UCLA
UCLA Previously Published Works

Title

Validation of a Wireless, Self-Application, Ambulatory Electroencephalographic Sleep Monitoring Device in Healthy Volunteers.

Permalink

Journal

ISSN

Authors

Finan, Patrick H
Richards, Jessica M
Gamaldo, Charlene E
et al.

Publication Date

DOI

Peer reviewed

JCSM
Journal of Clinical Sleep Medicine

SCIENTIFIC INVESTIGATIONS

Validation of a Wireless, Self-Application, Ambulatory Electroencephalographic Sleep Monitoring Device in Healthy Volunteers

Patrick H. Finan, PhD[1]; Jessica M. Richards[2]; Charlene E. Gamaldo[1]; Dingfen Han[1]; Jeannie Marie Leoutsakos[1]; Rachel Salas[1]; Michael R. Irwin[3]; Michael T. Smith[1]

[1]Johns Hopkins University School of Medicine, Baltimore, MD; [2]The Sandra and Malcolm Berman Brain & Spine Institute, Baltimore, MD; [3]Cousins Center for Psychoneuroimmunology, UCLA, Los Angeles, CA

Study Objectives: To evaluate the validity of an ambulatory electroencephalographic (EEG) monitor for the estimation of sleep continuity and architecture in healthy adults.

Methods: Healthy, good sleeping participants (n = 14) were fit with both an ambulatory EEG monitor (Sleep Profiler) and a full polysomnography (PSG) montage. EEG recordings were gathered from both devices on the same night, during which sleep was permitted uninterrupted for eight hours. The study was set in an inpatient clinical research suite. PSG and Sleep Profiler records were scored by a neurologist board certified in sleep medicine, blinded to record identification. Agreement between the scored PSG record, the physician-scored Sleep Profiler record, and the Sleep Profiler record scored by an automatic algorithm was evaluated for each sleep stage, with the PSG record serving as the reference.

Results: Results indicated strong percent agreement across stages. Kappa was strongest for Stage N3 and REM. Specificity was high for all stages; sensitivity was low for Wake and Stage N1, and high for Stage N2, Stage N3, and REM. Agreement indices improved for the manually scored Sleep Profiler record relative to the autoscore record.

Conclusions: Overall, the Sleep Profiler yields an EEG record with comparable sleep architecture estimates to PSG. Future studies should evaluate agreement between devices with a clinical sample that has greater periods of wake in order to better understand utility of this device for estimating sleep continuity indices, such as sleep onset latency and wake after sleep onset.

Keywords: sleep, ambulatory EEG, validation, polysomnography

Citation: Finan PH, Richards JM, Gamaldo CE, Han D, Leoutsakos JM, Salas R, Irwin MR, Smith MT. Validation of a wireless, self-application, ambulatory electroencephalographic sleep monitoring device in healthy volunteers. *J Clin Sleep Med* 2016;12(11):1443–1451.

INTRODUCTION

Polysomnography (PSG) is the gold standard assessment of sleep continuity and architecture, which are derived from continuous recordings of multiple scalp electroencephalographic (EEG) electrodes, electroculographic (EOG) recordings of eye movements, and electromyographic (EMG) recordings of muscle activity. Historically PSG has been conducted within an in-lab sleep facility. It is a costly, labor intensive procedure, requiring skilled technicians to apply and monitor, which can be obtrusive and alter sleep architecture and continuity. This intensive "brick and mortar" procedure has caused financial and resource challenges for both academic and clinical sleep programs.[1] Furthermore, because patients must sleep outside of their home environment, at least one adaptation night is typically required prior to the testing night, further increasing patient burden and expense.[2] These factors have limited the use of PSG for essential clinical and research purposes, despite its utility in objectively quantifying sleep.

Ambulatory PSG offers a reasonable alternative to objectively monitor sleep in the comfort of a patient's home.[3] However, it requires that a sleep technician place a full montage of scalp and body electrodes and, thus, suffers from similar resource-related limitations as PSG. Given these limitations,

BRIEF SUMMARY

Current Knowledge/Study Rationale: This study was conducted to evaluate the validity of a wireless self-application electroencephalographic device for the assessment of sleep architecture and continuity. For this validation study, we compared the Sleep Profiler to the gold standard polysomnography.

Study Impact: This study impacts the field by demonstrating that an ambulatory sleep EEG device can be used to assess sleep architecture and continuity with acceptable comparability to the gold standard laboratory procedure, polysomnography. As the field moves away from brick-and-mortar sleep labs and toward ambulatory assessment, a self-application device like the Sleep Profiler holds promise for both research and clinical care.

there is considerable interest in valid sleep EEG devices that require fewer resources than PSG or ambulatory PSG.

The Sleep Profiler (Advanced Brain Monitoring, Inc.) is an ambulatory, self-application sleep EEG device developed for patients to use in their home sleep environment without the assistance of a professional technician. The Sleep Profiler runs on a continuous charge for approximately 16 hours and provides 3 channels of frontal EEG, a pulse rate, and a sensor to detect head movement. The device is light weight (2.5 oz) and utilizes three proprietary, pre-gelled, self-adhesive, snap electrodes.

The device can be comfortably applied and strapped to the forehead in 5 minutes and features a built in impedance check, which prompts the user via interactive voice response to reaffix the electrodes when impedance is high. The accompanying software automatically scores the resulting EEG record using an algorithm that spectrally decomposes the EEG signal, computes descriptors of sleep macro- and microstructure, and categorizes each epoch as one of the 5 sleep stages.[4] The software is also designed to facilitate manual override of the autoscored epochs if, for example, a scorer disagrees with a particular stage determination on review. The automatic scoring algorithm that the Sleep Profiler software employs has been validated using EEG signals acquired from PSG.[4] However, to date, the Sleep Profiler device has not been directly validated against PSG. To our knowledge, no other self-application ambulatory sleep EEG monitor has been validated against PSG, leaving open the question of the clinical and research utility of this generation of devices.

