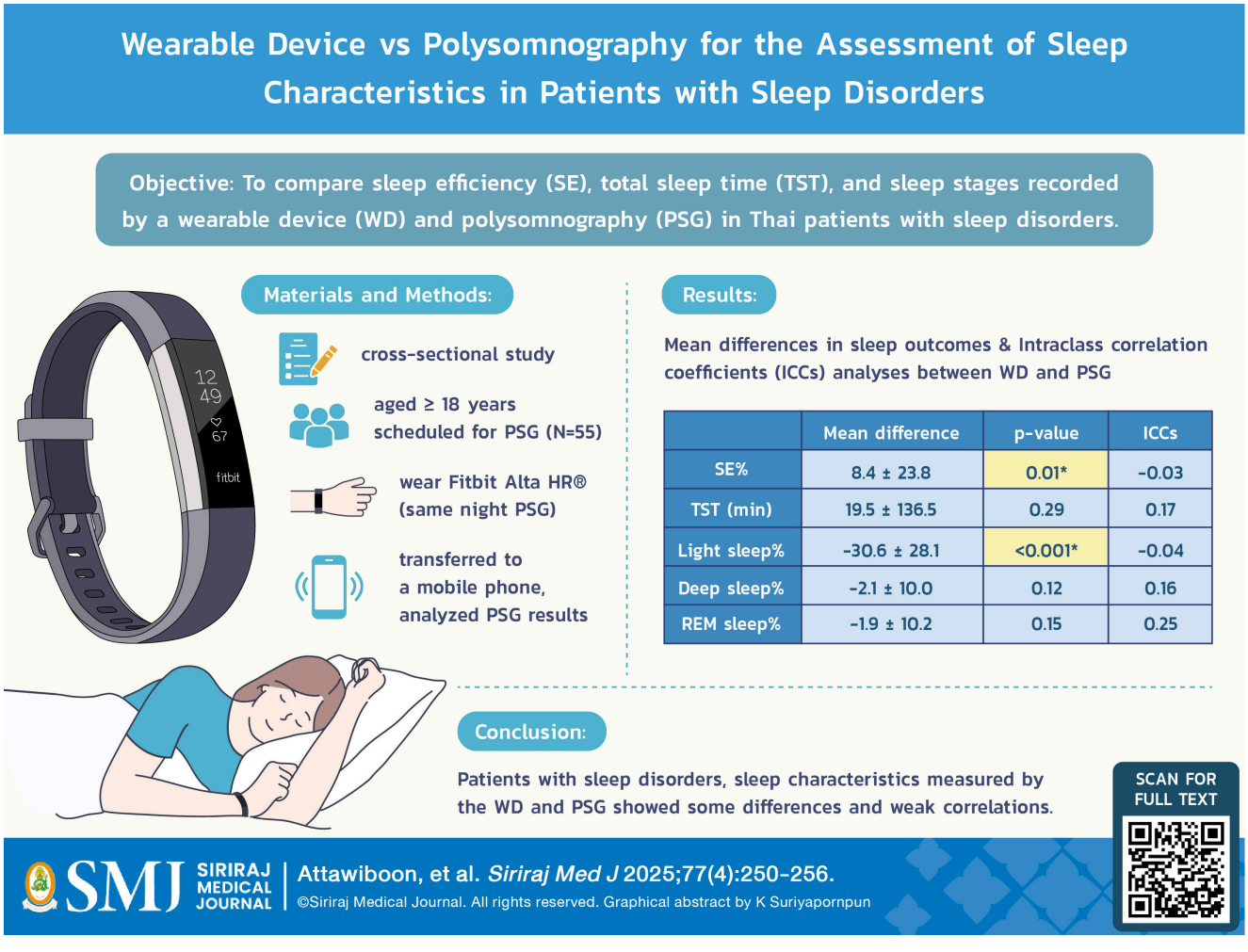
Wearable Device versus Polysomnography for the Assessment of Sleep Characteristics in Patients with Sleep Disorders

Kankanok Attawiboon, M.D.[1], Wish Banhiran, M.D.[1,*], Phawin Keskool, M.D.[1], Wattanachai Chotinaiwattarakul, M.D.[2], Sarin Rungmanee, M.D.[1]

[1]Department of Otorhinolaryngology, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, 10700 Thailand, [2]Department of Medicine, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, 10700, Thailand.



