Sleep and Biological Rhythms
https://doi.org/10.1007/s41105-022-00421-5

ORIGINAL ARTICLE

Comparison of the usability of an automatic sleep staging program via portable 1-channel electroencephalograph and manual sleep staging with traditional polysomnography

Aoi Kawamura[1,2] · Takuya Yoshiike[2] · Masahiro Matsuo[1] · Hiroshi Kadotani[1] · Yuki Oike[1] · Midori Kawasaki[1] · Yuichi Kurumai[1] · Kentaro Nagao[2] · Masanori Takami[1] · Naoto Yamada[1] · Kenichi Kuriyama[1,2]

Received: 7 June 2022 / Accepted: 16 August 2022

Abstract

Automatic algorithms are a proposed alternative to manual assessment of polysomnography data for analyzing sleep structure; however, none are acceptably accurate for clinical use. We investigated the feasibility of an automated sleep stage scoring system called Sleep Scope, which is intended for use with portable 1-channel electroencephalograph, and compared it with the traditional polysomnography scoring method. Twenty-six outpatients and fourteen healthy volunteers underwent Sleep Scope and polysomnography assessments simultaneously. Polysomnography records were manually scored by three sleep experts. Sleep Scope records were scored using a dedicated auto-staging algorithm. Sleep parameters, including total sleep time, sleep latency, wake after sleep onset, and sleep efficiency, were calculated. The epoch-by-epoch pairwise concordance based on the classification of sleep into five stages (i.e., wake, rapid eye movement, N1, N2, and N3) was also evaluated after validating homogeneity and bias between Sleep Scope and polysomnography. Compared with polysomnography, Sleep Scope seemed to overestimate sleep latency by approximately 3 min, but there was no consistent tendency in bias in other sleep parameters. The *K* values ranged from 0.66 to 0.75 for experts' inter-rater polysomnography scores and from 0.62 to 0.67 for Sleep Scope versus polysomnography scores, which indicated sufficient agreement in the determination of sleep stages based on the Landis and Koch criteria. We observed sufficient concordance between Sleep Scope and polysomnography despite lower concordance in sleep disorder patients. Thus, this auto-staging system might serve as a novel clinical tool for reducing the time and expenses required of medical staff and patients.

Keywords Sleep structure · Neurophysiology · Medical economy · Clinical measurement · Measurement equipment · Diagnostic marker

Introduction

An electroencephalograph (EEG) evaluates brain neurophysiology and is useful for diagnosing some types of epilepsy and consciousness disturbances [1, 2]. It is also used to assess sleep physiology and structure. Sleep stage is a core index in evaluating sleep structure, and determining sleep stage is important in clinical measurement and basic sleep research [3]. Trained experts visually determine sleep stages based on the electrophysiological profile measured during polysomnography (PSG). PSG typically records multimodal physiological information from biometric sensors, which include six-channel (Ch) EEG (i.e., the bilateral frontal, central, and occipital positions), electrooculography, electromyography, electrocardiography, and respiratory event monitoring sensors. PSG is useful for the differential diagnosis and assessment of severity in various sleep disorders [3], although its precise measurement environment is confined only to hospitals or sleep laboratories under experts' supervision. Thus, it is difficult to eliminate the influences of an unfamiliar sleeping environment on sleep physiology (e.g., the first-night effect) in addition to the temporal and economic constraints of hospitalization [4].

✉ Kenichi Kuriyama
kenichik@ncnp.go.jp

[1] Department of Psychiatry, Shiga University of Medical Science, Otsu, Japan

[2] Department of Sleep-Wake Disorders, National Institute of Mental Health, National Center of Neurology and Psychiatry, 4-1-1 Ogawa-Higashi, Kodaira, Tokyo 187-8551, Japan

Published online: 26 August 2022

To address these issues, activity-based sleep monitoring devices that gather sleep information based on physical activity measurements have been used to promote clinical research in sleep science [5–9]. However, this modality is less accurate in estimating sleep structure than PSG and in individuals with sleep disorders than those without [7–9]. Thus, it is unsuitable for elucidating the pathophysiology of sleep disorders and is only appropriate for screening these disorders. Recent studies have examined the utility of portable EEGs for confirming sleep structure [10–12] and overcoming the drawbacks of activity-based sleep monitoring devices while maintaining portability. Although some studies have attempted to apply an automatic sleep staging algorithm to portable EEG data, few have compared the results with those of conventional methods, and most studies targeted primarily young healthy participants [13, 14].

Sleep Scope (SS) by SleepWell Co., Ltd. Osaka, Japan [15] is portable and capable of sampling a sleep-recording EEG. It is a 1-Ch EEG device with a dedicated automatic sleep stage analysis program. The accuracy of the automatic analysis system of SS in healthy adults has been compared with that of traditional PSG scoring by a single PSG expert [16], and the agreement rate of sleep stage determination between SS and PSG has been reported; however, SS-measured EEG data were analyzed using a previous version of the dedicated automatic sleep staging program (version 1.0-r3). Furthermore, since only young healthy participants were examined, the different features of sleep architecture between patients and healthy participants were not considered.

We aimed to estimate the agreement rate for determination of sleep stage between an updated SS version (version 2.0) of the automated sleep stage analysis system and traditional PSG-based scoring by PSG experts in patients with sleep disorders and healthy volunteers. Moreover, we examined the agreement rate for sleep stage determination of the SS system using manual PSG scorings of sleep stage by several experts, considering inter-rater variability.

Materials and methods

Participants

We recruited 26 adult outpatients aged < 80 years who had undergone PSG at the Sleep Disorders Center of the Shiga University of Medical Science Hospital (Otsu, Japan) and 15 healthy adult volunteers without any sleep or neuropsychiatric problems based on a clinical interview and a self-completed questionnaire score of ≤ 3 on the Athens Insomnia Scale [17] between February 2018 and March 2018. Participants with neuropsychiatric disorders such as dementia or epilepsy and those who engaged in any type of night-shift work (0:00–5:00) were excluded. Participants with sleep latency (SL) ≥ 180 min were also excluded to eliminate the effects of inadequate sleep hygiene or possible technical errors of SS and PSG measurements. The flow diagram of the study participants is shown in Fig. 1.

