

SCIENTIFIC INVESTIGATIONS

Accuracy of Automatic Polysomnography Scoring Using Frontal Electrodes

Magdy Younes, MD^{1,2}; Mark Younes, BMu²; Eleni Giannouli, MD¹

¹Sleep Disorders Centre, University of Manitoba, Winnipeg, Manitoba, Canada; ²YRT Ltd, Winnipeg, Canada

Study Objectives: The economic cost of performing sleep monitoring at home is a major deterrent to adding sleep data during home studies for investigation of sleep apnea and to investigating non-respiratory sleep complaints. Michele Sleep Scoring System (MSS) is a validated automatic system that utilizes central electroencephalography (EEG) derivations and requires minimal editing. We wished to determine if MSS' accuracy is maintained if frontal derivations are used instead. If confirmed, home sleep monitoring would not require home setup or lengthy manual scoring by technologists.

Methods: One hundred two polysomnograms (PSGs) previously recorded from patients with assorted sleep disorders were scored using MSS once with central and once with frontal derivations. Total sleep time, sleep/stage R sleep onset latencies, awake time, time in different sleep stages, arousal/awakening index and apnea-hypopnea index were compared. In addition, odds ratio product (ORP), a continuous index of sleep depth/quality (Sleep 2015;38:641–54), was generated for every 30-sec epoch in each PSG and epoch-by-epoch comparison of ORP was performed.

Results: Intraclass correlation coefficients (ICCs) ranged from 0.89 to 1.0 for the various sleep variables (0.96 ± 0.03). For epoch-by-epoch comparisons of ORP, ICC was > 0.85 in 96 PSGs. Lower values in the other six PSGs were related to signal artifacts in either derivation. ICC for whole-record average ORP was 0.98.

Conclusions: MSS is as accurate with frontal as with central EEG derivations. The use of frontal electrodes along with MSS should make it possible to obtain high-quality sleep data without requiring home setup or lengthy scoring time by expert technologists.

Keywords: frontal EEG, home sleep studies (HST), odds ratio product (ORP)

Citation: Younes M, Younes M, Giannouli E. Accuracy of automatic polysomnography scoring using frontal electrodes. *J Clin Sleep Med* 2016;12(5):735–746.

INTRODUCTION

It is currently expensive to obtain proper evaluation of sleep in home studies. Central encephalography (EEG) electrodes need to be placed and this cannot be done by the patient. Manual scoring of the data is also time-consuming and expensive. These difficulties limit home studies to investigation of respiratory sleep disorders and, even then, lack of sleep data limits eligible patients to those with high pretest probability,¹ and presents problems in calculating the respiratory disturbance index and in evaluating the effect of disordered breathing on sleep quality.²

Frontal electrodes can be easily applied by the patient. Furthermore, there has been much progress in automatic scoring of sleep with dozens of publications in this area (Penzel et al.³) and three free-standing validated automatic scoring systems have become commercially available.^{4–8} Thus, automatic scoring of frontal EEG signals can potentially simplify home sleep monitoring. A number of investigations reported on the agreement between automatic scoring using facial electrodes (electro-ocular or frontal) and manual scoring.^{9–17} With few exceptions^{11,15,16} these studies involved a small number of normal subjects and in all but one¹⁶ validation was against the scoring of one or two local technologists. Although agreement with manual scoring was acceptable in normal subjects, it deteriorated when significant sleep-disordered breathing was present.^{11,15,16} More importantly, measuring agreement with manual

BRIEF SUMMARY

Current Knowledge/Study Rationale: It is currently not known whether the accuracy of automatic scoring of sleep using frontal electroencephalography electrodes is comparable to that obtained when using the standard central electrodes. Michele Sleep Scoring System (MSS) is a well validated automatic scoring system that was developed using central electrodes, but has not been validated if frontal electrodes were used instead.

Study Impact: This study has shown that results of MSS when using frontal electrodes are comparable to those when using central electrodes. This makes it possible to obtain reliable scoring from frontal electrodes, which can be easily applied by the patient at home, thereby making it less complicated and expensive to obtain information about sleep in home studies.

scoring fails to distinguish between disagreements related to different electrode sites, different scoring techniques (manual versus automatic) or scoring bias by the specific technologists used in validation. Interscorer variability is now well recognized so that the scoring of one or few local technologists to validate automatic systems is no longer satisfactory. Thus, it is currently not clear whether automatic scoring of sleep stages and arousals from frontal EEG derivations alone provides acceptable results in patients with sleep disorders.

Michele Sleep Scoring System (MSS) is a recently introduced automatic system (YRT Ltd, Winnipeg, Canada). In a completely independent multicenter study,⁷ agreement

between its results and the average of 10 experienced scorers in five academic institutions was comparable to or better than between-site agreement and well within the range of agreement between the two scorers in each institution. In addition to providing sleep stages and arousals, MSS also provides the odds ratio product (ORP), a continuous index of sleep depth that is highly correlated with arousability.¹⁸ MSS' algorithms were developed using central signals. Distinction between sleep and wakefulness with MSS is primarily based on average ORP in each epoch.¹⁸ Because ORP is not particularly sensitive to delta power of the EEG,¹⁸ which is the main difference between frontal and central signals, we hypothesized that MSS' results while using frontal signals will be comparable to those obtained from central signals.

The purpose of this study is to compare results of MSS when using frontal versus central EEG signals in a large number (102) of clinical sleep studies. The use of the same automated system for both types of signals eliminates the confounding variables related to use of different scoring systems (manual versus automatic) and, by extension, scoring bias by the validating technologists. Furthermore, because the results of scoring with central signals have been adequately validated by a large number of highly experienced technologists,⁷ agreement between the results with central or frontal electrodes, particularly in a large number of patients with clinical sleep disorders, would provide adequate assurance of the validity of using frontal electrodes.

METHODS

The 102 polysomnogram (PSG) records were the same used in a recent study of MSS.⁸ The studies were recorded using a Sandman system (Natus, San Carlos, CA) and included two central, two frontal, and two occipital EEG channels; two electro-oculograms, electrocardiogram (EKG), chin electromyogram (EMG), and two leg electromyograms; and nasal pressure, thermister, chest, and abdomen bands; oximetry; and audio signals for respiratory monitoring. Manual scoring was performed by one of three very senior technologists according to the 2007 American Academy of Sleep Medicine guidelines.¹⁹

The Sandman records were exported in the European data format (EDF) format and sent electronically to YRT for automatic scoring by MSS. Differences between the version of MSS used in the current study and the version used earlier⁷ were minor and both versions had the same agreement level for five-stage sleep scoring against the stock files used for validating new versions of the software (82.5% in both cases). Scoring was performed twice, once mapping the central electrodes (C3 and C4) and once mapping the frontal electrodes (F3 and F4). All other mapped channels were identical and included chin EMG, two oculograms, EKG, leg EMG, and the various respiratory channels. Two reports were generated that included times in different sleep stages, arousal and awakening index, periodic limb movement (PLM) index and the number, type, and indexes of respiratory events. In addition, an Excel file was generated that listed the average ORP

value in each 30-sec epoch for each of the two central and two frontal electrodes.

Analyses

The automatic scoring was not edited. Of the 102 records, 26 were split studies. In split studies clinical report data (sleep stages, indexes, etc.) were calculated separately for the PSG sections before and after institution of continuous positive airway pressure. For statistical analysis, data obtained from either section were treated as a separate study provided the duration of the section was > 3 h. In 14 split studies the two sections of the PSG were > 3 h. As a result, there were 116 pairs for comparison for each variable. Intraclass correlation coefficients (ICCs) were calculated to determine agreement between central and frontal scoring results and between each automatic scoring and manual scoring. Analysis of variance for repeated measures with Tukey test for multiple comparisons (ANOVA-R) was used to evaluate differences in the average values obtained by the three methods of scoring.

For the ORP data, 30-sec ORP values of the two central signals and the two frontal signals were separately calculated after removing data identified by the scoring system as invalid (no other data were excluded). ICC for all data pairs in each file (approximately 800 epochs/pairs per record) was calculated. Average, standard deviation (SD) and range of ICCs of the 102 records will be reported. In addition, average ORP within each sleep stage, within total sleep time, and for the whole record were calculated. ICC was used to compare these averages with central and frontal signals.