*Corresponding Author: Wish Banhiran
E-mail: wishbanh@gmail.com
Received 4 December 2024 Revised 23 January 2025 Accepted 25 January 2025
ORCID ID:http://orcid.org/0000-0002-4029-6657
https://doi.org/10.33192/smj.v77i4.272562

ABSTRACT

Objective: To compare sleep efficiency (SE), total sleep time (TST), and sleep stages recorded by a wearable device (WD) and polysomnography (PSG) in Thai patients with sleep disorders.

Materials and Methods: Patients aged ≥18 years scheduled for PSG were included in this cross-sectional study. All research subjects completed questionnaires and wore a WD (Fitbit Alta HR®) on the same night they underwent PSG study. The data from the WD were transferred to a mobile phone and analyzed independently of PSG results, which were scored by sleep technicians. Bland-Altman plots and intraclass correlation coefficients (ICCs) were used for the analyses.

Results: Data from 55 patients (33 males, 22 females) were analyzed, with four patients excluded due to data errors. The mean differences between WD and PSG for SE (%) and light sleep were 8.4±23.8 and 43.6±26.4, respectively, both statistically significantly ($p<0.05$). The ICCs for SE and light sleep were -0.03 and -0.04, indicating poor reliability. However, the mean differences for TST, deep sleep, and REM sleep between the two methods were not statistically significant ($p>0.05$), with ICC values of 0.17, 0.16, and 0.25, respectively, all considered poor correlations.

Conclusion: In patients with sleep disorders, sleep characteristics measured by the WD and PSG showed some differences and weak correlations. As technology advances, the accuracy of wearable devices may improve. Further studies are needed to evaluate different devices and populations.

Keywords: Wearable device; Fitbit Alta HR®; sleep disorder; Thai (Siriraj Med J 2025; 77: 250-256)

INTRODUCTION

Sleep is an essential part of life, and maintaining good sleep hygiene can help reduce the risk of serious physical and mental illnesses linked to several sleep disorders affecting people worldwide. While polysomnography (PSG) is currently considered the standard diagnostic method for several sleep disorders, it has some limitations, such as its high cost, labor intensity, and complexity, which makes it less accessible to a large portion of general population.[1-3] In recent years, electronic wearable devices —often in the form of wristbands or smartwatches— have been introduced to help individuals track and record their sleep patterns. These devices measure sleep activity using mechanisms like actigraphy, which syncs with mobile phones or personal computers.[3-6] Their convenience and compatibility with modern lifestyles make them widely accepted as inexpensive and accessible tools for improving quality of life.

By facilitating remote patient monitoring enabling assessment and management of patients' sleep health without necessitating in-person visits, wearable devices can extend the reach of healthcare providers, enabling assessment and management of patients' sleep health without necessitating in-person visits. This approach not only reduces the strain on sleep centers but also enhances access to care for individuals who might otherwise face obstacles due to geographical or socioeconomic factors. Furthermore, the continuous data collection afforded by wearables allows for longitudinal monitoring, offering insights into sleep patterns over time, which is advantageous for both diagnosis and ongoing management.

Previous studies have shown that wristband-like wearable devices may be used to evaluate sleep, yielding varied results in young, healthy adults and some children with sleep disorders.[7-13] Much of the literature suggests that these devices may serve as alternative tools for sleep evaluation, demonstrating high sensitivity, but low specificity for distinguishing different sleep–wake stages.[7] However, studies comparing the performance of wearable devices with standard PSG in patients with sleep disorders remain limited, and none have been conducted on Thai populations. Therefore, the purpose of this study was to evaluate the sleep characteristics measured by a wearable device, namely the Fitbit Alta HR®, a popular and inexpensive wristband, and compare them with PSG measurements in Thai patients with sleep disorders.