For all participants, full-night sleep EEG was simultaneously measured using the SS and traditional PSG once during overnight hospitalization. The conclusive diagnosis of sleep disorders in 26 adult outpatients were made by sleep medicine specialists certified by the Japanese Society of Sleep Research (JSSR) based on a clinical interview and PSG results, followed by multiple SL tests when required, according to the criteria of the International Classification of Sleep Disorders, Third Edition [18].

All study procedures were approved by the Ethics Committee of Shiga University of Medical Science (Approval No. 29-266). The study was conducted in accordance with the Helsinki Declaration, and written informed consent was obtained from all participants.

Polysomnography

PSG recordings were obtained using the Alice-5 system (Respironics Inc., Murrysville, PA, USA) with the following set of measurements: six-electrode scalp-encephalography based on the international 10–20 system (i.e., central [C]3–auricular [A]2, C4–A1, occipital [O]1–A2, O2–A1, frontal [F]3–A2, and F4–A1); two-electrode electrooculography (placed near the outer canthus of the eyes), electrocardiography, and electromyography for the chin and bilateral anterior tibialis; and sensors to detect oral/nasal airflow and chest/abdominal movements [19]. Oxygen saturation was monitored using pulse oximetry (SpO_2). The EEG sampling rate and filter settings were 200 Hz and 0.3–35 Hz,

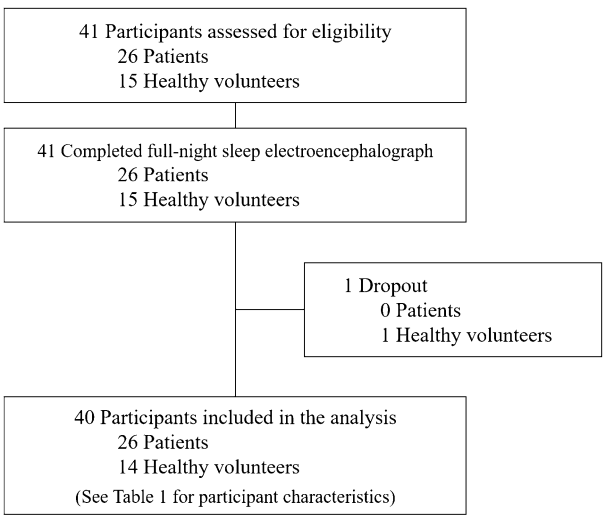

Fig. 1 Flowchart of the study participants

respectively. Other specifications met the EEG standard of the Japanese Industrial Standard (JIS; T1203) [20]. PSG data were recorded online using Alice Sleepware (version 2.8, Respironics Inc., Murrysville, PA, USA). Based on the EEG data, sleep was classified into five stages (wake (WK); rapid eye movement (REM); and stage N1 (N1), N2, and N3) and visually scored for each 30-s epoch by PSG experts once per epoch of data according to the American Academy of Sleep Medicine (AASM) scoring manual, version 2.5 [19]. The experts scored any epoch as "undeterminable" when more than one-half of an EEG epoch contained artifacts due to unspecified physiological or electrical sources, such as nocturnal behaviors with arousals. Sleep parameters such as total sleep time (TST), SL, wake after sleep onset (WASO), and sleep efficiency (SE) were calculated according to AASM definition [19]. For strict verification, three experts (Experts 1–3) with > 5 years of experience and blinded to participants' conditions separately evaluated the data. They were also blinded to the results of sleep staging scored by the other experts, including via the SS system. All experts were certified by the JSSR, and two of the experts were also registered polysomnographic technologists certified by the Board of Registered Polysomnographic Technologists (Arlington, VA, USA).

The apnea–hypopnea index was estimated by counting the hourly number of apnea (i.e., cessation in breathing lasting at least 10 s) and hypopnea events (i.e., reduction in airflow amplitude or respiratory effort by at least 30% in association with > 3% reduction in SpO_2 for at least 10 s). A sleep-onset REM period was defined as REM latency < 15 min after sleep onset. REM sleep without atonia was defined as excessive tonic muscular activity (i.e., submental electromyographic activity) exceeding twice that of the background activity for > 50% of the 30-s epoch or as excessive phasic muscular activity (i.e., submental electromyographic activity burst lasting 0.1–5.0 s, which is four times higher than the background activity in 3-s mini-epochs). These sleep metrics were referenced in the diagnosis of sleep apnea, hypersomnia, and REM sleep behavior disorder based on the criteria of the International Classification of Sleep Disorders, Third Edition [18].

Portable 1-Ch EEG (SS system)

The SS consists of a portable 1-Ch EEG device and a dedicated automatic sleep staging program version 2.0 (SleepWell Co.). The 1-Ch EEG device is palm-sized (6.3-cm wide, 9.4-cm long, 3.4-cm deep), light-weight (125 g), and powered by two chargeable AA batteries. One electrode (self-adhesive disposable Ag/AgCl surface Blue Sensor M; Ambu, St Ives, Cambridgeshire, UK) was placed over the middle of the forehead (2 cm below the midline between the two frontal electrodes of PSG), while the other was placed at the left mastoid. In this position, hair resistance can be avoided, and the device is easily worn by the individual. The 1-Ch EEG obtained from these two electrodes reflects the activity of the entire brain by bipolar derivation, instead of collecting site-specific EEG by monopolar derivation using a reference electrode, as does a multi-channel EEG. All data were recorded digitally in European Data Format. The sampling rate and filter settings were 128 Hz and 0.5–64 Hz, respectively. Other specifications met JIS T1203 criteria.

The 1-Ch EEG recordings and PSG recordings were strictly synchronized by a PSG technician. Therefore, the temporal error in recording initiation was within 6 s (mean ± standard deviation: 2 ± 1 s) for each participant. The data obtained by the device were forwarded to cloud services (SEAS-G, SleepWell Co., Ltd, Osaka, Japan), in which the spectral analysis of the EEG data was conducted.

