Agreement in the Scoring of Respiratory Events and Sleep Among International Sleep Centers

Ulysses J. Magalang, MD[1]; Ning-Hung Chen, MD[2]; Peter A. Cistulli, MD, PhD[3]; Annette C. Fedson, PhD[4]; Thorarinn Gíslason, MD, PhD[5]; David Hillman, MBBS[6]; Thomas Penzel, PhD[7]; Renaud Tamisier, MD, PhD[8]; Sergio Tufik, MD, PhD[9]; Gary Phillips, MAS[10]; Allan I. Pack, MBChB, PhD[4]; for the SAGIC Investigators

[1]Division of Pulmonary, Allergy, Critical Care, and Sleep Medicine, The Ohio State University Wexner Medical Center, Columbus, OH; [2]Division of Pulmonary, Critical Care, and Sleep Medicine, Chang Gung Memorial Hospital, Taoyuan, Taiwan; [3]Centre for Sleep Health and Research, Department of Respiratory Medicine, Royal North Shore Hospital, and University of Sydney, Sydney, Australia; [4]Center for Sleep and Circadian Neurobiology, Division of Sleep Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA; [5]Department of Respiratory Medicine and Sleep, Landspitali University Hospital, Reykjavik, Iceland; [6]West Australian Sleep Disorders Research Institute, Sir Charles Gairdner Hospital, Nedlands, Australia; [7]Center of Sleep Medicine, Charité University Hospital, Berlin, Germany; [8]Inserm U1042-HP2 Laboratory (Hypoxia & Pathophysiology), Joseph Fourier University, Grenoble, France; [9]Disciplina de Medicina e Biologia do Sono, Departamento de Psicobiologia, Universidade Federal de São Paulo, São Paulo, Brazil; [10]Center for Biostatistics, The Ohio State University, Columbus, OH

Study Objectives: The American Academy of Sleep Medicine (AASM) guidelines for polysomnography (PSG) scoring are increasingly being adopted worldwide, but the agreement among international centers in scoring respiratory events and sleep stages using these guidelines is unknown. We sought to determine the interrater agreement of PSG scoring among international sleep centers.

Design: Prospective study of interrater agreement of PSG scoring.

Setting: Nine center-members of the Sleep Apnea Genetics International Consortium (SAGIC).

Measurements and Results: Fifteen previously recorded deidentified PSGs, in European Data Format, were scored by an experienced technologist at each site after they were imported into the locally used analysis software. Each 30-sec epoch was manually scored for sleep stage, arousals, apneas, and hypopneas using the AASM recommended criteria. The computer-derived oxygen desaturation index (ODI) was also recorded. The primary outcome for analysis was the intraclass correlation coefficient (ICC) of the apnea-hypopnea index (AHI). The ICCs of the respiratory variables were: AHI = 0.95 (95% confidence interval: 0.91-0.98), total apneas = 0.77 (0.56-0.87), total hypopneas = 0.80 (0.66-0.91), and ODI = 0.97 (0.93-0.99). The kappa statistics for sleep stages were: wake = 0.78 (0.77-0.79), nonrapid eye movement = 0.77 (0.76-0.78), N1 = 0.31 (0.30-0.32), N2 = 0.60 (0.59-0.61), N3 = 0.67 (0.65-0.69), and rapid eye movement = 0.78 (0.77-0.79). The ICC of the arousal index was 0.68 (0.50-0.85).

Conclusion: There is strong agreement in the scoring of respiratory events among the SAGIC centers. There is also substantial epoch-by-epoch agreement in scoring sleep variables. Our results suggest that centralized scoring of PSGs may not be necessary in future research collaboration among international sites where experienced, well-trained scorers are involved.

Keywords: General, polysomnography, scoring

Citation: Magalang UJ; Chen NH; Cistulli PA; Fedson AC; Gíslason T; Hillman D; Penzel T; Tamisier R; Tufik S; Phillips G; Pack AI. Agreement in the scoring of respiratory events and sleep among international sleep centers. *SLEEP* 2013;36(4):591-596.

INTRODUCTION

Polysomnography (PSG) is the current gold standard for defining the presence and severity of obstructive sleep apnea (OSA).[1,2] In 2007, the American Academy of Sleep Medicine (AASM) published revised guidelines for the scoring of PSG, including the definitions of respiratory events and sleep stages.[3] Since its publication, these guidelines have been increasingly adopted internationally for scoring PSGs.