The goal of the present study was to compare sleep continuity and architecture between PSG and Sleep Profiler records collected on the same night. For this initial validity study, we recruited a sample of healthy, good sleeping adults and evaluated differences in both aggregated sleep continuity parameters (e.g., sleep onset latency [SOL]; wake after sleep onset [WASO]; total sleep time [TST]; sleep efficiency [SE]) and the degree of epoch-by-epoch sleep stage concordance between PSG, the autoscored Sleep Profiler record, and a manually scored Sleep Profiler record.

METHODS

Participants

Healthy participants (n = 14) were recruited from and conjointly participated in a parent research project studying the mechanisms of sleep disruption hyperalgesia (R01 DA032922, MTS, MRI).

Eligibility was assessed first with a phone screen and subsequently through three screening visits. A clinical interview[5] was conducted to rule out the full range of sleep disorders with provisional diagnoses for sleep disorders, such as sleep disordered breathing, requiring PSG determination. Sleep diaries and actigraphy were obtained to verify that participants were sleeping on average within a window of 21:00 to 10:00 one week prior to the study. A history and physical was obtained by a physician or nurse practitioner. Eligible participants were scheduled for an overnight sleep study, which formed the basis for the present study, in the Johns Hopkins Bayview Medical Center, Clinical Research Unit (CRU).

Participants were included if they were healthy, between 18 and 48 years of age, met Research Diagnostic Criteria for normal sleepers,[6] a non-smoker/nicotine user, reported low levels of caffeine intake (≤ 2 cups of coffee or equivalent per day), reported a stable sleep phase between 21:00 and 10:00, scored < 5 on the Pittsburgh Sleep Quality Index Total Score, averaged between 6.5 and 8.5 h/night TST (assessed via one week of diary and actigraphy, demonstrated a sleep efficiency ≥ 85%), and scored < 10 on the Epworth Sleepiness Scale.

Participants were excluded if they had a BMI ≥ 35; a history of chronic pain (≥ 6 months); acute pain; significant medical/psychiatric morbidity within the last 6 months; lifetime history of bipolar disorder, psychotic disorder, recurrent major depression, posttraumatic stress disorder, traumatic brain injury, or seizures; significant symptoms of psychological distress; respiratory, hepatic, renal, or cardiac conditions that would contraindicate opioid administration; lifetime history of substance abuse or dependence, including alcohol; lifetime opioid use > 36 doses or > 7 days consecutive use; prior adverse reactions to general anesthetics/opioids or capsaicin; clinically significant abnormal complete blood count or comprehensive metabolic profile; positive urine toxicology screen for recreational drugs (THC), stimulants, opioids, or benzodiazepines; psychotropic medications; were pregnant or lactating.

The mean age was 26.43 (SD 3.74; range: 22–34). Eight participants identified as female and six identified as male. The self-identified racial and ethnic composition of the sample was as follows: 50% White (n = 7), 29% African American (n = 4), 14% Asian (n = 2), and 7% mixed race (n = 1). All participants reported some college, with 12 achieving a bachelor's degree or higher.

Measures

Polysomnography (PSG)

One of two registered PSG technicians (RPSGT) applied a standard recording montage recommended by the American Academy of Sleep Medicine guidelines.[7] Six referential EEG leads were placed: 2 frontal (F3-M2 and F4-M1), 2 central (C3-M2 and C4-M1), and 2 occipital (O1-M2 and O2-M1). In addition, standard EOG (E1-M2, E2-M1), 3 bipolar mentalis EMGs, ECG, and 2 bipolar tibial EMGs were also placed. Respiratory function and effort were measured via oronasal thermistor, nasal air pressure transducer, pulse oximetry, and abdominal and thoracic plethysmography belts. Electrophysiologic signals were acquired using an Embla N7000 polygraph using standard sampling rates (EEG, EOG, EMG, 500 Hz). A routine biocalibration procedure was performed to ensure signal integrity and facilitate scoring, Impedances were required to be below 5 K Ω.

Records were scored according to AASM guidelines[8] by a board certified sleep neurologist (CEG).

Sleep Profiler

The X4 Sleep Profiler was used for the present study. Technical details regarding the dynamic ranges, sampling, resolution, processing, and filtering are publically available on the company's website (https://cportal.b-alert.com/sleep-profiler/documentshelp). The EEG signal was acquired from the differential recording between AF7 and AF8. The signals were acquired at 256 samples per second, with a gain of ± 1000 μV, with a high pass filter of 0.1 Hz and low pass filter of 80 Hz. Head movement and position are acquired through a triaxial accelerometer. In addition, the device has the capability to attach a chin EMG lead for assistance in scoring. We elected not to use the chin lead in the present study in order to reduce subject burden.

