ORIGINAL RESEARCH

Exploratory Validation of Sleep-Tracking Devices in Patients with Psychiatric Disorders

Masaya Ogasawara [1], Masahiro Takeshima [1], Shumpei Kosaka [2], Aya Imanishi [1], Yu Itoh [1], Dai Fujiwara [1], Kazuhisa Yoshizawa [1], Norio Ozaki [3], Kazuyuki Nakagome [4], Kazuo Mishima [1]

[1]Department of Neuropsychiatry, Akita University Graduate School of Medicine, Akita, Japan; [2]Department of Psychiatry, Akita Prefectural Center for Rehabilitation and Psychiatric Medicine, Daisen, Japan; [3]Department of Pathophysiology of Mental Disorders, Nagoya University Graduate School of Medicine, Nagoya, Japan; [4]Department of Psychiatry, National Center of Neurology and Psychiatry, Tokyo, Japan

Correspondence: Kazuo Mishima, Department of Neuropsychiatry, Akita University Graduate School of Medicine, 1-1-1 Hondo, Akita, 010-8543, Japan, Tel +81-18-884-6122, Fax +81-18-884-6445, Email clocksclub@gmail.com

Purpose: Sleep-tracking devices have performed well in recent studies that evaluated their use in healthy adults by comparing them with the gold standard sleep assessment technique, polysomnography (PSG). These devices have not been validated for use in patients with psychiatric disorders. Therefore, we tested the performance of three sleep-tracking devices against PSG in patients with psychiatric disorders.

Patients and methods: In total, 52 patients (32 women; 48.1 ± 17.2 years, mean ± SD; 18 patients diagnosed with schizophrenia, 19 with depressive disorder, 3 with bipolar disorder, and 12 with sleep disorder cases) were tested in a sleep laboratory with PSG, along with portable electroencephalography (EEG) device (Sleepgraph), actigraphy (MTN-220/221) and consumer sleep-tracking device (Fitbit Sense).

Results: Epoch-by-epoch sensitivity (for sleep) and specificity (for wake), respectively, were as follows: Sleepgraph (0.95, 0.76), Fitbit Sense (0.95, 0.45) and MTN-220/221 (0.93, 0.40). Portable EEG (Sleepgraph) had the best sleep stage-tracking performance. Sleep-wake summary metrics demonstrated lower performance on poor sleep (ice, shorter total sleep time, lower sleep efficiency, longer sleep latency, longer wake after sleep onset).

Conclusion: Devices demonstrated similar sleep-wake detecting performance as compared with previous studies that evaluated sleep in healthy adults. Consumer sleep device may exhibit poor sleep stage-tracking performance in patients with psychiatric disorders due to factors that affect the sleep determination algorithm, such as changes in autonomic nervous system activity. However, Sleepgraph, a portable EEG device, demonstrated higher performance in mental disorders than the Fitbit Sense and actigraphy.

Keywords: polysomnography, actigraphy, portable EEG, consumer sleep technologies, psychiatric patients

Introduction

Assessment of sleep problems is an important clinical component of psychiatric treatment, as sleep problems coexist with a high percentage of psychiatric disorders. A cross-sectional study reported insomnia symptoms in 50% of patients with schizophrenia spectrum disorders and generalized anxiety disorder and in more than 80% of patients with depressive disorder.[1–3] In addition, sleep disorders in psychiatric conditions are not limited to the acute phase of the disease but could also persist for longer periods of time. This could impact the patients' quality of life, leading to impaired cognitive and social functioning, and increasing relapse and recurrence. In the STAR*D trial, which was conducted in the United States, insomnia was the most common residual symptom among patients with major depression who responded to antidepressant therapy but did not achieve remission. It was observed in 94.6% of these patients.[4] A study evaluating quality of life measurements reported that patients with post-remission residual insomnia exhibited lower physical and mental daily role functioning compared to those without insomnia.[5] An association has also been reported between sleep disturbances and the risk of recurrence of depressive disorders, bipolar disorders, and schizophrenia, suggesting that maintaining good sleep status is important for good clinical outcomes.[6–12] Alternatively, cases of sleep state

misrecognition have also been described, in which subjective insomnia symptoms persist despite objective sleep maintenance.[13] Insomnia frequently persists after the psychiatric disease that triggered the sleep problems has resolved, hence becoming the main symptom and primary theme in these patients' lives. Prolonged experiences of insomnia could lead to escalated anxiety. Thus, it is necessary to understand and evaluate both subjective and objective symptoms when assessing sleep in patients with psychiatric disorders.

In particular, it is essential to observe the relationship between psychiatric symptoms and sleep architecture in studies of the pathophysiology and clinical outcomes of mental disorders. For example, previous polysomnography (PSG) studies have reported rapid eye movement (REM) sleep abnormalities and alterations in slow wave sleep in schizophrenia patients, which are associated with psychiatric symptoms.[14] It has also been reported that PSG REM sleep abnormalities are found in patients with depressive disorder and change after drug treatment.[15] Continuous measurement of these sleep abnormalities by wearable devices is valuable, but the validity of the measurements needs to be verified.

The recent development of wearable sleep-tracking devices has made it possible to measure sleep parameters and sleep stages easily at home. Various types of devices that measure sleep, including wristband-type, ring-type, and under-mattress devices, are available. Portable electroencephalograph (EEG) devices that can be used to measure sleep at home are also being further developed. The American Academy of Sleep Medicine (AASM) stated that validation of device accuracy is needed for these devices to be considered part of medical evaluation and treatment.[16]

PSG is the main reference used to validate the performance of sleep-tracking devices. Several studies have demonstrated its use in healthy subjects. The validity of sleep-wake determination with sleep-tracking devices has been reported to be higher than that with actigraphy, which has been used to measure sleep at home.[17,18] On the contrary, few studies have focused on the performance of sleep-tracking devices in patients with psychiatric illnesses, and this remains to be elucidated. Patients with psychiatric disorders have different sleep architecture compared to healthy subjects.[19,20] Since consumer sleep technology, such as Fitbit, evaluates sleep based on machine learning using healthy subjects as training data, the performance of these devices may be reduced when patients with psychiatric disorders use them. We therefore proposed an exploratory evaluation of the sleep-tracking performance of three wearable devices against PSG in patients with psychiatric disorders and aimed to compare performance across the devices. We tested a portable EEG and two devices relying primarily on an accelerometer and hypothesized that the portable EEG would show better agreement than the other devices.

Methods

The flow diagram of the study is shown in Figure 1.

