SCIENTIFIC INVESTIGATIONS

The Accuracy, Night-to-Night Variability, and Stability of Frontopolar Sleep Electroencephalography Biomarkers

Daniel J. Levendowski, MBA[1]; Luigi Ferini-Strambi, MD[2]; Charlene Gamaldo, MD[3]; Mindy Cetel, MD[4]; Robert Rosenberg, DO[5]; Philip R. Westbrook, MD[1]

[1]Advanced Brain Monitoring, Carlsbad, California; [2]Department of Clinical Neurosciences, San Raffaele Scientific Institute, Sleep Disorders Center, Università Vita-Salute San Raffaele, Milan, Italy; [3]Department of Neurology, Johns Hopkins University School of Medicine, Baltimore, Maryland; [4]Integrative Insomnia and Sleep Health Center, San Diego, California; [5]Sleep Disorders Center of Prescott Valley, Prescott Valley, Arizona

Study Objectives: To assess the validity of sleep architecture and sleep continuity biomarkers obtained from a portable, multichannel forehead electroencephalography (EEG) recorder.

Methods: Forty-seven subjects simultaneously underwent polysomnography (PSG) while wearing a multichannel frontopolar EEG recording device (Sleep Profiler). The PSG recordings independently staged by 5 registered polysomnographic technologists were compared for agreement with the autoscored sleep EEG before and after expert review. To assess the night-to-night variability and first night bias, 2 nights of self-applied, in-home EEG recordings obtained from a clinical cohort of 63 patients were used (41% with a diagnosis of insomnia/depression, 35% with insomnia/obstructive sleep apnea, and 17.5% with all three). The between-night stability of abnormal sleep biomarkers was determined by comparing each night's data to normative reference values.

Results: The mean overall interscorer agreements between the 5 technologists were 75.9%, and the mean kappa score was 0.70. After visual review, the mean kappa score between the autostaging and five raters was 0.67, and staging agreed with a majority of scorers in at least 80% of the epochs for all stages except stage N1. Sleep spindles, autonomic activation, and stage N3 exhibited the least between-night variability ($P < .0001$) and strongest between-night stability. Antihypertensive medications were found to have a significant effect on sleep quality biomarkers ($P < .02$).

Conclusions: A strong agreement was observed between the automated sleep staging and human-scored PSG. One night's recording appeared sufficient to characterize abnormal slow wave sleep, sleep spindle activity, and heart rate variability in patients, but a 2-night average improved the assessment of all other sleep biomarkers.

Commentary: Two commentaries on this article appear in this issue on pages 771 and 773.

Keywords: biomarkers, in-home, scoring, sleep architecture, sleep continuity, sleep EEG, validity, variability

Citation: Levendowski DJ, Ferini-Strambi L, Gamaldo C, Cetel M, Rosenberg R, Westbrook PR. The accuracy, night-to-night variability, and stability of frontopolar sleep electroencephalography biomarkers. *J Clin Sleep Med.* 2017;13(6):791–803.

INTRODUCTION

The use of cardiovascular, tissue, and blood biomarkers are quite common in clinical research and clinical practice in large part because these measures characterize and repeatedly demonstrate a context for the interpretation of clinical outcomes across treatments and conditions.[1] Mounting evidence has linked sleep phenomenology with the manifestation of a number of chronic diseases such as hypertension, heart disease, and diabetes.[2] Emerging evidence also suggests that sleep architecture and sleep continuity (ie, sleep biomarkers) may be beneficial in monitoring brain health in the setting of management of neurodegenerative disorders. For example, sleep spindle characteristics during non-rapid eye movement (NREM) sleep have been associated with cognitive decline in patients with Parkinson disease,[3–5] and reduced slow wave sleep has been associated with increased beta amyloid (directly linked to Alzheimer disease) concentrations in the cerebrospinal fluid.[6] Because sleep spindles and slow wave sleep are believed to be associated with the metabolic clearance systems of the brain,[7] it is now argued that the routine monitoring of change in these measures are useful in evaluating the risk for, or progression of neurodegeneration.[8]

BRIEF SUMMARY

Current Knowledge/Study Rationale: The accuracy of autostaged frontopolar electroencephalography (EEG) compared to human-scored polysomnography (PSG) in patients with sleep disorders has not been established. The repeatability of sleep biomarkers acquired in the home of patients with insomnia is not currently known.

Study Impact: This study demonstrates that autoscored multichannel frontopolar EEG is as accurate as human-staged PSG, and ranks sleep biomarkers by variability and stability. The results point toward the use of sleep biomarkers for directing patient care and for assessing the effect of pharmacological interventions on sleep quality.

Although sleep biomarkers are commonly used as outcome measures and covariables in clinical trials and clinical research,[2] these biomarkers are rarely used in clinical practice to predict outcomes or direct therapies that could improve morbidity and mortality. For example, the diagnosis of insomnia is commonly based solely on a patient's self-report. However, there is clear evidence of an objectively measured insomnia subtype (ie, insomnia with short sleep time) which has been associated with increased inflammatory state profiles and significant morbidity

and mortality.[9,10] Evidence also suggests that electroencephalography (EEG) based sleep time can be used to identify a "normal sleep duration" insomnia phenotype and positively predict those individuals most likely to respond to cognitive behavioral treatment intervention (the now recognized first-line therapy for insomnia).[11] Additional phenotypes are emerging through the use of sleep biomarkers related to depressive and anxiety disorders and/or stress-related insomnia.[12,13]

Critical cost-benefit analysis discussion remains paramount when the practitioner and patient decides if and what type of pharmacotherapeutic sleep agent is most appropriate for treating their respective sleep complaint.[14,15] Moreover, the United States Food and Drug Administration has continued to stress the importance of monitoring all classes of sleep aides due to concerns over safety and habituation[16,17] calling for the potential clinical value for repeated-measures dose-response EEG studies. Although classes of drugs are currently available to enhance slow wave sleep (ie, carbamazipine, gabapentin, and tiagabine),[2] the relative cost of acquiring the physiological data (primarily via in-laboratory polysomnography [PSG]) needed to objectively evaluate the long-term health benefits of drugs that deepen sleep, and to identify patients who may benefit from this type of intervention has been a limiting factor.

A prerequisite for establishing the validity and clinical utility of a sleep biomarker(s) is to demonstrate its accuracy and its reproducibility.[1] Two previous reports evaluated the accuracy of the Sleep Profiler device used in this study, a portable, self-applied, multichannel recorder that could potentially meet the requirements for simple and routine assessment of sleep architecture and sleep continuity (ie, sleep biomarkers). One study compared the agreement between the autoscoring before and after manual editing to human-scored PSG in healthy subjects.[18] A second study evaluated the agreement with PSG in a group that included elderly control patients for an Alzheimer disease investigation.[19] The first part of the current study was intended to demonstrate the validity of sleep biomarkers by first evaluating the accuracy based on the agreement in the autostaging from frontopolar sites to simultaneously acquired PSG in a group of patients mainly referred for probable obstructive sleep apnea. In the second part of this study, we evaluated the night-to-night consistency of sleep biomarkers identified by this device in a clinical population of insomnia patients vulnerable to a first-night effect. Variability assessments were conducted to identify those biomarkers, which provide a high degree of between-night stability such that a medication dose effect might be measurable in a single night. The stability of an abnormality across nights was used to identify the sleep biomarkers that may be useful in predicting a phenotypic trait, or in confirming a disease state.

METHODS

Biomarker Accuracy: Part 1

For this Institutional Review Board-approved prospective accuracy study, 65 subjects underwent laboratory PSG at Complete Sleep Solutions in Murrieta, California, United States using Alice 5 systems (Philips Respironics, Monroeville, Pennsylvania, United States) while simultaneous multichannel frontopolar EEG recordings were acquired with a forehead-worn recording device (Sleep Profiler, Advanced Brain Monitoring, Carlsbad, California, United States). The sleep laboratory study technician who conducted the study assisted the subject apply the recording device.

After the PSG biocalculations were performed, the recording device was turned on by the technician. Subjects were instructed to sit up in bed and slowly count to 10 so that the sound to be used for signal synchronization could be captured by both the PSG and the forehead EEG device followed by "lights out" and instructions to recline. The 2 sets of recordings were synchronized by first identifying the clock time associated with the start of a PSG epoch just prior to the 10-count. This clock time established the start of PSG recording time. The Sleep Profiler record was then cut so each epoch started at the same clock time as the PSG.