Using the AASM guidelines, a previous retrospective study found substantial interrater agreement in the scoring of sleep stages among six scorers from three sleep centers.[4] A major revision within the new guidelines is a recommended defini-

tion for a hypopnea that incorporates a ≥ 4% oxyhemoglobin desaturation associated with the respiratory event.[3] A previous multi-center study showed that the interrater agreement for scoring respiratory events by three sleep technologists at one centralized scoring center was excellent when respiratory indices during sleep were defined by association with an oxyhemoglobin desaturation rather than by presence of associated electroencephalographic (EEG) arousals.[5]

Although centralized scoring of PSGs in multicenter studies has been shown to be an excellent approach in standardizing their analysis,[5] it is also more resource-intensive in terms of infrastructure (sharing of PSGs and server space) and adequate scoring personnel, particularly in research involving patients who are to have PSGs for clinical reasons. The need for these additional resources is likely to be magnified where the multicenter studies include different countries in various continents. Such studies will be increasingly needed with proliferation of large multicenter clinical trials and studies of the genetics of sleep disorders. One alternative approach would be local scoring of the sleep studies by experienced technologists in the sleep centers where the PSGs were acquired as is done in routine clinical practice. However, the inter-rater agreement of scoring respiratory events and sleep stages using the AASM guidelines among international sleep centers is unknown.

We sought to determine the agreement of scoring respiratory events and sleep stages in the nine center-members of the Sleep Apnea Genetics International Consortium (SAGIC) [http://www.med.upenn.edu/sleepctr/SAGIC.shtml]. Our primary hypothesis was that there would be good agreement in the scoring of the apnea-hypopnea index (AHI) among experienced scorers in the SAGIC centers using the new AASM definitions.

METHODS

The SAGIC centers are located in the following cities: Perth, Australia; Sydney, Australia; São Paulo, Brazil; Grenoble, France; Berlin, Germany; Reykjavik, Iceland; Taipei, Taiwan; and Columbus, Ohio and Philadelphia, Pennsylvania in the United States of America. The study was approved by the Institutional Review Board of The Ohio State University Medical Center. Informed consent was waived for this study given that PSG was previously recorded and deidentified.

Polysomnography

Fifteen previously recorded attended, in-laboratory PSGs in the Columbus, Ohio site were chosen by a SAGIC investigator (UM). The studies were randomly selected from a database of clinical PSGs recorded during one quarter, in order to represent a wide spectrum of OSA severity based on AHI: AHI 0-20/h: n = 5; AHI 21-30/h: n = 5; AHI > 30/h: n = 5. Once the required number of studies in each level of OSA severity has been selected, no further studies in the specified AHI range were included. Exclusion criteria included split-night and titration studies, those done while on oxygen supplementation, and less than 4 h of recording time. Once the 15th sleep study was selected, no further sleep studies were evaluated. During this process, there were no studies excluded on the basis of poor signal quality.

All PSGs were originally recorded using an N-7000 amplifier (Embla Systems, LLC, Ontario, Canada). The sampling rates of the channels were: EEG, electrooculogram (EOG), electromyogram (EMG), electrocardiogram (ECG), snore, thermocouple airflow, and nasal pressure: 200 Hz; chest and abdominal movement: 50 Hz; and pulse oximetry: 10 Hz. The pulse oximeter (Nonin Medical, Inc, Plymouth, MN) used in acquiring the PSGs used an averaging algorithm of 3 sec or faster for pulse rates of 60 beats per min or greater. The PSGs were deidentified and then converted into European Data Format (EDF). EDF conversion removed all previous scoring of sleep stages, respiratory events, and sleep technologist notations because the format contains only the biosignals and calibration information.