Participants

Fifty-two patients (20 males, 32 females) aged 16 years or older (mean age 48.1 ± 17.2 years, range 16–83 years) diagnosed with schizophrenia, depressive disorder, bipolar disorder, or sleep disorder who were admitted to Akita University Hospital for a sleep study between December 2020 to November 2021 were included in the study. The admissions were specifically for sleep tests and were of short duration. The diagnosis of psychiatric disorders was established by clinically experienced psychiatrists based on the Diagnostic and Statistical Manual of Mental Disorders-5 (DSM-5). Schizophrenia was diagnosed in 18 patients, depression in 19 patients, bipolar disorder in 3 patients, and sleep disorder in 12 patients. Exclusion criteria were as follows: patients under 16 years of age, and patients whose participation in the study was deemed inappropriate by the treating physician (eg, due to anticipated worsening of medical condition). Since sleep apnea is a common comorbidity in psychiatric patients, we did not exclude these patients.[21,22] Data regarding body mass index (BMI) and the medications taken by the patient on the day of recording were collected. Medications were compared and classified: chlorpromazine equivalents for antipsychotics, imipramine equivalents for antidepressants, and diazepam equivalents for sleeping pills and anti-anxiety medications.[23] All patients were considered fit to provide informed consent for participation in research by their treating physician. Informed consent was obtained prior to participation in the study, and all patients received compensation. Parental or legal guardian consent was obtained for all study participants under the age of 18 years. The Ethics Commission of the National Center of Neurology and Psychiatry, set up as the Central Ethics Review Board, approved all the study protocols and experimental

302

DovePress

Nature and Science of Sleep 2023:15
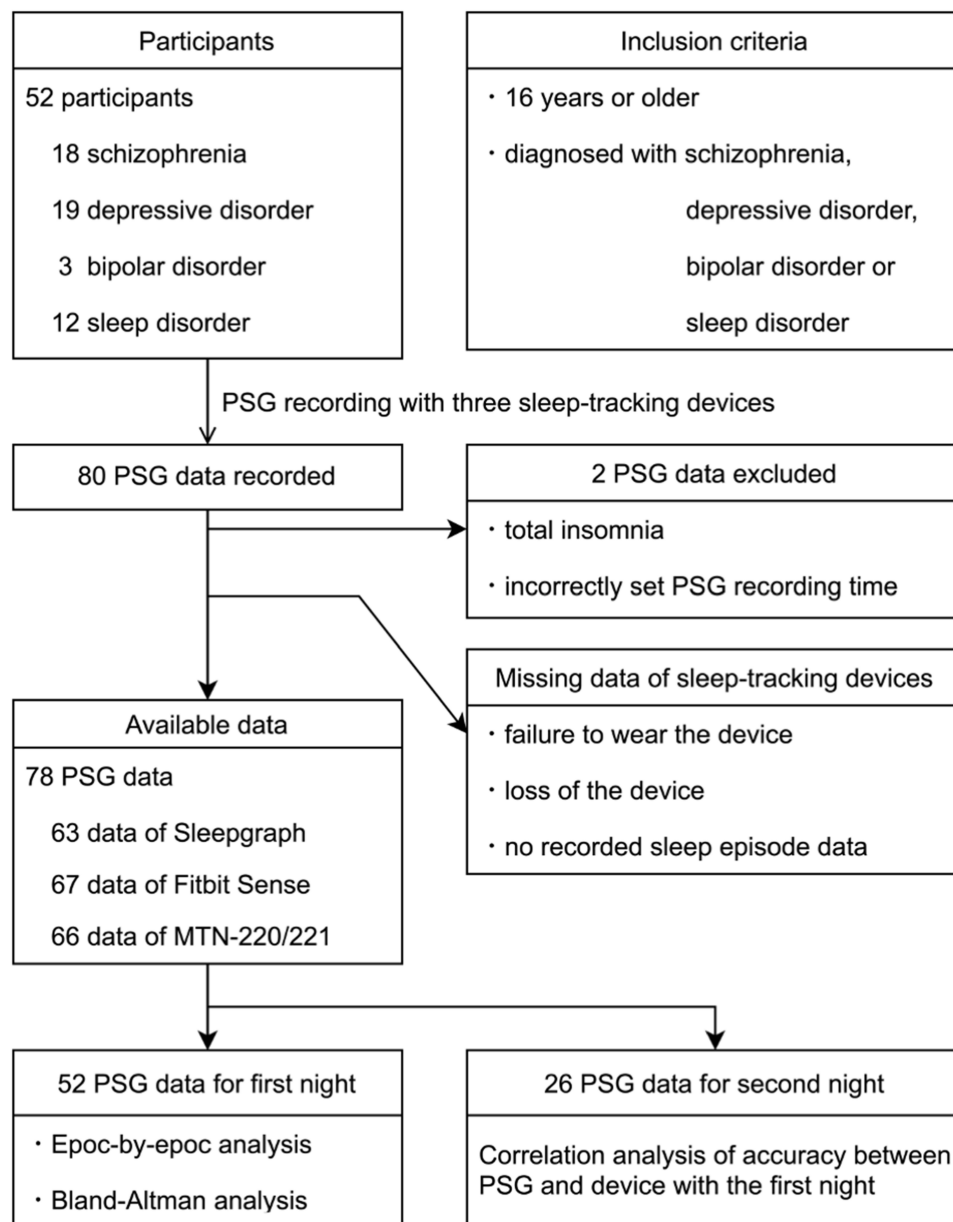
Figure 1 Flow diagram of the study.
Notes: Fifty-two patients aged 16 years or older, diagnosed with schizophrenia, depressive disorder, bipolar disorder, or sleep disorder, who were admitted to Akita University Hospital for a sleep study between December 2020 to November 2021 were included in the study. Three sleep-tracking devices were tested with PSG as a reference: Sleepgraph, Fitbit Sense, and MTN-220/221. PSG recordings were performed for two nights in 26 patients who gave consent. When two nights of data were available, the first night was used in the epoch-by-epoch analysis and Bland-Altman analysis, and the association of accuracy between the two nights was analyzed using Spearman correlation coefficient.

procedures (approval no. B2019-001). The ethics commission of the Akita University gave permission to conduct the study. The study was conducted in accordance with the Declaration of Helsinki.

Sleep-Tracking Devices

Three sleep-tracking devices were tested: Sleepgraph (Proassist, Ltd.; Osaka-shi, Osaka, Japan),[24] Fitbit Sense (Fitbit, Inc.; San Francisco, CA, USA),[25] and MTN-220/221 (Kissei Comtec Co., Ltd.; Matsumoto-shi, Nagano, Japan).[26] The basic characteristics of each device have been demonstrated in previous studies. The Sleepgraph is a portable EEG device, which requires the attachment of electrodes at four locations: right frontal region, below the left eye, right

mandible, and left mastoid process. EEG data were recorded at two channels: right frontal (Fp2)-mastoid (M1) and left ocular (EOG)-right mandible (EMG). Each sleep EEG was scored by a technologist skilled in the evaluation of minority-channel sleep EEG. The same technologist also scored the PSG. The sleep technologist was blinded to the study design. Sleep stages by Sleepgraph were used to score sleep at four stages, in accordance with the AASM criteria for PSG scoring.[27] However, N1 and N2 were integrated as light sleep to compare scoring with the data obtained from Fitbit Sense that scores at three stages: light sleep, deep sleep and REM sleep. Fitbit Sense is a device worn on the wrist that uses machine learning to detect sleep using acceleration sensors and heart rate estimations obtained by photoplethysmography (PPG).[25] It was worn on the non-dominant arm and measured simultaneously with PSG. MTN-220/221 is an accelerometer worn on the waist. It determines sleep/wake status based on the activity intensity calculated from the activity value.[26] Epoch duration for sleep data output by Sleepgraph and Fitbit Sense was 30 seconds. However, MTN-220/221 only produced output 2 minutes of sleep data per epoch.