Forty-seven records from 35 males and 12 females (ages 23–77 years) with a minimum of 3 hours of PSG-based recording time and ≥ 85% good signal quality across all 3 frontopolar EEG channels were submitted for staging by 5 independent scorers. Eleven records did not meet the minimum recording time criteria (ie, split-night studies) and 7 had ≤ 85% good EEG quality. Forty records were from subjects referred for a PSG for an assessment of sleep-disordered breathing and the rest were from presumably healthy controls. Scorer 1 was a registered polysomnographic technologist (RPSGT) affiliated with the American Academy of Sleep Medicine (AASM)-accredited site where the PSG studies took place. Scorers 2 and 3 were independent RPSGTs, and scorers 4 and 5 were technologists specialized in sleep staging for research studies conducted at New York University. All of the PSG records were scored according to the AASM-recommended criteria. Scorer 1 staged sleep with a screen view that presented all of the PSG signals.[20] Scorers 2 through 5 staged sleep using only the EEG, electroocular (EOG), and electromyographic (EMG) signals (ie, blinded to the cardiorespiratory signals). The staged epoch was exported and pooled into a single cross-tabulation for all studies. Interrater comparisons between the 5 scorers were made using Cohen kappa scores and by- and across-sleep stage agreements. Each scorer's staging was compared to the autostaging, before and after expert editing. Accuracy was also assessed with comparisons between those epochs where a majority of scorers agreed with the autostaging, both before and after review.

Biomarker Variability and Stability: Part 2

A retrospective study was conducted on studies acquired between August 2013 and July 2016 from the Sleep Disorders Center of Prescott Valley (Prescott Valley, Arizona, United States) and the Integrative Insomnia and Sleep Health Center (San Diego, California, United States). Patients were briefly instructed on how to self-apply the Sleep Profiler based on the written instructions. With in-home, self-application, patients acquired 2 nights of recordings and completed a medical questionnaire and sleep diary. An internet-based portal was used to enter the questionnaire responses, process the studies, enable expert review of the studies, and export the variables used for the analyses.

Selected records required 2 nights of data with no more than a 3-hour difference between each night's recording time, and with at least 85% of the epochs staged with the EEG channel.

Records were further excluded based on missing entries in the medical questionnaire. These responses were used to characterize the cohort (**Table 1**) and to assess the effect of self-reported prescription sleeping aids and antidepressant medications on sleep architecture and sleep continuity. The responses were also used to select records that met the study's inclusion criteria of clinical insomnia. The sleep diary responses enabled confirmation that the use of prescription sleeping aids on the nights of the study matched the frequency of prescription sleep aids' use reported on the medical questionnaire.

Inclusion into the insomnia cohort required an Insomnia Severity Index (ISI) score ≥ 14 (n = 34),[21] administration of prescription sleep aids (n = 23), or having an ISI ≥ 12 with either a reported clinician diagnosis of insomnia (n = 4) or a concurrent diagnosis of depression and anxiety (n = 2). Sleep biomarkers were extracted from 63 patient records, after expert review. To reduce the influence of outliers on the first-night reliability analyses, 3 subjects were excluded from rapid eye movement (REM) latency analysis due to less than 3 minutes of REM across an entire night. For the analysis of autonomic activation, 16 subjects were excluded due to less than 75% of reliable pulse rate recordings in each night.

Sleep biomarkers obtained from each night and averaged across both nights were compared to clinical thresholds to identify the stability of abnormal sleep characteristics. When available, abnormality was defined by the 10th or 90th percentile age- and sex-matched reference values reported for the Sleep Heart Health Study.[22] For the remaining sleep measures, reference thresholds were obtained using the first or third interquartile cutoffs from a group of healthy controls. Inclusion in this healthy control group (n = 48) required not taking prescription sleeping aids or antidepressants, no diagnosis of obstructive sleep apnea, having an ISI score ≤ 12, and daytime somnolence (Epworth Sleepiness Scale), depression (Patient Health Questionnaire-9) and anxiety (Generalized Anxiety Disorder 7-Item) scores ≤ 10. The controls included records acquired by Washington University for the Knight Alzheimer's Disease Research Center (16 males and 18 females, ages 65 to 89 years) and by Advanced Brain Monitoring (9 males and 5 females, ages 24 to 44 years). Subjects 70 years or older were excluded from the interquartile range calculations for sleep spindle activity.

Intraclass correlations were used to compare the reliability of the first-night sleep quality measures. Bland-Altman plots were used to assess potential bias resulting from a first-night effect; *t* tests and Fisher exact tests were applied to the normally distributed/transformed variables.

Biomarker Measurement

Hardware

The Sleep Profiler used in this study was a battery-powered recorder designed to acquire 3 frontopolar EEG signals between AF7-AF8, AF7-Fpz, and AF8-Fpz (**Figure 1**). The EEG signals were sampled at 256 Hz with a gain of ± 1000 µV, and filtered with a 0.1-Hz high-pass and 80-Hz low-pass filter. The scalp/electrode impedances were obtained at each sensor site at the

Table 1—Characteristics of the night-to-night variability study patient cohort.

Demographic	
Age, years, mean ± SD	60.7 ± 14.7
Females, % (n)	62% (39)
BMI, kg/m², mean ± SD	27.3 ± 5.7
Questionnaire	
Epworth Sleepiness Scale score, mean ± SD	6.0 ± 4.9
Insomnia Severity Index, mean ± SD	18.4 ± 4.7
Depression score (PHQ-9), mean ± SD	10.0 ± 5.1
Anxiety score (GAD-7), mean ± SD	7.4 ± 5.2
Medications	
Blood pressure, % (n)	40% (25)
Hypnotic sleeping aids, % (n)	37% (23)
Antidepressant, % (n)	35% (22)
Antianxiety, % (n)	18% (11)
Narcotic pain, % (n)	16% (10)
Stimulant, % (n)	6% (4)
Comorbidities	
Insomnia, % (n)	65% (41)
Depression, % (n)	59% (37)
Obstructive sleep apnea, % (n)	52% (33)
Anxiety, % (n)	44% (28)
Hypertension, % (n)	43% (27)
Restless leg syndrome, % (n)	16% (10)
Diabetes, % (n)	13% (8)
Heart disease, % (n)	10% (6)

BMI = body mass index, GAD-7 = Generalized Anxiety Disorder 7-Item, PHQ-9 = Patient Health Questionnaire-9, SD = standard deviation.

Figure 1

Left: Depiction of Sleep Profiler worn in a home setting. Right: Sensor strip electrode configuration for the Sleep Profiler.

Figure 2—Presentation of a 10-minute screen of the Sleep Profiler signals.
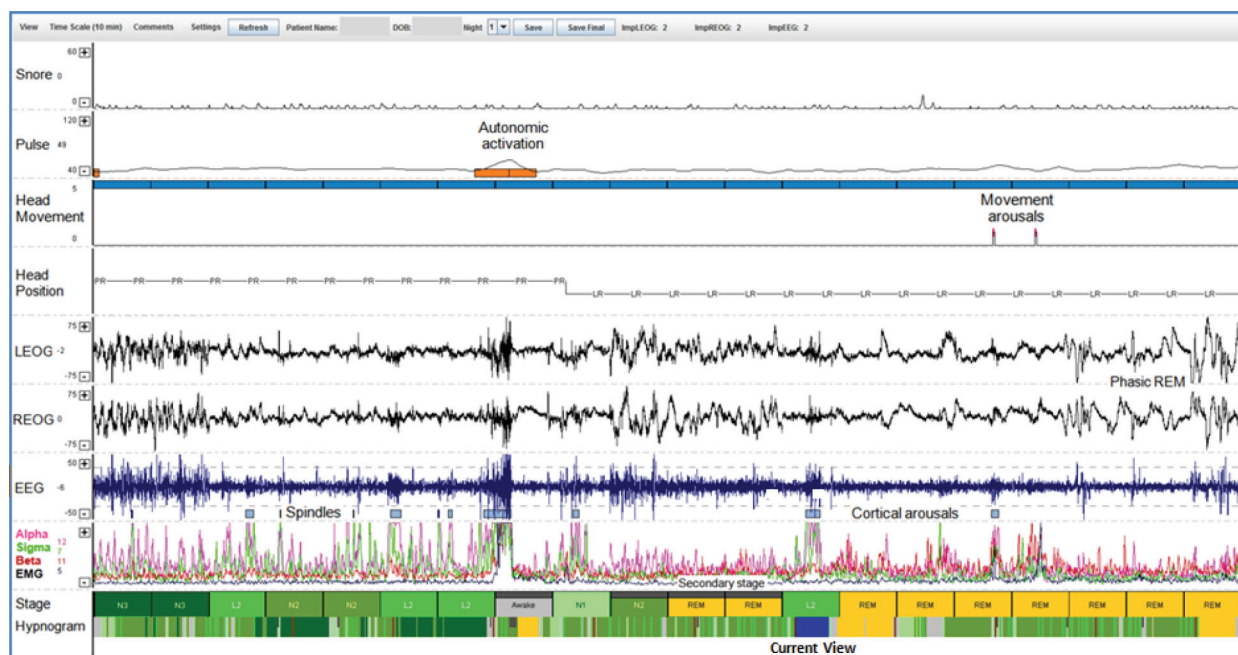
Assigned primary and secondary staging of 30-second epochs marked [Awake] [Stage N1] [Light N2] [Stage N2] [Stage N3] [REM], upper gray secondary stripes [Awake] identify epochs to be reviewed, with stripes identifying spindles, cortical arousals, and autonomic activations. Note changes in amplitudes in the frontopolar EEG signals labeled LEOG, REOG, and EEG and reduction in sigma power in transitions from NREM to REM. In this example secondary stripes identify the need to review the transition between NREM and REM. Low amplitude EEG in combination with phasic REM patterns to the right of the transition confirm the REM period. EEG = electroencephalography, EMG = electromyography, EOG = electrooculography, NREM = non-rapid eye movement; REM = rapid eye movement.

