Nature and Science of Sleep

Dovepress
open access to scientific and medical research

Open Access Full Text Article

ORIGINAL RESEARCH

Validation of Fitbit Inspire 2™ Against Polysomnography in Adults Considering Adaptation for Use

Su Eun Lim*, Ho Seok Kim*, Si Woo Lee, Kwang-Ho Bae, Young Hwa Baek

KM Data Division, Korea Institute of Oriental Medicine, Daejeon, Republic of Korea

*These authors contributed equally to this work

Correspondence: Young Hwa Baek, KM Data Division Korea Institute of Oriental Medicine, 1672 Yuseong-daero, Yuseong-gu, Daejeon, 34054, Republic of Korea, Tel +82 10-2555-2138, Email aori79@kiom.re.kr

Purpose: The commercialization of sleep activity tracking devices has made it possible to manage sleep quality at home. However, it is necessary to verify the reliability and accuracy of wearable devices through comparison with polysomnography (PSG), which is the standard for tracking sleep activity. This study aimed to monitor overall sleep activity using Fitbit Inspire 2™ (FBI2) and to evaluate its performance and effectiveness through PSG under the same conditions.

Patients and Methods: We compared the FBI2 and PSG data of nine participants (four male and five female participants; average age, 39 years) without severe sleeping problems. The participants wore FBI2 continuously for 14 days, considering the period of adaptation to the device. FBI2 and PSG sleep data were compared using paired *t*-tests, Bland–Altman plots, and epoch-by-epoch analysis for 18 samples by pooling data from two replicates.

Results: The average values for each sleep stage obtained from FBI2 and PSG showed significant differences in the total sleep time (TST), deep sleep, and rapid eye motion (REM). In the Bland–Altman analysis, TST ($P = 0.02$), deep sleep ($P = 0.05$), and REM ($P = 0.03$) were significantly overstated in FBI2 compared to PSG. In addition, time in bed, sleep efficiency, and wake after sleep onset were overestimated, while light sleep was underestimated. However, these differences were not statistically significant. FBI2 showed a high sensitivity (93.9%) and low specificity (13.1%), with an accuracy of 76%. The sensitivity and specificity of each sleep stage was 54.3% and 62.3%, respectively, for light sleep, 84.8% and 50.1%, respectively, for deep sleep, and 86.4% and 59.1%, respectively for REM sleep.

Conclusion: The use of FBI2 as an objective tool for measuring sleep in daily life can be considered appropriate. However, further research is warranted on its application in participants with sleep-wake problems.

Keywords: wearable, sleep, tracking, polysomnography, validation study

Introduction

Sleep is essential for maintaining a healthy life. Inadequate sleep increases the risk of cardiovascular disease, type 2 diabetes, and hypertension.[1] In a recent study, shorter sleep duration was associated with 1.28 times higher risk of metabolic syndrome.[2] The average sleep time of Koreans is 6.76 hours, which shows a general trend of getting shorter over time.[3] The number of patients with sleep disorders has also increased by 8% annually over the past five years, and the number of patients with sleep disorders was 650,000 as of 2020.[4] Warning values for these sleep states lead to an interest in sleep products that help healthy sleep. Therefore, the use of wearable devices that easily measure sleep status in daily life has increased.[5] Moreover, as the interest in sleep health has increased, research focused on verifying the reliability and accuracy of wearable devices has expanded.[6]

Polysomnography (PSG), an accurate standard method for measuring sleep, has been widely used in research on objective sleep assessments.[7] However, the measurement of PSG takes eight to ten hours, and there is a limit to

Received: 15 October 2022
Accepted: 15 February 2023
Published: 28 February 2023

measuring the usual sleep time by conducting the test in an unfamiliar environment. Therefore, the test requires time, effort, and payment.[5] To overcome these limitations, a wrist-worn Fitbit, which is a non-invasive method certified by the Food and Drug Administration, is used.[8]

The latest Fitbit, manufactured after 2017, uses a sleep stage evaluation method that monitors data at regular intervals for waking and sleeping times through a motion sensor and heart rate measurement. Compared with the initial model, it is possible to evaluate not only sleep parameters but also sleep stages.[9] Therefore, it is possible to effectively use Fitbit for routine sleep monitoring and clinical sleep evaluation in patients with sleep or mental disorders. In addition, the continuous management of sleep habits is possible.[10,11] However, Fitbit has a limitation in that it only uses information collected after measuring sleep based on movement and heart rate, unlike PSG, which detects sleep stages through biosignals.