MATERIALS AND METHODS

Study design

This cross-sectional study was conducted on Thai patients with sleep disorders at the snoring clinic, Department of Otorhinolaryngology, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand, between June 2019 and June 2020. It was approved by the Siriraj Institutional Review Board (SiRB), COA no. Si 451/2019. Written consent was obtained from all participants before enrollment.

Subjects

The inclusion criteria for the study were patients aged ≥ 18 years who were being treated at the snoring clinic and scheduled for diagnostic PSG. The exclusion criteria included patients with severe or unstable medical problems, such as recent myocardial infarction, stroke, epilepsy, neuromuscular disorders, Parkinsonism, schizophrenia, and sleep-related movement disorders. The final analysis included data from 55 patients (33 males, 22 females), aged 25 to 78 years.

Interventions

All participants completed pre-treatment questionnaires and wore the designated wearable device while undergoing routine diagnostic PSG throughout the night. As one sleep technician scored the PSG results, the data simultaneously recorded by the wearable device were transferred to a mobile phone. The data were analyzed by a research assistant blinded to the PSG results and by an independent assessor not involved in the clinical trial.

Wearable device

The wearable device used in this study was the Fitbit Alta HR® (Fitbit Inc., San Francisco, CA, USA) designed to track various personal metrics such as heart rate, motion, and sleep activity (Fig 1). Recording began automatically once the device was worn and stopped upon removal. Examples of the extracted data are shown in Fig 2.

Polysomnography

Overnight technician-attended PSG (SOMNOmedics, SOMNO HD PSG, DOMINO 3.0.0.3; Randersacker, Germany) was routinely performed at the sleep center in Siriraj Hospital. The recording channels included electroencephalogram (EEG), electro-oculogram (EOG), electromyogram (EMG), electrocardiogram (ECG), nasal pressure transducer, airflow thermistor, respiratory effort measurement, body position sensor, pulse oximetry, and real-time video recordings. The recording began when patients indicated they were ready to sleep (lights off) and concluded when they woke up (lights on). All PSG parameters were manually scored by well-trained sleep technologists and reviewed by certified sleep specialists.

Sleep-related parameters

The outcomes of this study included several important sleep-related parameters defined as follows: Sleep efficiency (SE), expressed as a percentage and calculated by dividing the total sleep time (TST) by the total time in bed (TIB) or total recording time (TRT). TIB or TRT refers to the time-period from light-offs to lights-on (from the beginning to the end of the recording). Light sleep, deep sleep, and REM sleep are also expressed as percentages. Light sleep is calculated from the time spent in sleep stages N1 and N2, deep sleep from stage N3, and REM sleep from the time spent in REM, all divided by the TST. In general, good sleep quality in a young adult population is characterized by a TST of ≥7 hours, SE of 80%, light sleep constituting 50%–55%, deep sleep of 20%, and REM sleep making up 20%–25%.

Statistical analysis

Categorical data are presented as numbers and

Fig 1. Model of Fitbit Alta HR® used in this study.