start of the study and every one-half hour throughout the night. The photoplethysmography (sampled at 100 Hz) obtained from the forehead was used to calculate the pulse rate at 1 Hz. Snoring sounds acquired with an acoustic microphone were sampled with firmware at 2 kHz, root mean square filtered to create a sound envelope that was downsampled to 10 Hz and saved in the study record. A triaxial accelerometer was sampled at 100 Hz, with the X|Y|Z signals converted to 360 degree angles and saved at 10 Hz.

For the in-home studies, patients were instructed to wipe their forehead thoroughly with an alcohol wipe prior to affixing the device to obtain acceptable skin-sensor impedances. Voice messages alerted patients when the impedances were too high at the beginning of the night. Patients replaced the forehead sensors prior to night 2.

Software

Automated sleep staging was applied to sleep markers extracted from each 30-second epoch for the 3 frontopolar EEG channels (**Figure 2**). After rejection of periods when the absolute amplitude is ≥ 500 µV, the signals were notch filtered, and then infinite impulse response band pass-filtered to obtain 16 Hz samples of the power values for delta (1–3.5 Hz), Delta^C (delta power corrected for ocular activity), theta (4–6.5 Hz), alpha (8–12 Hz), sigma (12–16 Hz), beta (18–28 Hz), and EMG bands (> 40 Hz with a 80 Hz, 3 dB rolloff). A second set of power values was obtained after application of a 0.75-Hz high-pass filter. Both filtered and unfiltered power spectral data

from the three frontopolar EEG channels were used to characterize and stage sleep.

If at least 15 seconds of valid data were available, the AF7-AF8 channel was used for autostaging, followed by the AF7-Fpz and AF8-Fpz channels. When either the AF7-Fpz or the AF8-Fpz signal was used for staging, the power spectra were increased to compensate for signal attenuation attributed to amplifier common mode rejection resulting from the substantially shorter interelectrode distances.

The power spectra values, averaged from 16 to 4 Hz, were used to detect sleep spindles, characterized by spikes in the absolute and relative alpha and sigma power that met empirically derived thresholds designed to ensure there was a sufficient sigma component (**Figure 3**). The minimum spindle length was 0.25 Hz with no maximum spindle length. To reduce the likelihood of misclassifying pseudospindles, the beta and EMG power bands required simultaneous suppression relative to the alpha and sigma power. When the sigma power peaked prior to the alpha power, the spindles were classified as fast-dominant. The spindle length (ie, from the start to the end of the spindle) was determined by either the alpha or sigma power crossing a minimum absolute power threshold. Spindle duration was tallied as the sum of all spindle lengths. Cortical arousals were detected when 3 or more seconds of absolute and relative alpha power exceeded the median alpha power from the preceding 2 minutes. Increased absolute and relative EMG powers were similarly compared to median values to detect micro-arousals. The duration of detected spindles and arousals were marked

Figure 3—Three 30-second epoch screenshots of frontopolar EEG signals AF7-Fpz, AF8-Fpz and AF7-AF8 labeled LEOG, REOG and EEG from two elderly healthy controls.
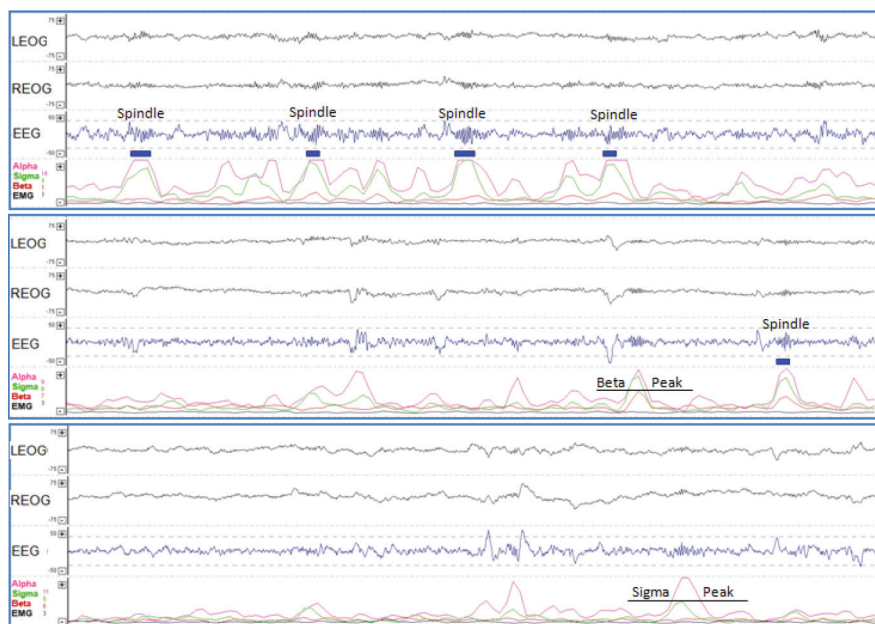
Top: Epoch screenshot from a 74-year-old woman with total of 10.9 seconds of sleep spindle activity (sum of all blue spindle stripe durations). Middle and bottom: Screenshot from a 74-year-old man with a very low total sleep spindle duration < 5 seconds with one potential event not marked due to excessive beta power relative to the alpha and sigma (middle), and one potential sleep spindle with insufficient sigma power accompanying the alpha burst (bottom). EEG = electroencephalography, EMG = electromyography, EOG = electrooculography.

with a stripe in the staged channel. If a spindle occurred during a cortical arousal, the spindle was not marked.

Patterns in the AF7-Fpz and AF8-Fpz signals that characterized and distinguished slow eye rolls from phasic REMs were recognized by computing Pearson correlations of the 6.5 Hz zero phase infinite impulse response filtered outputs. The distinction between elevated delta power resulting from phasic REM versus slow wave sleep was made by comparisons of the filtered and unfiltered Delta[c].

Autonomic activation events were detected when the pulse rate changed by 6 or more beats per minute compared to the pulse rate 10 seconds prior and/or 10 seconds subsequent to the current second. Movement intensities from 0 to 5 were assigned to each second based on the sum of the actigraphy changes across the X|Y|Z axes. Position changes typically resulted in a movement ranking of 3 or more. For each epoch, the average magnitude (dB) and the percentage of time snoring were calculated. Crescendo and loud snoring events terminating with a significant decrease in sound or a single loud snore (ie, gasp) were identified as likely indicators of sleep-disordered breathing for visual inspection.

For each 30-second epoch, the power spectra values were averaged, and the number of arousals, spindles, movements, snoring, and other patterns tallied. These data in combination with ratios of the mean power spectra were used to assign sleep stages. A discriminant function analysis was used to differentiate stages N2 and N3. The number of seconds (ie, slow wave seconds) that delta power exceeded a threshold equivalent to ± 30 μV was primarily used to stage N3. Epochs

with relatively high beta and low EMG power were classified as tonic REM and distinguished from tonic stage N1, based on temporal proximity to epochs with phasic REM. Brief but intense periods (eg, < 5 seconds) of increased EMG power typically resulted in an epoch being staged awake. A sub-stage labeled Light N2 was assigned to N2 epochs with no spindles and characterized with either a K-complex or dominant theta activity with relatively elevated levels of alpha or EMG power.