PSG Recording and Scoring

PSG recordings were acquired using a digital PSG recorder (Somno HD; Fukuda Denshi, Bunkyo-ku, Tokyo, Japan or Polymate Pro; Miyuki giken, Bunkyo-ku, Tokyo, Japan) in a hospital room. PSG sleep stages (N1, N2, N3, and REM) and wake status were manually scored by a polysomnographic technologist certified by the Japanese Society of Sleep Research. They were scored at another facility in accordance with AASM manual.[27] Light-off time and light-on time were determined based on position sensors. PSG recordings were performed for one night on all patients, and for two consecutive nights on 26 patients who gave consent.

Recordings were started at 0 or 30 seconds for comparison with other sleep-tracking devices. PSG data were scored in 30-s epochs and were compared epoch-by-epoch (EBE) with sleep data obtained from Sleepgraph and Fitbit Sense devices. To align with the MTN-220/221 sleep data, the PSG data were additionally scored in 2-minutes epochs according to the following rule: If there were two or more occurrences of stage wake in four continuous epochs, this was defined as wake.

Statistical Analysis

For demographic data and data following a normal distribution, multiple comparisons were performed by one-way analysis of variance. The Scheffe method was used for between-group comparisons. The Steel-Dwass method was used for data not following a normal distribution.

Analysis of the sleep data was performed using the open source R code by Menghini et al (https://github.com/SRI-human-sleep/sleep-trackers-performance).[28] The following EBE agreement statistics were calculated between PSG and each individual device for every night; sensitivity (the percentage of epochs scored as sleep by PSG that were also detected as sleep by the device), specificity (the proportion of epochs scored as wake by PSG that were also detected as wake by the device), positive predictive value (PPV; the proportion of epochs detected as sleep by the device that were also scored as sleep by PSG), negative predictive value (NPV; the proportion of epochs detected as wake by the device that were also scored as wake by PSG), accuracy (proportion of all PSG epochs correctly detected by the device), and bias-adjusted kappa (PABAK; Cohen's kappa weighted to account for the amount of inequality between the number of sleep and wake epochs). For the Sleepgraph and Fitbit Sense devices, an error matrix was created that exhibited sensitivity and misclassification for each sleep stage. The mean, standard deviation, and 95% confidence interval for each parameter were calculated. The following sleep parameters were calculated for each device; total sleep time (TST; total time spent in all sleep stages), sleep efficiency (SE; percentage of TST divided by time in bed), sleep latency (SL; time taken to fall asleep from bedtime), and wake after sleep onset (WASO; total time awake after sleep onset), and REM latency (REML; time from sleep onset to first determination of REM). The light off/light on time for each device was set at the same time as the PSG. Sleep parameters were subjected to Bland-Altman analysis to estimate systematic and chance errors between PSG and each device. Bland-Altman plots were created to visualize discrepancies between the PSG and sleep measurement devices. For patients whose sleep data were recorded over two nights, the association of accuracy between the two nights was analyzed using Spearman correlation coefficient. All analyses were conducted by R version 4.2.0 (R Foundation, Vienna, Austria, 2022). Significance level was set at p<0.05.

Results

Fifty-two patients participated in the study, with a combined total of 80 nights of recorded sleep data. After excluding total insomnia (1 sleep data) and incorrectly set PSG recording time (1 sleep data), PSG data for 78 nights in 52 patients was included in the analysis. Data loss occurred due to reasons such as failure to wear the device or loss of the device: Sleepgraph (15, 19.2%), Fitbit Sense (11, 14.1%), and MTN-220/221 (12, 15.4%). PSG recording was performed for two nights in 26 patients who gave consent. When two nights of data were available, the first night was used in the analysis. The disease distribution and characteristics of the participants are demonstrated in Table 1. The patients included in the depression group were significantly older than the patients included in the schizophrenia and sleep disorder groups. Individual clinical information and medications at the time of study participation are listed in Supplementary Table S1.

Epoch-by-Epoch Analysis

Table 2 and Table 3 demonstrate the results of the EBE analysis. In the sleep-wake assessment, the sensitivity was high for all devices, and the specificity of Sleepgraph was higher than that of Fitbit Sense and MTN-220/221. In the EBE analysis of sleep stages, Sleepgraph's REM sleep demonstrated high sensitivity and specificity, while the sensitivity of deep sleep and light sleep was relatively low. Fitbit Sense exhibited lower sensitivity than Sleepgraph for all sleep stages. The specificity of Fitbit Sense was high for deep sleep and REM sleep, but low for light sleep. Fitbit Sense over-judged light sleep and demonstrated decreasing sensitivity for other sleep stages. The error matrix is displayed in Table 4. Sleepgraph incorrectly classified stage of wake and deep sleep as light sleep, and Fitbit Sense also incorrectly classified stage of deep sleep and REM sleep as light sleep.

Sleep Parameters

Bland-Altman plots of sleep parameters are demonstrated in Figure 2A–E. Detailed results of Bland-Altman analysis are displayed in Supplementary Tables S2 and S3. The TST measured by Sleepgraph exhibited a low mean difference from