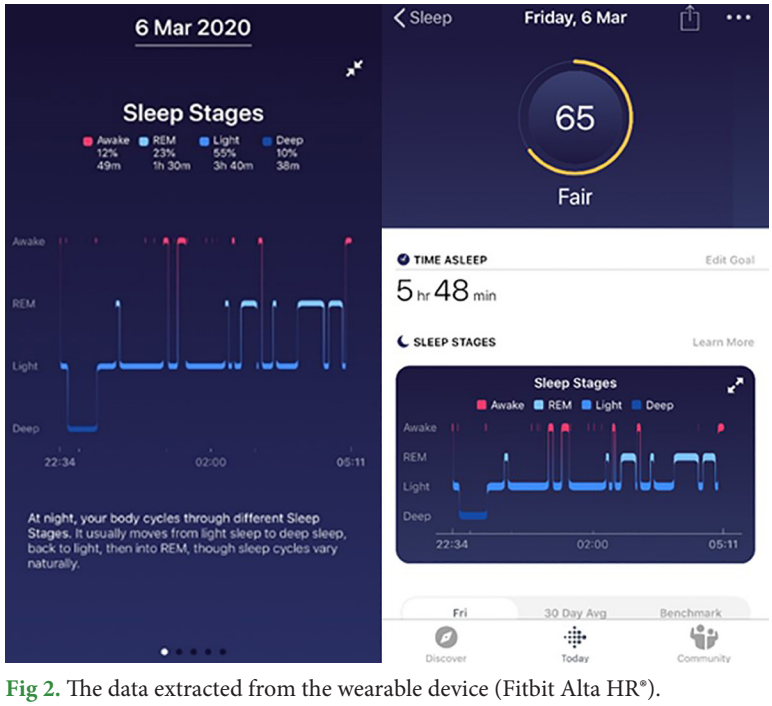


Fig 2. The data extracted from the wearable device (Fitbit Alta HR®).

percentages, while continuous data are expressed as means ± standard deviations (SD) or medians. Comparisons of SE, TST, and sleep stages between the Fitbit Alta HR® and PSG were analyzed using paired t-tests. Intraclass correlation coefficients (ICCs) with a two-way random-effects model with absolute agreement and single measures, and Bland-Altman plots to depict any systematic bias and identify outliers were used to evaluate the agreement between the two assessments. An ICC of <0.4 was considered indicative of poor reliability, while an ICC of 0.4–0.74 indicated a moderate level of reliability, and an ICC of ≥0.75 indicated excellent reliability.[14] All statistical analyses were performed using the Statistical Package for the Social Sciences (SPSS)® software, version 18.0, with significance set at $p < 0.05$ for two-tailed tests.

RESULTS

Initially, 59 patients were enrolled in the study, but 4 patients were excluded due to technical issues (data recording errors). Further details about the patients' characteristics are provided in Table 1. The primary outcomes, mean differences, and agreement (ICC) of SE between the wearable device and PSG are shown in Table 2. The secondary outcomes, including mean differences, and agreement (ICC) of TST, light sleep, deep sleep, and REM sleep are presented in Tables 2 and Table 3. The Bland–Altman plots comparing SE between the wearable device and PSG are displayed in Fig 3, showing the mean differences ± standard deviations (SD) and lower and upper limits of agreement as 8.4 ± 23.8, -38.2, and 55, respectively. It reveals a mean bias, indicating that Fitbit's SE measurements systematically

TABLE 1. Baseline characteristics of study participants.

Characteristic	Data
Age, year	50.4 ± 13.2
BMI, kg/m²	24.9 ± 4.1
AHI, events/h	32.8 ± 23.6
Non-OSA	5 (9.1)
Mild OSA	10 (18.2)
Moderate OSA	13 (23.6)
Severe OSA	27 (49.1)
Nighttime symptoms	
Sleep apnea	16 (28.1)
Snoring	49 (86)
Light sleeper	27 (47.4)
Chocking	18 (31.6)
Daytime symptoms	
Headache	28 (49.1)
Daytime sleepiness	39 (68.4)
Loss of productivity	30 (52.6)
Depression	7 (12.3)
Underlying disease	
Coronary artery disease	5 (8.5)
Diabetes	3 (5.1)
Dyslipidemia	22 (37.3)
Hypertension	18 (30.5)

Continuous data are presented as mean ± standard deviation; categorical data are presented as numbers (percentages); AHI = apnea-hypopnea index (events/h), Non-OSA = AHI <5, Mild OSA = AHI 5 to <15, Moderate OSA = AHI 15 to <30, and Severe OSA = AHI >30 events/h