Figure 1—Sensor strip electrode configuration for the Sleep Profiler.
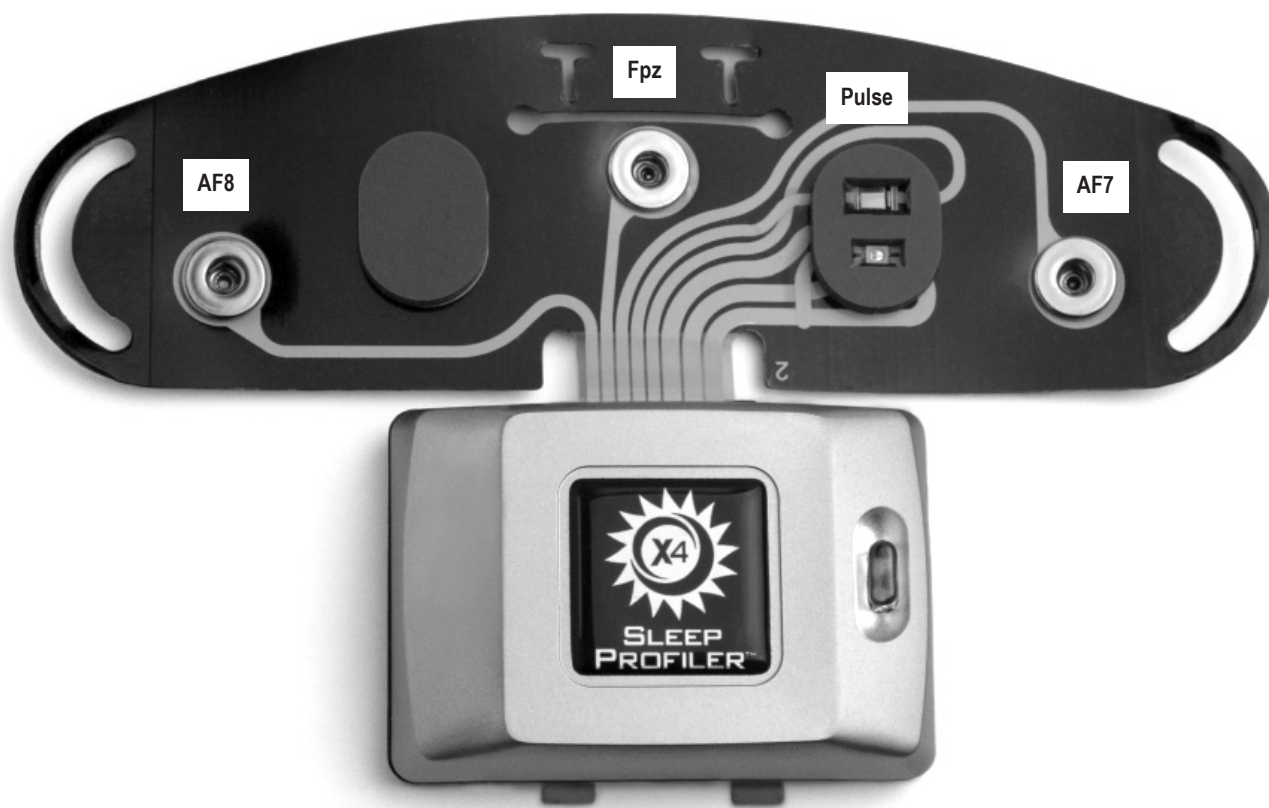
This graphic of the Sleep Profiler X4 is used with permission from Advanced Brain Monitoring, Inc.

The Sleep Profiler was applied to the forehead by the RPSGT after the standard PSG montage was affixed, according to manufacturer instructions. A picture of the Sleep Profiler electrode configuration is presented in **Figure 1**. Impedance checks were automatically initiated when the device was powered on and entered into record mode. If impedances were unacceptably high, a voice response was initiated asking the subject to recalibrate. These features are standard on the Sleep Profiler, and not specific to our study design.

After each record was acquired (procedure described below), it was uploaded onto the Sleep Profiler Portal, an internet-based software application. In the Sleep Profiler Portal, each record was scored by the automatic staging algorithm, and subsequently edited by 2 scorers. The automatic scoring algorithm (SP Auto) was guided by general staging rules consistent with the AASM scoring rules,[8] with the following exceptions: (1) Four consecutive non-wake epochs are necessary to identify sleep onset in the first 5 minutes of recording; (2) If a wake period is identified within a block of REM epochs, the first subsequent epoch is staged N1; (3) If an arousal occurs within an epoch in which the majority is characterized by a particular stage, the next lighter stage is automatically staged; (4) N2 stages are marked as "Light N2" if elevated alpha activity is observed, elevated baseline EMG, or EMG excursion is observed, or a cortical arousal is observed. For the purposes of this study, all "Light N2" epochs were subsequently recoded to N2 to permit comparison with PSG, which does not include a "Light N2" designation.

The manual scoring process (SP Manual) required a scorer to determine whether she agreed with the SP Auto staging based on visual inspection. In the event of a disagreement, the scorer changed the SP Auto stage to the stage she determined to be most appropriate. All other epochs were left unchanged. The primary SP Manual scorer was the same board certified sleep neurologist (CEG) who scored the PSG records. All Sleep Profiler and PSG records were de-identified, though the subject identification number was embedded in each of the records. To mitigate the possibility of scorer bias, the primary scorer first scored the Sleep Profiler records in a batch in ascending order by date. Approximately 3 months later, the primary scorer was provided with PSG records to score in a batch in ascending order by ID (which was in a different order than the date). The Sleep Profiler and PSG records were housed on completely separate servers, and the scoring was conducted on distinct software platforms that could not communicate with each other. At no point in scoring either set of records did the primary scorer have access to the other set of records. The secondary scorer (JMR) was a licensed clinical psychologist with training in PSG scoring. The primary scorer's records were used for direct comparison with PSG; the secondary scorer's records were only used for interrater reliability.

Procedure

Data for the present study were collected exclusively during the PSG sleep disorders screening/adaptation night (i.e., the

Table 1—Means, standard deviations, and correlations between PSG, SP Auto, and SP Manual on sleep continuity and sleep architecture.