Table 1 Participant characteristics

	All ($N=40$)	Patient group ($N=26$)		Healthy group ($N=14$)
Age (years) median (25th–75th percentile)	24.00 (21.50–44.25)	39.5 (24.00–54.00)*		21.00 (20.00–22.00)*
Male, *N* (%)	29 (72.5%)	21 (80.8%)[n.s]		8 (57.1%)[n.s]
Diagnosis		Primary diagnosis	Secondary diagnosis	
Obstructive sleep apnea		8	0	N/A
Insomnia		1	0	
Hypersomnia		9	1	
Narcolepsy		2	0	
Other		7	1	
Circadian sleep–wake rhythm disorder		3	1	
Rapid eye movement sleep behavior disorder		1	0	
Restless legs syndrome		4	0	

N, number; N/A; not applicable; n.s.; the sex distribution difference between the two groups was not significant confirmed by Fisher's Exact test

*$P < 0.001$; the age difference between the two groups was tested using the Mann–Whitney *U* test

Table 2 Sleep outcome measures for the patient and healthy groups based on the Sleep Scope system and polysomnography findings

Outcome measure	Sleep Scope	Polysomnography (expert)		
		1	2	3
All participants ($N=40$)				
Total sleep time (min)	408 (69)	425 (69)	414 (72)	400 (74)
Sleep onset latency (min)	24 (18)	21 (16)	21 (17)	21 (16)
WASO (min)	70 (58)	55 (58)	67 (61)	81 (63)
SE (%)	81 (12)	85 (12)	82 (13)	80 (13)
Number of awakenings, ≤2 min	131 (115)	188 (69)	119 (77)	111 (57)
Sleep stage				
N1 (min)	56 (50)	83 (49)	85 (60)	80 (54)
N2 (min)	222 (64)	205 (57)	193 (72)	223 (86)
N3 (min)	36 (34)	62 (40)	47 (33)	24 (23)
REM (min)	93 (38)	74 (27)	88 (31)	73 (29)
Undeterminable (% of total epochs of all participants)	0.214	1.11	0.546	0.0224
Patient group ($N=26$)				
Total sleep time (min)	385 (71)	403 (72)	391 (76)	374 (75)
Sleep onset latency (min)	21 (12)	18 (11)	18 (11)	19 (11)
WASO (min)	88 (63)	72 (65)	86 (68)	101 (69)
SE (%)	78 (14)	82 (14)	79 (15)	76 (15)
Number of awakenings, ≤2 min	165 (126)	197 (58)	135 (83)	123 (61)
Sleep stage				
N1 (min)	56 (50)	83 (49)	85 (60)	80 (54)
N2 (min)	222 (64)	205 (57)	193 (72)	223 (86)
N3 (min)	36 (34)	62 (40)	47 (33)	24 (23)
REM (min)	93 (38)	74 (27)	88 (31)	73 (29)
Undeterminable (% of total epochs of all participants)	0.327	1.19	0.642	0.00389
Healthy group ($N=14$)				
Total sleep time (min)	450 (41)	466 (38)	456 (35)	448 (40)
Sleep onset latency (min)	30 (24)	26 (22)	27 (22)	26 (22)
WASO (min)	36 (19)	24 (17)	33 (18)	43 (20)
SE (%)	87 (5)	90 (5)	88 (5)	87 (5)
Number of awakenings, ≤2 min	69 (39)	169 (81)	89 (48)	90 (38)
Sleep stage				
N1 (min)	56 (50)	83 (49)	85 (60)	80 (54)
N2 (min)	222 (64)	205 (57)	193 (72)	223 (86)
N3 (min)	36 (34)	62 (40)	47 (33)	24 (23)
REM (min)	93 (38)	74 (27)	88 (31)	73 (29)
Undeterminable (% of total epochs of all participants)	0.0138	0.976	0.374	0.0554

All participants ($N=40$)

All data are presented as the mean value and standard deviation [mean (SD)]

N1, sleep stage N1; N2, sleep stage N2; N3, sleep stage N3; REM, rapid eye movement; WASO, wake after sleep onset; SE, sleep efficiency

Note (for all tables): Sleep Scope is manufactured by SleepWell Co., Ltd. (Osaka, Japan; https://sleepwell.co.jp/sleepscope)

Automated scoring

Data obtained from the SS system were analyzed using an auto-staging algorithm for every 30-s epoch and classified automatically into five sleep stages: WK, REM, N1, N2, and N3. The unanalyzable epochs due to poor EEG data quality were classified as "undeterminable." When more than one-half of an epoch consisted of a detectable EEG pattern, the epoch was sleep staged. The details of this auto-staging algorithm are not open to the public; however, SleepWell Co., Ltd. outlines it as follows: for the "correct answer" in

Table 3 Mean differences between the Sleep Scope system and polysomnography (i.e., Expert 1–3) measurements for each sleep parameter

Sleep parameter	Friedman test			Wilcoxon signed rank test					
	Chi-square	*df*	*P*	Ex1 versus Ex2	Ex1 versus Ex3	Ex2 versus Ex3	Ex1 versus SS	Ex2 versus SS	Ex3 versus SS
Total sleep time	57.861	3	< 0.001*	< 0.001*	< 0.001*	< 0.001*	< 0.001*	0.080	0.557
Sleep onset latency	37.877	3	< 0.001*	5.418	0.373	0.616	< 0.001*	< 0.001*	< 0.001*
WASO	57.023	3	< 0.001*	< 0.001*	< 0.001*	< 0.001*	< 0.001*	1.331	0.032*
SE	57.861	3	< 0.001*	< 0.001*	< 0.001*	< 0.001*	< 0.001*	0.069	0.558

All participants ($N = 40$)

WASO, wake after sleep onset; SE, sleep efficiency; Ex, expert; SS, Sleep Scope

*$P < 0.05$; for the Wilcoxon signed rank test, the Bonferroni correction was applied to correct for multiple testing

the supervised learning, which is a reference to SS data, stage determination by JSSR-certified experts was adopted using the AASM criteria (version 2.0; October 2012) based on simultaneously measured PSG data (EEG, electrooculography, and electromyography). The stage determination algorithm was obtained by examining the sleep staging mechanism based on stage-specific waveform patterns of the "correct answer." The sample data (i.e., teacher) were increased from 234 to 1,199 in the latest version (2.0), compared to the previous version, to improve accuracy. This version upgrade of SS was simply due to the increased number of teacher data. The increase in teacher data led to an increase in the number of parameters, which are the stage-specific waveform patterns referenced in machine learning, from 46 to 67 over the previous version.