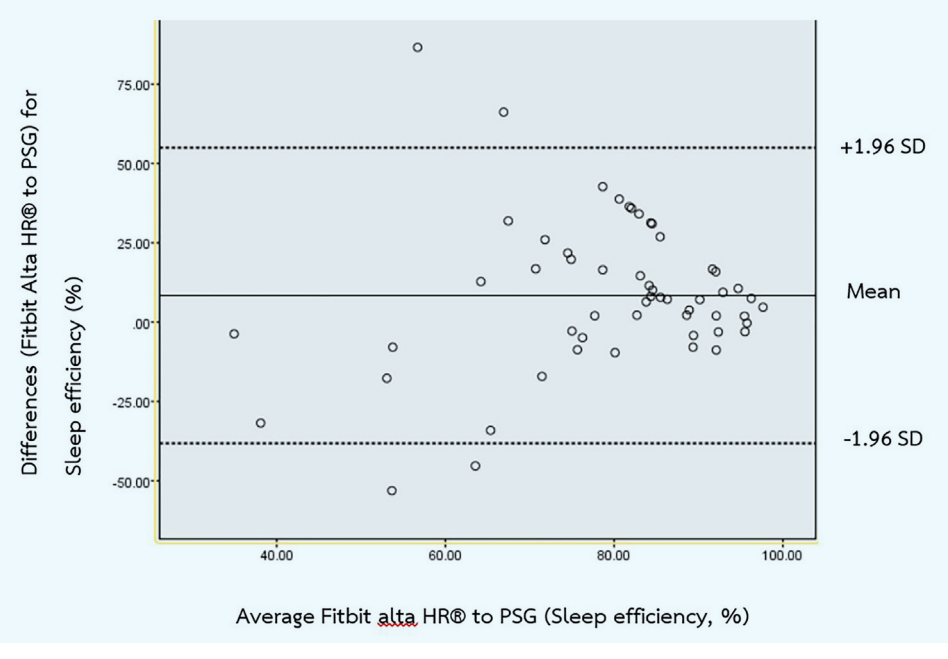


Fig 3. The Bland–Altman plots comparing sleep efficiency (SE) between the wearable device and polysomnography (PSG): Mean bias, the wide limits of agreement, the presence of outliers are shown.

TABLE 2. Mean differences in sleep outcomes between the wearable device (Fitbit Alta HR®) and polysomnography.

	PSG	Fitbit Alta HR®	Difference (Mean ± SD)	Lower limit	Upper limit	p-value
Sleep efficiency, %	74.9 ± 16.9	83.3 ± 19.9	8.4 ± 23.8	1.9	14.8	0.01*
Total sleep time, min	357.5 ± 82.1	377.1 ± 115.8	19.5 ± 136.5	-17.3	56.4	0.29
Light stage, %	74.3 ± 10.3	43.6 ± 26.4	-30.6 ± 28.1	-38.3	-23.0	<0.001*
Deep stage, %	10.2 ± 7.2	8.0 ± 6.3	-2.1 ± 10.0	-4.8	0.5	0.12
REM stage, %	10.2 ± 7.2	12.8 ± 9.0	-1.9 ± 10.2	-4.7	0.7	0.15

Abbreviations: PSG: Polysomnography, SD: Standard deviation.
**The p-value <0.05 indicates statistical significance.*

TABLE 3. Agreement of sleep outcomes between the wearable device (Fitbit Alta HR®) and polysomnography.

	ICC	95% CI
Sleep efficiency	-0.03	-0.22, 0.18
Total sleep time	0.17	-0.06, 0.39
Light stage	-0.04	-0.16, 0.10
Deep stage	0.16	-0.06, 0.38
REM stage	0.25	-0.06, 0.47

Abbreviations: ICC: intraclass correlation coefficient, CI: confidence interval

differ from PSG's. The wide limits of agreement suggest significant variability between the two methods, with discrepancies increasing at higher average SE levels, implying proportional bias. Additionally, the presence of outliers indicates instances where Fitbit's SE estimates substantially deviate from PSG measurements.

