP, McDevitt EA, Ricker AA, Mednick SC. Direct comparison of two actigraphy devices with polysomnographically recorded naps in healthy young adults. *Chronobiol Int.* 2013;30(5):691–698.

3. Shambroom JR, Fábregas SE, Johnstone J. Validation of an automated wireless system to monitor sleep in healthy adults. *J Sleep Res.* 2012;21(2):221–230.

4. Slater JA, Botsis T, Walsh J, King S, Straker LM, Eastwood PR. Assessing sleep using hip and wrist actigraphy. *Sleep Biol Rhythms.* 2015;13(2):172–180.

5. Van de Water ATM, Holmes A, Hurley DA. Objective measurements of sleep for non-laboratory settings as alternatives to polysomnography--a systematic review. *J Sleep Res.* 2011;20(1 Pt 2):183–200.

6. McNamara P, McLaren D, Smith D, Brown A, Stickgold R. A "Jekyll and Hyde" within aggressive versus friendly interactions in REM and non-REM dreams. *Psychol Sci.* 2005;16(2):130–136.

7. Herscovici S, Pe'er A, Papyan S, Lavie P. Detecting REM sleep from the finger: an automatic REM sleep algorithm based on peripheral arterial tone (PAT) and actigraphy. *Physiol Meas.* 2007;28(2):129–140.

8. Ko PR, Kientz JA, Choe EK, Kay M, Landis CA, Watson NF. Consumer sleep technology: a review of the landscape. *J Clin Sleep Med.* 2015;11(12):1455–1461.

9. Driller M, McQuillan J, O'Donnell S. Inter-device reliability of an automatic-scoring actigraph for measuring sleep in healthy adults. *Sleep Sci.* 2016;9(3):198–201.

10. Cantero JL, Atienza M, Stickgold R, Hobson JA. Nightcap: a reliable system for determining sleep onset latency. *Sleep.* 2002;25(2):238–245.

11. Paalasmaa J, Toivonen H, Partinen M. Adaptive heartbeat modelling for beat-to-beat heart rate measurement in ballistocardiograms. *IEEE J Biomed Health Inform.* 2015;19(6):1945–1952.

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

13. Kosmadopoulos A, Sargent C, Darwent D, Zhou X, Roach GD. Alternatives to polysomnography (PSG): a validation of wrist actigraphy and a partial-PSG system. *Behav Res Methods.* 2014;46(4):1032–1041.

14. de Zambotti M, Rosas L, Colrain IM, Baker FC. The sleep of the ring: comparison of the ōura sleep tracker against polysomnography. *Behav Sleep Med.* 2017:1–15.

15. Alihanka J. Vaahtoranta K, Saarikivi I. A new method for long-term monitoring of ballistocardiogram, heart rate and respiration. *Am J Physiol.* 1981;240(5):R384–R392

16. Salmi T, Leinonen L. Automatic analysis of sleep records with static charge sensitive bed. *Electroencephalogr Clin Neurophysiol.* 1986;64(1):84–87.

17. Paalasmaa J, Waris M, Toivonen H, Leppäkorpi L, Partinen M. Unobtrusive online monitoring of sleep at home. *Conf Proc IEEE Eng Med Biol Soc.* 2012;2012:3784–3788.

18. Paalasmaa J, Leppäkorpi L, Partinen M. Quantifying respiratory variation with force sensor measurements. *Conf Proc IEEE Eng Med Biol Soc.* 2011;2011:3812–3815.

19. Buysse DJ, Reynolds CF 3rd, Monk TH, Berman SR, Kupfer DJ. The Pittsburg Sleep Quality Index: a new instrument for psychiatric practice and research. *Psychiatry Res.* 1989;28(2):193–213.

20. Iber C, Ancoli-Israel S, Chesson AL Jr, Quan SF; for the American Academy of Sleep Medicine. *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications.* 1st ed. Westchester, IL: American Academy of Sleep Medicine; 2007.

21. Kolla BP, Mansukhani S, Mansukhani MP. Consumer sleep tracking devices: a review of mechanisms, validity and utility. *Expert Rev Med Devices.* 2016;13(5):497–506.

22. de Zambotti M, Baker FC, Colrain IM. Validation of sleep-tracking technology compared with polysomnography in adolescents. *Sleep.* 2015;38(9):1461–1468.

23. Mantua J, Gravel N, Spencer RM. Reliability of sleep measures from four personal health monitoring devices compared to research-based actigraphy and polysomnography. *Sensors.* 2016;16(5):E646.

SUBMISSION & CORRESPONDENCE INFORMATION

DISCLOSURE STATEMENT

EDITOR'S NOTE

The Emerging Technologies section focuses on new tools and techniques of potential utility in the diagnosis and management of any and all sleep disorders. The technologies may not yet be marketed, and indeed may only exist in prototype form. Some preliminary evidence of efficacy must be available, which can consist of small pilot studies or even data from animal studies, but definitive evidence of efficacy will not be required, and the submissions will be reviewed according to this standard. The intent is to alert readers of *Journal of Clinical Sleep Medicine* of promising technology that is in early stages of development. With this information, the reader may wish to (1) contact the author(s) in order to offer assistance in more definitive studies of the technology; (2) use the ideas underlying the technology to develop novel approaches of their own (with due respect for any patent issues); and (3) focus on subsequent publications involving the technology in order to determine when and if it is suitable for application to their own clinical practice. The *Journal of Clinical Sleep Medicine* and the American Academy of Sleep Medicine expressly do not endorse or represent that any of the technology described in the Emerging Technologies section has proven efficacy or effectiveness in the treatment of human disease, nor that any required regulatory approval has been obtained.