Table 1 Participants' Demographic Data

	All Patients	**Schizophrenia**	**Depressive Disorder**	**Bipolar Disorders**	**Sleep Disorders**
Number of subjects	52	18	19	3	12
Age	48.1 ± 17.2	41.8 ± 12.4 *	59.3 ± 16.4 *†	37.7 ± 13.5	42.4 ± 18.6†
Sex					
Male	20 (38.5%)	6 (33.3%)	8 (42.1%)	0 (0%)	6 (50.0%)
Female	32 (61.5%)	12 (66.7%)	11 (57.9%)	3 (100%)	6 (50.0%)
BMI (kg/m^2)	22.6 (13.7–42.2)	22.3 (16.8–37.9)	23.9 ± 4.7	31.5 ± 9.3	24.0 ± 6.6
Drugs					
Antipsychotics (mg, CPZE)	87.5 (0–880)	491.8 ± 240.3*,**	0 (0–200)*	367.2 ± 239.8†	0 (0–100)**†
Antidepressants (mg, IMPE)	25 (0–350)	0 (0–350)	112.5 (0–300)	0 (0–0)	0 (0–200)
Hypnotics (mg, DZPE)	6.45 (0–50)	11.3 ± 8.5	7.5 (0–28.5)	8.3 ± 7.6	0 (0–50)
Sleep parameters					
TST (min)	453.8 ± 86.7	455.4 ± 85.1	453.8 ± 84.5	436.8 ± 104.0	435.5 ± 98.4
SE (%)	83.6 (55.4–97.3)	85.4 (56.3–97.3)	87.2 (55.4–96.9)	75.9 ± 14.5	77.4 ± 12.2
SL (min)	12 (0.5–185.5)	17.5 (3–168)	10.5 (0.5–75.5)†	104.3 ± 76.7†	8 (1.5–57.5)
WASO (min)	60 (2.5–241)	40.3 (2.5–219)	60.5 (9.5–241)	26.5 (26.5–32)	105.5 ± 63.9
AHI (events/h)	2.5 (0–96.1)	1.9 (0–72.0)	5.1 (0.4–62.4)	3.9 ± 4.1	2.2 (0–96.1)
Stage N1 (%TST)	9.5 (2.3–72.2)	9.1 (2.3–72.2)	10.1 ± 5.4	4.7 ± 2.0	12.0 (3.6–36.4)
Stage N2 (%TST)	50.0 (24.7–74.8)	54.6 ± 9.9	43.0 ± 12.5	55.4 ± 2.8	37.1 ± 10.3
Stage N3 (%TST)	16.4 ± 8.0	16.6 ± 7.4	17.9 ± 8.2	24.2 ± 5.7	11.7 ± 6.9
Stage REM (%TST)	14.1 ± 7.4	15.2 ± 7.4	12.5 ± 6.8	15.8 ± 5.9	14.6 ± 9.1
REM latency (min)	153.5 (31.5–455.5)	182.2 ± 96.6	204.9 ± 95.7	145.5 ± 36.3	66.5 (53.0–317.0)

Notes: Values are presented as mean ± standard deviation, median (min–max) or number (%). *, **p<0.01 †p<0.05.
Abbreviations: BMI, body mass index; CPZE, chlorpromazine equivalents; IMPE, imipramine equivalents; DZPE, diazepam equivalents; TST, total sleep time; SE, sleep efficiency; SL, sleep latency; WASO, wake after sleep onset; AHI, apnea hypopnea index; REM, rapid eye movement.

Table 2 Epoch-by-Epoch Analysis for Sleep/Wake Detection

Device	Sensitivity	Specificity	PPV	NPV	Accuracy	PABAK
Sleepgraph	0.95 (0.07)	0.76 (0.17)	0.95 (0.05)	0.78 (0.17)	0.92 (0.07)	0.85 (0.14)
	[0.93, 0.97]	[0.71, 0.81]	[0.93, 0.97]	[0.73, 0.83]	[0.9, 0.94]	[0.81, 0.89]
Fitbit Sense	0.95 (0.04)	0.45 (0.23)	0.88 (0.12)	0.58 (0.29)	0.85 (0.09)	0.70 (0.18)
	[0.94, 0.96]	[0.38, 0.51]	[0.84, 0.91]	[0.49, 0.66]	[0.82, 0.88]	[0.64, 0.75]
MTN-220/221	0.93 (0.09)	0.40 (0.24)	0.87 (0.10)	0.63 (0.25)	0.84 (0.09)	0.67 (0.18)
	[0.90, 0.96]	[0.33, 0.47]	[0.84, 0.90]	[0.55, 0.70]	[0.81, 0.86]	[0.62, 0.72]

Notes: Sensitivity is the proportion of "true" sleep epochs that are correctly classified as sleep by the device. Specificity is the proportion of 'true' wake epochs that are correctly classified as wake by the device. Accuracy is the proportion of correctly classified sleep and wake epochs over the total number of epochs. PPV is the proportion of epochs classified as sleep by the device that are 'true' sleep epochs. NPV is the proportion of epochs classified as wake by the device that are 'true' wake epochs. PABAK is Cohen's kappa weighted to account for the amount of inequality between the number of sleep and wake epochs. Values are presented as mean (standard deviation) [95% confidence interval].
Abbreviations: PPV, positive predictive value; NPV, negative predictive value; PABAK, prevalence and bias-adjusted kappa.

Table 3 Epoch-by-Epoch Analysis for Sleep Stages Detection

Sleep Stage	Device	Sensitivity	Specificity	PPV	NPV	Accuracy	PABAK
Light sleep	Sleepgraph	0.79 (0.11)	0.82 (0.10)	0.83 (0.08)	0.77 (0.14)	0.80 (0.08)	0.61 (0.17)
		[0.75, 0.82]	[0.79, 0.85]	[0.81, 0.86]	[0.73, 0.81]	[0.78, 0.83]	[0.56, 0.65]
	Fitbit Sense	0.68 (0.12)	0.55 (0.17)	0.64 (0.15)	0.59 (0.15)	0.61 (0.11)	0.23 (0.22)
		[0.64, 0.71]	[0.51, 0.60]	[0.60, 0.68]	[0.55, 0.63]	[0.58, 0.65]	[0.17, 0.30]
Deep sleep	Sleepgraph	0.69 (0.18)	0.97 (0.05)	0.79 (0.22)	0.95 (0.04)	0.93 (0.05)	0.85 (0.09)
		[0.64, 0.74]	[0.95, 0.98]	[0.73, 0.86]	[0.94, 0.96]	[0.91, 0.94]	[0.83, 0.88]
	Fitbit Sense	0.44 (0.31)	0.92 (0.05)	0.45 (0.28)	0.90 (0.09)	0.84 (0.10)	0.68 (0.19)
		[0.35, 0.54]	[0.91, 0.94]	[0.37, 0.53]	[0.87, 0.92]	[0.82, 0.87]	[0.63, 0.74]
REM sleep	Sleepgraph	0.91 (0.12)	0.93 (0.07)	0.66 (0.25)	0.99 (0.03)	0.92 (0.07)	0.85 (0.13)
		[0.88, 0.95]	[0.91, 0.95]	[0.59, 0.73]	[0.98, 0.99]	[0.91, 0.94]	[0.82, 0.89]
	Fitbit Sense	0.49 (0.30)	0.86 (0.07)	0.33 (0.24)	0.92 (0.06)	0.82 (0.09)	0.64 (0.17)
		[0.40, 0.58]	[0.84, 0.88]	[0.26, 0.40]	[0.91, 0.94]	[0.80, 0.85]	[0.59, 0.69]

Notes: Light sleep is the equivalent of PSG N1 + N2 sleep. Deep sleep is the equivalent of PSG N3 sleep. Sensitivity is the proportion of epochs classified as a given sleep stage by the reference that are correctly classified as that stage by the device. Specificity is the proportion of epochs not classified as a given stage by the reference that are correctly not classified as that stage by the device. PPV is the proportion of epochs classified as a target stage by the device that are classified as that stage by the reference. NPV is the proportion of epochs not classified as a target stage by the device that are not classified as that stage by the reference. Accuracy is the proportion of correctly classified epochs for a given stage over the total number of epochs. Values are presented as mean (standard deviation) [95% confidence interval].
**Abbreviations**: PSG, polysomnography; REM, rapid eye movement; PPV, positive predictive value; NPV, negative predictive value; PABAK, prevalence and bias-adjusted kappa.