RESULTS

The PSG records included 24 PSGs with mild obstructive sleep apnea (OSA) (apnea-hypopnea index [AHI], 5–15 h⁻¹), 14 with moderate OSA (AHI, 15–35 h⁻¹), 11 with severe OSA (AHI > 35 h⁻¹), 14 with insomnia, and 27 with no pathology. Twenty-three PSGs contained PLMs (> 15 h⁻¹). Of these, 11 also had OSA and are included in the aforementioned OSA groups.

Agreement in Scoring Common PSG Variables from Frontal Versus Central Electrodes

Table 1 shows average results of common PSG variables when scored manually and automatically from central and frontal electrodes. The only significant difference between the two automatic scores was in N3 time, which was overscored by 12 ± 13 min from the frontal electrodes ($p < 0.0001$). Intraclass correlation showed excellent agreement between the two automatic scores with coefficients ≥ 0.95 in eight variables and ≥ 0.88 in the other four (**Table 2**). **Figures 1** and **2** are scatterplots of the values in individual patients.

As reported previously,⁸ there were several significant differences between the average values obtained from manual scoring and from automatic scoring using central signals (**Table 1**). With the exception of a greater difference in N3 time, the differences between manual and automatic scoring using frontal signals were essentially the same as those observed using central signals (**Table 1**). ICCs for manual versus

Table 1—Agreement Between Polysomnography Variables When Scored Manually and Automatically by Central or Frontal EEG Signals (n = 116).

Variable	Central EEG, mean (SD)	Frontal EEG, mean (SD)	Manual Scoring
Total sleep time (min)	246 (95)	249 (94)	262 (95) ^b
Sleep efficiency (%)	68 (18)	69 (18)	73 (16) ^b
Sleep onset latency (min)	23 (30)	23 (34)	21 (23)
REM onset latency (min)	129 (81)	135 (82)	131 (82)
Awake time (min)	112 (71)	109 (70)	95 (64) ^b
Stage non-REM 1 (min)	33 (21)	31 (19)	35 (22) ^b
Stage non-REM 2 (min)	147 (64)	142 (61)	166 (68) ^b
Stage non-REM 3 (min)	32 (37)	44 (44) ^a	21 (31) ^b
Stage REM (min)	34 (30)	32 (30)	41 (30) ^b
Arousal/awakening index (h ⁻¹)	29 (17)	29 (18)	23 (21) ^b
PLM index (h ⁻¹)	12 (21)	12 (21)	12 (20)
AHI (h ⁻¹)	15 (21)	15 (22)	11 (21) ^b

^aSignificantly different from central electrodes by analysis of variance for repeated measures. ^bSignificantly different from both automatic scores by analysis of variance for repeated measures. AHI, apnea-hypopnea index; PLM, periodic limb movement; REM, rapid eye movement.

Table 2—Intraclass correlation coefficient for comparisons between manual scores, auto-scoring using central electrodes and using frontal electrodes (n = 116).

Variable	Manual vs. Unedited Central	Manual vs. Unedited Frontal	Central vs. Frontal
Total sleep time (min)	0.92	0.91	0.98
Sleep efficiency (%)	0.76	0.73	0.96
Sleep onset latency (min)	0.76	0.63	0.94
REM onset latency (min)	0.65	0.66	0.95
Awake time (min)	0.83	0.82	0.97
Stage non-REM 1 (min)	0.63	0.62	0.88
Stage non-REM 2 (min)	0.80	0.76	0.96
Stage non-REM 3 (min)	0.74	0.56	0.90
Stage REM (min)	0.84	0.82	0.94
Arousal/awakening index (h ⁻¹)	0.72	0.74	0.96
PLM index (h ⁻¹)	0.93	0.92	1.00
AHI (h ⁻¹)	0.97	0.96	0.99
Average	0.80	0.76	0.95
SD	0.11	0.13	0.03

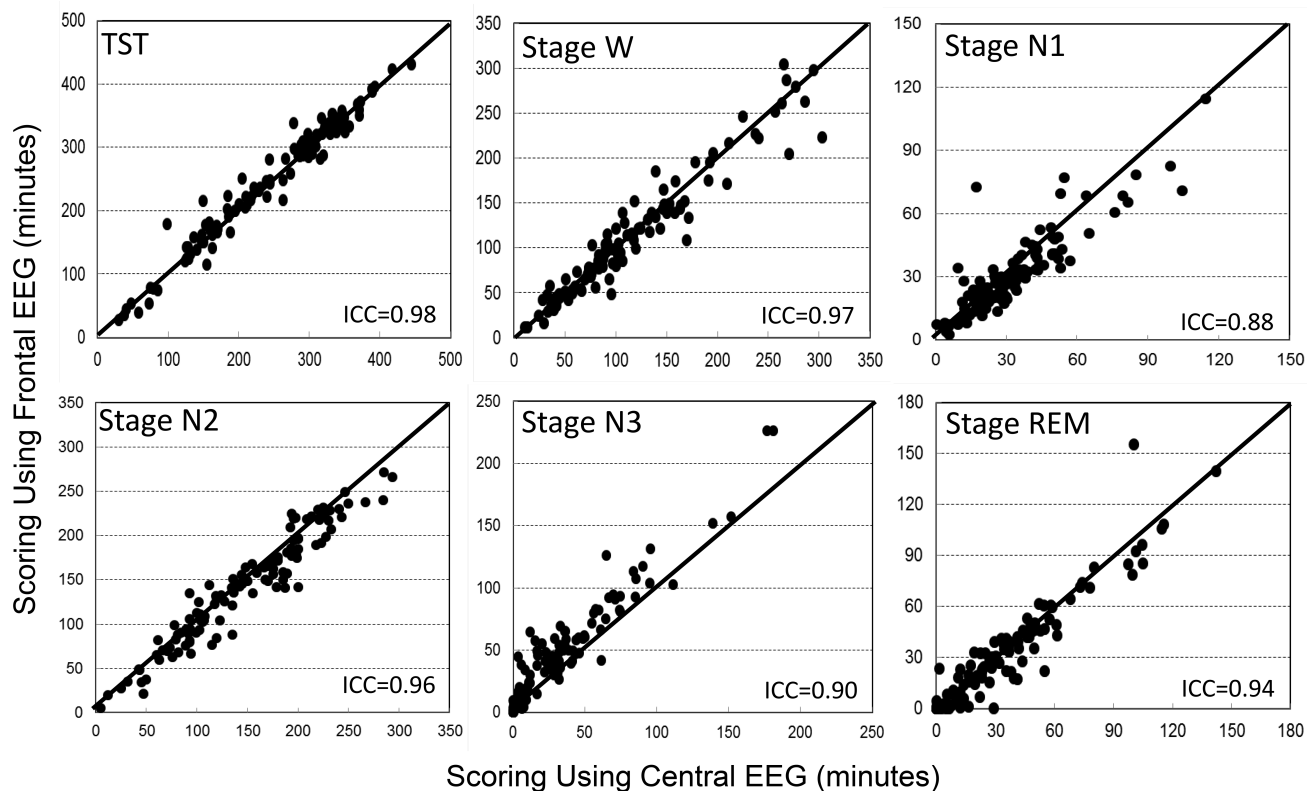
AHI, apnea hypopnea index; PLM, periodic limb movements; REM, rapid eye movement sleep; SD, standard deviation.

frontal scoring were also essentially the same as for manual versus central scoring (**Table 2**) except for some deterioration in agreement in N2 and N3 times, due to N3 time being overestimated at the expense of N2 (**Figure 1**), and a reduction in ICC for sleep onset latency (SOL) from 0.76 to 0.63. The latter change was because of two patients in whom SOL differed substantially between central and frontal scoring (**Figure 2**, SOL). In one, SOL using frontal electrodes was overestimated by 97 min due to missing several epochs of sleep in a patient with severe insomnia (sleep efficiency = 18%). The frontal leads in this patient were noisy. In the other patient, the opposite occurred; SOL was underestimated by 59 min using the frontals due to a clear scoring error (three epochs scored asleep early in the study when the patient was clearly awake and continued to be awake for 1 h after). This was due to a coding bug exposed with the frontal EEG pattern that has since been corrected.