Certain differences exist in the reliability and accuracy based on the Fitbit model; hence, many comparative studies using PSG have been reported to evaluate the effectiveness of Fitbit.[12] Overall, Fitbit showed high sensitivity (0.95–0.96) and relatively low specificity (0.58–0.69) for sleep detection in a meta-analysis of the accuracy of Fitbit and PSG, which can evaluate sleep stages. Wake time after sleep onset (WASO), total sleep time (TST), and sleep efficiency (SE) did not differ from those of PSG. However, the sleep onset latency is underestimated.[13] Thus, Fitbit is considered an alternative to collecting objective sleep data, which was previously only available through PSG in the laboratory.[14]

Nevertheless, studies comparing the validity of Fitbit and PSG are limited because most studies were one-time experiments.[15] In addition, the first-night effect occurred in 50% of the participants, as the experiment was conducted in an unfamiliar laboratory different from the familiar environment at night. Moreover, 66–81% of participants reported discomfort during sleep due to the simultaneous use of several unfamiliar devices.[16] These problems decreased the participants' sleep quality during the experiment, which affected the interpretation of the sleep study results.

Therefore, it is necessary to compare the results obtained from the everyday wear of a device and repeated measurements with PSG to verify the reliability of the Fitbit. This study aimed to evaluate the accuracy and reliability of Fitbit by comparing the results obtained from the Fitbit Inspire 2™ (FBI2) and PSG with repeated measurements after using the device for a certain period.

Materials and Methods

Study Participants and Study Design

Study Participants

This study was conducted from October 2021 to December 2021, targeting 12 participants aged between 20 and 55 years living in Daejeon without severe sleep disorders. Based on the mean difference and standard deviation of TST measured using Fitbit and PSG in previous studies, 9 participants were selected with a significance level of 5% and a power of 80% and a total of 12 participants were calculated by applying a dropout rate of 20%.[17,18] The participants were recruited through a recruitment notice at D Hospital in Daejeon. Written informed consent was obtained from all the participants. The criteria for selecting participants were adults between the ages of 20 and 55 years and those with mild to moderate insomnia, as indicated by a score of 8–21 on the insomnia severity index (ISI).[19] The following participants were excluded: those with sleep disorders, psychiatric disorders, internal and external medical disorders, and musculoskeletal disorders; those taking psychiatric/neurological drugs, including sleeping pills; and those with irregular sleep-wake cycles, such as shift work. The subject's conformity assessment was performed through a doctor's examination, questionnaire, and blood tests. The study was approved by the institutional review board of the Daejeon Korean Medicine Hospital of Daejeon University (approval number: DJDSKH-21-BM-15), and conducted in accordance with the Declaration of Helsinki.

Study Design

In this study, a total of three visits to the hospital were performed: subject registration (visit one), the first comparison of the Fitbit (FitBit® Inc, San Francisco, California, USA) and PSG (one week after visit one), and the second comparison of the Fitbit and PSG (two weeks after visit one). The Fitbit was worn continuously for 14 days, which was the study period, considering the participant's adaptation period to the Fitbit. After being admitted to the hospital for two days and