Sleep Parameter	PSG Mean (SD)	SP Auto Mean (SD)	SP Manual Mean (SD)	PSG vs. SP Auto rho	PSG vs. SP Manual rho
SOL	8.04 (7.84)	5.00 (5.52)*	4.82 (6.18)*	0.59*	0.45
WASO	26.86 (22.98)	26.96 (13.30)	19.14 (10.89)	0.39	0.24
TST	445.36 (28.03)	439.96 (13.75)	448.29 (14.09)	0.43	0.26
SE	0.93 (0.06)	0.93 (0.03)	0.95 (0.03)	0.58*	0.41
N1	12.43 (5.56)	46.57 (29.75)***	27.21 (11.64)***	0.13	0.17
N2	225.36 (35.51)	200.96 (27.57)	206.93 (43.11)	0.31	0.39
N3	107.32 (38.61)	78.14 (47.53)**	97.14 (46.80)	0.64**	0.72***
REM	100.25 (24.31)	114.29 (62.40)	117.00 (44.64)	0.21	0.53*

Means are presented across all epochs and participants. Means were compared with paired t-test versus PSG: *p < 0.05; **p < 0.01; ***p < 0.001. SD, standard deviation

Table 2—Agreement analysis of PSG versus SP Auto.

Stage	Agreement	κ (95% CI)	Sensitivity % (95% CI)	Specificity % (95% CI)
Wake	0.92	0.42 (0.30, 0.53)	44.24 (41.07, 47.44)	96.17 (95.82, 96.50)
N1	0.89	0.07 (0.04, 0.10)	27.03 (22.33, 32.14)	90.58 (90.06, 91.07)
N2	0.78	0.55 (0.47, 0.64)	71.83 (70.69, 72.95)	83.34 (82.45, 84.21)
N3	0.88	0.62 (0.57, 0.67)	60.28 (58.49, 62.05)	96.00 (95.60, 96.37)
REM	0.85	0.58 (0.45, 0.70)	72.44 (70.73, 74.10)	88.52 (87.89, 89.13)

Agreement, κ, sensitivity, and specificity, along with 95% confidence intervals, are presented for each stage for the comparison of average of the scored polysomnography records to the average of the autoscore of the Sleep Profiler.

first night of the parent sleep deprivation research project [R01 DA032922, MTS/MRI]). Participants were first fit with the PSG montage by an RPSGT at the CRU. After the PSG montage was successfully applied, the RPSGT assisted the participant in applying the Sleep Profiler to the forehead according to manufacture instructions (https://advancedbrainmonitoring.app.box.com/s/7qmxxxa80nnxp5w2lbkg). The lower end of middle of the 3 sensors is placed along the nasion reference midline just above the top of the eyebrows. Although the Sleep Profiler is designed to be self-applied without the assistance of a technician, for the purposes of this study we elected to have the RPSGT assist in the Sleep Profiler application in order to ensure that the PSG montage was not altered by the application of the Sleep Profiler. Subject were permitted to sleep undisturbed in a private room designed for sleep research studies.

Lights out was at 23:00, at which time both the Sleep Profiler and PSG recordings were initiated. *Lights on* was 07:00, at which time both devices were removed. If a subject awoke prior to 07:00, they were instructed to remain in bed. The standardization of lights out and lights on time across participants was a necessary component of the parent study, which included time-sensitive daytime measures requiring participants to be awake by 07:00.

Although we attempted to initiate recording at the same time between devices, exact initial synchronization was not possible because the 2 devices were not designed to communicate with each other and because it was not possible to exactly calibrate the internal clocks of the 2 devices a priori. In order to precisely co-register the 30-s epochs between the Sleep

Profiler and PSG, the first scored epoch of PSG was used as the "start time benchmark." Two raters (PHF and JMR) separately reviewed the physiological signals between the PSG and Sleep Profiler epoch-by-epoch until a clear physiological concordance was identified (e.g., a sleep spindle or K complex, frank awakening was observed). Epochs were then aligned between records and the "start time" and "end time" of the Sleep Profiler were synchronized accordingly to match those of the PSG.

Data Analysis

Three records were compared on an epoch-by-epoch basis per subject: PSG; Sleep Profiler Autoscore (SP Auto); Sleep Profiler Manual Score (SP Manual). Sleep Profiler epochs identified as invalid (e.g., high impedance) were discarded. We present two primary sets of comparisons: PSG vs. SP Auto; PSG vs. SP Manual. In both comparisons, the PSG record serves as the reference standard. For interrater reliability, a comparison between SP Manual Scorer 1 vs. SP Manual Scorer 2 is also presented.

Data are presented in **Tables 1–5**, aggregated across epochs and study participants. For each sleep stage (Wake, Stage N1, Stage N2, Stage N3, and rapid eye movement [REM]), percent agreement, kappa, sensitivity, and specificity are presented. Percent agreement is calculated as the percent of all epochs within each sleep stage in which the Sleep Profiler (auto or manual) identifies the same sleep stage. Kappa (κ)[9] provides an estimate for agreement correcting for the chance probability of agreement. Sensitivity is defined as the probability of the Sleep Profiler (auto or manual) to positively identify a sleep stage

Table 3—Stage by stage percent of SP Auto misidentifications.

SP Auto		PSG				
		Wake	N1	N2	N3	REM
Dichotomized % Agreement		*92*	*89*	*78*	*88*	*85*
	Wake	44.24	15.62	3.40	3.05	4.20
	N1	31.67	27.02	6.31	1.73	16.88
Cross Tabulation	N2	7.58	35.74	71.84	27.27	6.19
	N3	1.66	2.10	6.12	60.28	0.29
	REM	14.85	19.52	12.34	7.67	72.44

"Dichotomized % agreement" is the percentage of epochs where both PSG and SP Auto rate the stage as present, or where both PSG and SP Auto agree that the stage is not present. The "Cross Tabulation" provides the percentage of epochs of each stage rated by SP Auto when PSG rated a given stage.