Table 4 Error Matrix of Epoch-by-Epoch Analysis

| Sleep Stage	Device	Wake (PSG)	Light Sleep (PSG)	Deep Sleep (PSG)	REM Sleep (PSG)
Wake (device)	Sleepgraph	0.76 (0.18) [0.71, 0.81]	0.06 (0.08) [0.04, 0.08]	0 (0.01) [0, 0.01]	0.02 (0.03) [0.01, 0.03]
	Fitbit Sense	0.43 (0.22) [0.37, 0.50]	0.06 (0.05) [0.05, 0.07]	0.03 (0.05) [0.01, 0.04]	0.06 (0.12) [0.02, 0.09]
Light sleep (device)	Sleepgraph	0.17 (0.15) [0.12, 0.21]	0.79 (0.11) [0.75, 0.82]	0.30 (0.17) [0.25, 0.35]	0.07 (0.09) [0.04, 0.09]
	Fitbit Sense	0.37 (0.21) [0.31, 0.43]	0.67 (0.12) [0.64, 0.71]	0.48 (0.25) [0.41, 0.56]	0.40 (0.26) [0.32, 0.48]
Deep sleep (device)	Sleepgraph	0.01 (0.02) [0, 0.01]	0.05 (0.07) [0.03, 0.07]	0.69 (0.18) [0.63, 0.74]	0 (0) [0, 0]
	Fitbit Sense	0.05 (0.10) [0.02, 0.08]	0.10 (0.06) [0.09, 0.12]	0.43 (0.30) [0.34, 0.52]	0.04 (0.11) [0, 0.07]
REM sleep (device)	Sleepgraph	0.07 (0.08) [0.04, 0.09]	0.10 (0.09) [0.07, 0.12]	0.01 (0.03) [0, 0.01]	0.91 (0.12) [0.88, 0.95]
	Fitbit Sense	0.14 (0.13) [0.10, 0.18]	0.16 (0.10) [0.13, 0.19]	0.05 (0.09) [0.02, 0.08]	0.49 (0.30) [0.41, 0.58]

Notes: Proportions for EBE agreement metrics are demonstrated for each sleep stage (versus the combination of all other possible classifications) for the devices compared with the corresponding epochs from the reference PSG. Values are presented as mean (SD) [95% confidence interval]. Higher values (closer to 1.0) indicate better performance on that metric. Results are displayed for all devices that output sleep stage classifications.
Abbreviations: PSG, polysomnography; REM, rapid eye movement.
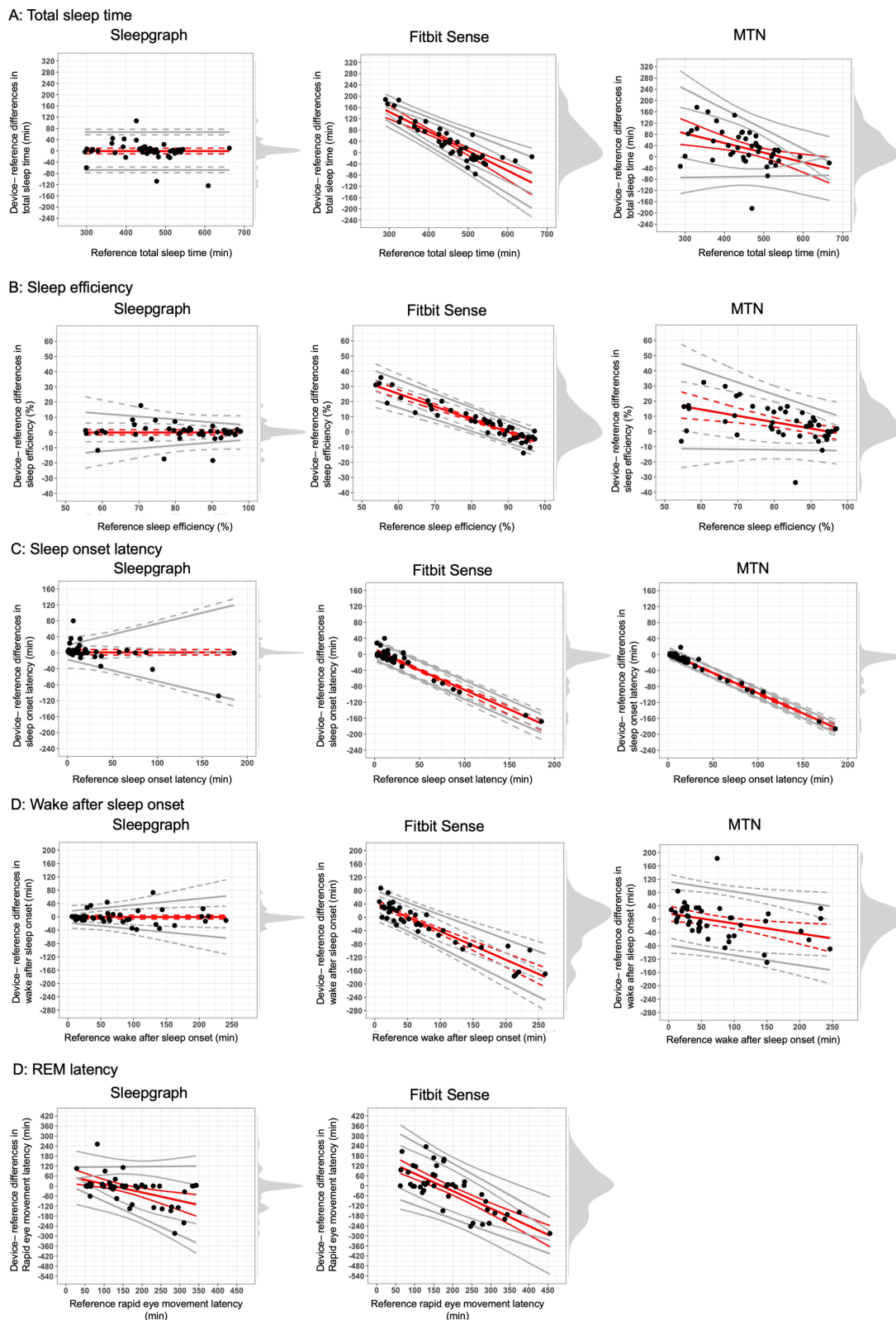
Figure 2 Bland-Altman plots.
Notes: Plots depict the mean bias (solid red line) and upper and lower limits of agreement (solid gray lines) for deviation in sleep parameters for the devices compared with the reference PSG. Black circles depict individual nights. Dashed lines represent the 95% confidence intervals around the bias and limits of agreement lines. Gray shaded regions on the right y-axis are density plots demonstrate the distribution of individual night biases. Zero on the y-axis represents no difference, with positive and negative y-axis values indicating an overestimation or underestimation, respectively, compared with the reference. Diagonal mean bias lines indicate significant proportional bias. Non-parallel limits of agreement lines indicate significant heteroscedasticity.
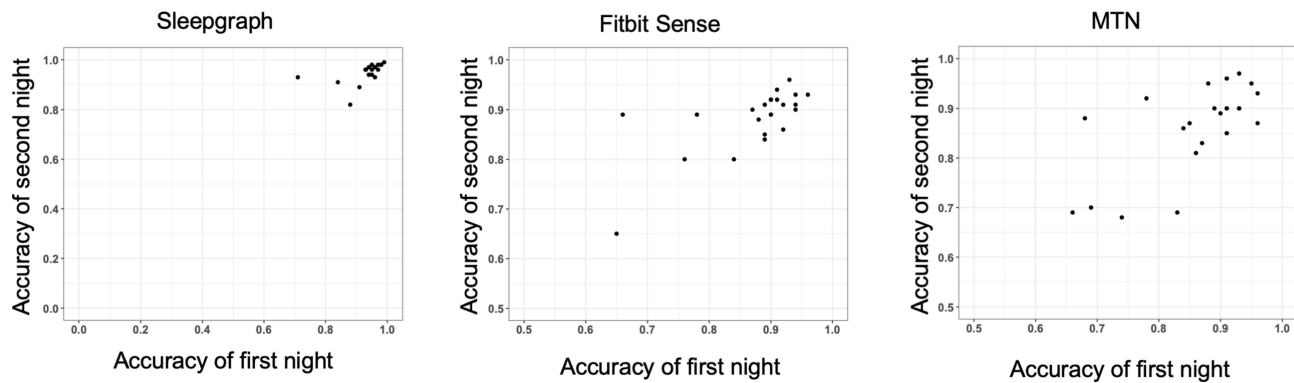
Figure 3 Plots of accuracy within the same patient measured over 2 nights.
Notes: The x-axis indicates the first night's accuracy, and the y-axis indicates the second night's accuracy. Higher values for accuracy indicate higher agreement of sleep/wake detection.