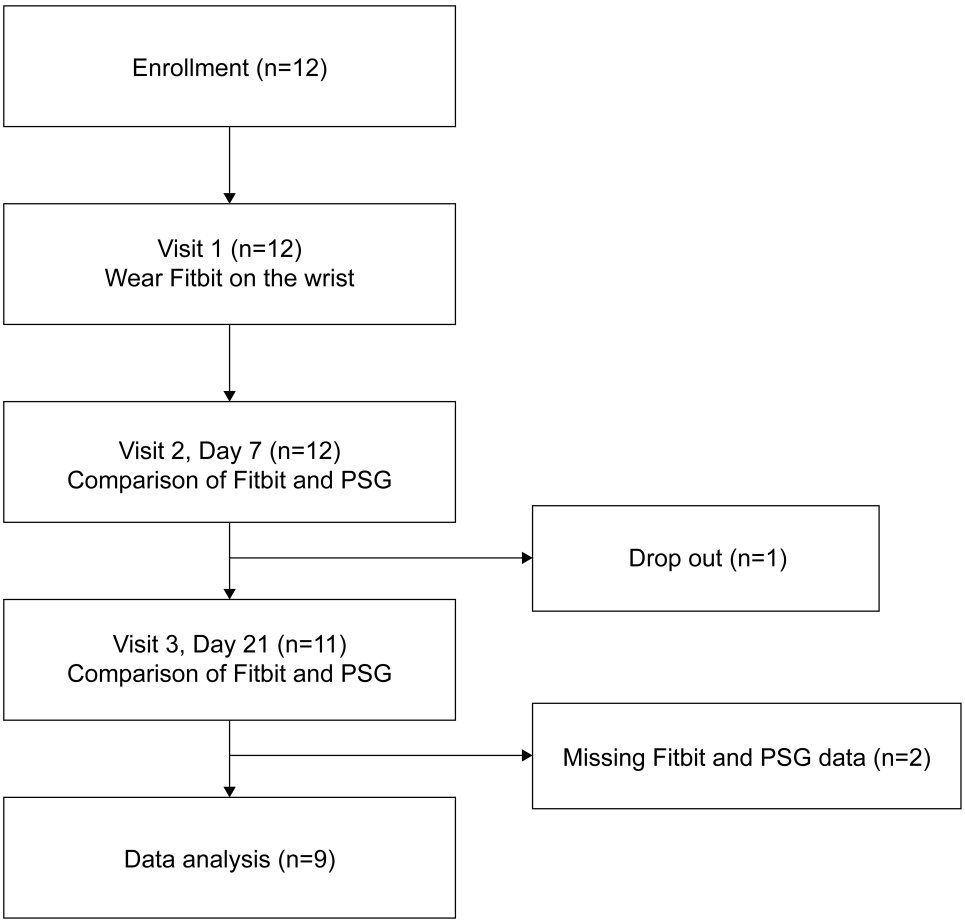

Figure 1 Flow diagram for selection of study participants. PSG, polysomnography.

one night, PSG was measured during sleep by PSG professional medical staff. At this time, the measurement was conducted while wearing the Fitbit. During the PSG measurement, the researcher recorded the sleep diaries before and after the participants slept, and continuously monitored the safety status of the participants. Twelve participants were enrolled, and data from nine participants who completed the three visits were included in the final analysis (excluding data from one dropout and two missing data from the Fitbit) (Figure 1).

Study Tools

PSG

The patients underwent PSG twice. Routine activity was maintained before the start of PSG. However, consumption of caffeine, tobacco, and alcohol was discouraged. The temperature, humidity, and illuminance of all PSG rooms were kept the same, and all the participants wore the same lab coat. The measurement was conducted at 8 p.m. in the sleep PSG test room before the test, and two professional sleep PSG technicians took turns conducting the test while the subject slept. Rechtschaffen and Kales sleep phase scoring was calculated based on a 30-second epoch in accordance with standard guidelines.[20] Compumedics/Siesta 802a was used as the PSG instrument, and a certified polysomnographic technician performed a standardized method. The sensors were connected to measurement locations, such as the head, face, trunk, and lower extremities. Various biosignals such as respiration, snoring, electroencephalogram, electrooculogram, electromyogram, electrocardiogram, body position, and blood oxygen concentration of the participants were detected through sensors. In addition, the sleep stages were identified by recording the entire night's sleep time. The variables used for PSG analysis were as follows: time in bed (TIB; min), TST (min), SE (%), WASO (min), Stages 1–3 (min), and total rapid eye motion (REM) sleep time (min).

FBI2

The collected sleep activities and patterns were accessible through a mobile app with the subject's consent. The device was worn before bedtime to the following morning in the laboratory, excluding shower time. A fully charged device was used in this study. Fitbit data were extracted by assigning sleep and wake stages of 30-seconds for epoch-by-epoch (EBE) analysis. Sleep information obtained through FBI2 was matched with PSG variables as follows by referring to sleep parameters on the Fitbit website:[21] TIB (min), TST (min), SE (%), WASO (min), total REM sleep time (min). For comparison against FBI2 staging classifications, a summation of stages one and two was calculated to represent light sleep and stages three and four was considered deep sleep.

Data Analysis and Statistical Methods

Data from nine participants were included in the final analysis. Eighteen cases were analyzed by integrating the data of the two visits with both Fitbit and PSG data. The mean and standard deviation (SD) values are presented for continuous variables, such as the demographic characteristics of the study participants. The frequency and percentage values are presented for categorical data. These were used for analysis by referring to the standard methods for evaluating the performance of sleep PSG and sleep trackers.[22] A paired *t*-test was used to compare the results of the sleep parameters. The overall agreement of the results was analyzed using the Bland–Altman plot and EBE comparisons. Statistical data analysis was performed using SPSS Version 24 (SPSS, IBM Corp.) and R version 4.1.0 (The R Foundation for Statistical Computing). A two-sided test was performed, and statistical significance was set at a P value of < 0.05.

Bland–Altman Method

The Bland–Altman method displays the difference between the two measurements and compares the distribution of the differences. In this study, the difference and SD between the two measurements, the 95% confidence interval for the difference, and the lower and upper limits of agreement (difference of mean [1.96 SD]) for TIB, SE, TST, WASO, light sleep, deep sleep, and REM presented by PSG and Fitbit were calculated and visually displayed on a plot.