Scoring

At each SAGIC site, the studies in EDF format were imported into the local software used for scoring. The scoring software included Remlogic (Embla Systems, LLC, Ontario, Canada) [Grenoble, Berlin, and São Paulo, Brazil sites]; Compumedics (Compumedics Limited, Victoria, Australia) [Perth, Sydney, and Taipei, Taiwan sites]; Sandman (Embla Systems, LLC, Ontario, Canada) [Philadelphia, PA site]; Rembrandt (Embla Systems, LLC, Ontario, Canada) [Columbus, OH site]; and Somnologica (Embla Systems, LLC, Ontario, Canada) [Reykjavik, Iceland site]. A protocol including a review of the AASM guidelines on scoring of sleep and respiratory events (in Microsoft Power Point) was provided to one experienced

scorer from each SAGIC site who performed the scoring of the imported PSGs. All scorers have at least 5 yr of experience in scoring clinical PSGs and were designated by the investigator at each site. The recording start time, recording end time, lights-off time, and lights-on time were provided for each study. The channel names of the different signals as well as their derivations were also provided: EEG: F4, C4, O2 all referenced to M1 and F3, C3, O1 all referenced to M2; EOG: E1-M2, E2-M2; chin EMG; snore channels, nasal pressure; thermocouple oronasal airflow; chest and abdominal movements; oxyhemoglobin saturation and plethysmographic signal by pulse oximetry; electrocardiogram; and leg EMG. Each 30-sec epoch was manually scored for sleep stage, apneas, hypopneas, and arousals using the AASM recommended criteria.[3] Briefly, an apnea is scored when all of the following criteria are met: a drop in thermal sensor excursion by $\geq$ 90% of baseline, the duration of the event lasts at least 10 sec, and at least 90% of the event's duration meets the amplitude reduction criteria. An apnea is classified as obstructive if it is associated with continued or increased respiratory effort throughout the entire period of absent airflow, central if it is associated with absent respiratory effort throughout the entire period of absent airflow, and mixed if it is associated with absent respiratory effort in the initial portion of the event, followed by resumption of respiratory effort in the second portion of the event. A hypopnea is scored if the following criteria are met: the nasal pressure signal excursions drop by $\geq$ 30% of baseline, there is a $\geq$ 4% desaturation from pre-event baseline, and at least 90% of the event's duration meets the amplitude reduction criteria. The computer-derived oxygen desaturation index (ODI), defined as the number of oxygen desaturations $\geq$ 4% per h of sleep, was also obtained. All scorers recorded the following variables for each PSG: total sleep time (TST); min Stages N1, N2, N3, nonrapid eye movement (NREM) and Stage R (rapid eye movement, REM); arousal index; AHI; number of apneas; number of obstructive, central, and mixed apneas; number of hypopneas; and ODI. Scoring of sleep stages and arousals all followed the new AASM guidelines.[3] Using the local scoring software available at each site, the epoch-by-epoch sleep stage scoring of each PSG was exported into a text file and then compiled into a Microsoft Excel spreadsheet by a sleep scoring quality assurance program (QSleep, Victoria, Australia).

Sample Size

The primary outcome for the interrater agreement analysis was the intraclass correlation coefficient (ICC) of the AHI. Given nine sleep scorers (one at each site), the 15 PSGs had a power of 83% to detect an ICC of at least 0.90, assuming a null hypothesis of ICC = 0.70.

Statistical Analysis

The interrater reliability measures used to examine the agreement among the nine different scorers were the ICC for continuous variables (respiratory indices, duration of sleep stages, and arousal index) and the kappa (κ) statistic for multiple raters for the categorical variables (sleep stages). The levels of agreement using the ICCs of respiratory indices, duration of sleep stages, and arousal index were classified as follows: 0.00-0.25 = little, 0.26-0.49 = low, 0.50-0.69 = moderate,