Agreement in Scoring Odds Ratio Product from Frontal Versus Central Electrodes

Whole-PSG average ORP ranged from 0.50 (deep sleep)¹⁸ to 2.15 (nearly awake continuously)¹⁸ in different patients. There was no significant difference between averages of frontal and central ORP (1.11 ± 0.35 versus 1.12 ± 0.35 , $n = 102$ PSGs). **Figure 3** shows the range of ICCs for comparisons between 30-sec ORP values determined from central versus frontal electrodes in individual PSGs. For each PSG, all 30-sec epochs with at least one valid electrode from each EEG pair were included in the comparisons. In more than half the records (60 of 102) ICC was > 0.95. ICC was < 0.90 in 13 records and < 0.80 in only three. An example of the relation between the two values in a record with an ICC of 0.98 is shown in **Figure 4A**. A review of the raw EEG data was undertaken to determine the reasons for differences in the 13 PSGs where ICC was < 0.90.

Figure 1—Scatterplots of the relation between values obtained with use of frontal versus central derivations for different sleep variables.



ICC, intraclass correlation coefficient; N1 to N3, stages non-REM 1 to 3; REM, rapid eye movement sleep; TST, total sleep time; W, stage awake.

In four PSGs, although the results for most epochs were very similar, frontal ORP was appreciably higher (difference > 0.5) than central ORP in an important fraction (**Figure 4B**). EEG in these epochs showed a visibly higher frontal beta activity (15–35 Hz). The increased beta activity was continuous, lasting from a few epochs to 20 min, and involved both frontal signals but not any other signal. An example is shown in **Figure 5**. Because ORP is very sensitive to beta activity,¹⁸ this resulted in a higher ORP value in these epochs.

In two patients central ORP was substantially higher than frontal ORP because of brief (0.5 to 2 sec) bursts in the beta2/gamma range (> 20 Hz) that, again, affected only both central EEG signals. An example of these bursts is shown in **Figure 6**. Three-second epochs including such bursts displayed a high ORP, which raised the average 30-sec ORP by an amount that varied with the frequency and intensity of the bursts. In these two patients the bursts were present throughout sleep time but their frequency varied from zero to six per epoch. This phenomenon accounted for the lowest ICC among the 102 records (0.36). A scatterplot of the data in this patient is shown in **Figure 4C**.

In four of the remaining seven patients, review of the EEG data revealed nothing remarkable. ICCs in these patients were 0.84, 0.88, 0.89, and 0.89. In the last three patients, technical differences between frontal and central electrodes were present in the affected epochs. In one case C3 and F4 were invalid such that the comparison was between C4 and F3 but M1 had more beta noise than M2. In two other files there were large

sweat artifacts in the affected epochs. The ICCs in these three patients were 0.68, 0.79, and 0.82.

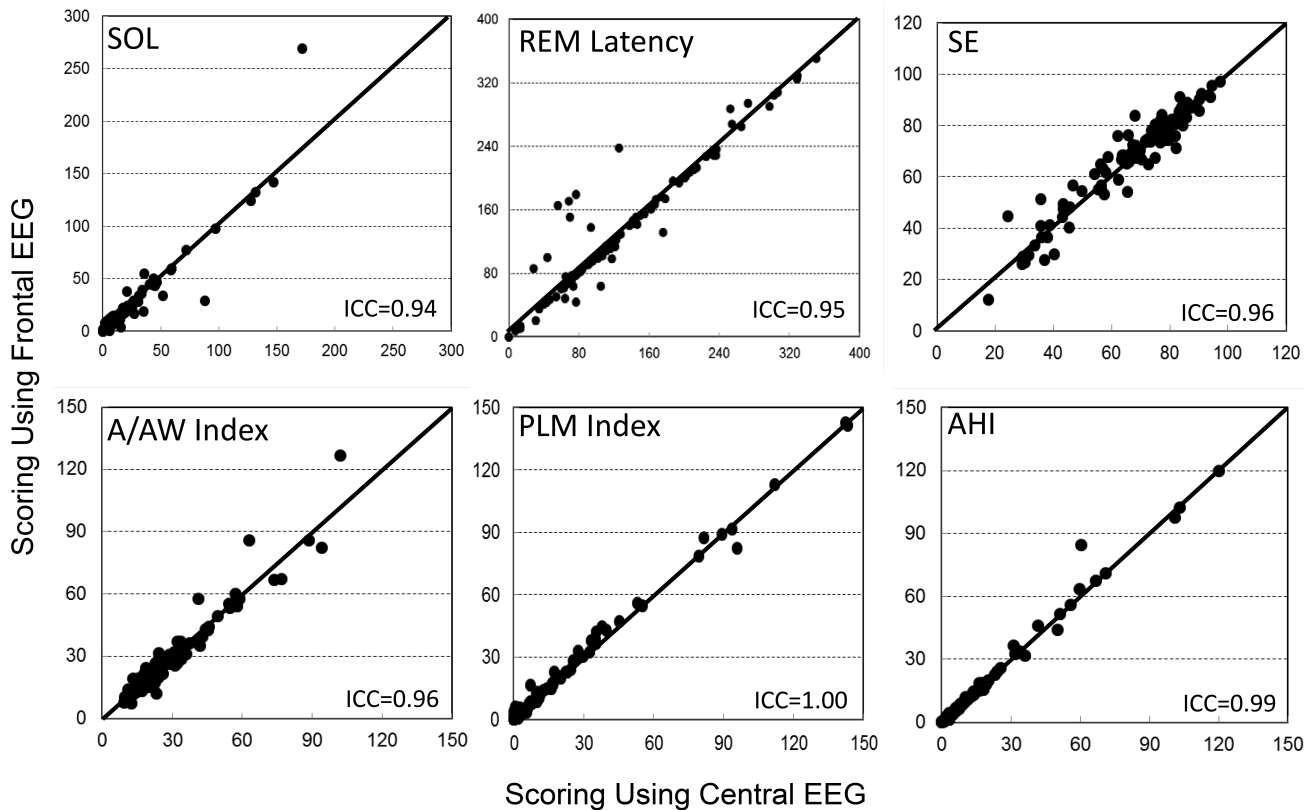
Sleep Stage-Specific ORP Values with Central Versus Frontal EEG

Figure 7 shows scatterplots of stage-specific ORP values calculated from central and frontal EEG signals. There was an excellent correlation between the two values in all stages and in total recording time. **Table 3** shows the average values. There were only very small bidirectional differences. Note that average ORP decreased as stage progressed from wakefulness to stage N3 but that within each stage there was a range of ORP values in different patients. For the dominant stage N2, ORP ranged from 0.3 to 1.6. ORP in stage R sleep was comparable to that in stage N1. **Table 3** also shows the variability index measured from both signals. This index is the average of absolute epoch to next epoch difference in the entire PSG and is a measure of sleep stability. The index ranged from 0.07 to 0.33 in different records and there was excellent agreement between frontal and central estimates.

DISCUSSION

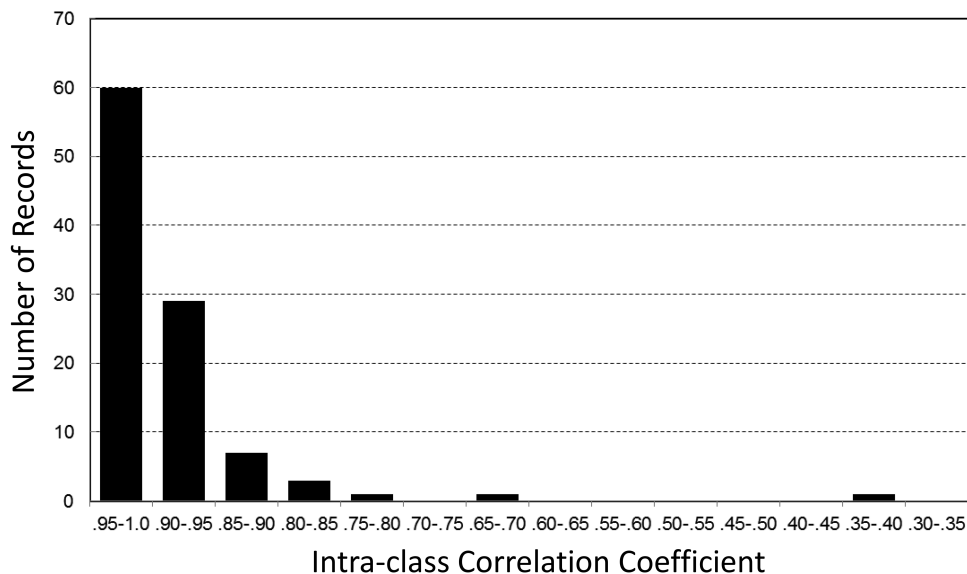
The main finding from this study is that sleep stages, arousal scores, and ORP values scored by MSS are similar when either frontal or central EEG derivations are used. This should

Figure 2—Scatterplots of the relation between values obtained with use of frontal versus central derivations for different polysomnography variables.