Sleep stages were scored based on SS EEG data for each 30-s epoch. These timing data allowed us to easily synchronize SS and PSG results. Sleep parameters (i.e., TST, SL, WASO, and SE) were calculated from the obtained data according to the definition of the AASM criteria [19].

Statistical analysis

Each statistical method was applied after confirming distribution normality using the Shapiro–Wilk tests. The occurrence ratio of sleep stage was calculated by dividing the number of epochs of each sleep stage by the total number of epochs. Fisher's exact test was used to examine the difference in sex distribution between patient and healthy groups. The Mann–Whitney *U* test was used to examine differences in age and ratio of each sleep stage between patient and healthy groups, as well as the difference in the ratio of matches between PSG and SS determinations for each expert between patient and healthy groups. Friedman test followed by Wilcoxon signed-rank test with Bonferroni correction was used to compare the means between SS and the three PSG-expert measurements among sleep parameters (TST, SL, WASO, and SE). Bland–Altman plot analysis was used to evaluate bias between SS and each PSG measurement on sleep parameters. For statistical assessment, each mean bias was tested with a one-sample *t* test (null hypothesis was defined as the mean bias between SS and PSG measurement, which is zero). Bland–Altman analysis was also separately performed for patient and healthy groups. To evaluate the pairwise concordance in sleep staging between SS and PSG data, κ value and percent agreement were calculated. The statistical plan for validating the feasibility of SS against PSG was guided by reference to recommendations from the International Biomarkers Workshop on Wearables in Sleep and Circadian Science [21]. All statistical analyses were conducted using SPSS (version 25; IBM, Armonk, NY, USA). A p value < 0.05 was considered statistically significant. To evaluate κ values, the Landis and Koch criteria were adopted [22].

Results

Participant characteristics are summarized in Table 1. Data from one healthy volunteer were excluded from the analysis, because SL was ≥ 180 min. The healthy group was significantly younger than the patient group (median age, 21.00 years; $P < 0.001$). The sex distribution difference between the two groups was not significant ($P = 0.147$).

Outcomes of the one-night sleep EEG obtained from patient and healthy groups are shown in Table 2. The percentage of undeterminable epochs in each sleep stage for SS and PSG was within the level of 1% in both groups (Table 2). TST of all participants obtained via PSG exceeded 240 min, which suggested that measurements were conducted precisely with reference to minimum hours required to diagnose apnea (i.e., 180 min) and quality assessment of sleep recordings by Redline et al. [23]. The mean differences between the SS system and three PSG-expert measurements for each sleep parameter are shown in Table 3. The significant mean