PSG. Alternatively, Fitbit Sense and MTN-220/221 demonstrated a proportional bias, overestimating TST on nights when PSG TST was low. SE, SL and WASO demonstrated a proportional bias for Fitbit Sense and MTN-220/221, overestimating SE and underestimating SL and WASO in cases of low SE and long SL and WASO as determined by PSG, respectively. REML exhibited a proportional bias with Sleepgraph and Fitbit Sense, underestimating REML in cases of long PSG REML.

Intra-Subject Comparison

A positive correlation for all devices was observed when the accuracy of PSG and each device was compared for the same patient, with measurements obtained over 2 nights (Figure 3, Sleepgraph: r = 0.78, p<0.001; Fitbit Sense: r = 0.70 p<0.001; MTN-220/221: r = 0.68, p<0.01).

Discussion

In the present study, we validated the performance of sleep tracking devices by using PSG as a reference. This allowed us to directly compare the validity of each device's sleep assessment in patients with psychiatric disorders. The sleep-tracking devices tested in this study demonstrated high sensitivity and moderate to high specificity in determining sleep/wake. Among the devices assessed, Sleepgraph, which is a portable EEG device, exhibited higher specificity than the other devices and was able to reliably determine sleep/wake in patients with psychiatric disorders. MTN-220/221 demonstrated a relatively low specificity, indicating that it underestimates wake status.

In studies validating sleep tracking devices for healthy adults,[18] the sensitivity of actigraphy, which determines sleep using only an accelerometer, was 0.91 to 0.96 and the specificity was 0.34 to 0.63. The sensitivity of devices that utilize PPG in addition to using accelerometers such as the Fitbit, Apple Watch, and Oura Ring is reported to be 0.93–0.98, and the specificity is 0.41–0.69. In the present study, Sleepgraph, Fitbit Sense, and MTN-220/221 validated in patients with psychiatric disorders also demonstrated similar or higher agreement than that of actigraphy. Our results support the at home use of these wearable devices for determining sleep/wake for patients with psychiatric disorders.

In the EBE analysis of sleep stages, both devices often misclassified other sleep stages as light sleep (Table 4). In a given sleep period, REM sleep, light sleep, and deep sleep periods compete with each other. Therefore, the sensitivity and specificity of each sleep stage in a device may be related, and each device may have certain associations, such as a higher sensitivity for REM sleep and a lower specificity of light sleep. Fitbit has been found to over-score light sleep in previous study, which may be a feature of the scoring algorithm.[25]

The performance of the Fitbit Sense in determining each sleep stage in psychiatric patients was relatively low compared to the results of previous study conducted in healthy adults.[17] Most wearable devices developed in the recent years use data obtained from heart rate variability (HRV) and acceleration sensors to determine sleep status. In healthy adults, HRV is useful for estimating sleep stages.[29] HRV information is also used in Fitbit's sleep assessment.[25] There are

several factors in patients with psychiatric disorders that physiologically affect the functional correlation between sleep stage and HRV. For example, HF power, which reflects parasympathetic activity in HRV analysis, is decreased in patients with depression and schizophrenia. It is also associated with disease severity.[30,31] In addition, some antidepressants, such as tricyclic antidepressants, modify HRV through their anticholinergic and other effects.[31] Although the characteristics of training data and features of classifiers on sleep stage of Fitbit Sense determination have not been disclosed, the accuracy of sleep stage determination was measured in this study using existing algorithms and found to be lower for patients with psychiatric disorders, as compared with studies that included healthy subjects. This could have resulted from the clinical factors modifying the HRV parameters. However, we have not included a healthy comparison group, so this hypothesis needs to be tested in future studies.

The performance of all devices in measuring sleep parameters decreased with poor sleep: short TST, low SE, long SL, and long WASO. Similarly, low sleep efficiency decreased the validity of sleep parameters in a validation study that included healthy adults.[32] REML demonstrated a proportional bias for both devices, and a study comparing PSG and Fitbit Alta HR in patients with hypersomnia exhibited a similar trend.[33] A meta-analysis of PSG for psychiatric disorders reported that patients with depression and schizophrenia had lower TST and SE and longer SL and WASO compared with healthy controls.[34] Participants in the present study included patients with poor sleep quality, which may have reduced the validity of sleep parameters related to sleep/wake.

Intra-subject accuracy of PSG and each device was positively correlated for all devices. This suggests that for patients with low sleep-wake accuracy on the first night, it is also low on the second night. Patient-related factors led to the reduction in accuracy.