A/AW index, arousal and awakening index; AHI, apnea-hypopnea index; ICC, intraclass correlation coefficient; PLM, periodic limb movement; REM, rapid eye movement sleep; SE, sleep efficiency; SOL, sleep onset latency.

Figure 3—Distribution of intraclass correlation coefficients for epoch-by-epoch comparison of the odds-ratio-product in individual polysomnography records.

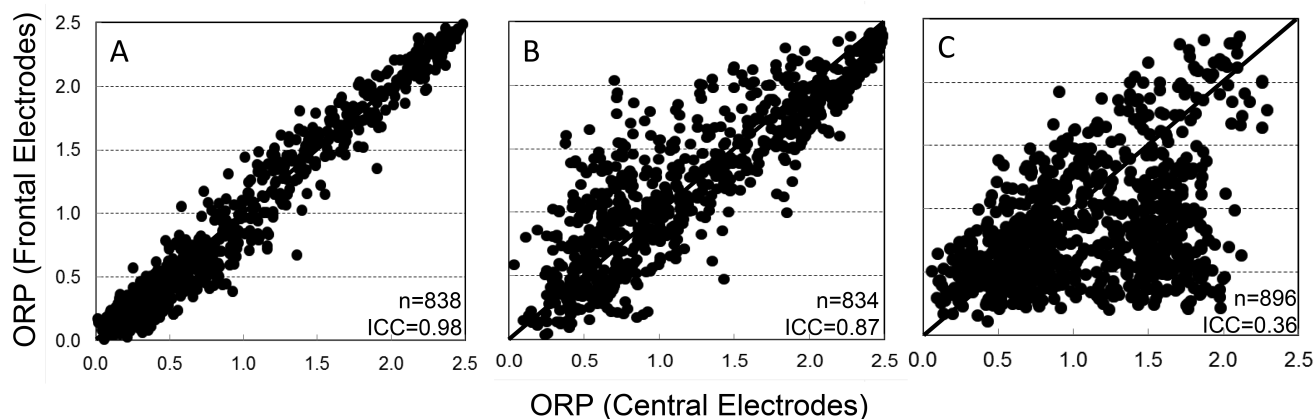


enhance level 3 studies for the diagnosis of sleep-disordered breathing and facilitate home studies for the investigation of non-respiratory sleep disorders.

The Need for Sleep Monitoring in the Home

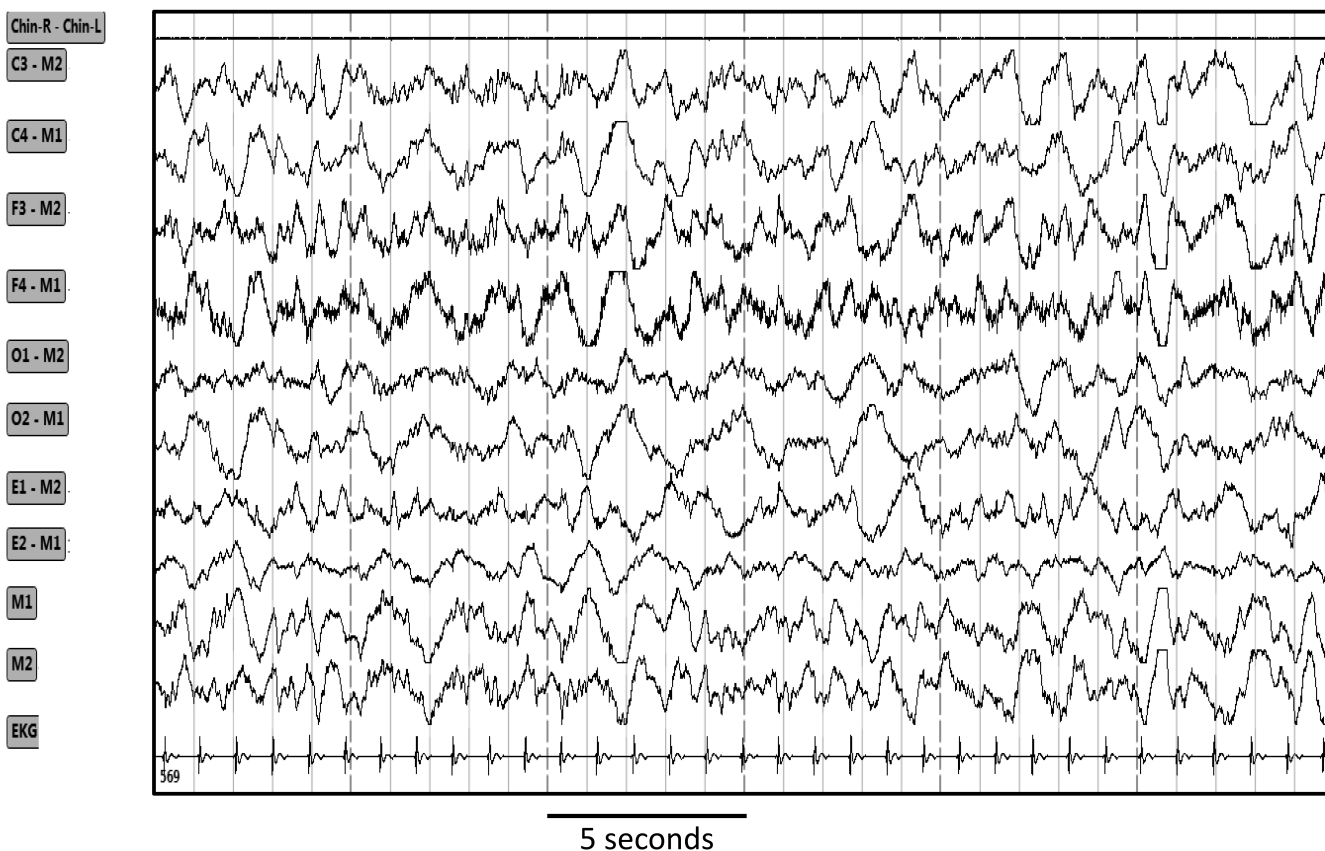
The prevalence of sleep-disordered breathing is very high²⁰ and increasing.²¹ Clinically significant OSA is undiagnosed in most

Figure 4—Examples of the relation between 30-sec values of odds ratio product (ORP) scored with frontal and central derivations in three polysomnography records.



ICC, intraclass correlation coefficient; N, number of 30-sec epochs in each record.

Figure 5—Tracings from one 30-sec epoch showing selective increase in high frequency activity in the frontal EEG electrodes (F3 and F4).