In the absence of a sleep spindle, the first sleep epoch following an awake epoch was staged N1. Epochs with at least 1 sleep spindle and no arousal, 2 sleep spindles and 1 arousal, or at least 3 sleep spindles with 2 cortical arousals were staged N2. Epochs with a cortical arousal or microarousal and no sleep spindle, or 2 arousals with 2 spindles were staged N1.

A number of additional rules and thresholds were developed to identify epochs that should be visually inspected. During visual inspection, these epochs were identified by the presentation of "primary" and "secondary" stage stripes (**Figure 2**). The greatest number of epochs assigned primary and secondary strips were those transitioning between awake and sleep. The software enabled selection of a setting, applied to the entire record that biased staging toward an increased classification of sleep. This awake = low setting reduced the editing needed for individuals with severe disruptions attributed to sleep-disordered breathing. The default setting applied a sequence of discriminant function analyses (awake versus N1, awake versus N2, and awake versus REM) to reassign the primary stage as awake, and provide a secondary stripe based on the original sleep stage. Epochs staged REM during the first 10 minutes after

sleep onset were assigned a primary stage N1 and a secondary stage of REM. Intermittent NREM epochs within a block of REM epochs were assigned REM as the primary and NREM as the secondary stage. Intermittent REM epochs in proximity to multiple awake epochs were assigned primary and secondary stages of N1 and REM, respectively. Epochs with large phasic REM but staged NREM due to the incomplete correction for ocular activity were assigned an REM secondary stage.

Sleep onset was based on 4 consecutive sleep epochs during the initial 5 minutes of recording time, and 3 consecutive epochs after 5 minutes of recording time. All epochs were staged awake prior to sleep onset. The cortical and autonomic indexes were based on the total number of events divided by sleep time. Autonomic activation indexes were additionally computed for events that occurred during NREM and REM sleep time. Spindle and slow wave events were divided by the time staged N2 and N3.

For visual inspection, the signal acquired from AF7-AF8 was labeled EEG, default scaled to ± 50 µV and presented with the ocular activity removed. The signals from AF7-Fpz and AF8-Fpz were labeled LEOG and REOG and default scaled to ± 75 µV.

Expert Review

One rater visually inspected the frontopolar EEG signal waveforms along with the presentations of the alpha, sigma, beta, and EMG power to confirm the veracity of the autostaging near REM transitions and when epochs were assigned secondary stripes (**Figure 2**). Minimal editing was made when the primary/secondary stripes were awake/N1 in the absence of snoring with the goal to: (1) adjust the beginning of the study when necessary if the device was turned on prior to lying in bed; (2) review the autoscoring to accurately identify sleep onset, and the EMG power should be lower than the alpha, sigma, and beta; (3) review epochs with gray "secondary" stripes; (4) check the start and end of all REM periods, and look for missed REM at the end of the first sleep cycle; (5) inspect the dark blue signal used for staging when it includes segments marked red (rejected signals), ie, gross differences between the LEOG and REOG indicate artifact that may cause stage N3 or REM to be incorrectly staged; and (6) adjust the end of the study if the device was inadvertently left on (signals are flat). A detailed description of the sleep staging rules and application of expert editing is provided in the training video.[23]

RESULTS

Accuracy

After pooling of all staged PSG epochs, the mean interscorer agreements between the 5 technologists were 75.9% overall, with 90.1%, 51.3%, 75.5%, 67.2%, and 91.1% for stages awake, N1, N2, N3, and REM, respectively. The mean kappa score across the 10 comparisons was 0.70 (range 0.61 to 0.78). Scorers 1, 4, and 5 staged over 4.5 times more N1 epochs than scorer 2, and over 2.5 more N1 epochs than scorer 3. As compared to scorers 1, 4, and 5, scorers 2 and 3 staged over 2.5 times more N3 epochs.

When unedited default autostaged epochs were compared to each of the 5 scorers, the mean overall agreement was 71.3%, and 80.9%, 22.9%, 79.7%, 74.9%, and 71.5% for stages awake, N1, N2, N3, and REM, respectively. After expert review, the mean overall agreement improved to 73.9%, and to 85.3%, 27.9%, 80.6%, 75.3%, and 77.6% for stages awake, N1, N2, N3, and REM, respectively. The mean autostaging kappa score increased from 0.63 (range 0.62–0.65) to 0.67 (range 0.65–0.68) after visual review.

Table 2 and **Table 3** compare the majority agreement in staging among the 5 scorers to the autostaging without and with the increased classification of awake time (which affected 3% of the epochs). **Table 4** presents comparisons after the autostaging was expert reviewed. The primary benefit of technical review was improved sensitivity and specificity between REM and awake, and transitions between REM and N2. With each step, the overall agreements increased from 75.8% to 77.1 and 80.0%. As a result of the expert review, 3.3% of the total number of epochs was changed.

In contrast to the contingency tables, which compared exact agreement of each epoch, **Table 5** compares the agreement among sleep architecture biomarkers while allowing for different epochs to contribute to the method-specific percentage of staged time, and incorporating method specific differences in the staging of sleep and wake. Although minimal differences were observed across stages awake and REM, there was a threefold difference across scorers in the percent time staged N1 and N3. Differences resulting from the 3 autoscoring methods were far less than that of PSG interscorer variability.

The scatterplots in **Figure 4** show a strong concordance between majority agreement and the autostaging after expert review for all stages except N1. The mean biases point to the autoscoring overreporting awake time, and underreporting stages N1 and REM, compared to majority agreement.

Variability

Figure 5 presents scatterplots displaying the night-to-night variability and between-night bias for biomarkers useful in characterizing sleep. Sleep spindles, autonomic activation index, and slow wave sleep had the most consistent patterns, whereas sleep time, sleep latency, and wake after sleep onset had the greatest night-to-night variability. The mean alpha and sigma power across stages N2 and N3 exhibited very limited night-to-night variability (intraclass correlation of 0.97 and 0.98) given the relative prominence of these frequencies during sleep spindles and slow wave sleep.

There was significantly greater sleep spindle activity, and greater sigma and beta power across stages N2/N3 in those taking antidepressants ($P < .01$). Patients taking antihypertensive medications exhibited significantly less stage N3, coupled with increased Light N2, and lower delta, theta, alpha, and sigma power across stages N2/N3 ($P < .001$). Additionally, patients taking antihypertensive medications exhibited lower autonomic activation during REM sleep ($P < .03$).

Stability

The stability of the sleep biomarkers was evaluated by comparing individual results to clinical thresholds used to assist with clinical interpretation (**Table 6**). Biomarkers with the strongest trait characteristic would be consistently normal or abnormal

Table 2—Contingency table of agreement between majority PSG staging by human scorers versus Sleep Profiler autostaging of pooled epochs (awake = low).

Majority of scorers staged as:	Autostaging					
	Awake (%)	N1 (%)	N2 (%)	N3 (%)	REM (%)	No. Epochs
Awake	75.2	34.2%	3.5%	0.4%	11.5%	7,749
N1	12.6	27.6%	7.2%	0.1%	4.1%	3,057
N2	5.6	20.4%	80.4%	20.0%	3.2%	14,190
N3	0.2	0.4%	3.9%	79.1%	0.7%	3,705
REM	2.9	11.6%	3.3%	0.0%	77.4%	4,118
No consensus	3.5	5.9%	1.6%	0.4%	3.0%	816
No. epochs	7,590	3,032	15,232	3,861	3,920	33,635

PSG = polysomnography.

Table 3—Contingency table of agreement between majority PSG staging by human scorers versus Sleep Profiler autostaging of pooled epochs (awake = default).

Majority of scorers staged as:	Autostaging				
	Awake (%)	N1 (%)	N2 (%)	N3 (%)	REM (%)
Awake	82.2	35.9	4.0	0.5	5.8
N1	9.6	26.1	4.7	0.4	8.7
N2	3.8	33.2	85.0	16.0	12.3
N3	0.2	0.2	5.4	82.3	0.0
REM	4.2	4.7	0.8	0.8	73.2
No consensus	3.8	5.6	1.5	0.4	2.4
No. epochs	8,052	3,238	14,404	3,721	4,220

PSG = polysomnography.

Table 4—Contingency table of agreement between majority PSG staging by human scorers versus Sleep Profiler expert edited default autostaging of pooled epochs.