Table 4—Agreement analysis of PSG versus SP Manual.

Stage	Agreement	κ (95% CI)	Sensitivity % (95% CI)	Specificity % (95% CI)
Wake	0.93	0.44 (0.32, 0.56)	40.08 (36.97, 43.26)	97.68 (97.39, 97.93)
N1	0.93	0.13 (0.10, 0.17)	27.03 (22.33, 32.14)	94.79 (94.39, 95.16)
N2	0.81	0.61 (0.52, 0.70)	75.91 (74.83, 76.97)	84.63 (83.77, 85.47)
N3	0.90	0.71 (0.67, 0.74)	73.92 (72.30, 75.50)	94.74 (94.29, 95.17)
REM	0.90	0.73 (0.65, 0.80)	86.03 (84.69, 87.31)	91.41 (90.86, 91.94)

Agreement, κ, sensitivity, and specificity, along with 95% confidence intervals, are presented for each stage for the comparison of average of the scored polysomnography records to the average of the primary manual score of the Sleep Profiler.

Table 5—Stage by stage percent of SP Manual misidentifications.

SP Manual		PSG				
		Wake	N1	N2	N3	REM
Dichotomized % Agreement		*93*	*93*	*81*	*90*	*90*
	Wake	40.08	13.21	2.24	1.83	1.74
	N1	27.52	27.03	3.94	0.78	5.03
Cross Tabulation	N2	8.83	37.84	75.90	22.99	6.80
	N3	2.39	3.30	7.97	73.93	0.40
	REM	21.18	18.62	9.95	0.47	85.12

"Dichotomized % agreement" is the percentage of epochs where both PSG and SP Manual rate the stage as present, or where both PSG and SP Manual agree that the stage is not present. The "Cross Tabulation" provides the percentage of epochs of each stage rated by SP Manual when PSG rated a given stage.

when it is present on PSG. Specificity is defined as the probability of the Sleep Profiler (auto or manual) to identify when a sleep stage is not present on PSG.

All analyses were conducted in STATA version 12.1 (Stata Corp LP, 2011) and R version 3.1.3 (R Foundation for Statistical Computing, 2015).

RESULTS

PSG Versus Sleep Profiler Autoscore

Table 1 provides means and standard deviations, and significance tests, between PSG and the Sleep Profiler on measures of sleep continuity and sleep architecture aggregated for all epochs across the night; 1.67% of all epochs across participants were identified as invalid (range: 0.83–5.60%). Overall, the data from **Table 1** indicate that the Sleep Profiler estimates comparable levels of sleep continuity relative to PSG on

aggregate, but overestimates N1, likely leading to differences in Wake estimation that may have contributed to the underestimation of sleep onset latency, and the variable correlations observed between devices.

Table 2 provides results for the comparisons of PSG to SP Auto within each sleep stage. Notably, agreement was strong across stages, but κ substantially varied from stage to stage. The strongest κ values were for Stage 3 (slow wave sleep) and REM (κ = 0.62 and κ = 0.58, respectively). The poor κ values for Wake and Stage N1 were accompanied by poor sensitivity, which was likely influenced by the low overall base rate of Wake and Stage N1 epochs (7.3% and 2.6%, respectively). Specificity was strong across stages (83% to 96%).

Table 3 shows the percent of epochs of each stage type erroneously identified by SP Auto as a function of PSG stage. For example, when PSG identified an epoch as Wake, SP Auto erroneously identified it as Stage N1 31.67% of the time. PSG Stage N1 was most commonly misidentified by SP Auto as Stage N2

Figure 2—Subject by subject visualization of sleep architecture across modalities.
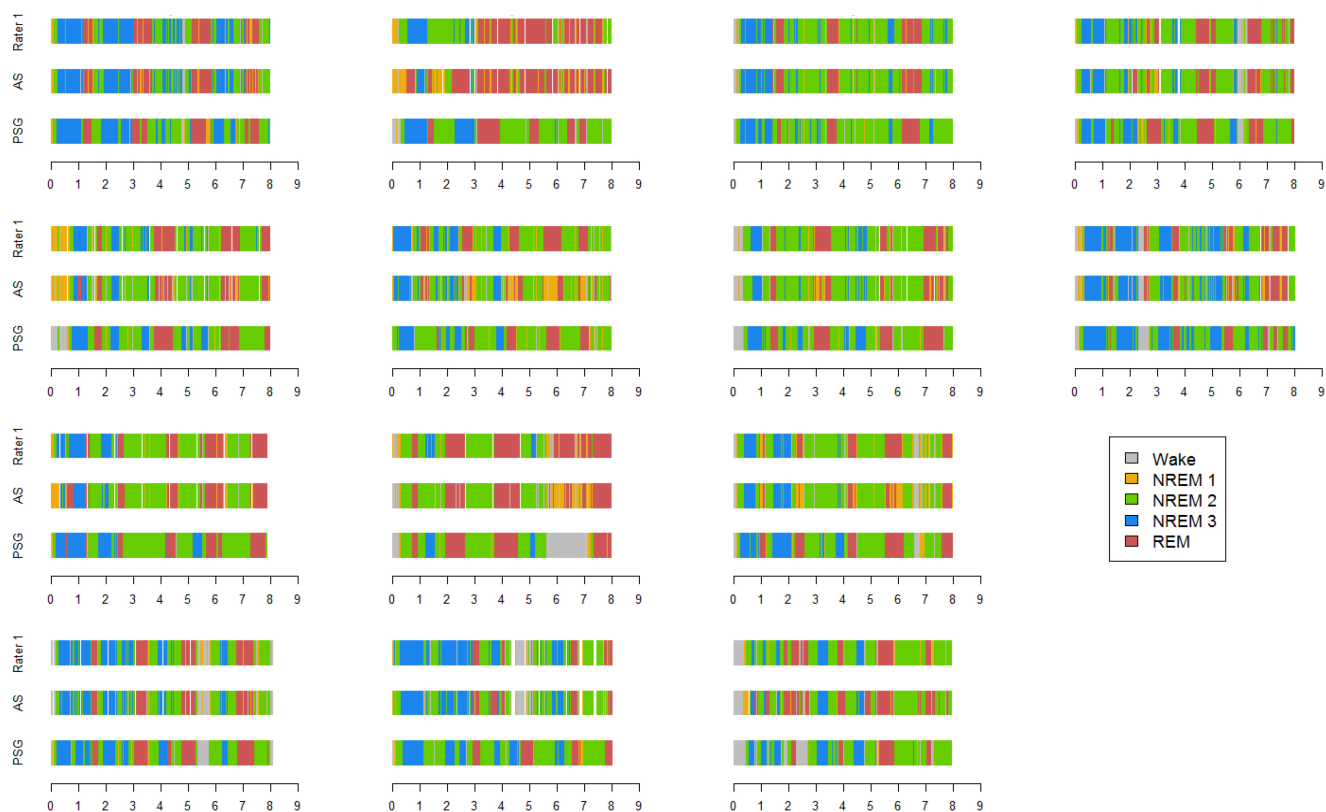
Sleep stages are color coded for visual comparison across modalities. PSG, polysomnography; AS, SP Auto; Rater 1, SP Manual.