Sensor accuracy is a major concern in consumer sleep technology, as inaccurate data and inaccurate recommendations for improving sleep from sleep-tracking devices have been reported to cause orthosomnia.[35] Compared with Fitbit Sense and MTN-220/221, Sleepgraph demonstrated higher accuracy and greater sensitivity and specificity in determining each sleep stage, even in patients with psychiatric disorders. It also had a higher agreement rate for sleep parameters. This indicates the superiority of using a portable EEG device with a small number of channels for determining sleep quality in psychiatric patients. It is also less burdensome than PSG and can be performed at home. A study comparing portable EEG and actigraphy in healthy adults also demonstrated that portable EEG was more reliable for assessing sleep parameters, similar to our results, although different portable EEG devices were used.[36] Portable EEG devices that can determine sleep using automatic determination are now available. They have been reported to be highly reliable in validation studies including psychiatric patients.[37] These EEG-based sleep assessment devices are expected to facilitate clinical and epidemiological studies that seek to assess natural sleep over consecutive nights in the home setting in patients with psychiatric disorders.

This study had the following limitations: (1) As the measurements were obtained in a laboratory, they do not reflect the validity of measurements acquired in a natural environment, such as at home. The hospital environment may reduce the validity of these assessments by being more disruptive than normal everyday conditions. (2) As with many previous studies evaluating device performance, there were measurement failures for each device and not all measurements could be analyzed. (3) Due to the lack of drug restrictions, EEG and HRV could be impacted by pharmacological effects. (4) Consumer sleep technologies such as Fitbit may update their device models, apps, and sleep assessment algorithms, so these results only correspond to the versions available during the study period.

Conclusion

Our validation results indicate that sleep-tracking devices can be a promising tool for objective sleep state assessment in patients with psychiatric disorders. However, the validity of these devices may be compromised in patients with psychiatric disorders due to factors that affect the wearable device's sleep assessment algorithm, such as changes in autonomic nervous system activity and drugs/medications taken by the patient. Sleepgraph, a portable EEG device with a few channels, demonstrated higher validity in sleep assessment than actigraphy and the Fitbit Sense.

Abbreviations

EEG, electroencephalography; PSG, polysomnography; DSM-5, Diagnostic and Statistical Manual of Mental Disorders-5; BMI, body mass index; AASM, American Academy of Sleep Medicine; REM, rapid eye movement; PPG, photoplethysmography; EBE, epoch-by-epoch; PPV, positive predictive value; NPV, negative predictive value; PABAK, prevalence and bias adjusted kappa; TST, total sleep time; SE, sleep efficiency; SL, sleep latency; WASO, wake after sleep onset; REML, rapid eye movement latency; HRV, heart rate variability.

Data Sharing Statement

Data will be made available by the corresponding author upon reasonable request.

Ethics Approval and Informed Consent

The Ethics Commission of the National Center of Neurology and Psychiatry, set up as the Central Ethics Review Board, approved all study protocols and experimental procedures (approval no. B2019-001). The ethics commission of the Akita University gave permission to conduct the study. Informed consent was obtained from patients prior to their participation in the study.

Consent for Publication

As there is no patient identifying data, consent for publication was not required.

Acknowledgments

We would like to thank Dr. Mari Oba of the National Center of Neurology and Psychiatry in Japan for her support with the statistical analysis. We would like to thank Editage (www.editage.com) for English language editing.

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding

Disclosure

Dr. Mishima has received speaker's honoraria from EISAI Co, Ltd; Nobelpharma Co, Ltd; Takeda Pharmaceutical Co, Ltd; and MSD Inc and research grants from The Japan Agency for Medical Research and Development (AMED) (JP21dk0307103KM), grants from The Japanese Ministry of Education, Science, and Technology (22568378), grants from The National Center of Neurology and Psychiatry (2-1); Eisai Co, Ltd; Sumitomo Pharma Co, Ltd; and Takeda Pharmaceutical Co, Ltd.

Dr. Ozaki has received research support or speakers' honoraria from, or has served as a consultant to, Sumitomo Pharma Co, Ltd; Eisai Co, Ltd; Otsuka Pharmaceutical Co, Ltd; KAITEKI Mitsubishi Chemical Group Corporation; Mitsubishi Tanabe Pharma Corp; Shionogi & Co, Ltd; Eli Lilly KK; MOCHIDA PHARMACEUTICAL CO, LTD; DAIICHI SANKYO Co, Ltd; Nihon Medi-Physics Co, Ltd; Takeda Pharmaceutical Co, Ltd; Meiji-Seika Pharma Co, Ltd; Kyowa Co, Ltd; EA Pharma Co, Ltd; Viatris Inc; Kyowa Kirin Co, Ltd; MSD Inc; Janssen Pharmaceutical KK; and Nippon Boehringer Ingelheim Co, Ltd, outside the submitted work.

Dr. Nakagome has received speaker's honoraria from Otsuka Pharmaceutical Co, Ltd; Sumitomo Pharma Co, Ltd; Meiji-Seika Pharma Co, Ltd; Janssen Pharmaceutical KK; MOCHIDA PHARMACEUTICAL CO, LTD; Mitsubishi Tanabe Pharma Corp; Viatris Inc; and Eisai Co, Ltd. He has received payment for advising from Takeda Pharmaceutical

References

1. Sunderajan P, Gaynes BN, Wisniewski SR, et al. Insomnia in patients with depression: a STAR*D report. *CNS Spectr.* 2010;15(6):394–404. doi:10.1017/S1092852900029266

2. Brenes GA, Miller ME, Stanley MA, Williamson JD, Knudson M, McCall WV. Insomnia in older adults with generalized anxiety disorder. *Am J Geriatr Psychiatry.* 2009;17(6):465–472. doi:10.1097/JGP.0b013e3181987747

3. Laskemoen JF, Simonsen C, Büchmann C, et al. Sleep disturbances in schizophrenia spectrum and bipolar disorders - a transdiagnostic perspective. *Compr Psychiatry.* 2019;91:6–12. doi:10.1016/j.comppsych.2019.02.006

4. McClintock SM, Husain MM, Wisniewski SR, et al. Residual symptoms in depressed outpatients who respond by 50% but do not remit to antidepressant medication. *J Clin Psychopharmacol.* 2011;31(2):180–186. doi:10.1097/JCP.0b013e31820ebd2c

5. Nierenberg AA, Husain MM, Trivedi MH, et al. Residual symptoms after remission of major depressive disorder with citalopram and risk of relapse: a STAR*D report. *Psychol Med.* 2010;40(1):41–50. doi:10.1017/S0033291709006011

6. Herz MI, Melville C. Relapse in schizophrenia. *Am J Psychiatry.* 1980;137(7):801–805.

7. Sylvia LG, Dupuy JM, Ostacher MJ, et al. Sleep disturbance in euthymic bipolar patients. *J Psychopharmacol.* 2012;26(8):1108–1112. doi:10.1177/0269881111421973