Majority of scorers staged as:	Expert edited autostaging				
	Awake (%)	N1 (%)	N2 (%)	N3 (%)	REM (%)
Awake	86.8	32.2	3.0	0.3	1.9
N1	9.3	32.6	5.3	0.6	8.8
N2	2.7	33.3	86.0	16.2	9.3
N3	0.1	0.1	5.4	82.9	0.0
REM	1.2	1.8	0.3	0.0	80.0
No consensus	3.4	6.3	1.6	0.5	2.0
No. epochs	8,018	3,262	14,425	3,723	4,201

PSG = polysomnography.

on nights 1 and 2 (ie, stable) despite the night-to-night variability. The biomarkers with the highest stability were stage N3/slow wave sleep, spindle activity, and autonomic activation; sleep biomarkers were associated with chronic disease and neurodegeneration.[2,8] For measures with lower stabilities, 2-night averages were needed to determine a normal/abnormal state in more than 25% of patients. Short sleep time, based on fewer than 6 hours, had the least night-to-night stability, and there were substantial differences in the proportion of those classified as abnormal based on the 10th percentile versus fewer than 6 hours of sleep across 2-night averages (14% versus 52%, respectively). A greater number of patients were identified as long sleepers based on the 90th percentile compared to the more than 10-hour threshold; however, neither measure could be thoroughly evaluated, because the inclusion criteria required complaints of chronic insomnia. Use of prescription sleeping aids did not influence the distributions of cases classified as either normal or abnormal. Based on the stability of the majority of sleep biomarkers, a 2-night study appears necessary for accurate profiling or assessment of treatment outcomes.

A framework for interpreting abnormal sleep biomarkers within the construct of a clinical evaluation or research investigation is presented in **Table 7**.

Table 5—Comparisons of the percentages of sleep staging time across pooled epochs by- and across-human scorers, and by autoscoring.

	Awake (%)	N1 (%)	N2 (%)	N3 (%)	REM (%)
Human Scorers					
Scorer 1 – RPSGT	22.6	19.9	55.4	10.6	14.1
Scorer 2 – RPSGT	22.1	4.0	59.6	22.4	13.9
Scorer 3 – RPSGT	25.4	7.1	48.6	26.5	17.8
Scorer 4 – Research tech	21.5	18.2	57.4	7.5	16.9
Scorer 5 – Research tech	23.2	19.6	51.4	12.9	16.1
Majority of Scorers	23.6	12.2	56.6	14.8	16.4
Autoscored					
Awake-low	22.3	11.2	58.8	15.1	15.4
Awake-default	25.3	10.5	59.0	15.7	15.2
Expert-edited	25.0	11.6	59.6	15.7	14.5

RPSGT = registered polysomnographic technologist.

Figure 4—Scatterplots of the relationship between minutes of sleep and awake measured by majority agreement of staging from PSG versus staging from default autoscored and expert edited frontopolar EEG.
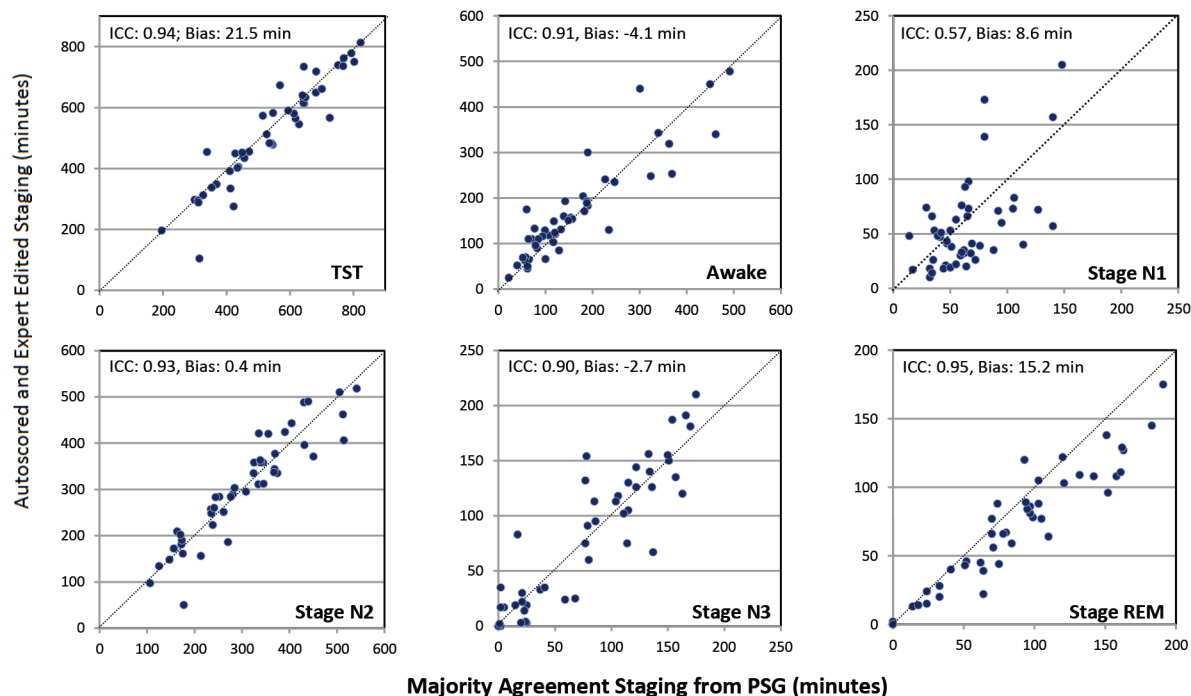
EEG = electroencephalography, ICC = intraclass correlation; PSG = polysomnography, TST = total sleep time

DISCUSSION

This study demonstrated that sleep biomarkers obtained from the multichannel frontopolar EEG recording device (Sleep Profiler) are valid, accurate, and reproducible. Three analyses were applied to assess the autostaging accuracy of frontopolar recordings compared to human scored PSG-based sleep architecture measures: direct agreement of epoch-by-epoch staging, total staged sleep times, and proportional agreement based on the percentages of staged time. Our study supports previously published studies reporting a fairly broad range of interscorer variability.[24,25] Thus, our findings also underscore the difficulty

in validating the accuracy of autostaging software against the current gold standard practice of technician-based visual scoring[24,26–28] or by RPSGT certification,[29–31] even when the same equipment and software is used.

The interscorer variability in staging N1 and N3 observed in this study is consistent with reports from the AASM Inter-Scorer Reliability Program.[25] One of the challenges of validating autostaging is overcoming the inherent scorer bias. If this study had been conducted where all of the scorers were trained to stage sleep in a manner similar to that of scorers 2 and 3, the autoscoring would have underreported stage N3 and overreported stage N1. Conversely, if the training reflected the