(35.74%). PSG Stage N2 was most commonly misidentified by SP Auto as REM (12.34%). PSG Stage N3 was most commonly misidentified by SP Auto as Stage N2 (27.27%). Finally, PSG REM was most commonly misidentified by SP Auto as Stage N1 (16.88%).

Overall, the data indicate that there is strong agreement between PSG and the autoscore feature of the Sleep Profiler, and that sensitivity (and κ) is heavily influenced by the number of epochs within a given stage identified by PSG.

PSG Versus Sleep Profiler Manual Score

Table 4 shows the results of the comparisons of PSG to SP Manual within each stage. Overall, the results paralleled those from the PSG/SP Auto comparison, with increases in agreement, κ, sensitivity, and specificity in most of the stages. Notably, the biggest improvements in κ from SP Auto to SP Manual were observed for Stage N3 and REM, which reached a conventionally accepted definition of "good" agreement[10] in the PSG vs. SP Manual comparison (κ = 0.71 and κ = 0.73, respectively).

Table 5 shows the percent of misidentified epochs as a function of each stage of SP Manual. PSG Wake was most commonly misidentified by SP Manual as Stage N1 (27.52%). PSG Stage N1 was most commonly misidentified by SP Manual as Stage N2 (37.84%). PSG Stage N2 was most commonly misidentified by SP Manual as REM (9.95%). PSG Stage N3 was most commonly misidentified by SP Manual as Stage N2

(22.99%). Finally, PSG REM was most commonly misidentified by SP Manual as Stage N2 (6.80%).

In sum, the SP Manual provided a stronger level of agreement with PSG than SP Auto, though the results were largely parallel in form. **Figure 2** provides a subject-by-subject visualization of each stage across the 3 measurement modalities throughout the entire 8-h sleep opportunity period. The graphs comprising **Figure 2** largely support the agreement statistics, with the exception of one subject who appears to be an outlier with a systematic overestimation of REM in both the SP Auto and SP Manual records.

Interrater Reliability

Table 6 presents interrater reliability data. Across stages, the raw agreement percentages between our primary rater and secondary rater were high (agreement ≥ 88%). Other than Stage N1 (κ = 0.48), the κ values were all "good" or "excellent" according to published conventions.[10]

DISCUSSION

The present study compared sleep architecture acquired by the Sleep Profiler, a self-application ambulatory sleep EEG device, to PSG. The data largely support the validity of the Sleep Profiler in estimating sleep stages. The Sleep Profiler software includes a default to automatically stage the raw

Table 6—Sleep Profiler Interrater Reliability.

Stage	Agreement	κ (95% CI)	Sensitivity % (95% CI)	Specificity % (95% CI)
Wake	0.98	0.75 (0.69, 0.81)	70.80 (67.02, 74.37)	99.23 (99.06, 99.37)
N1	0.93	0.48 (0.39, 0.57)	63.18 (59.58, 66.67)	95.10 (94.71, 95.48)
N2	0.88	0.77 (0.70, 0.84)	91.61 (90.85, 92.31)	85.98 (85.17, 86.77)
N3	0.94	0.81 (0.77, 0.84)	76.67 (75.03, 78.25)	98.64 (98.40, 98.85)
REM	0.93	0.80 (0.70, 0.90)	80.77 (79.37, 82.10)	96.67 (96.30, 97.02)

Agreement, κ, sensitivity, and specificity, along with 95% confidence intervals, are presented for each stage for the comparison of Sleep Profiler Scorer 1 to Sleep Profiler Scorer 2.

EEG files, which can then be manually edited as necessary by a user. Interrater reliability for manual scoring was found to be good. We compared both the autoscored and the manually scored records to PSG acquired in the same participants on the same nights. Percent agreement was generally excellent in both autoscored and manually scored records. The κ statistic, which corrects for chance agreement, was attenuated relative to percent agreement in both sets of Sleep Profiler records, but was strongest for the estimation of slow wave sleep (Stage 3) and REM sleep in the manually scored Sleep Profiler record.