EBE Comparisons

To compare the epoch between the two measurement methods, only the data measured simultaneously for each method were used for the analysis. The stages of each measurement were changed as follows: In the PSG, awake was changed to WASO, and Stages 1, 2, 3, and REM were changed to sleep. In the Fitbit, wake was retained as wake, and light, deep, and REM were changed to sleep. In addition, because the time interval between the epoch measurements was not the same, the epoch measurement time interval for each method was changed to 30s before recording. When both wake and sleep stages were measured in the 30-second block, they were coded as wake stages. Agreement, sensitivity, specificity, positive predictive value, and negative predictive value were evaluated using the classified data.[23]

Results

General Characteristics of Participants

The general characteristics of the nine participants are presented in Table 1. There were four males (36.4%) and five females (45.5%), and the mean ISI score was 13.7 (4.4).

Comparison Between FBI2 and PSG

Table 2 shows the results of the comparison between the FBI2 and PSG using 18 data samples combined from visits two and three. Statistically significant differences were observed between the Fitbit and PSG in TST, deep sleep, and REM sleep ($P < 0.05$). The Fitbit and PSG TST showed a difference of 17.91 minutes. Deep sleep in the Fitbit was 14.67 minutes longer than in the PSG. Other variables showed no differences between the Fitbit and PSG.

Bland–Altman Mean Difference Analysis

Figure 2 shows the comparison of differences and distributions between the Fitbit and PSG measurements by Bland–Altman analysis. From the results of the Bland–Altman comparison, TST ($P = 0.02$), deep sleep ($P = 0.05$), and REM

Table 1 Participant Characteristics

Variables	n=9	(%)
Gender		
Male	4	44.4
Female	5	55.6
Age (years)	39.0±8.0	
Height (cm)	167.5±8.3	
Weight (kg)	66.8±15.2	
BMI (kg/m²)	23.6±3.3	
Drinking		
Yes	3	33.3
No	6	66.7
Smoking		
Yes	6	66.7
No	3	33.3
ISI		
Mean	13.7±4.4	
Mild (8–14)	5	55.6
Moderate (15–21)	4	44.4

Abbreviations: BMI, body mass index, ISI, insomnia severity index.

Table 2 PSG and Fitbit Inspire 2 Sleep Outcomes Using Full Sample (n=18)

Variables	Fitbit		PSG		t	*p*-value
	Mean±SD	±95% CI	Mean±SD	±95% CI		
TIB (min)	414.94±88.93	370.72~459.17	413.81±52.25	387.82~439.79	0.076	0.940
TST (min)	364.33±79.98	324.56~404.11	346.42±73.89	309.67~383.16	2.487	0.024
SE (%)	87.83±4.32	85.68~89.98	83.68±14.03	76.70~90.66	1.364	0.190
WASO (min)	50.61±21.85	39.74~61.48	40.25±49.55	15.61~64.89	0.757	0.459
Light sleep (min)	225.39±52.45	199.30~251.47	231.75±47.37	208.19~255.31	−0.485	0.634
Deep sleep (min)	64.28±27.27	50.72~77.84	49.61±35.10	32.16~67.07	2.104	0.051
REM (min)	74.67±24.70	62.38~86.95	65.06±27.70	51.28~78.83	2.308	0.034

Notes: Light sleep (Stage 1+2); deep sleep (Stage 3); *p*-value calculated by paired *t*-test.
Abbreviations: PSG, polysomnography; CI, confidence interval; SD, standard deviation; TIB, time in bed; TST, total sleep time; SE, sleep efficiency; WASO, wake after sleep onset; REM, rapid eye movement.

($P = 0.03$) measured by FBI2 were significantly overestimated. TIB ($P = 0.94$), SE ($P = 0.19$), and WASO ($P = 0.46$) tended to be overestimated but were not statistically significant. In addition, light sleep ($P = 0.63$) was underestimated but was not statistically significant. In the Bland–Altman plot, all measurements except one case were in the 95% confidence interval, indicating a high degree of agreement.