Figure 5—Scatterplots displaying night-to-night variability of sleep architecture and sleep continuity measures.
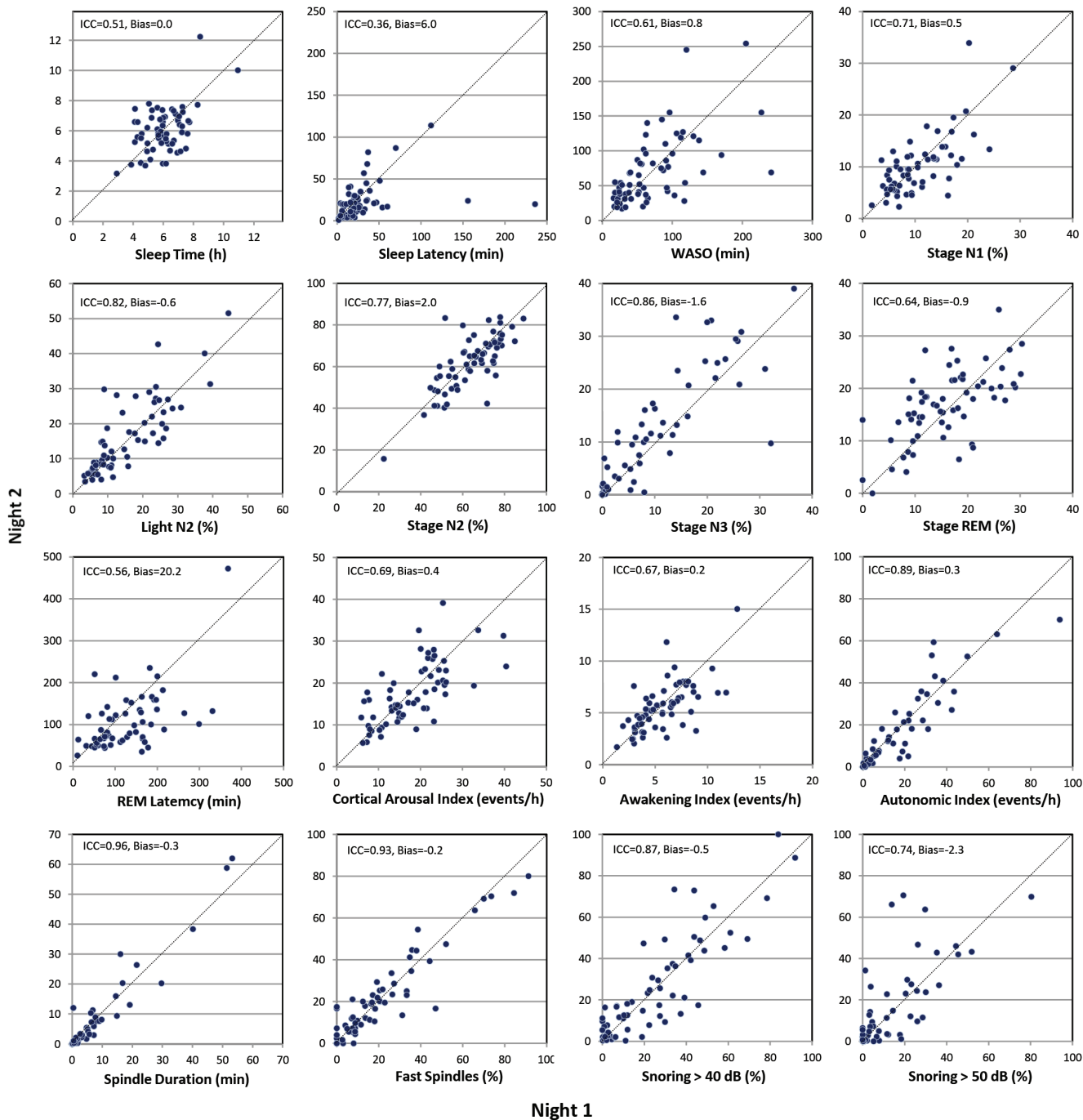
All intraclass correlation (ICC) *P* < .0001. WASO = wake after sleep onset.

styles of scorers 1, 4, and 5, the autoscoring would overreport stage N3 and underreport stage N1. In the absence of a true gold standard, the autoscoring achieved a level of "Goldilocks" accuracy.[32]

Autoscoring combines speed and consistency, with the capability to further improve results with a focused, relatively brief technical review.[18,26] In this study, changes in alpha, sigma, beta, and EMG power viewed on a 10-minute screen enable recognition of 30-second epochs that should be manually inspected due to transitions between NREM and REM or between sleep and wake. Autodetected sleep spindles and arousals were visually marked in the record, a technique shown to reduce interscorer variability.[29] The presentation of secondary stripes assisted in the identification of epochs that might benefit from visual inspection, a concept similar to the editing helper feature described by Younes et al.[33] A total of 3.3% of autostaged epochs were changed, with the greatest effect noted in the improved staging of REM, N1, and awake, specifically

Table 6—Stability of abnormal sleep biomarkers in the patient cohort when compared to reference values.

Sleep Quality Metrics	Clinical Thresholds		N1 and N2 Both Night		Stability N1 & N2	N1 and N2 Averaged	
	Abnormal based on	Abnormal range	Normal	Abnormal		Normal	Abnormal
Highest stability							
Sleep spindle duration, min	< IQR1 < 70	1.6	46%	44%	90%	54%	46%
Stage N3, %	< 10th	1.3–9.5	47%	40%	87%	49%	51%
Light N2, %	> IQR3	21	64%	22%	86%	73%	27%
Autonomic activation index	< IQR1	11.4	51%	34%	85%	62%	38%
Snoring > 50 dB, %	> IQR3	9	52%	33%	85%	60%	40%
Moderate stability							
Snoring > 40 dB, %	> IQR3	18	41%	43%	84%	49%	51%
Awakening index	> 90th	4.6–5.9	33%	49%	82%	43%	57%
Wake after sleep onset, min	> IQR3	95	65%	16%	81%	75%	25%
Sleep latency, min	< 10th	31–47	73%	6%	79%	86%	14%
Lowest stability							
Cortical arousal index	> 90th	19–27	62%	11%	73%	71%	29%
Stage N2, % long	> 90th	64–76	46%	27%	73%	63%	37%
Stage N1, %	> 90th	6.4–10.6	24%	49%	73%	35%	65%
Sleep time, h short	< 10th	4.2–5.3	64%	6%	70%	86%	14%
REM latency, min	> 90th	119–150	48%	20%	68%	67%	33%
Sleep efficiency, %	< 10th	69–79	54%	14%	68%	71%	29%
Stage REM, % short	< 10th	12.6–14.6	48%	19%	67%	65%	35%
Sleep time, h short	<	6.0	24%	32%	56%	48%	52%
Biased toward normalcy							
Sleep time, h long	>	10.0	97%	2%	99%	97%	3%
Stage REM, % long	> 90th	27–30	95%	0%	95%	97%	3%
Stage N2, % short	< 10th	35–48	94%	3%	97%	97%	3%
Sleep time, h long	> 90th	7.1–7.5	67%	8%	75%	95%	5%

IQR = interquartile range.

Table 7—Clinical and research framework for the evaluation/investigation of associations between sleep biomarkers and pharmacotherapy, chronic medical, neurodegenerative, and psychiatric conditions.

	Insomnia	Cardio-vascular	Diabetes	Obesity	Hyper-somnia	Neurode-generation	Pain	Psychiatric or Mood	Pharmaco-therapy
Overall Sleep Patterns									
Sleep time	✓	✓	✓	✓	✓		✓	✓	✓
Sleep latency	✓			✓					✓
WASO	✓								✓
Sleep Architecture									
Slow wave sleep	✓	✓	✓	✓	✓	✓	✓	✓	✓
Sleep spindles	✓	✓	✓		✓	✓		✓	✓
Stage 1	✓	✓	✓	✓	✓		✓		✓
K-complex/light N2	✓	✓	✓	✓	✓		✓		✓
REM						✓	✓	✓	✓
Sleep Continuity									
Cortical arousals	✓	✓		✓	✓		✓		✓
Autonomic activation	✓	✓	✓	✓	✓	✓	✓	✓	✓
Awakenings	✓	✓	✓	✓	✓		✓		✓
Other Sleep Measures									
Loud snoring	✓	✓	✓	✓	✓	✓			
Movement	✓				✓		✓		✓

WASO = wake after sleep onset.

between awake and REM and transitions between REM and stage N2, consistent with a previous report.[18] The Sleep Profiler provides the capability to optionally acquire submental EMG; however, the accuracy results in this study were achieved without use of this signal. Studies are underway to evaluate the benefit of including this signal for the in-home assessment of REM behavioral disorder, another sleep biomarker associated with neurodegeneration.[34] After expert editing, the kappa score between frontopolar sleep EEG and majority agreement from PSG increased to 0.67, approximately the same as the mean kappa score from the 10 comparisons between the 5 scorers (0.70). **Table 6** results suggest that accuracy may be further improved with more targeted editing during transitions between REM and N1.

Due to the challenges of human application of the N1 staging rules, a 5-fold difference in the percentage of epochs staged N1 was observed across scorers. Although there was relatively poor agreement between epochs scored by majority agreement and the autostaging, the percent times staged N1 for the 2 methods were quite similar, suggesting the discrepancies were related to the timing rather than inaccuracy in the recognition of light NREM sleep. Given the difficulty in human staging N1 and N3, it is likely that computer-assisted scoring is required to further characterize sleep biomarkers based on the depth of NREM.[35,36] For this study, a subtype of stage N2 (ie, Light N2) was used to classify epochs with K-complex or dominant theta activity with relatively elevated levels of alpha or EMG power, and absent spindle activity. The staging of Light N2 exhibited less night-to-night variability than the overall stage N2 (intraclass correlation 0.82 versus 0.77). Light N2 was highly stable across both nights and it increased significantly in those taking antihypertensive medications, whereas deeper stage N2 did not. This finding not only validated Light N2 as a unique biomarker for depth of NREM sleep, but also highlighted the benefit of subcharacterizing NREM sleep. The user option that increased autoscored sleep time resulted in improved concordance between apnea-hypopnea indexes obtained by PSG and with Sleep Profiler with cardiorespiratory signals was also evaluated.