The performance of the Sleep Profiler in identifying Wake epochs was not consistent across analyses. One metric, percent agreement, indicated excellent overlap (> 90%) between the Sleep Profiler (both SP Auto and SP Manual records) and PSG on the identification of Wake epochs. Specificity was also quite high for Sleep Profiler identification of Wake. In contrast, the κ statistic and sensitivity were markedly lower for Wake than the Stage N2, N3, and REM comparisons. Taken as a whole, the data suggest that the Sleep Profiler has moderate difficulty identifying a wake epoch when one is present, but is quite accurate in the wake epochs it does identify.

However, interpretations about the performance of Sleep Profiler in identifying Wake should be balanced against several considerations. First, the sample consisted of healthy, good sleepers, and consequently produced very few wake epochs from which to draw comparisons. The dearth of wake epochs in this sample likely biased the κ estimate and measure of sensitivity downward, as the penalties for chance agreement and failing to identify an epoch are higher when there are relatively few true positives. Second, it is possible that PSG may have inaccurately estimated the number of "true positives" for those stages. Although PSG is our "gold standard" comparator, interrater reliability on PSG is lowest for the Wake and Stage N1 stages[11] due to the similarity of the wake alpha to both theta in Stage N1 and high frequency alpha in REM. Indeed, PSG Wake was most commonly misidentified by the Sleep Profiler as Stage N1 or REM. Furthermore, **Figure 2** demonstrates that the Sleep Profiler had difficulty tracking the PSG in the few instances of wake after sleep onset into and out of REM periods. Whereas PSG offers several features for differentiating REM from wake, such as arousals and ocular movements, the Sleep Profiler relies primarily on the power of each stage's frequency band. Because we did not use the chin EMG lead, we may have limited our ability to more accurately detect REM from Wake. Despite these mixed findings, minutes of TST and

WASO, which are reliant upon the differentiation of wake from sleep, were not significantly different between PSG and the Sleep Profiler. Taking these factors into consideration, we conclude that the Sleep Profiler provides a reasonable alternative to PSG to objectively assess sleep continuity, but it is clear that further research is necessary to determine if agreement would increase with a larger sample of Wake epochs.

The present findings suggest the Sleep Profiler, and potentially other self-application ambulatory sleep EEG devices of similar design, have promise for enhancing both research and clinical practice. It is of great value to acquire valid and reliable objective sleep data that can be recorded from an individual's home environment without the assistance of a technician. Although the hardware and software capabilities are evolving, the device we tested in the present study had the capacity to store two nights' worth of data. Current hardware and software features permit recording up to seven nights of data on a single device before uploading to a storage site for scoring. Based on the results of the present study, the quality of the Sleep Profiler EEG signals from which estimates of sleep continuity and sleep architecture are generated are comparable to PSG, yet the Sleep Profiler has the capability to obtain multiple in-home objective sleep EEG recordings at a fraction of the cost of a single laboratory PSG study.

The ability to obtain consecutive nights of objective sleep data in one's home environment could open up numerous possibilities for research designs that heretofore have not been possible. For example, slow wave sleep and REM, which were both very well characterized by the Sleep Profiler, have been linked to depression[12,13] and are associated with individual differences in the experience of positive and negative emotions.[14–16] Although the laboratory studies that have generated those findings have been informative, it is not clear whether natural variation in slow wave sleep and REM predicts day-to-day changes in emotions and emotional reactivity to natural stressors, and vice versa. Ambulatory sleep EEG provides the opportunity to evaluate such questions prospectively.

In the present study, we showed that the Sleep Profiler was comparable to PSG on estimates of sleep continuity. Future research should compare actigraphy versus sleep profiler estimates of sleep continuity, particularly in patients with insomnia, for whom actigraphy has been found to overestimate sleep versus wake.[17] The Sleep Profiler has several advantages over actigraphy, including lack of reliance on self-report for accurate assessment of sleep onset, and its ability to stage

sleep—particularly REM and slow wave sleep. From the standpoint of total resources required for administration, the Sleep Profiler may prove less resource demanding than actigraphy, as it does not require concomitant completion of sleep diaries for validity. Future studies should directly evaluate usability and subject burden between these device classes.

This study had several limitations that should be addressed in future research. First, we only recruited healthy, good sleepers, which we felt was an important first step in validating basic sleep continuity and architecture. The downside of that decision is that because most participants slept without interruption throughout the night, we had a very small sample of Wake epochs from which to draw comparisons. Future studies should recruit clinical samples with prominent sleep maintenance problems to not only provide a better test of the Sleep Profiler's ability to detect wake, but also to evaluate the diagnostic validity of the Sleep Profiler relative to PSG. A second limitation is the lack of perfect synchrony in start times between the Sleep Profiler and PSG within-subject. Despite this limitation, we feel confident based on visual inspection of the records, as displayed in **Figure 2**, that we arrived at a reasonable concordance of start time after matching epochs for distinct physiological phenomena. Third, our sample size was relatively small, requiring replication in a larger sample in future research. Notably, there was a substantial logistical challenge in fitting the Sleep Profiler on the same head as a full montage of PSG electrodes. Nonetheless, the records we acquired were typically considered of adequate signal quality for staging by both the primary and secondary scorer. Third, we only conducted PSG versus Sleep Profiler comparisons on a single night, and therefore were unable to assess test-retest reliability. Fourth, all sleep studies were conducted on the adaptation night of a larger study. Although sleep associated with adaptation nights may not generalize to sleep in a "normal" state, that concern may be mitigated by the relatively low amount of wakefulness observed among the participants in this study. Finally, although the Sleep Profiler is designed to be self-applied, due to logistical concerns in the present study, we elected to have the RPSGT apply them. As such, the data recorded in the present study may not reflect data that would have been obtained in a real-world setting.