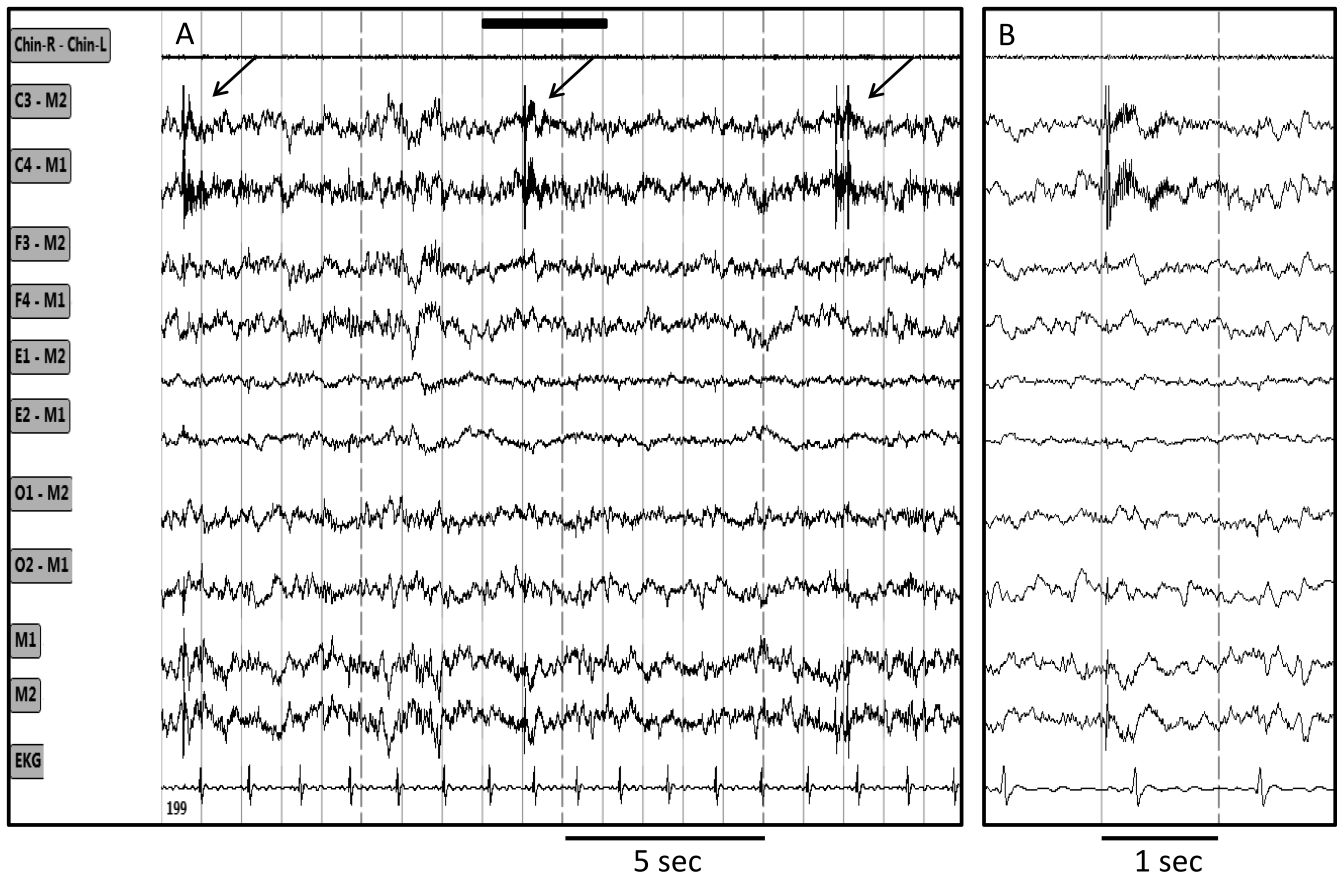


Chin R-Chin L, chin electromyogram; C3, C4, F3, F4, O1, O2, M1, M2 are electroencephalography tracings from left and right central, frontal, occipital and mastoid electrodes; E1 and E2, left and right eye electrodes. EKG, electrocardiogram.

patients.^{20,22} Given the well-established link between OSA and cognitive impairment,²³ and the increasingly convincing evidence that OSA is a risk factor for cardiovascular disease,^{24–26} diabetes,^{27,28} cognitive impairment,²⁹ and mortality,³⁰ there is a serious need to simplify the diagnostic procedure and make it less costly.

Overnight polysomnography remains the gold standard for diagnosis because it evaluates all the variables associated with OSA morbidity, namely effect on sleep quality/architecture and frequency/severity of hypoxemia. There has recently been a dramatic shift from using polysomnography to using home

Figure 6—Tracings from one epoch showing intermittent very high frequency bursts limited to the central electroencephalograph (EEG) derivations.



(A) 20-sec epoch. (B) Faster tracings in the region outlined by the solid bar in A. Chin R-Chin L, chin electromyogram; C3, C4, F3, F4, O1, O2, M1, M2 are electroencephalography tracings from left and right central, frontal, occipital and mastoid electrodes; E1 and E2, left and right eye electrodes. EKG, electrocardiogram.

Table 3—Odds ratio product determined from central and frontal signals in different sleep/wake stages (n = 102).

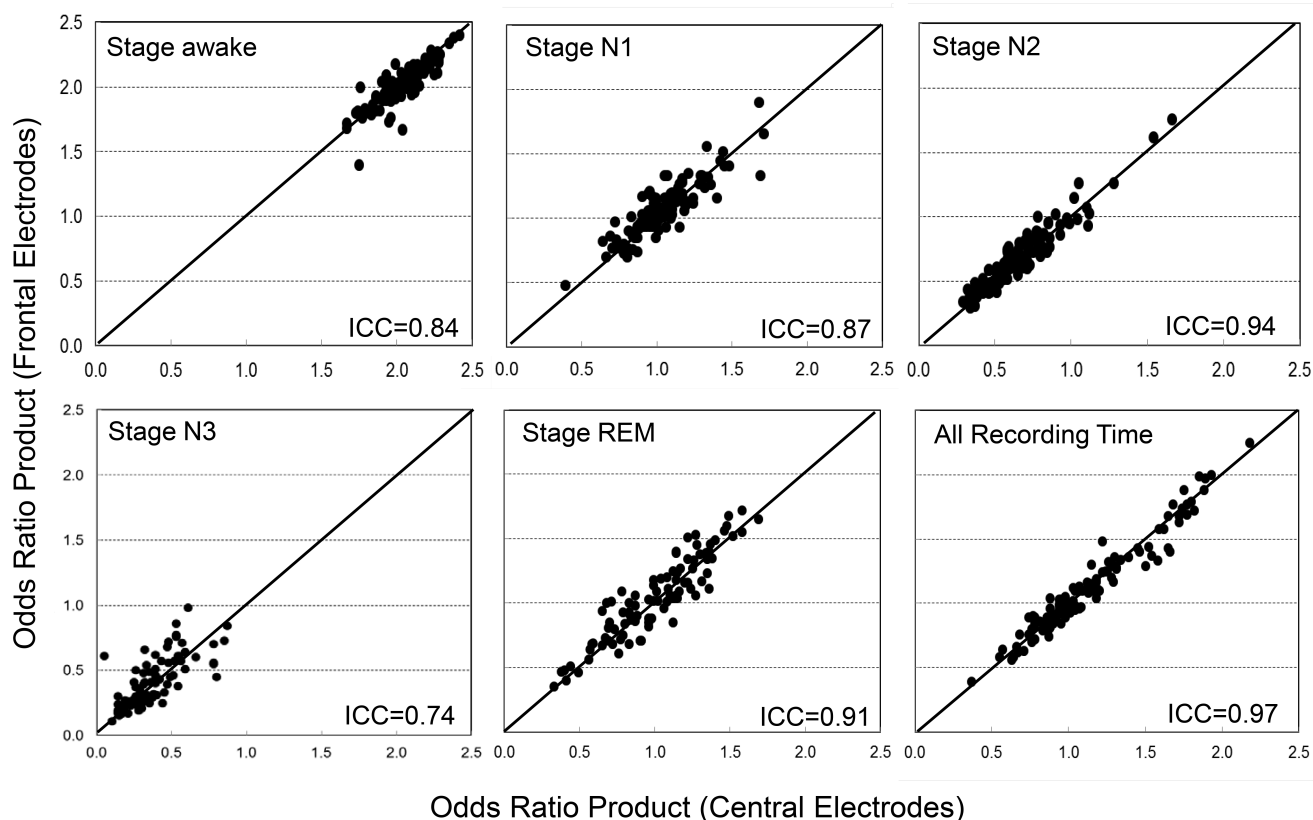
Stage	Central EEG, mean (SD)	Frontal EEG, mean (SD)	Difference, mean (SD)	ICC
Awake	2.06 (0.15)	2.03 (0.16)	0.03 (0.10) ^a	0.84
Non-REM 1	1.04 (0.22)	1.08 (0.22)	-0.04 (0.12) ^b	0.87
Non-REM 2	0.68 (0.23)	0.70 (0.24)	-0.02 (0.08) ^a	0.94
Non-REM 3	0.36 (0.18)	0.40 (0.22)	-0.03 (0.13) ^a	0.74
REM	1.01 (0.32)	1.05 (0.32)	-0.03 (0.13) ^a	0.91
All non-REM	0.70 (0.24)	0.71 (0.25)	-0.01 (0.08)	0.95
All sleep	0.75 (0.26)	0.77 (0.29)	-0.02 (0.19)	0.96
All stages	1.15 (0.38)	1.14 (0.38)	0.01 (0.08)	0.97
Variability index	0.24 (0.04)	0.24 (0.04)	0.00 (0.02)	0.89