DISCUSSION

This is likely the first study to compare the performance of a wearable device (Fitbit Alta HR®) with PSG performed in adult Thai patients with sleep disorders, primarily obstructive sleep-disordered breathing (OSDB), addressing a knowledge gap in wearable device applicability. The study results revealed a statistically significant difference ($p < 0.05$) between the two devices in terms of mean sleep efficiency and light sleep, but not for mean TST, deep sleep, or REM sleep ($p > 0.05$). On average, the Fitbit Alta HR® overestimated SE and TST compared to PSG, with differences of 8.4 ± 23.8% and 19.5 ± 136.5 minutes, respectively. However, all sleep parameters measured by both methods showed poor agreement overall.

The findings of this study differ from those of *de Zambotti et al.*[10] who compared another wearable fitness-tracker device, the Jawbone UP®, with PSG in 18 healthy adults. Their study demonstrated good agreement between the wearable device and PSG for sleep estimation, including for TST and sleep onset latency, but less accuracy in detecting wake time after sleep onset (i.e., a poor ability to detect being awake). Similarly, *de Zambotti et al.*[8] conducted another study in 65 healthy adolescents using the same device and found good agreement with PSG. A systemic review by *Evenson et al.*,[7] reported a pilot study comparing the Fitbit Flex® with PSG in OSDB children, which showed high sensitivity for detecting sleep but low specificity for detecting wakefulness. The differences in results between our study and others are likely due to variations in subject age groups, health conditions, device models, and methods for measuring sleep parameters. In addition, factors such as dietary habits (high consumption of caffeinated beverages, particularly energy drinks), environmental conditions (co-sleeping), and cultural sleep patterns (varying sleep schedules) of Thai population might also influence findings.[16] However, these were not well recorded in our study.

Wearable devices like the Fitbit utilize proprietary algorithms to monitor sleep patterns, primarily through motion-based actigraphy and heart rate variability (HRV). These algorithms estimate sleep stages—light, deep, and REM sleep—by analyzing movement and physiological

signals, however, has certain limitations. It may inaccurately detect periods of wakefulness as sleep, leading to errors in tracking sleep duration and quality. Additionally, the device might struggle to accurately identify short naps or power naps, resulting in incomplete sleep data. Furthermore, heart rate can vary due to factors unrelated to sleep, such as stress or illness, which (may) potentially confoundsleep stage estimation.[12,15]

There are some potential limitations of this study. First, the results are based on a single model of a commercially available wearable device capable of detecting sleep characteristics in patients with sleep disorders. Therefore, the results may not apply to newer models or the latest technology from Fitbit or other smartwatch manufacturers, as devices are continuously evolving. Second, the study recorded sleep data from only a single night. Measuring over longer periods or across multiple nights might yield more practical outcomes. Third, the Bland–Altman plots are not true quantitative analyses. Bias and limits of agreement may be superimposed on its visual presentation of the data. Finally, the study did not include any healthy subjects as a control group. For future direction, we recommend that subsequent studies should focus on comparing multiple devices, integrating multi-sensor data, and employing machine learning models to enhance the accuracy of wearable devices. They should also include control groups and incorporate the latest technology to provide more comprehensive insights.

CONCLUSION

Wearable devices (Fitbit Alta HR®) offer a convenient means to monitor sleep patterns over extended periods in naturalistic settings. However, in patients with sleep disorders, it exhibited some discrepancies and weak correlations in sleep measurements compared to PSG. In real-world practice, integrating data from these devices can enhance such patients monitoring and engagement but should be interpreted with caution and in conjunction with comprehensive clinical evaluations. As technology continues to advance and future studies are coming, the outcomes of wearable devices for sleep monitoring may significantly improve.

Data Availability Statement

The data that support the findings of this study are available upon request from the corresponding author, [W.B.]. The data are not publicly available due to containing information that could compromise the privacy of research participants.

ACKNOWLEDGEMENTS

We would like to express our gratitude to Miss Julaporn Pooliam, our consulting statistician, and all the patients who took part in this study. We also extend our thanks to Miss Jeerapa Kerdnoppakhun and Miss Ngamrat Treerassapanich for their assistance with manuscript preparation. Additionally, we thank the staff of the Department of Otorhinolaryngology as well as the sleep technicians at Siriraj Sleep Center, Faculty of Medicine Siriraj Hospital, Mahidol University.