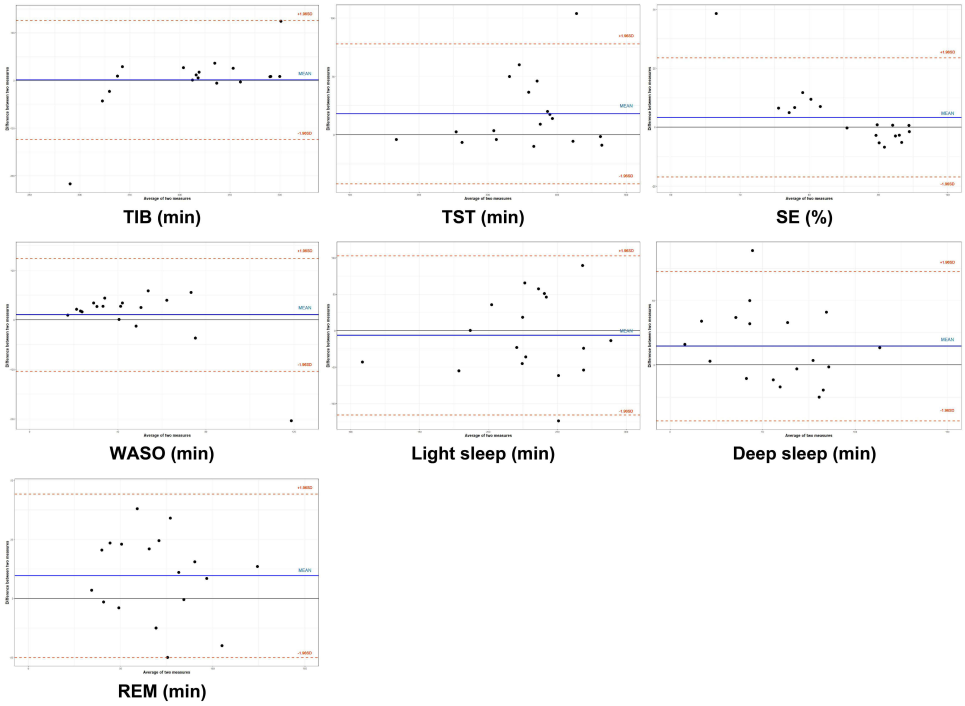

Figure 2 Bland–Altman plots of the FBI2 versus PSG. Bland–Altman plots presenting the different values of the FBI2 and PSG on the y-axis against PSG values on the x-axis across TIB, TST, SE, WASO, light sleep (Stage 1+2), deep sleep (Stage 3) and REM. The horizontal solid blue line denotes the average mean difference, while the dashed lines represent the 95% confidence interval (or lower-upper agreement limit). FBI2, Fitbit Inspire 2™.
Abbreviations: PSG, polysomnography; TIB, time in bed; TST, total sleep time; SE, sleep efficiency; WASO, wake after sleep onset; REM, rapid eye movement.

Fitbit and PSG EBE Comparisons

The sensitivity and specificity of total sleep, light sleep, deep sleep, and REM sleep were investigated for each sleep stage in the FBI2 and PSG by EBE analysis. For total sleep, FBI2 had low specificity (13.1%), high sensitivity (93.9%), and moderate-to-low accuracy (76.0%). For light sleep, FBI2 showed a relatively low sensitivity (54.3%), specificity (62.3%), and accuracy (59.1%), while sensitivity to deep sleep (84.8%) and REM sleep (86.4%) was moderate. However, the accuracy was high for deep (98.2%) and REM sleep (92.3%). Deep sleep (50.1%) and REM sleep (59.1%) showed a moderate-to-low specificity (Table 3).

Discussion

In this study, we compared the results of FBI2 with those of PSG through two repeated experiments at intervals of one week under the same conditions to evaluate the effectiveness of sleep and wake measurement function of FBI2 in nine participants without sleep disorders. This is a supplementary study verifying Fitbit's accuracy. The two main results of

Table 3 Sensitivity, Specificity, and Accuracy of Fitbit Inspire 2 Compared with PSG Using Full Sample (n=18)

Category	Sensitivity (%)	Specificity (%)	Accuracy (%)	PPV (%)	NPV (%)
Total sleep	93.9%	13.1%	76.0%	79.2%	38.0%
Light sleep Deep sleep	54.3% 84.8%	62.3% 50.1%	59.1% 83.7%	49.2% 98.2%	67.0% 9.5%
REM	86.4%	59.1%	82.3%	92.3%	43.2%

Notes: Light sleep (Stage 1+2); deep sleep (Stage 3).
Abbreviations: PSG, polysomnography; PPV, positive predictive value; NPV, negative predictive value; REM, rapid eye movement.

this study are as follows: First, the FBI2 variables TST, deep sleep, and REM sleep were significantly overestimated. Second, the total sleep result showed a sensitivity of 93.9%, specificity of 13.1%, and accuracy of 76.0%. Therefore, the accuracy of each sleep parameter was determined.

Bland–Altman's comparison evaluated the performance of FBI2 through sleep variables for PSG, which was consistent with the results of Bland–Altman's paired *t*-test. The study results showed significant differences in TST, deep sleep, and REM sleep among the sleep variables measured twice using PSG and FBI2.