ICC, intraclass correlation coefficient; REM, rapid eye movement sleep. ^ap < 0.01; ^bp < 0.001.

monitors with limited channels (respiratory excursions, snoring, saturation of peripheral oxygen [SpO₂], and heart rate). This shift is largely driven by economic considerations with the economy arising primarily from elimination of sleep monitoring because, without sleep monitoring, the patient can apply the device himself or herself at home and there is no need for costly visual sleep scoring. There are several drawbacks to excluding sleep monitoring in home studies²:

1. Respiratory-only monitoring does not permit the diagnosis of a respiratory disorder that is primarily associated with sleep fragmentation (e.g., upper airway resistance syndrome³¹ or hypopneas with arousals without significant desaturation). With such patients, the reason for daytime symptoms would be missed and the patient would not be treated.

Figure 7—Scatterplots of the relation between odds ratio product (ORP) obtained with use of frontal versus central derivations during wakefulness, different stages of sleep, and for the whole record.



Each dot is the average of all ORP values in the indicated stage in one record. N1 to N3, stages non-REM 1 to 3; REM, rapid eye movement sleep; ICC, intra-class correlation coefficient. For reference, ORP > 2.0 is found during wakefulness, ORP between 1.0 and 2.0 represents epochs with both awake and sleep features, and ORP < 1.0 represents stable sleep but with gradually decreasing arousability.¹⁸ Note that the range of ORP decreases progressively as state changes from wakefulness to deeper non-REM stages but that there is a wide range among different subjects within each stage. Note also the very wide range of average ORP during REM sleep and for all recording time (see Table 2 for summary data).

- Limited monitoring becomes economically justifiable only if pretest probability of moderate or severe OSA is high (> 50%).^{1,32} Thus, a patient whose chance of having significant OSA is 4 or 5 in 10 would not be eligible for a home study.
- The absence of a measure of total sleep time leads to variable underestimation of the AHI (the denominator is total recording time and not total sleep time).
- Coexisting sleep disorders such as insomnia or poor quality sleep (unrelated to OSA) would be missed while they may be the main reason for the patient's symptoms.
- Approximately 8% of home studies for OSA are inadequate because of poor signals (oximeter and/or flow signals).³³ Because fluctuations in SpO₂ occur frequently during wakefulness and since, without sleep data, reductions in flow can be interpreted only in light of SpO₂, with limited sensors both oximeter and flow signals must be of good quality for proper interpretation. Addition of sleep monitoring would salvage most of these inadequate studies, obviating the need for a full PSG, because either a decrease in SpO₂ during confirmed sleep or amplitude reduction associated with

arousal can be scored as a hypopnea.³⁴ Although frontal EEG signals may also fail, with addition of two frontal electrodes three of four signals (two EEG electrodes, SpO₂ and flow) instead of one in two signals (SpO₂ and flow) must fail for the study to be written off.

OSA-unrelated sleep disorders such as insomnia (which includes idiopathic sleep fragmentation) are also highly prevalent³⁵ and have been linked to cognitive and memory impairment,^{36,37} mood disorders,^{38,39} weight gain,^{40,41} diabetes,^{42,43} and increased overall mortality.^{44,45} These patients can only be investigated with devices that include sleep monitors. Sleep monitoring that does not require expert technologists to apply electrodes and score the results would make it possible to investigate patients with significant complaints of insomnia or nonrestorative sleep at home. It may also be argued that home studies in such disorders (with easy arousability) may be more representative than in-laboratory PSGs.

Scoring Sleep with Frontal Versus Central Derivations

Since the inception of formalized sleep scoring⁴⁶ central derivations (C3 and/or C4) have been the only or principal

derivations used for scoring sleep and arousals, and the accepted scoring guidelines invariably require the presence of these derivations.^{19,46–48} As a result, it is generally accepted that in-home sleep monitoring must include at least one central derivation. Placement of central electrodes requires home setup by a technologist. In addition to the cost involved in home setup, use of central electrodes necessitates that the monitor is placed remotely from the electrodes. Thus, lengthy wires are required, which adds awkwardness to the procedure and increases the chance of entanglement and dislodging of the electrode. If it were possible to reliably score sleep and arousals from frontal derivations, the patient can easily apply the electrodes to the forehead and, given the advances in microelectronics, the monitor can be small enough to be mounted directly on the forehead (e.g., Sleep Profiler [Advanced Brain Monitoring, Carlsbad, CA]). However, we are not aware of any studies that validated visual sleep scoring using frontal derivations only and such shift from use of frontal instead of central electrodes for visual scoring would require retraining of technologists. Furthermore, continued complete reliance on visual scoring of sleep in home studies would not help reduce the cost of such studies.

It must be emphasized that this study was not done to validate automatic scoring using frontal signals against manual scoring; such validation would require the use of a consensus score of several highly experienced technologists as the reference. This was not the case here. Manual data are given in this report simply as a reference to determine whether agreement between MSS and a single scorer is affected by the EEG derivation used.

As indicated in the Introduction a number of previous studies evaluated the agreement between automatic scoring using frontal signals and manual scoring by one to three scorers.^{9–17} Although some of the reported results are promising, these approaches are far from being available for widespread use in clinical home testing for a number of reasons:

1. With few exceptions,^{11,15,16} these studies used normal subjects and in most of these the number of subjects studied was small. For these automatic techniques to be adopted, large studies involving patients with clinical sleep disorders need to be performed. Data from the three studies that included patients with sleep-disordered breathing^{11,15,16} are not reassuring in this respect.
2. In all but one study¹⁶ only one or two scorers were used as the gold standard. Given the large interscorer variability, such demonstrated agreement simply reflects agreement with one or two local technologists and may not be confirmed when tested against a larger number of independent technologists. Thus, completely different validation studies need to be performed. The study by Stepnowsky et al.¹⁶ used the majority decision of three technologists as a reference, but the number of patients with moderate/severe OSA was small (12 of 44) and the results in these patients were not described in sufficient detail.
3. These studies utilized automatic scoring systems that are built into specific commercial products not certified

for diagnosis of sleep disorders^{9–11,13,14} or were applied using general analytical tools (e.g., MATLAB^{15,17} or unspecified^{12,16}). Assuming the good results are confirmed in larger studies, much development work would need to be done to adapt such automatic systems for use by standard EEG recording equipment.

4. There are no published data on the ability of any of these systems to score respiratory or motor sleep abnormalities. Hence, they must still be adapted to do that if they are to be used for scoring PSGs.
5. Assuming any of these systems passes all the aforementioned tests, the manufacturer still needs to pass all the regulatory requirements.