DECLARATION

Grants and Funding Information

This study was funded by a grant from the Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand (Grant Number R016231067).

Conflict of Interest

The authors declare they have no conflicts of interest. All authors have seen and approved this manuscript.

Registration Number of Clinical Trials

TCTR20250204008

Author Contributions

K.A., W.B.; Conceptualization, Methodology. W.B.; Original draft preparation, K.A. K.A., P.W., W.C. and S.R.; Data Curation. W.B.; Writing - Review & Editing. All authors have accepted responsibility for the entire content of this manuscript and have approved its submission.

Use of Artificial Intelligence

None

REFERENCES

1. Practice parameters for the use of portable recording in the assessment of obstructive sleep apnea. Standards of Practice Committee of the American Sleep Disorders Association. Sleep. 1994;17:372-7.
2. Lee SA, Paek JH, Han SH. Sleep hygiene and its association with daytime sleepiness, depressive symptoms, and quality of life in patients with mild obstructive sleep apnea. J Neurol Sci. 2015;359:445-9.
3. Mehra R, Moul DE, Strohl KP. In: Kryger M, Roth T, Dement WC, editors. Principles and Practice of Sleep Medicine. 6th ed. Elsevier, Philadelphia; 2017.p.167-73.
4. Marino M, Li Y, Rueschman MN, Winkelman JW, Ellenbogen JM, Solet JM, et al. Measuring Sleep: Accuracy, Sensitivity, and Specificity of Wrist Actigraphy Compared to Polysomnography. Sleep. 2013;36:1747-55.
5. Martin JL, Hakim AD. Wrist actigraphy. Chest. 2011;139:1514-27.
6. Morgenthaler T, Alessi C, Friedman L, Owens J, Kapur V,

Boehlecke B, et al. Standards of Practice Committee; 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;30:519-29.
7. Evenson KR, Goto MM, Furberg RD. Systematic review of the validity and reliability of consumer-wearable activity trackers. Int J Behav Nutr Phys Act. 2015;12:159.
8. de Zambotti M, Baker FC, Colrain IM. Validation of Sleep-Tracking Technology Compared with Polysomnography in Adolescents. Sleep. 2015;38:1461-8.
9. 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. 2018;28(4): e12789.
10. de Zambotti M, Claudatos S, Inkelis S, Colrain IM, Baker FC. Evaluation of a consumer fitness-tracking device to assess sleep in adults. Chronobiol Int. 2015;32:1024-8.
11. Cook JD, Prairie ML, Plante DT. Ability of the Multisensory Jawbone UP3 to Quantify and Classify Sleep in Patients With Suspected Central Disorders of Hypersomnolence: A Comparison Against Polysomnography and Actigraphy. J Clin Sleep Med. 2018;14:841-8.
12. Lee HA, Lee HJ, Moon JH, Lee T, Kim MG, In H, et al. Comparison of Wearable Activity Tracker with Actigraphy for Sleep Evaluation and Circadian Rest-Activity Rhythm Measurement in Healthy Young Adults. Psychiatry Investig. 2017;14:179-85.
13. Osterbauer B, Koempel AJ, Davidson Ward SL, Fisher LM, Don DM, Ver Halen J. A Comparison Study of The Fitbit Activity Monitor And PSG For Assessing Sleep Patterns And Movement In Children. J Otolaryngol Adv. 2016;1:24-35.
14. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research [published correction appears in J Chiropr Med. 2017;16:346.
15. de Zambotti M, Cellini N, Goldstone A, Colrain IM, Baker FC. Wearable sleep technology in clinical and research settings. Med Sci Sports Exerc [Internet]. 2019;51(7):1538-57.
16. Banhiran W. Characteristics and clinical presentations of patients at the Siriraj snoring clinic. Siriraj Med J [Internet]. 2020 [cited 2025 Jan 22];72(3):202-8.