8. Dombrovski AY, Mulsant BH, Houck PR, et al. Residual symptoms and recurrence during maintenance treatment of late-life depression. *J Affect Disord.* 2007;103(1–3):77–82. doi:10.1016/j.jad.2007.01.020

9. Emslie GJ, Armitage R, Weinberg WA, Rush AJ, Mayes TL, Hoffmann RF. Sleep polysomnography as a predictor of recurrence in children and adolescents with major depressive disorder. *Int J Neuropsychopharmacol.* 2001;4(2):159–168. doi:10.1017/S1461145701002383

10. Cho HJ, Lavretsky H, Olmstead R, Levin MJ, Oxman MN, Irwin MR. Sleep disturbance and depression recurrence in community-dwelling older adults: a prospective study. *Am J Psychiatry.* 2008;165(12):1543–1550. doi:10.1176/appi.ajp.2008.07121882

11. Sakurai H, Suzuki T, Yoshimura K, Mimura M, Uchida H. Predicting relapse with individual residual symptoms in major depressive disorder: a reanalysis of the STAR*D data. *Psychopharmacology.* 2017;234(16):2453–2461. doi:10.1007/s00213-017-4634-5

12. Hiranyatheb T, Nakawiro D, Wongpakaran T, et al. The impact of residual symptoms on relapse and quality of life among Thai depressive patients. *Neuropsychiatr Dis Treat.* 2016;12:3175–3181. doi:10.2147/NDT.S124277

13. Edinger JD, Fins AI, Glenn DM, et al. Insomnia and the eye of the beholder: are there clinical markers of objective sleep disturbances among adults with and without insomnia complaints? *J Consult Clin Psychol.* 2000;68(4):586–593. doi:10.1037/0022-006X.68.4.586

14. Reeve S, Sheaves B, Freeman D. The role of sleep dysfunction in the occurrence of delusions and hallucinations: a systematic review. *Clin Psychol Rev.* 2015;42:96–115. doi:10.1016/j.cpr.2015.09.001

15. Riemann D, Krone LB, Wulff K, Nissen C. Sleep, insomnia, and depression. *Neuropsychopharmacology.* 2020;45(1):74–89. doi:10.1038/s41386-019-0411-y

16. Khosla S, Deak MC, Gault D, et al. Consumer sleep technology: an American academy of sleep medicine position statement. *J Clin Sleep Med.* 2018;14(5):877–880. doi:10.5664/jcsm.7128

17. 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

18. 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. doi:10.3389/fdgth.2021.721919

19. Benca RM, Obermeyer WH, Thisted RA, Gillin JC. Sleep and psychiatric disorders. A meta-analysis. *Arch Gen Psychiatry.* 1992;49(8):651–668. doi:10.1001/archpsyc.1992.01820080059010

20. Chan MS, Chung KF, Yung KP, Yeung WF. Sleep in schizophrenia: a systematic review and meta-analysis of polysomnographic findings in case-control studies. *Sleep Med Rev.* 2017;32:69–84. doi:10.1016/j.smrv.2016.03.001

21. Stubbs B, Vancampfort D, Veronese N, et al. The prevalence and predictors of obstructive sleep apnea in major depressive disorder, bipolar disorder and schizophrenia: a systematic review and meta-analysis. *J Affect Disord.* 2016;197:259–267. doi:10.1016/j.jad.2016.02.060

22. Kalucy MJ, Grunstein R, Lambert T, Glozier N. Obstructive sleep apnoea and schizophrenia--a research agenda. *Sleep Med Rev.* 2013;17 (5):357–365. doi:10.1016/j.smrv.2012.10.003

23. Inada T, Inagaki A. Psychotropic dose equivalence in Japan. *Psychiatry Clin Neurosci.* 2015;69(8):440–447. doi:10.1111/pcn.12275

24. Kataoka H, Takatani T, Sugie K. Two-channel portable biopotential recording system can detect REM sleep behavioral disorder: validation study with a comparison of polysomnography. *Parkinsons Dis.* 2022;2022:1888682. doi:10.1155/2022/1888682

25. Beattie Z, Oyang Y, Statan A, et al. Estimation of sleep stages in a healthy adult population from optical plethysmography and accelerometer signals. *Physiol Meas.* 2017;38(11):1968–1979. doi:10.1088/1361-6579/aa9047

26. Nakazaki K, Kitamura S, Motomura Y, et al. Validity of an algorithm for determining sleep/wake states using a new actigraph. *J Physiol Anthropol.* 2014;33(1):31. doi:10.1186/1880-6805-33-31

27. Berry RB, Albertario CL, Harding SM, et al. *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications. Version 2.5.* Darien (IL): American Academy of Sleep Medicine; 2018.

28. 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

29. de Zambotti M, Trinder J, Silvani A, Colrain IM, Baker FC. Dynamic coupling between the central and autonomic nervous systems during sleep: a review. *Neurosci Biobehav Rev.* 2018;90:84–103. doi:10.1016/j.neubiorev.2018.03.027

30. Jung W, Jang KI, Lee SH. Heart and brain interaction of psychiatric illness: a review focused on heart rate variability, cognitive function, and quantitative electroencephalography. *Clin Psychopharmacol Neurosci.* 2019;17(4):459–474. doi:10.9758/cpn.2019.17.4.459

31. Kemp AH, Quintana DS, Gray MA, Felmingham KL, Brown K, Gatt JM. Impact of depression and antidepressant treatment on heart rate variability: a review and meta-analysis. *Biol Psychiatry.* 2010;67(11):1067–1074. doi:10.1016/j.biopsych.2009.12.012

32. Chinoy ED, Cuellar JA, Jameson JT, Markwald RR. Performance of four commercial wearable sleep-tracking devices tested under unrestricted conditions at home in healthy young adults. *Nat Sci Sleep.* 2022;14:493–516. doi:10.2147/NSS.S348795

33. 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

34. Baglioni C, Nanovska S, Regen W, et al. Sleep and mental disorders: a meta-analysis of polysomnographic research. *Psychol Bull.* 2016;142 (9):969–990. doi:10.1037/bul0000053

35. Baron KG, Abbott S, Jao N, Manalo N, Mullen R. Orthosomnia: are some patients taking the quantified self too far? *J Clin Sleep Med.* 2017;13 (2):351–354. doi:10.5664/jcsm.6472

36. Matsuo M, Masuda F, Sumi Y, et al. Comparisons of portable sleep monitors of different modalities: potential as naturalistic sleep recorders. *Front Neurol.* 2016;7:110. doi:10.3389/fneur.2016.00110

37. Miyata S, Iwamoto K, Banno M, et al. Performance of an ambulatory electroencephalogram sleep monitor in patients with psychiatric disorders. *J Sleep Res.* 2021;30(4):e13273. doi:10.1111/jsr.13273