By contrast, three stand-alone comprehensive automatic PSG scoring systems are currently available commercially and have been shown to provide acceptable sleep scoring.^{4–8} They can be used immediately for scoring from frontal signals if it can be shown that their accuracy is maintained if frontal signals are used. We thought that MSS has this potential for a number of reasons:

1. MSS requires only one derivation (but a second is typically used for back-up).
2. Although the scoring algorithms were developed using central derivations, the algorithms were adjusted empirically so that the results agreed with visual scoring by technologists using the full PSG EEG montage. The use of 10,000 different spectral patterns to identify sleep ensured that any epoch classified as sleep by the technologist using the full montage is accounted for in the database and is recognized as such. In a large multicenter study that was completely independent of the developer,⁷ agreement between unedited total sleep time by MSS and average scores of 10 academic scorers was excellent and similar to agreement between different centers scoring the same PSGs (ICC = 0.87 in both cases).
3. The first step of distinguishing sleep from awake epochs in MSS is based on the ORP score,¹⁸ which is not very sensitive to alpha or delta powers, the main differences between occipital, central, and frontal derivations.⁴⁸ Thus, prominence of delta waves in frontal electrodes is less likely to affect total sleep time if frontal electrodes are used.
4. Identification of the different non-stage R sleep stages (N1 to N3) from central derivations by MSS does not appear to have suffered from lack of occipital and frontal derivations; agreement between MSS and average scores of 10 academic scorers was higher than agreement between scorers in different centers using the full PSG montage (ICC 0.56 versus 0.44 for N1, 0.84 versus 0.61 for N2, and 0.47 versus 0.40 for N3).⁷ Thus, MSS' algorithms appear to be insensitive to the regional differences in the various features used for staging non-stage R sleep.⁴⁸
5. Clinically important scoring weaknesses of MSS have been identified and an editing helper was introduced that identifies potential errors and suggests editing actions.⁸

Exclusively following these suggestions reduced editing time to 6 min, whereas results of the briefly edited records were not different in any important way from those with full editing.⁸ Performance of this editing helper is not affected by the EEG derivation because its assessment is based on the final clinical report (not the EEG pattern), and the editing actions most commonly recommended can be effectively implemented with either central or frontal derivations.

Figures 1 and 2 and **Tables 1 and 2** show clearly that in 116 comparisons results of sleep staging and arousals by MSS using frontal derivations are substantially the same as when the records are scored using central derivations. The main difference is a modest increase in N3 time at the expense of N2 time. As a result, agreement with manual scoring was similar except for some reduction in agreement for N2 and N3 times (**Table 2**). This is unlikely to influence clinical decisions, particularly given the well-known poor interrater agreement in scoring this stage.^{7,49,50} Furthermore, excessive N3 scoring can be corrected, if necessary, with minor adjustment in the algorithm.

SOL was substantially inaccurate in two patients while using the frontal signals (**Figure 2**). However, the reasons for these errors were technical and not specific to the frontal electrodes. Furthermore, the editing helper feature associated with MSS routinely asks the technologist to confirm SOL.⁸

Comparison of ORP Values

It was also important to establish whether ORP values determined from frontal derivations differed from centrally determined values: First, we wished to confirm that the fundamental step used by MSS in sleep scoring is not importantly altered. Second, it was necessary to determine whether the interpretation of ORP values that was based on central derivations¹⁸ still applies when frontal derivations were used. Based on visual sleep scoring and the likelihood of arousals occurring within 30 sec (i.e., arousability) this interpretation is as follows¹⁸: an ORP > 2.0 is almost invariably associated with stage W, an ORP between 1.0 and 2.0 reflects an unstable state with awake and sleep features, whereas an ORP < 1.0 reflects stable sleep but with graded arousability (50% probability of an arousal occurring within 30 seconds at ORP = 1.0 versus < 10% probability at ORP < 0.2). Thus, average ORP during sleep can be used as a reflection of average sleep depth/quality. This average ranges widely among different patients even in the same Rechtschaffen and Kales stage¹⁸ (see also **Figure 7**). Accordingly, ORP results may be helpful in distinguishing reduced quantity versus reduced quality of sleep in patients with nonrestorative sleep. The excellent agreement found in most records in epoch-by-epoch (**Figure 3**) and stage-by-stage comparisons of ORP (**Figure 7**) indicates that interpretations established from central derivations¹⁸ are applicable when using frontal derivations.

Epoch-by-epoch agreement in ORP was < 0.90 in 13 records. In four of these the difference was due to higher beta power limited to the frontal electrodes in some sections of the record (**Figure 5**). Although technical reasons cannot be excluded entirely, this may represent examples of regional differences in

sleep state as described before.^{51–55} In four other records less pronounced differences in ORP were present but were not obvious visually. These may also be less pronounced examples of regional differences in sleep depth. In the remaining five records, the differences were related to technical problems or artifacts affecting either or both derivations. None of these necessarily reflects inferiority of frontal derivations relative to central derivations. Furthermore, these differences lasted for only a fraction of the night and there was excellent agreement, with no evidence of bias, between frontal and central ORP averages in different stages, or in the total recording (**Figure 7**).

CONCLUSIONS

In summary, this study has shown that sleep staging, arousal index, and ORP values obtained from frontal or central EEG derivations are basically similar when the record is scored by MSS. This should make it possible to monitor sleep in the home with devices that can be applied by the patient while the scoring results require minimal editing. Because of anticipated great reduction in the cost of obtaining such information, this may make it possible to economically perform sleep monitoring in the home as an improvement to home studies for diagnosis of OSA or for investigation of nonrespiratory sleep disorders.

ABBREVIATIONS

A/AW index, arousal and awakening index
 AHI, apnea-hypopnea index
 ANOVA-R, analysis of variance for repeated measures
 EDF, European data format
 EEG, electroencephalogram
 EKG, electrocardiogram
 EMG, electromyogram
 HST, home sleep study
 ICC, intraclass correlation coefficient
 MSS, Michele Sleep Scoring
 N1, non-rapid eye movement stage 1
 N2, non-rapid eye movement stage 2
 N3, non-rapid eye movement stage 3
 ORP, odds ratio product
 OSA, obstructive sleep apnea
 PLM, periodic limb movement
 PSG, polysomnogram
 SOL, sleep onset latency
 SpO₂, oxyhemoglobin saturation
 stage R, stage rapid eye movement sleep
 stage W, stage awake

REFERENCES

1. Collop NA, Tracy SL, Kapur V, et al. Obstructive sleep apnea devices for out-of-center (OOC) testing: technology evaluation. *J Clin Sleep Med* 2011;7:531–48.