The mean difference between FBI2 and PSG in the TST was 17.9 minutes and 14.67 minutes in deep sleep, which was significantly overestimated. Previous studies have shown that there is a significant mean difference between the PSG and sleep measurement wearable devices, wherein sleep measurement wearable devices make that overestimation. In a previous study that performed PSG using the Fitbit Alta HR on 49 participants, TST and deep sleep were significantly overestimated, similar to the results obtained in this study.18 In addition, SE was 4.15%, and WASO was 10.36 minutes, which was overestimated. WASO is mainly found in patients with sleep disorders; in previous studies when comparing WASO with insomnia and healthy individuals, in the healthy individual group, the standard deviation was found to be greater than the average value.[24,25] In addition, other previous studies have demonstrated the standard deviation to be large; moreover, no significant results were explained as a limitation of the small number of participants.[26] Considering the limitations of this study, the number of participants was small, and the average value of WASO was determined to be greater than the standard deviation by analyzing the general public.

The Bland–Altman plot was included in the 95% confidence interval, indicating a constant value. This is contrary to the results of a previous study in which the agreement between the Fitbit and PSG was lower as WASO increased in the Bland–Altman plot when the Fitbit Alta HR device was used for patients with insomnia. This conflicting result might be because the participants in this study did not have sleep disorders.[27] Therefore, the accuracy was confirmed to be high when FBI2 was used in participants with sleep disturbance. Light sleep was underestimated by −6.36. In a previous study that compared Withings Pulse O2 with PSG in 40 healthy adults, light sleep (average bias of 79 min) was also underestimated, which is similar to the results of this study.[28]

The sensitivity of the participants for the total sleep stage of FBI2 was 93.9%, indicating that sleep was accurately detected. However, the specificity for detecting total sleep stage was relatively low (13.1%). This is consistent with the results of the evaluation of the sleep effectiveness of the wrist-worn wearable device based on PSG results. Therefore, it was verified that these devices could accurately measure sleep owing to their high overall sensitivity and low specificity.[29] The low specificity of the Fitbit is related to the perception of sleep when there is no movement. However, the latest Fitbit detected more accurately the wake epoch, or WASO, during sleep.[13] Nevertheless, compared with the results of other studies using Fitbit with a similar algorithm, the sensitivity in this study was similar. However, the specificity (0.58 to 0.69) is slightly lower.[13] Therefore, although repetitive measurements were performed in this study, it can be inferred that the subject slept with minimal movement in an uncomfortable state in hospital-based laboratories, unlike sleep at home.

The first-night effect is mainly observed in patients with insomnia, and this study was conducted in the general public; thus, the first night effect was not substantially applied.[30] However, the results were compared by repeated measurements to collect accurate sleep information using sleep polysomnography based on previous studies that demonstrated the sleep polymorphism test record to be critical for two consecutive days.[31] The test results of the first and second visits were similar in the participants of this study; thus, the first night effect did not appear (the results are not presented).

To supplement this, we propose making the laboratory environment similar to that of a home. In this study, the accuracies of light sleep, deep sleep, and REM sleep were 59.1%, 83.7%, and 82.3%, respectively. In a comparative study of Fitbit Charge 2 and PSG for shift workers, the accuracy of light sleep was 0.49, deep sleep was 0.78, and REM sleep was 0.86, which was similar to the results of previous studies.[32] In a systematic literature review of each Fitbit model, the accuracy of sleep parameters through EBE analysis was 0.69~0.81 for light sleep, 0.36~0.89 for deep sleep, and 0.62~0.89 for REM sleep.[13] Compared with the results of this study, the accuracy of light sleep was low, whereas deep sleep and REM showed high accuracy.

In a recent study, large-scale data collection became possible as the demand for Fitbit increased. Therefore, a rapid growth trend may occur.[33] Fitbit devices can measure not only sleep, but also heart rate and physical activity. The validity

of various Fitbit device models was evaluated through meta-analysis and discovered as an important factor of validity events in heart rate function and physical activity measurement function.[34] Sleep disorders or changes in sleep status can be monitored to predict health by validating the Fitbit devices. Considering this, this study confirmed the possibility of using FBI2 and FBI2 as screening tools for sleep stage measurement in daily life.

The advantages of this study are as follows: First, by comparing Fitbit and PSG for Koreans, the sleep characteristics of Koreans were reflected in the same environmental conditions. Therefore, the condition for increasing the reliability of the Fitbit was satisfied. Second, the recently standardized criteria to confirm the accuracy of the sleep tracking device through the EBE analysis method were applied using the main sleep parameters.[22] Third, there was a period of adaptation to the laboratory environment using the data measured twice with a one-week interval.