These limitations may be weighed against several notable strengths of this study. First, the within-subject, within-night epoch-by-epoch comparison of the Sleep Profiler and PSG provides the strongest possible test of one EEG device relative to another. Second, our sample was rigorously screened, well-characterized as healthy sleepers, and demographically diverse. Finally, we demonstrated strong interrater reliability of the manually scored Sleep Profiler records, supporting the integrity of our primary scorer's staging determinations.

CONCLUSIONS

Clinically, a validated ambulatory self-application sleep EEG device could substantively assist in the determination of diagnosis and treatment plan for a variety of sleep disorders, including obstructive sleep apnea and chronic insomnia by offering a low burden, relatively inexpensive way to assess sleep continuity and architecture from the comforts of a patient's home. The present data provide initial validity data supporting the use of the Sleep Profiler as a stand-alone sleep EEG device in healthy participants. Future studies will be needed to validate the Sleep Profiler in the evaluation of sleep disorders such as insomnia, hypersomnia, or obstructive sleep apnea. Although the autoscoring algorithm was reasonably concordant with PSG, manual editing of the autoscored files enhanced the level of agreement across stages. We defer making firm conclusions about the autoscore algorithm, however, because it is in a state of evolution, and future iterations of the algorithm may prove more accurate. In summary, the Sleep Profiler, an ambulatory sleep EEG device, has the potential to advance research and clinical practice by providing an opportunity to objectively assess sleep continuity and architecture over multiple nights in a patient's home environment.

ABBREVIATIONS

BMI, body mass index
CRU, clinical research unit
EEG, electroencephalographic
EMG, electromyographic
EOG, electrooculographic
N1, stage non-REM 1
N2, stage non-REM 2
N3, stage non-REM 3
PSG, polysomnography
REM, rapid eye movement
RPSGT, registered polysomnography technician
SE, sleep efficiency
SOL, sleep onset latency
SP Auto, Automatic Sleep Profiler Scoring
SP Manual, Manual Sleep Profiler Scoring
TST, total sleep time
WASO, wake after sleep onset

REFERENCES

1. Whitelaw WA, Brant RF, Flemons WW. Clinical usefulness of home oximetry compared with polysomnography for assessment of sleep apnea. Am J Respir Crit Care Med 2005;171:188–93.

2. Agnew HW, Webb WB, Williams RL. The first night effect: an EEG study of sleep. Psychophysiology 1966;2:263–6.

3. Collop, N. A., Anderson, W. M., Boehlecke, B., Claman, D., Goldberg, R., Gottlieb, D. J., Hudgel, D., Satteia, M. & Schwab, R. Clinical guidelines for the use of unattended portable monitors in the diagnosis of obstructive sleep apnea in adult patients. J Clin Sleep Med 2007;3:737–747.

4. Stepnowsky C, Levendowski D, Popovic D, Ayappa I, Rapoport DM. Scoring accuracy of automated sleep staging from a bipolar electroocular recording compared to manual scoring by multiple raters. Sleep Med 2013;14:1199–207.

5. Schramm E, Hohagen F, Grasshoff U, et al. Test-retest reliability and validity of the Structured Interview for Sleep Disorders according to DSM-III-R. Am J Psychiatry 1993;150:867–72.

6. Edinger JD, Bonnet MH, Bootzin RR, et al. Derivation of research diagnostic criteria for insomnia: report of an American Academy of Sleep Medicine Work Group. Sleep 2004;27:1567–96.

7. Iber C, Ancoli-Israel S, Chesson A, Quan S. The AASM manual for the scoring of sleep and associated events: rules, terminology and technical specifications, 1st ed. Westchester, IL: American Academy of Sleep Medicine, 2007.

8. Berry RB, Brooks R, Gamaldo CE, Hardling SM, Marcus CL, Vaughn BV. The AASM manual for the scoring of sleep and associated events: rules, terminology and technical specifications, version 2.0. www.aasmnet.org. Darien, IL: American Academy of Sleep Medicine, 2012.

9. Cohen J. Nominal scale agreement provision for scaled disagreement or partial credit. Psychol Bull 1960;70:213–20.

10. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159–74.

11. Rosenberg RS, Van Hout S. The American Academy of Sleep Medicine inter-scorer reliability program: sleep stage scoring. J Clin Sleep Med 2013;9:81–7.

12. Armitage R, Hoffmann R, Trivedi M, Rush AJ. Slow-wave activity in NREM sleep: sex and age effects in depressed outpatients and healthy controls. Psychiatry Res 2000;95:201–13.

13. Palagini L, Baglioni C, Ciapparelli A, Gemignani A, Riemann D. REM sleep dysregulation in depression: state of the art. Sleep Med Rev 2013;17:377–90.

14. Finan PH, Quartana PJ, Smith MT. The effects of sleep continuity disruption on positive mood and sleep architecture in healthy adults. Sleep 2015;38:1735–42.

15. Gujar N, Yoo SS, Hu P, Walker MP. Sleep deprivation amplifies reactivity of brain reward networks, biasing the appraisal of positive emotional experiences. J Neurosci 2011;31:4466–74.

16. Nishida M, Pearsall J, Buckner RL, Walker MP. REM sleep, prefrontal theta, and the consolidation of human emotional memory. Cereb Cortex 2009;19:1158–66.

17. Hauri PJ, Wisbey J. Wrist actigraphy in insomnia. Sleep 1992;15:293–301.

SUBMISSION & CORRESPONDENCE INFORMATION

DISCLOSURE STATEMENT