2. Parthasarathy S. CON: Thoughtful steps informed by more comparative effectiveness research is needed in home testing. *J Clin Sleep Med* 2013;9:9–12.
3. Penzel T, Hirshkowitz M, Harsh J, et al. Digital analysis and technical specifications. *J Clin Sleep Med* 2007;3:109–20.
4. Pittman SD, MacDonald MM, Fogel RB, et al. Assessment of automated scoring of polysomnographic recordings in a population with suspected sleep-disordered breathing. *Sleep* 2004;27:1394–403.
5. Anderer P, Gruber G, Parapatics S, et al. An E-health solution for automatic sleep classification according to Rechtschaffen and Kales: validation study of the Somnolyzer 24 x 7 utilizing the Siesta database. *Neuropsychobiology* 2005;51:115–33.
6. Punjabi NM, Shifa N, Doffner G, Patil S, Pien G, Aurora RN. Computer-assisted automated scoring of polysomnograms using the Somnolyzer System. *Sleep* 2015;38:1555–66.
7. Malhotra A, Younes M, Kuna ST, et al. Performance of an automated polysomnography scoring system versus computer-assisted manual scoring. *Sleep* 2013;36:573–82.
8. Younes M, Thompson W, Leslie C, Egan T, Giannouli E. Utility of technologist editing of polysomnography scoring performed by a validated automatic system. *Ann Am Thorac Soc* 2015;12:1206–18.
9. Sleigh JW, Andrzejowski J, Steyn-Ross A, Steyn-Ross M. The bispectral index: a measure of depth of sleep? *Anesth Analg* 1999;88:659–61.
10. Tung A, Lynch JP, Roizen MF. Use of the BIS monitor to detect onset of naturally occurring sleep. *J Clin Monit Comput* 2002;17:37–42.
11. Nieuwenhuijs D, Coleman EL, Douglas NJ, Drummond GB, Dahan A. Bispectral index values and spectral edge frequency at different stages of physiologic sleep. *Anesth Analg* 2002;94:125–9.
12. Virkkala J, Hasan J, Väri A, Himanen SL, Müller K. Automatic sleep stage classification using two-channel electro-oculography. *J Neurosci Meth* 2007;166:109–15.
13. Agrawal G, Modarres M, Zikov T, Bibian S. NREM sleep staging using WAV(CNS) index. *J Clin Monit Comput* 2011;25:137–42.
14. Shambroom JR, Fábregas SE, Johnstone J. Validation of an automated wireless system to monitor sleep in healthy adults. *J Sleep Res* 2012;21:221–30.
15. Levendowski DJ, Popovic D, Berka C, Westbrook PR. Retrospective cross-validation of automated sleep staging using electroocular recording in patients with and without sleep disordered breathing. *Int Arch Med* 2012;5:21.
16. 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.
17. Popovic D, Khoo M, Westbrook P. Automatic scoring of sleep stages and cortical arousals using two electrodes on the forehead: validation in healthy adults. *J Sleep Res* 2014;23:211–21.
18. Younes M, Ostrowski M, Soiferman M, et al. Odds ratio product of sleep EEG as a continuous measure of sleep state. *Sleep* 2015;38:641–54.
19. Iber C, Ancoli-Israel S, Chesson AL Jr, Quan SF. 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.
20. Young T, Peppard PE, Gottlieb DJ. Epidemiology of obstructive sleep apnea: a population health perspective. *Am J Respir Crit Care Med* 2002;165:1217–39.
21. Peppard PE, Young T, Barnet JH, et al. Increased prevalence of sleep disordered breathing in adults. *Am J Epidemiol* 2013;177:1006–14.
22. Badran M, Yassin BA, Fox N, Laher I, Ayas N. Epidemiology of sleep disturbances and cardiovascular consequences. *Can J Cardiol* 2015;31:873–9.
23. Epstein L, Weiss W. Clinical consequences of obstructive sleep apnea. *Semin Respir Crit Care Med* 1998;19:123–32.
24. Peppard PE, Young T, Palta M, Skatrud J. Prospective study of the association between sleep-disordered breathing and hypertension. *N Engl J Med* 2000;342:1378–84.
25. Gottlieb DJ, Punjabi NM, Mehra R, et al. CPAP versus oxygen in obstructive sleep apnea. *N Engl J Med* 2014;370:2276–85.
26. Haentjens P, Van Meerhaeghe A, Moscariello A, et al. The impact of continuous positive airway pressure on blood pressure in patients with obstructive sleep apnea syndrome: evidence from a meta-analysis of placebo-controlled randomized trials. *Arch Intern Med* 2007;167:757–64.
27. Drager LF, Li J, Reinke C, et al. Intermittent hypoxia exacerbates metabolic effects of diet-induced obesity. *Obesity (Silver Spring)* 2011;19:2167–74.
28. Louis M, Punjabi NM. Effects of acute intermittent hypoxia on glucose metabolism in awake healthy volunteers. *J Appl Physiol* 2009;106:1538–44.
29. Lim DC, Pack AI. Obstructive sleep apnea and cognitive impairment: addressing the blood-brain barrier. *Sleep Med Rev* 2014;18:35–48.
30. Marin JM, Carrizo SJ, Vicente E, Agustí AG. Long-term cardiovascular outcomes in men with obstructive sleep apnoea-hypopnoea with or without treatment with continuous positive airway pressure: an observational study. *Lancet* 2005;365:1046–53.
31. Guilleminault C, Kim YD, Palombini L, Li K, Powell N. Upper airway resistance syndrome and its treatment. *Sleep* 2000;23 Suppl 4:S197–200.
32. Ayas NT, Fox J, Epstein L, et al. Initial use of portable monitoring versus polysomnography to confirm obstructive sleep apnea in symptomatic patients: an economic decision model. *Sleep Med* 2010;11:320–4.
33. Zeidler MR, Santiago V, Dzierzewski JM, Mitchell MN, Santiago S, Martin JL. Predictors of obstructive sleep apnea on polysomnography after a technically inadequate or normal home sleep test. *J Clin Sleep Med* 2015;11:1313–8.
34. Berry RB, Budhiraja R, Gottlieb DJ, et al. Rules for scoring respiratory events in sleep: update of the 2007 AASM Manual for the Scoring of Sleep and Associated Events. Deliberations of the Sleep Apnea Definitions Task Force of the American Academy of Sleep Medicine. *J Clin Sleep Med* 2012;8:597–619.
35. Ohayon M, Reynolds CF, III. Epidemiological and clinical relevance of insomnia diagnoses algorithms according to the DSM-IV and the International Classification of Sleep Disorders (ICSD). *Sleep Med* 2009;10:952–60.
36. Alena A, van der Werf YD, Strijers RL, et al. Sleep loss affects vigilance: effects of chronic insomnia and sleep therapy. *J Sleep Res* 2008;17:335–43.
37. Nissen C, Kloepfer C, Nofzinger EA, et al. Impaired sleep-related memory consolidation in primary insomnia. *Sleep* 2006;29:1068–73.
38. Riemann D, Voderholzer U. Primary insomnia: a risk factor to develop depression? *J Affect Disorders* 2003;76:255–59.
39. Baglioni C, Spiegelhalder K, Lombardo C, et al. Sleep and emotions: focus on insomnia. *Sleep Med Rev* 2010;14:227–38.
40. Patel S, Hu FB. Short sleep duration and weight gain: a systematic review. *Obesity* 2008;16:643–53.
41. Spiegel K, Tasali E, Leproult R, et al. Effects of poor and short sleep on glucose metabolism and obesity risk. *Nat Rev Endocr* 2009;5:253–61.
42. Nilsson PM, Röst M, Engström G, et al. Incidence of diabetes in middle-aged men is related to sleep disturbances. *Diabetes Care* 2004;27:2464–69.
43. Mallon L, Broman JE, Hetta J. High incidence of diabetes in men with sleep complaints or short sleep duration. *Diabetes Care* 2005;28:2762–7.
44. Gallicchio L, Kalesan B. Sleep duration and mortality: a systematic review and meta-analysis. *J Sleep Res* 2009;18:148–58.
45. Cappuccio FP, D'Elia L, Strazzullo P, Miller MA. Sleep duration and all-cause mortality: a systematic review and meta-analysis of prospective studies. *Sleep* 2010;33:585–92.
46. Rechtschaffen A, Kales A. A manual of standardized terminology, techniques, and scoring system for sleep stages of human subjects. NIH Publication. No. 204. Washington, DC: U S Government Printing Office, 1968.
47. American Sleep Disorders Association. EEG arousals: scoring rules and examples. A preliminary report from the Sleep Disorders Atlas Task Force of the American Sleep Disorders Association. *Sleep* 1992;15:174–84.
48. Silber MH, Ancoli-Israel S, Bonnet MH, et al. The visual scoring of sleep in adults. *J Clin Sleep Med* 2007;3:121–31.
49. Danker-Hopfe H, Kunz D, Gruber G, et al. Interrater reliability between scorers from eight European sleep laboratories in subjects with different sleep disorders. *J Sleep Res* 2004;13:63–9.
50. Whitney CW, Gottlieb DJ, Redline S, et al. Reliability of scoring respiratory disturbance indices and sleep staging. *Sleep* 1998;21:749–57.
51. Nobili L, De Gennaro L, Proserpio P, et al. Local aspects of sleep: observations from intracerebral recordings in humans. *Prog Brain Res* 2012;199:219–32.

52. Nir Y, Staba RJ, Andrillon T, et al. Regional slow waves and spindles in human sleep. *Neuron*. 2011;70:153–69.
53. Marzano C, Ferrara M, Moroni F, De Gennaro L. Electroencephalographic sleep inertia of the awakening brain. *Neuroscience* 2011;176:308–17.
54. Ferrara M1, De Gennaro L. Going local: insights from EEG and stereo-EEG studies of the human sleep-wake cycle. *Curr Top Med Chem* 2011;11:2423–37.
55. Marzano C, Moroni F, Gorgoni M, Nobili L, Ferrara M, De Gennaro L. How we fall asleep: regional and temporal differences in electroencephalographic synchronization at sleep onset. *Sleep Med* 2013;14:1112–22.

SUBMISSION & CORRESPONDENCE INFORMATION

Submitted for publication September, 2015

Submitted in final revised form December, 2015

Accepted for publication January, 2016

Address correspondence to: Magdy Younes, MD, 1001 Wellington Crescent, Winnipeg, Manitoba R3M 0A7; Tel: (204) 297-0978; Email: mkyounes@shaw.ca

DISCLOSURE STATEMENT

This was supported by YRT Limited, Winnipeg, Manitoba. Dr. Magdy Younes is the owner of YRT Ltd, the company that developed the automatic scoring system. Mark Younes is an employee of YRT. Dr. Giannouli has received research support from Respiroics and GSK. Mark Younes receives a salary from YRT Limited.