The limitations of this study are as follows: First, because this was an exploratory study with only nine participants, the sample size was small, which may have affected the accurate comparison of the FBI2 and PSG. Therefore, we analyzed the total number of participants measured per visit to compensate for this limitation. Second, the accuracy was relatively low because of the characteristics of the FBI2 device, the sleep measurement mode of which was not classified in detail. Third, the sleep parameters of the Fitbit were defined arbitrarily, unlike the interpretation of PSG polysomnographic technologists. Therefore, there were limitations to this interpretation.

Conclusion

This study compared sleep PSG and FBI2 with repeated measurements in the same sleep experimental environment in adults who felt uncomfortable with sleep. This study is meaningful in that it is the first exploratory study to verify the accuracy of the FBI2 and found the possibility of its use as an objective sleep indicator. By demonstrating nearly consistent results with previous studies upon comparison with Fitbit and PSG, the FBI2 can be used to monitor sleep health owing to its usability aspect, thereby supporting its use as an alternative to PSG.

Acknowledgments

The authors appreciate the Daejeon University Daejeon Korean Medical Hospital and Clinical Trial Center for their contribution towards data collection.

Funding

This work was supported by the Korea Institute of Oriental Medicine (Grant number: KSN1731121) and the Ministry of Trade, Industry and Energy (Grant number: NIN2013240).

Disclosure

The authors report no conflicts of interest in this work.

References

1. Cappuccio FP, Cooper D, D'Elia L, Strazzullo P, Miller MA. Sleep duration predicts cardiovascular outcomes: a systematic review and meta-analysis of prospective studies. *Eur Heart J*. 2011;32(12):1484–1492. doi:10.1093/eurheartj/ehr007
2. Xie J, Li Y, Zhang Y, et al. Sleep duration and metabolic syndrome: an updated systematic review and meta-analysis. *Sleep Med Rev*. 2021;59:101451. doi:10.1016/j.smrv.2021.101451
3. Shin D, Hur J, Cho KH, Cho EH. Trends of self-reported sleep duration in Korean adults: results from the Korea National Health and Nutrition Examination Survey 2007–2015. *Sleep Med*. 2018;52:103–106. doi:10.1016/j.sleep.2018.08.008
4. National Health Insurance Service. *Useful Health Life Statistics Information to Know [Internet]*. Wonju: National Health Insurance Sevice; 2021. Available from: https://www.nhis.or.kr/nhis/together/wbhaec06700m01.do?mode=view&articleNo=10811517. Accessed February 21, 2023.
5. Lee YJ, Kim DJ, Lee H. A study on sleep-wake assessment for substantiation of sleep products. *Sleep Med Psychophysiol*. 2020;27(2):51–55.
6. Depner CM, Cheng PC, Devine JK, et al. Wearable technologies for developing sleep and circadian biomarkers: a summary of workshop discussions. *Sleep*. 2020;43(2):zsz254. doi:10.1093/sleep/zsz254
7. Van de Water AT, 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. doi:10.1111/j.1365-2869.2009.00814.x
8. FDA. Device software functions including mobile medical applications U.S. food & drug administration website; 2022. Available from: https://www.fda.gov/medical-devices/digital-health-center-excellence/device-software-functions-including-mobile-medical-applications. Accessed September 28, 2022.