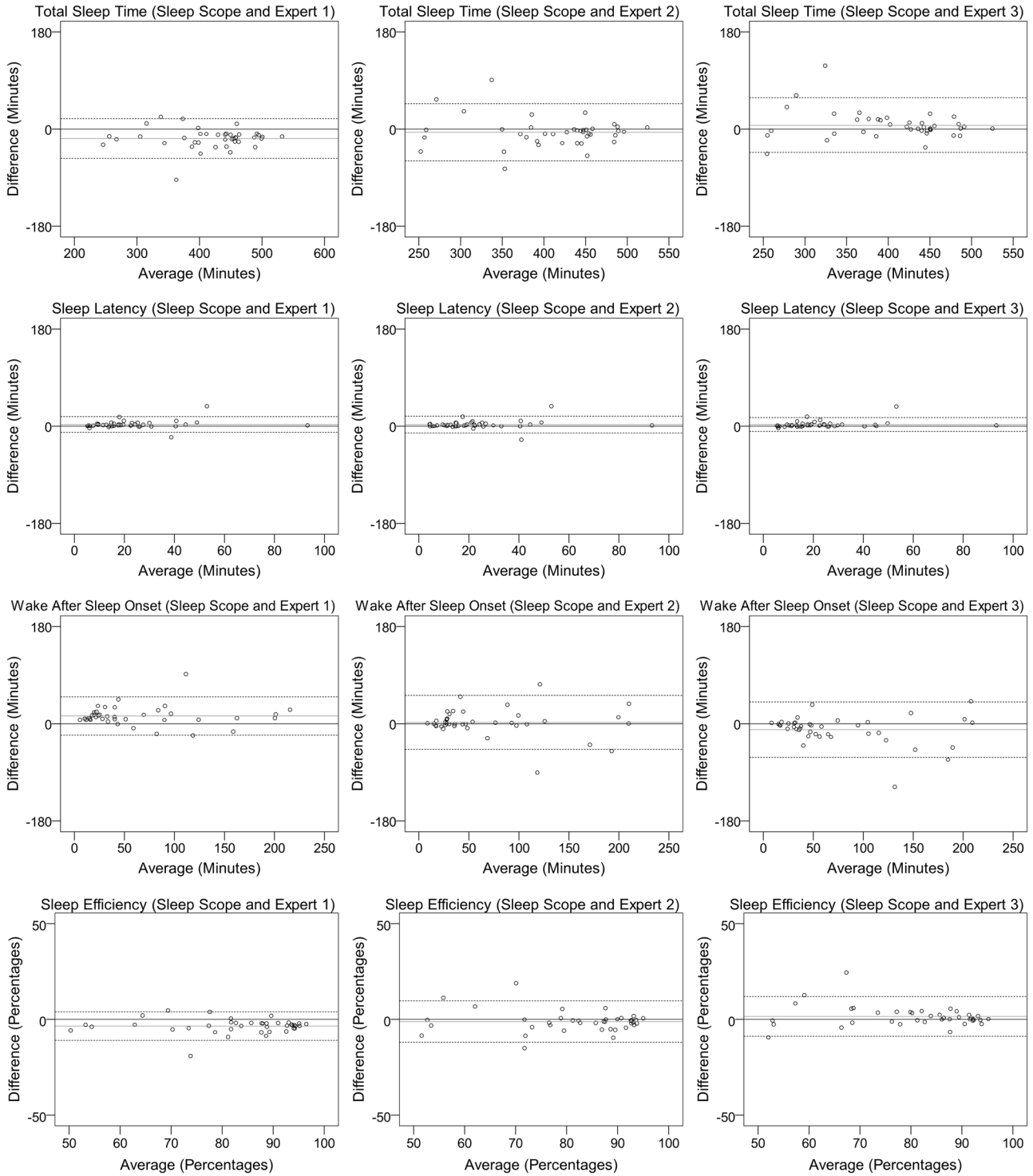

Fig. 2 Bland–Altman plot analysis of sleep parameters. Each plot represents the data for each participant's sleep parameters. The horizontal solid gray lines represent the means bias (i.e., the differences between the two measures), and the horizontal dotted lines represent the 95% limits of agreement. Positive plots on the vertical axis indicate an overestimation of each sleep parameter by Sleep Scope, and negative plots on the vertical axis indicate an underestimation of each sleep parameter by Sleep Scope, compared to the results from each polysomnography expert

Table 4 Agreement between experts in sleep stage determination based on polysomnography data

	Polysomnography (expert)		
	1 versus 2	1 versus 3	2 versus 3
All participants ($N=40$)			
Overall agreement (%)	80.8	75.1	81.4
Overall *kappa*	0.746	0.664	0.749
Patient group ($N=26$)			
Overall agreement (%)	79.6	73.5	80.1
Overall *kappa*	0.736	0.649	0.738
Healthy group ($N=14$)			
Overall agreement (%)	82.8	78.0	83.6
Overall *kappa*	0.760	0.682	0.762

Table 5 Agreement between the Sleep Scope system and the polysomnography experts in sleep stage determination

	Polysomnography (expert)		
	1 versus Sleep Scope	2 versus Sleep Scope	3 versus Sleep Scope
All participants ($N=40$)			
Overall agreement (%)	71.9	75.7	75.4
Overall *kappa*	0.620	0.672	0.657
Patient group ($N=26$)			
Overall agreement (%)	69.9	72.7	72.3
Overall *kappa*	0.599	0.638	0.623
Healthy group ($N=14$)			
Overall agreement (%)	75.5	81.0	80.9
Overall *kappa*	0.652	0.728	0.716

differences in TST, WASO, and SE among the four measurement groups, confirmed by the Friedman test, were not specifically dependent on the SS–PSG experts' measurement differences as direct comparisons both among PSG experts and between SS and each PSG expert (Wilcoxon signed-rank test with Bonferroni correction) showed statistical mean differences in these variables (Table 3). In contrast, the significant mean difference in SL, confirmed by the Friedman test, was found to be due to the difference between SS and PSG experts' measurements but not due to those among PSG experts' measurements (Table 3). The Bland–Altman plots of TST, SL, WASO, and SE indicated that they were not positively associated with measurement time (Fig. 2). The SS system had a slight systematic bias in measuring TST, SL, WASO, or SE (Fig. 2 and S1 Table in Online Resource 2). Based on mean bias values, only the SS system, compared with PSG Experts 1–3, significantly overestimated SL by about 3 min, while other sleep parameters showed no consistent tendency in bias (Fig. 2 and S1 Table in Online Resource 2). Furthermore, a similar trend was observed in measuring sleep parameters for patient and healthy groups based on Bland–Altman plot analyses, and the disappearance of the significance of difference of mean bias values may be attributed to the smaller sample size and higher variability in the healthy group (S1 and S2 Figs in Online Resource 1; S2 and S3 Tables in Online Resource 2).

The inter-rater agreement and SS versus PSG (i.e., Experts 1–3) agreement along with κ values are summarized for both groups in Tables 4 and 5, respectively. We found substantial agreement among all PSG expert combinations. Based on κ values comparing SS and PSG experts in all participants' data, all three SS–expert combinations had a substantial agreement. The variability in the PSG-matching ratio for each expert is shown as a box plot for each sleep stage for the two groups (Fig. 3). No sleep stage exhibited a significant intergroup difference in the matched ratio. Among the high agreement for sleep staging, relatively low agreement rates were found throughout N1.

The occurrence ratios of sleep stages are shown in Table 6. These occurrence ratios were calculated separately for each PSG expert and SS. The median of these four values (i.e., Experts 1–3 and SS) in each sleep stage was used to test for differences in the occurrence ratio of each sleep stage between patient and healthy groups (Mann–Whitney *U* test). Although the occurrence ratios of undeterminable epochs tended to be lower in the patient group than in the healthy group, differences between the two groups were not significant. Significantly higher occurrence ratios for N1 and WK and a significantly lower occurrence ratio for N2 occurred in the patient group than in the healthy group.

Discussion

Although there seemed to be lower concordance in sleep parameters except for SL in the patient than the healthy group, the agreement ratios for sleep parameters between the SS system and each PSG expert also suggest the potential utility of the SS system for clinical evaluation in sleep medicine.

While the overall agreement rate for sleep stage determination between the SS system and PSG was sufficient, the pairwise κ values of SS–PSG appeared to be smaller than those among the three PSG experts. The low SS–PSG concordance of sleep stages in the patient group may have been caused by more frequent concomitant sudden sleep stage transitions (e.g., awakening because of apnea or hypopnea), which is consistent with the greater incidence of WK and N1 in the patient group. Moreover, differences in sleep structure