Slow wave sleep has emerged as an important sleep biomarker, given its association with numerous chronic diseases and neurodegeneration.[2,7,8] When these patients were stratified by condition, reduced slow wave sleep was noted in those with hypertension, but not depression, sleeping aids, or obstructive sleep apnea. It is unlikely, however, that the clinical application of this sleep biomarker can be broadened when the human scoring of standard PSG is not only time consuming and expensive, but unreliable. In this study the manual staging of N3 ranged from 7.5% to 26.5% of the pooled epochs, approximately equivalent to the 10th and 90th percentile cutoffs for healthy adults.[22] For example, 2 sleep research technicians who underwent the same rigorous sleep staging training protocol as part of an academic research program still demonstrated relatively large differences in detected N3 sleep (8.1% and 12.9%, respectively). Conversely when PSG epochs staged N3 by majority agreement were compared to the autoscored forehead EEG, there was less than a 1% difference. Because the manual editing of autoscored N3 is only recommended for epochs with

artifact or arousals, the accuracies were similar for the unedited or edited results.

The benefit of multichannel frontopolar EEG is that all of the signal elements needed to visually stage sleep are present, including ocular, spindle, K-complex, slow wave, and cortical activity. The differential recordings acquired with Sleep Profiler were selected because of the simplicity with user application and reduced likelihood of study failure across multinight studies. Additionally, differential recordings do not require removal of heartbeat artifact that can contribute to artifact-induced increases in theta power, and result in the autostaged misclassification of stage N2. The disadvantage of differential recordings is that the signal amplitudes are attenuated and thus the expected amplitudes used to visually or autostage sleep must be scaled (eg, 75 versus 60 μV for stage N3). The Sleep Profiler sensor placements enabled acquisition of frontopolar EEG that can be staged as well as used to detect ocular activity for the differentiation of stages N1 from REM. As compared to the conventional EOG sensor sites, the frontopolar EEG signals include blink activity, but not saccades.

Despite the differences in sleep time, sleep spindle duration showed the greatest night-to-night concordance, suggesting a very strong trait effect. The interscorer agreement achieved by manual sleep spindle staging of 115-second segments in one study was superior to the reliability obtained by human scoring of cortical arousals in another study.[24,37] The accuracy of human sleep spindle staging under more realistic conditions, however, would be expected to decline as a result of scorer fatigue, given 5 times the number of sleep spindles were observed in our typical patient record as compared to cortical arousals (mean: 500 versus 100, maximum: 2900 versus 280, respectively). The approach used in this study to autodetect sleep spindles relied on patterns of the power spectral density, rather than extracting spindle patterns with filtering.[38] It is likely that the sleep spindle length/duration measured by this power spectra approach will be less than by filtering, because of the rule that marks the spindle length, and as a result of elimination of 11–13 Hz spindles (because of the requirement for peaks in both the alpha (8–12 Hz) and sigma (12–16 Hz) power bands. The sleep spindle detection algorithms used in this study were based on thresholds selected for differential frontopolar EEG. Further research is required to compare the automated sleep spindle detection by this approach with automated routines applied to more conventional EEG spindle detection sites,[38] and to evaluate the changes in this measure as it relates to age, sex, and neurodegeneration.[3–5,39]

A limitation of this study was the relatively dichotomous age representation of our healthy controls with very few participants in this cohort being between the ages of 40 and 60 years. Cognizant of the limitation, we excluded those older than 70 years when establishing the sleep spindle reference values. This enabled exclusion of all but one of the patients taking antihypertensive medication but may have excluded elderly women with sleep spindle activity appropriate for inclusion.[40]

This study introduced a highly stable sleep biomarker that combined sleep stage and cardiac autonomic tone. In the patients, the autonomic activation index (AAI), a measurement of brief but important episodic changes in pulse rate,

was found to be 2 times greater during NREM sleep as compared to REM sleep, a difference similar to the NREM and REM ratio of low- to high-frequency heart rate variability (HRV).[41,42] The night-to-night variability of the AAI and HRV were also similar during both NREM sleep (0.82 versus 0.90) and REM sleep (0.80 versus 0.89).[41] These findings suggest that both low HRV and low AAI are measuring increased sympathetic dominance,[43] a finding supported by the lower REM AAI in patients taking antihypertensive medications. Each autonomic activation event requires a 10-second detection window, thus tallying and computing an index for REM and NREM is possible in patients with limited REM time or who suffer from highly fragmented sleep. By comparison, obtaining a valid stage-dependent HRV measure is more challenging, given the need for a 5-minute detection window. Further investigations are needed to determine if the AAI during REM and NREM is associated with blood pressure, inflammation, and/or hypothalamic-pituitary-adrenal systems. If so, this biomarker may assist in subtyping patients with insomnia exacerbated by comorbid anxiety or depression. Longitudinal studies might benefit from the inclusion of the AAI biomarker combined with sleep spindle activity to monitor cognitive decline in old age, similar to the contributions made by measuring HRV.[44]

To achieve the study objectives, a prospectively acquired dataset was used to assess accuracy, and a retrospectively acquired dataset was used to assess sleep biomarker variability and stability. This study was limited by the fact that patients with different combinations of comorbidities, medications, etc. were included. For example, patients with depression disorders would be expected to have long REM sleep times, yet only 3% did, possibly because 35% were taking antidepressants, which suppresses REM sleep time. Many of the patients reported depressive or anxiety disorders, yet few had long sleep times (ie, > 10 hours), likely as the result of overlapping use of antidepressant and sleeping aids. At the same time, the heterogeneity of our sample size provides a more true to life sample of the relatively common complexity of comorbid conditions most sleep providers encounter in their sleep patient population.

Sleep biomarkers have been associated with a range of medical and neurodegenerative conditions and a number of medical therapies (**Table 7**); however, the neurobiological mechanisms underlying these associations remain poorly understood. Interpretive sleep biomarker profiling can serve as a critical step toward developing precision-based clinical protocols to improve the outcomes of patients frequently experiencing various combinations of these sleep and medical conditions. Treating patients with complex chronic medical conditions requires knowledge of the potential interplay between iatrogenic medications effects and sleep. For example, we found that antihypertensive medications were associated with increased sympathetic dominance during REM and suppressed slow wave sleep. Given the evidence suggesting learning, memory, and overall neuroprotective/health benefits of slow wave sleep, further studies are needed to examine which classes of antihypertensive medications (eg, alpha or beta blockers, angiotensin-converting-enzyme inhibitors, etc.) contribute most to the manifestation of these abnormal sleep biomarkers. Studies

utilizing this biomarker approach are also needed to determine whether medications associated with enhanced slow wave sleep or sleep spindle activity (or other positive sleep biomarker features) can provide the same neuroprotective and overall health benefits as natural sleep.

CONCLUSIONS

This study confirms the validity of multichannel frontopolar EEG recordings for use in clinical or research applications, and suggests that autoscoring may be superior to human scoring of sleep biomarkers. Furthermore, our findings indicate that 1 night of recording with Sleep Profiler is sufficient to characterize abnormal slow wave sleep, spindle activity, and heart rate variability in patients; however, a 2-night average would improve the assessment of abnormality for the balance of sleep architecture and sleep continuity biomarkers.

ABBREVIATIONS

AAI, autonomic activation index
AASM, American Academy of Sleep Medicine
BMI, body mass index
EEG, electroencephalography
EOG, electrooculography
EMG, electromyography
HRV, heart rate variability
ICC, intraclass correlation
NREM, non-rapid eye movement
PSG, polysomnography
REM, rapid eye movement
WASO, wake after sleep onset

REFERENCES

1. Strimbu K, Tavel JA. What are biomarkers? *Curr Opin HIV Aids.* 2010;5(6):463–466.

2. Javaheri S, Redline S. Sleep, slow-wave sleep and blood pressure. *Curr Hypertens Rep.* 2012;14(5):442–448.

3. Weiner OM, Thien TDV. Spindle oscillations in sleep disorders: a systematic review. *Neural Plast.* 2016;2016:7328725.

4. Latreille V, Carrier J, Lafortune M, et al. Sleep spindles in Parkinson's disease may predict the development of dementia. *Neurobiol Aging.* 2015;36(2):1083–1090.