9. Fitbit. What should I know about sleep stages? [homepage on the Internet]. Available from: https://help.fitbit.com/articles/en_US/Help_article/2163. Accessed July 26, 2022.
10. Cook JD, Prairie ML, Plante DT. Utility of the Fitbit Flex to evaluate sleep in major depressive disorder: a comparison against polysomnography and wrist-worn actigraphy. *J Affect Disord*. 2017;217:299–305. doi:10.1016/j.jad.2017.04.030
11. Moreno-Pino F, Porras-Segovia A, López-Esteban P, Artés A, Baca-García E. Validation of fitbit charge 2 and fitbit alta HR against polysomnography for assessing sleep in adults with obstructive sleep apnea. *J Clin Sleep Med*. 2019;15(11):1645–1653. doi:10.5664/jcsm.8032
12. Scott H, Lack L, Lovato N. A systematic review of the accuracy of sleep wearable devices for estimating sleep onset. *Sleep Med Rev*. 2020;49:101227. doi:10.1016/j.smrv.2019.101227
13. Haghayegh S, Khoshnevis S, Smolensky MH, Diller KR, Castriotta RJ. Accuracy of wristband fitbit models in assessing sleep: systematic review and meta-analysis. *J Med Internet Res*. 2019;21(11):e16273. doi:10.2196/16273
14. Munos B, Baker PC, Bot BM, et al. Mobile health: the power of wearables, sensors, and apps to transform clinical trials. *Ann N Y Acad Sci*. 2016;1375(1):3–18. doi:10.1111/nyas.13117
15. Cook JD, Eftekari SC, Dallmann E, Sippy M, Plante DT. Ability of the Fitbit Alta HR to quantify and classify sleep in patients with suspected central disorders of hypersomnolence: a comparison against polysomnography. *J Sleep Res*. 2019;28(4):e12789. doi:10.1111/jsr.12789
16. Lunsford-Avery JR, Keller C, Kollins SH, Krystal AD, Jackson L, Engelhard MM. Feasibility and acceptability of wearable sleep electroencephalogram device use in adolescents: observational study. *JMIR Mhealth Uhealth*. 2020;8(10):e20590. doi:10.2196/20590
17. 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. doi:10.5665/sleep.4918
18. de Zambotti M, Baker FC, Willoughby AR, et al. Measures of sleep and cardiac functioning during sleep using a multi-sensory commercially-available wristband in adolescents. *Physiol Behav*. 2016;158:143–149. doi:10.1016/j.physbeh.2016.03.006
19. Bastien CH, Vallières A, Morin CM. Validation of the Insomnia Severity Index as an outcome measure for insomnia research. *Sleep Med*. 2001;2(4):297–307. doi:10.1016/S1389-9457(00)00065-4
20. Hori T, Sugita Y, Koga E, Shirakawa S, Inoue K. Proposed supplements and amendments to 'A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects', the Rechtschaffen & Kales (1968) standard. *Psychiatry Clin Neurosci*. 2001;55:305–310.
21. Sleep API variable: sleep [homepage on the Internet]. Available from: https://dev.fitbit.com/. Accessed July 26, 2022.
22. Menghini L, Cellini N, Goldstone A, Baker FC, de Zambotti M. A standardized framework for testing the performance of sleep-tracking technology: step-by-step guidelines and open-source code. *Sleep*. 2021;44(2):zsaa170. doi:10.1093/sleep/zsaa170
23. 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. doi:10.3758/s13428-013-0438-7
24. Peters B. Why WASO has a negative effect on sleep quality; 2022. Available from: https://www.verywellhealth.com/understanding-and-treating-waso-3015394. Accessed February 21, 2023.
25. Scott H, Lovato N, Lack L. The development and accuracy of the THIM wearable device for estimating sleep and wakefulness. *Nat Sci Sleep*. 2021;13:39. doi:10.2147/NSS.S287048
26. Kim E-J, Ahn Y-M, Shin H-B, Kim J-W. Detrended fluctuation analysis of sleep electroencephalogram between obstructive sleep apnea syndrome and normal children. *Sleep Med Psychophysiol*. 2010;17(1):41–49.
27. Kahawage P, Jumabhoy R, Hamill K, de Zambotti M, Drummond SPA. Validity, potential clinical utility, and comparison of consumer and research-grade activity trackers in Insomnia Disorder I: in-lab validation against polysomnography. *J Sleep Res*. 2020;29(1):e12931. doi:10.1111/jsr.12931
28. 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):646. doi:10.3390/s16050646
29. Chinoy ED, Cuellar JA, Huwa KE, et al. Performance of seven consumer sleep-tracking devices compared with polysomnography. *Sleep*. 2021;44(5):zsaa291. doi:10.1093/sleep/zsaa291
30. Le Bon O, Staner L, Hoffmann G, et al. The first-night effect may last more than one night. *J Psychiatric Res*. 2001;35(3):165–172. doi:10.1016/S0022-3956(01)00019-X
31. Newell J, Mairesse O, Verbanck P, Neu D. Is a one-night stay in the lab really enough to conclude? First-night effect and night-to-night variability in polysomnographic recordings among different clinical population samples. *Psychiatr Res*. 2012;200(2–3):795–801. doi:10.1016/j.psychres.2012.07.045
32. Stucky B, Clark I, Azza Y, et al. Validation of Fitbit Charge 2 sleep and heart rate estimates against polysomnographic measures in shift workers: naturalistic study. *J Med Internet Res*. 2021;23(10):e26476. doi:10.2196/26476
33. Kawasaki Y, Kasai T, Sakurama Y, et al. Evaluation of sleep parameters and sleep staging (slow wave sleep) in athletes by fitbit alta HR, a consumer sleep tracking device. *Nat Sci Sleep*. 2022;14:819–827. doi:10.2147/NSS.S351274
34. Leung W, Case L, Sung MC, Jung J. A meta-analysis of Fitbit devices: same company, different models, different validity evidence. *J Med Eng Technol*. 2022;46(2):102–115. doi:10.1080/03091902.2021.2006350