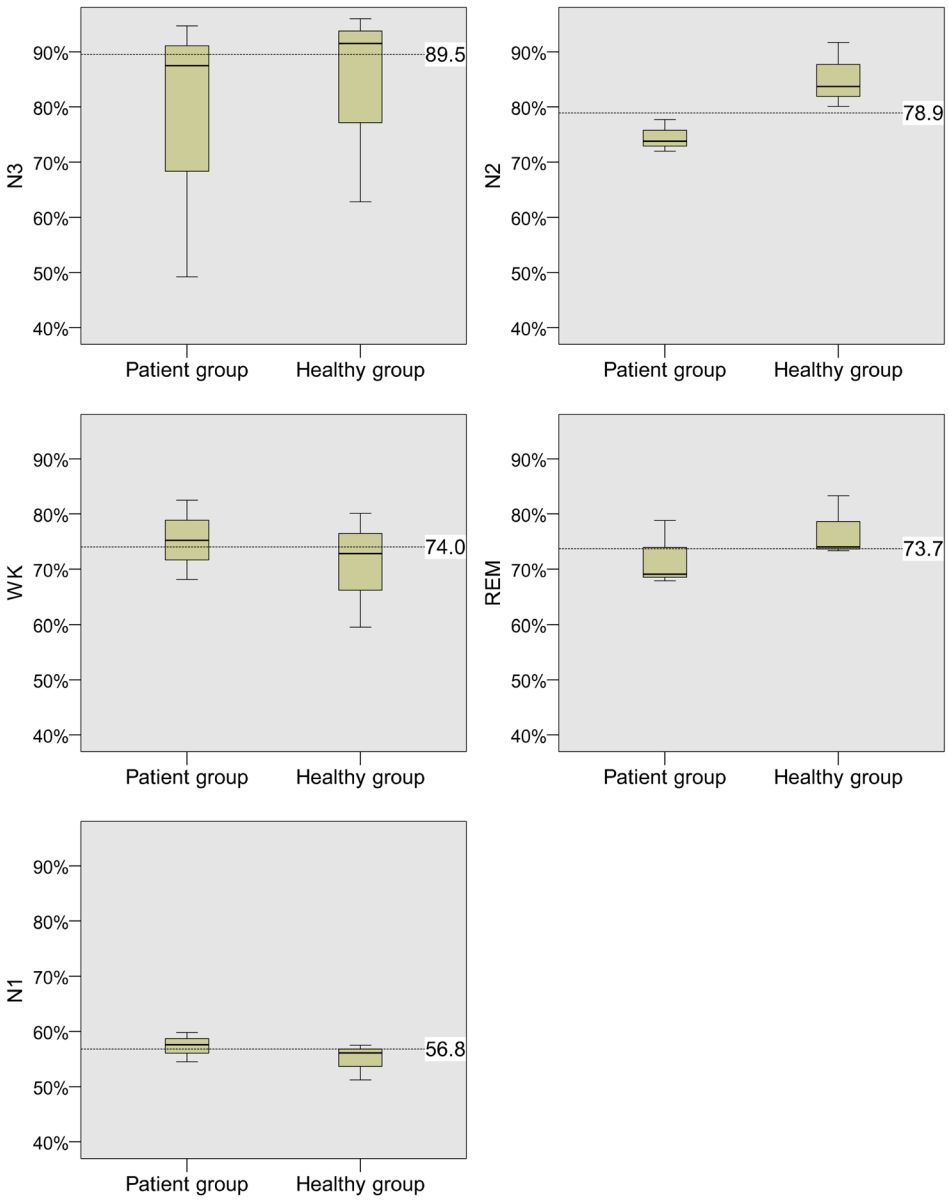

Fig. 3 Box plots of the matching ratio for the Sleep Scope system and each polysomnography expert's determination of each sleep stage. The boxes represent the interquartile range of matched ratios (the top and bottom lines of the box represent the first and third quartiles, and the thick line in the middle of the box represents the median). The whiskers show the maximum and minimum values. The horizontal dotted lines represent the median percentage of the combined data of the two groups. WK, wake; N1, sleep stage N1; N2, sleep stage N2; N3, sleep stage N3; REM, rapid eye movement

may have lowered the concordance because of the higher age distribution of patients than healthy participants [24]. The higher percentage of N1 and WK with a relatively low SS–PSG agreement rate and lower percentage of N2 with a relatively high SS–PSG agreement rate in the patient group may have decreased the κ value.

Overall differences in SS–PSG concordance may be intrinsically influenced by the difference in measurement sites, since the only middle frontal derivation of SS may detect alpha activity weaker than the occipital derivation, which is ascertainable by PSG [25, 26]. Determining the WK-N1 boundary depends on the frequency of alpha-wave entities based on the AASM criteria [19]; therefore, determination of N1 using the SS system seems to be a continued weakness of the current algorithm. Popovic et al. demonstrated that even if the same EEG data were referenced, the percentage of matched sleep stage between manual and automatic scoring is 78.7–76.8% for WK epochs and 51.3–51.1% for N1 epochs [13]. Despite the lack of an electrooculogram in the SS system, it seems nearly sufficient for detecting REM sleep. The median percentage of the SS system's

Table 6 Occurrence ratio of each sleep stage estimated by the Sleep Scope system and all PSG experts

Sleep stage (%)			Patient group ($N=26$)		Healthy group ($N=14$)	
N1	Expert	1	19.2	Median (25th–75th percentile) 19.2 (16.4–19.7)*	12.1	Median (25th–75th percentile) 10.8 (8.7–11.7)*
		2	20.2		11.3	
		3	19.1		10.3	
	Sleep Scope		13.6		7.1	
N2	Expert	1	37.9	Median (25th–75th percentile) 38.6 (36.2–40.4)*	46.2	Median (25th–75th percentile) 47.7 (46.1–51.3)*
		2	34.4		46.0	
		3	39.3		53.4	
	Sleep Scope		41.5		49.2	
N3	Expert	1	10.6	Median (25th–75th percentile) 6.9 (4.8–9.3)	15.6	Median (25th–75th percentile) 10.8 (8.1–13.8)
		2	8.0		12.0	
		3	3.7		6.7	
	Sleep scope		5.8		9.5	
WK	Expert	1	17.2	Median (25th–75th percentile) 21.1 (18.8–23.0)*	8.8	Median (25th–75th percentile) 12.0 (10.0–12.9)*
		2	20.3		11.2	
		3	24.2		13.1	
	Sleep Scope		21.8		12.7	
REM	Expert	1	13.9	Median (25th–75th percentile) 15.3 (13.8–16.8)	16.4	Median (25th–75th percentile) 17.7 (16.4–20.2)
		2	16.6		19.0	
		3	13.7		16.4	
	Sleep Scope		17.0		21.4	
Undeterminable	Expert	1	1.2	Median (25th–75th percentile) 0.5 (0.2–0.9)	1.0	Median (25th–75th percentile) 0.3 (0.1–0.7)
		2	0.6		0.4	
		3	0.0		0.1	
	Sleep Scope		0.3		0.0	

PSG, polysomnography; N1, sleep stage N1; N2, sleep stage N2; N3, sleep stage N3; WK, wake; REM, rapid eye movement

*$P<0.05$; the difference between the two groups (Patient vs. Healthy) was tested using the Mann–Whitney U test

match to expert determinations of REM based on PSG data exceeded 70% of the combined data of the two groups.