5. Christensen JA, Kempfner J, Zoetmulder M, et al. Decreased sleep spindle density in patients with idiopathic REM sleep behavior disorder and patients with Parkinson's disease. *Clin Neurophysiol.* 2014;125(3):512–519.

6. Christensen JA, Nikolic M, Warby SC, et al. Sleep spindle alterations in patients with Parkinson's disease. *Front Hum Neurosci.* 2015;9:233.

7. Varga AW, Wohlleber ME, Giménez S, et al. Reduced slow-wave sleep is associated with high cerebrospinal fluid Aβ42 levels in cognitively normal elderly. *Sleep.* 2016;39(11):2041–2048.

8. Pillai JA, Leverenz JB. Sleep and neurodegeneration: a critical appraisal. *Chest.* 2017 Jan 10. [Epub ahead of print].

9. Vgontzas AN, Fernandez-Mendoza J, Liau D, et al. Insomnia with objective short sleep duration: the most biologically severe phenotype of the disorder. *Sleep Med Rev.* 2013;17(4):241–254.

10. Hall MH, Smagula SF, Boudreau RM, et al. Association between sleep duration and mortality is mediated by markers of inflammation and health in older adults: the health, aging and body composition study. *Sleep*. 2015;38(2):189–195.

11. Fernandez-Mendoza J. The insomnia with short sleep duration phenotype: an update on its importance for health and prevention. *Curr Opin Psychiatry*. 2017;30(1):56–63.

12. Van Mill JG, Vogelzangs N, van Someren EJ, et al. Sleep duration, but not insomnia, predicts 2-year course of depressive and anxiety disorders. *J Clin Psychiatry*. 2014;75(2):119–126.

13. Dang-Vu TT, Salimi A, Boucetta S, et al. Sleep spindles predict stress-related increases in sleep disturbances. *J Clin Sleep Med*. 2013;9(1):81–87.

14. Coleman PJ, Gotter AL, Herring WJ, et al. The discovery of suvorexant, the first orexin receptor drug for insomnia. *Annu Rev Pharmacol Toxicol*. 2017;57:509–533.

15. Sateia MJ, Buysee D, Krystal AD, et al. Clinical practice guidelines for the pharmacological treatment of chronic insomnia in adults: an American Academy of Sleep Medicine clinical practice guideline. *J Clin Sleep Med*. 2017;13(2):307–349.

16. FDA Drug Safety Communication: FDA warns of next-day impairment with sleep aid Lunesta (eszopiclone) and lowers recommended dosage. United States Food and Drug Administration website. http://www.fda.gov/Drugs/DrugSafety/ucm397260.htm. Published May 15, 2014. Updated February 10, 2016. Accessed March 24, 2017.

17. FDA Drug Safety Communication: Risk of next-morning impairment after use of insomnia drugs; FDA requires lower recommended doses for certain drugs containing zolpidem (Ambien, Ambien CR, Edluar, Zolpimist). United States Food and Drug Administration website. http://www.fda.gov/Drugs/DrugSafety/ucm334033.htm. Published January 10, 2013. Updated January 16, 2016. Accessed March 24, 2017.

18. Finan PH, Richards JM, Gamaldo CE, et al. Validation of a wireless, self-application, ambulatory electroencephalographic sleep monitoring device in healthy volunteers. *J Clin Sleep Med*. 2016;12(11):1443–1451.

19. Lucey BP, McLeland JS, Toedebusch CD, et al. Comparison of a single-channel EEG sleep study to polysomnography. *J Sleep Res*. 2016;25(6):625–635.

20. Berry RB, Brooks R, Gamaldo CE, et al.; for the American Academy of Sleep Medicine. *The AASM Manual for Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specification*. Version 2.0. Darien, IL: American Academy of Sleep Medicine; 2012.

21. Gagnon C, Belanger L, Ivers H, et al. Validation of the Insomnia Severity Index in primary care. *J Am Board Fam Med*. 2013;26(6):701–710.

22. Walsleben JA, Vapur VK, Newman AB, et al. Sleep and reported daytime sleepiness in normal subjects: the Sleep Heart Health Study. *Sleep*. 2004;27(2):293–298.

23. Sleep Profiler: Application and editing of sleep staging. YouTube website. https://www.youtube.com/watch?v=CN-6vvvXdwI&feature=youtu.be. Published December 18, 2015. Accessed March 24, 2017.

24. Whitney CW, Gottlieb DJ, Redline S, et al. Reliability of scoring respiratory disturbance indices and sleep staging. *Sleep*. 1998;21(7):749–757.

25. Rosenberg RS, Van Hout S. The American Academy of Sleep Medicine Inter-Scorer Reliability program: sleep stage scoring. *J Clin Sleep Med*. 2013;9(10):81–87.

26. Moser D. Anderer P, Gruber G, et al. Sleep classification according to AASM and Rechtschaffen and Kales: effects on sleep scoring parameters. *Sleep*. 2009;32(2):139–149.

27. Levendowski DJ, Popovic D, Berka C, et al. Retrospective cross-validation of automated sleep staging using electroocular recordings in patients with and without sleep disordered breathing. *Int Arch Med*. 2012;5(1):21.

28. Stepnowsky C, Levendowski D, Popovic D, et al. Scoring accuracy of automated sleep staging from a bipolar electroocular recording compared to manual scoring by multiple raters. *Sleep Med*. 2013;14(11):1199–1207.

29. Kuna ST, Benca R, Kushida CA, et al. Agreement in computer-assisted manual scoring of polysomnograms across sleep centers. *Sleep*. 2013;36(4):583–539.

30. Punjabi NM, Shifa N, Dorffner G, et al. Computer-assisted automated scoring of polysomnography using the Somnolyzer system. *Sleep*. 2015;38(10):1555–1566.

31. Younes M, Raneri J, Hanly P. Staging sleep in polysomnograms: analysis of inter-scorer variability. *J Clin Sleep Med*. 2016;12(6):885–894.

32. Glassman PA, Good CB, Cunningham FE. Drug alerts and the Goldilocks principle: striving for "just right". *Am J Health Syst Pharm*. 2016;73(8):517.

33. Younes M, Soiferman M, Thompson W, et al. Performance of a new portable wireless sleep monitor. *J Clin Sleep Med*. 2017;13(2):245–258.

34. McCarter SJ, St Louis EK, Duwell EJ, et al. Diagnostic thresholds for quantitative REM sleep phasic burst duration, phasic and tonic muscle activity, and REM atonia index in REM sleep behavioral disorder with and without comorbid obstructive sleep apnea. *Sleep*. 2014;37(10):1649–1662.

35. Younes M, Ostrowski M, Soifeman M, et al. Odds ratio product of sleep EEG as a continuous measure of sleep state. *Sleep*. 2015;38(4):641–654.

36. Younes M, Hanly PJ. Minimizing interrater variability in staging sleep by use of computer-derived features. *J Clin Sleep Med*. 2016;12(10):1347–1356.

37. Wendt SL, Welinder P, Sorensen BD, et al. Inter-expert and intra-expert reliability in sleep spindle scoring. *Clin Neurophysiol*. 2015;126(8):1548–1556.

38. Parekh A, Selesnick IW, Rapoport DM, et al. Detection of K-complexes and sleep spindles (DETOKS) using sparse optimization. *J Neurosci Methods*. 2015;251:37–46.

39. Schwartz JF, Akerstedt T, Lindberg E, et al. Age affects sleep microstructure more than sleep macrostructure. *J Sleep Res*. 2017 Jan 17. [Epub ahead of print]

40. Clawson BC, Kurkin J, Aton SJ. Form and function of sleep spindles across the lifespan. *Neural Plast*. 2016;2016:6936381.

41. Israel B, Buysse DJ, Krafty RT, et al. Short-term stability of sleep and heart rate variability in good sleepers and patients with insomnia: for some measures, one night is enough. *Sleep*. 2012;35(9):1285–1291.

42. Jarrin DC, Chen IY, Ivers H et al. Nocturnal heart rate variability in patients treated with cognitive-behavioral therapy for insomnia. *Health Psychology*. 2016;35(6):638–641.

43. Elsenbruch S, Harnish MJ, Orr WC. Heart rate variability during waking and sleep in healthy males and females. *Sleep*. 1999;22(8):1067–1071.

44. Mahinrad S, Jukema JW, van Heemst D, et al. 10-sec heart rate variability and cognitive function in old age. *Neurology*. 2016;86(12):1120–1127.

ACKNOWLEDGMENTS

SUBMISSION & CORRESPONDENCE INFORMATION

DISCLOSURE STATEMENT