In healthy individuals, Matsuo et al. demonstrated the superiority of the SS system with regard to sleep-length estimation and sleep-epoch detection over two other activity-based sleep monitors (using PSG as a reference) [16]. The current results support this notion by showing good agreement between SS and PSG, regardless of the presence of sleep disorders.

Previous studies have suggested that activity-based sleep monitors have a low ability to identify WASO or wakefulness [6–9]. This limitation may be a potential issue in people whose sleep is highly fragmented, which occurs in older adults and patients with sleep disorders [7]. The high correlation of WASO between SS and PSG in patients and healthy participants may demonstrate the superiority of predicting sleep based on 1-Ch EEG over predictions based on an individual's movements.

PSG is generally utilized for diagnosing several sleep disorders. However, a portable EEG system that includes the SS system may have novel clinical utility for pre- and post-treatment evaluations regardless of whether the EEG is recorded in home or laboratory environments. Reducing time and economic restrictions of measurement is a substantial advantage of the SS system over traditional PSG. Furthermore, the ease of obtaining measurements using the SS system suggests that it is suitable for large data sampling, which may aid the development of novel sleep-related biomarkers

in patients with neuropsychiatric disorders [27–29]. Studies have suggested the vulnerability of N2 as a pathophysiological marker of insomnia, impaired REM sleep for subtyping major depressive disorder, and sleep stage transition from N2/N3 to REM as a biomarker of narcolepsy [30–32].Such evidence has been primarily obtained in PSG studies, but we expect that portable EEG devices will greatly contribute to future studies, owing to their good concordance with PSG, especially in discriminating between sleep stages, such as N2/3 and REM. In addition, the association between reduced REM sleep and increased risk of all-cause mortality in adults has recently drawn attention [33, 34]. Therefore, EEG-based portable devices, including the SS system, may facilitate public health promotion in the near future.

The current study has several limitations. First, we could not infer SS' ability to reliably measure an individual's sleep over multiple nights. Second, device synchronization with PSG where precision was within 6 s could decrease the agreement rate between them. Third, because fewer women than men participated in this study, caution is necessary when generalizing our findings to women. Fourth, while the usability of SS among a wide age range of individuals has been verified, the significant age difference between patient and healthy groups may have affected our results. Nevertheless, the significantly high concordance between SS and PSG outcomes suggested that the SS system is robust and can contribute to the development of new biomarkers and diagnostic or therapeutic techniques.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s41105-022-00421-5.

Acknowledgements The authors would like to thank Taeko Toyoda, Hiromi Mastuda, and Tomoko Yamada for their support in implementing the project, as well as Takashi Kanemura and Sachiko Sawada for their support in collecting the portable EEG and PSG recordings. We also thank Dr. Takashi Omori for his advice and support in conducting the statistical analysis. We would also like to thank Editage for English language editing.

Author contributions M.M., H.K., and K.K. designed the study protocol. A.K., M.K., and T.Y. encoded the data, participated in statistical analysis, interpreted the results, and wrote the article under the supervision of K.K. and N.Y. Y.O., Y.K., K.N., and M.T. recruited participants and conducted the study under the management of H.K. M.M. conducted a mathematical analysis. All authors read and approved the final article.

Funding This research was supported in part by the Japan Agency for Medical Research and Development (AMED: https://www.amed.go.jp/en/) under Grant Nos. 16hk0102041h0001, 21uk1024004h0001, and 21uk1024004s0201 and the Japan Society for the Promotion for Science (JSPS: https://www.jsps.go.jp/english/) KAKENHI under Grant Nos. 19K08016, 19H01047, and 16K17332. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Declarations

Conflict of interest The authors have declared that no competing interests exist.

Ethical committee permission This study was approved by the Ethics Committee of Shiga University of Medical Science (Approval No. 29-266).

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Informed consent This study was conducted in accordance with the Helsinki Declaration, and written informed consent was obtained from all participants.

References

1. Staba RJ, Stead M, Worrell GA. Electrophysiological biomarkers of epilepsy. Neurotherapeutics. 2014;11:334–46. https://doi.org/10.1007/s13311-014-0259-0.
2. Rosenow F, Klein KM, Hamer HM. Non-invasive EEG evaluation in epilepsy diagnosis. Expert Rev Neurother. 2015;15:425–44. https://doi.org/10.1586/14737175.2015.1025382.
3. Chokroverty S. Overview of sleep & sleep disorders. Indian J Med Res. 2010;131:126–40. https://doi.org/10.1016/B978-0-7506-9002-7.50007-0.
4. Agnew HW Jr, Webb WB, Williams RL. The first night effect: an EEG study of sleep. Psychophysiology. 1966;2:263–6. https://doi.org/10.1111/j.1469-8986.1966.tb02650.x.
5. Morgenthaler T, Alessi C, Friedman L, et al. Practice parameters for the use of actigraphy in the assessment of sleep and sleep disorders: an update for 2007. Sleep. 2007;30:519–29. https://doi.org/10.1093/sleep/30.4.519.
6. Weiss AR, Johnson NL, Berger NA, Redline S. Validity of activity-based devices to estimate sleep. J Clin Sleep Med. 2010;6:336–42. https://doi.org/10.5664/jcsm.27874.
7. Sadeh A. The role and validity of actigraphy in sleep medicine: an update. Sleep Med Rev. 2011;15:259–67. https://doi.org/10.1016/j.smrv.2010.10.001.
8. Meltzer LJ, Walsh CM, Traylor J, Westin AM. Direct comparison of two new actigraphs and polysomnography in children and adolescents. Sleep. 2012;35:159–66. https://doi.org/10.5665/sleep.1608.
9. Meltzer LJ, Hiruma LS, Avis K, et al. Comparison of a commercial accelerometer with polysomnography and actigraphy in children and adolescents. Sleep. 2015;38:1323–30. https://doi.org/10.5665/sleep.4918.
10. Johnstone SJ, Blackman R, Bruggemann JM. EEG from a single-channel dry-sensor recording device. Clin EEG Neurosci. 2012;43:112–20. https://doi.org/10.1177/1550059411435857.
11. Kaplan RF, Wang Y, Loparo KA, et al. Performance evaluation of an automated single-channel sleep–wake detection algorithm. Nat Sci Sleep. 2014;6:113–22. https://doi.org/10.2147/NSS.S71159.
12. Rogers JM, Johnstone SJ, Aminov A, et al. Test–retest reliability of a single-channel, wireless EEG system. Int J Psychophysiol. 2016;106:87–96. https://doi.org/10.1016/j.ijpsycho.2016.06.006.

13. Popovic D, Khoo M, Westbrook P. Automatic scoring of sleep stages and cortical arousals using two electrodes on the forehead: validation in healthy adults. J Sleep Res. 2014;23:211–21. https://doi.org/10.1111/jsr.12105.
14. Chen K, Zhang C, Ma J, et al. Sleep staging from single-channel EEG with multi-scale feature and contextual information. Sleep Breath. 2019;23:1159–67. https://doi.org/10.1007/s11325-019-01789-4.
15. Sleepwell C, Ltd. Sleep scope. Cited 2021 Nov 9 [Internet]. https://sleepwell.co.jp/sleepscope.
16. 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. https://doi.org/10.3389/fneur.2016.00110.
17. Soldatos CR, Dikeos DG, Paparrigopoulos TJ. Athens Insomnia Scale: validation of an instrument based on ICD-10 criteria. J Psychosom Res. 2000;48:555–60. https://doi.org/10.1016/s0022-3999(00)00095-7.
18. American Academy of Sleep Medicine. International classification of sleep disorders. 3rd ed. Darien: American Academy of Sleep Medicine; 2014.
19. Berry RB, Albertario CL, Harding SM, et al. The AASM manual for the scoring of sleep and associated events: rules, terminology and technical specifications. 2.5 version. Darien: American Academy of Sleep Medicine; 2018.
20. Japanese Standards Association. JIS T 1203:1998 Electroencephalographs. STANDARD by Japanese Industrial Standard; 1998.
21. Depner CM, Cheng PC, Devine JK, et al. Wearable technologies for developing sleep and circadian biomarkers: a summary of workshop discussions. Sleep. 2020;43:zsz254. https://doi.org/10.1093/sleep/zsz254.
22. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33:159–74. https://doi.org/10.2307/2529310.
23. Redline S, Sanders MH, Lind BK, et al. Methods for obtaining and analyzing unattended polysomnography data for a multicenter study. Sleep Heart Health Research Group. Sleep. 1998;21:759–67. https://doi.org/10.1093/sleep/21.7.759.
24. Ohayon MM, Carskadon MA, Guilleminault C, Vitiello MV. Meta-analysis of quantitative sleep parameters from childhood to old age in healthy individuals: developing normative sleep values across the human lifespan. Sleep. 2004;27:1255–73. https://doi.org/10.1093/sleep/27.7.1255.
25. De Gennaro L, Vecchio F, Ferrara M, et al. Changes in fronto-posterior functional coupling at sleep onset in humans. J Sleep Res. 2004;13:209–17. https://doi.org/10.1111/j.1365-2869.2004.00406.x.
26. Fernandez Guerrero A, Achermann P. Brain dynamics during the sleep onset transition: an EEG source localization study. Neurobiol Sleep Circadian Rhythms. 2019;6:24–34. https://doi.org/10.1016/j.nbscr.2018.11.001.
27. Benca RM, Obermeyer WH, Thisted RA, Gillin JC. Sleep and psychiatric disorders. A meta-analysis. Arch Gen Psychiatry. 1992;49:651–68. https://doi.org/10.1001/archpsyc.1992.01820080059010.
28. Hudson JI, Pope HG, Sullivan LE, et al. Good sleep, bad sleep: a meta-analysis of polysomnographic measures in insomnia, depression, and narcolepsy. Biol Psychiatry. 1992;32:958–75. https://doi.org/10.1016/0006-3223(92)90058-8.
29. Baglioni C, Nanovska S, Regen W, et al. Sleep and mental disorders: a meta-analysis of polysomnographic research. Psychol Bull. 2016;142:969–90. https://doi.org/10.1037/bul0000053.
30. Wei Y, Colombo MA, Ramautar JR, et al. Sleep stage transition dynamics reveal specific Stage 2 vulnerability in insomnia. Sleep. 2017. https://doi.org/10.1093/sleep/zsx117.
31. Hein M, Lanquart JP, Loas G, et al. Similar polysomnographic pattern in primary insomnia and major depression with objective insomnia: a sign of common pathophysiology? BMC Psychiatry. 2017;17:273. https://doi.org/10.1186/s12888-017-1438-4.
32. Liu Y, Zhang J, Lam V, et al. Altered sleep stage transitions of REM sleep: a novel and stable biomarker of narcolepsy. J Clin Sleep Med. 2015;11:885–94. https://doi.org/10.5664/jcsm.4940.
33. Zhang J, Jin X, Li R, et al. Influence of rapid eye movement sleep on all-cause mortality: a community-based cohort study. Aging (Albany, NY). 2019;11:1580–8. https://doi.org/10.18632/aging.101858.
34. Leary EB, Watson KT, Ancoli-Israel S, et al. Association of rapid eye movement sleep with mortality in middle-aged and older adults. JAMA Neurol. 2020;77:1241–51. https://doi.org/10.1001/jamaneurol.2020.2108.