Table 1—Interscorer reliability of respiratory events

| Variable | Sleep Center (mean) | | | | | | | | | ICC | 95%CI |
	1	2	3	4	5	6	7	8	9		
AHI (events/h)	25.8 ± 19.4	22.7 ± 18.4	25.7 ± 19.0	25.9 ± 19.8	24.3 ± 17.2	22.7 ± 17.4	23.0 ± 15.4	25.9 ± 21.3	23.6 ± 18.1	0.95	0.91-0.98
Total Apneas	54.1 ± 71.9	31.6 ± 68.2	45.3 ± 66.5	46.3 ± 63.9	42.3 ± 60.3	20.3 ± 41.8	58.1 ± 78.3	104.6 ± 127.7	51.1 ± 72.4	0.73	0.55-0.88
Obstructive	44.1 ± 61.3	28.7 ± 68.2	39.5 ± 67.1	35.9 ± 55.2	37.5 ± 57.1	17.7 ± 36.7	56.4 ± 76.6	100.7 ± 127.7	42.2 ± 68.9	0.70	0.51-0.86
Central	6.7 ± 9.1	2.9 ± 7.1	4.4 ± 7.8	7.3 ± 10.2	1.0 ± 2.0	0.9 ± 1.4	0.9 ± 1.9	3.3 ± 6.2	6.7 ± 7.8	0.46	0.27-0.70
Mixed	3.3 ± 9.5	0.1 ± 0.3	1.5 ± 3.0	3.1 ± 7.5	4.0 ± 7.5	1.7 ± 5.4	0.8 ± 2.3	0.6 ± 1.2	2.2 ± 5.6	0.42	0.24-0.67
Hypopneas	90.1 ± 76.6	100.7 ± 85.7	109.1 ± 85.4	94.8 ± 73.8	108.2 ± 79.4	109.2 ± 87.4	81.1 ± 58.1	51.0 ± 36.9	86.5 ± 75.1	0.80	0.65-0.91
ODI (events/h)	22.4 ± 17.6	22.0 ± 17.5	24.6 ± 18.3	21.6 ± 17.0	25.3 ± 17.5	22.4 ± 18.0	19.7 ± 14.2	22.9 ± 17.4	20.9 ± 16.6	0.97	0.93-0.99

AHI, apnea-hypopnea index; CI, confidence interval; ICC, intraclass correlation coefficient; ODI, oxygen desaturation index.

Table 2—Kappa (κ) statistics evaluating epoch-by epoch sleep staging (NREM combined)

	W	NREM	REM	All stages	Number of epochs
κ	0.78	0.77	0.78	0.78	12,712
95% CI	0.77-0.79	0.76-0.78	0.77-0.79	0.77-0.78	

CI, confidence interval; NREM, nonrapid eye movement; REM, rapid eye movement.

Table 3—Kappa (κ) statistics evaluating epoch-by-epoch staging (NREM separated)

	W	N1	N2	N3	R	All stages	Number of epochs
κ	0.78	0.31	0.60	0.67	0.78	0.63	12,712
95% CI	0.77-0.79	0.30-0.32	0.59-0.61	0.65-0.69	0.77-0.79	0.62-0.63	

CI, confidence interval; NREM, nonrapid eye movement.

0.70-0.89 = strong, 0.90-1.00 = very strong.[6-8] For the κ statistic for multiple raters, the levels of agreement were classified as follows: < 0 = no agreement, 0-0.20 = slight agreement, 0.21-0.40 = fair agreement, 0.41-0.60 = moderate agreement, 0.61-0.80 = substantial agreement, 0.81-1.0 = almost perfect agreement.[9] Data analyses were performed using Stata software version 12 (StataCorp, LP, College Station, TX).

RESULTS

Respiratory Event Scoring

The scored AHI ranged from 0 to 70.9 events/h of sleep. The interscorer agreement of AHI scoring (Table 1) among the SAGIC centers was very strong, with an ICC of 0.95 (95% confidence interval [CI] 0.91-0.98). There was strong agreement in the scoring of the total number of apneas as well as the total number of hypopneas. Table 1 displays the means and the standard deviations (SD) of the scoring of respiratory events in all 15 PSGs. One site (sleep center number six) tended to score the respiratory events as hypopneas, whereas another (sleep center number eight) tended to score the respiratory events as apneas (Table 1). The agreement of ODI scoring was also very strong with an ICC of 0.97 (95% CI 0.93-0.99).

Figure 1 displays all 15 PSG respiratory event scoring results at the different sites. As expected from the ICC values, there were more variations in the scoring of the total apneas and hypopneas compared with the AHI.

The AHI scores were highly correlated with the ODI scores (r = 0.96, P < 0.0001). Using the Bland-Altman method, the mean difference between AHI and ODI scores was +2.1 events/h (limits of agreement: +12.3 and -8.1 events/h).

Sleep Stage Scoring

Table 2 shows the epoch-by-epoch (n = 12,712) agreement in scoring Stages W, NREM, and REM sleep among the nine scorers. There was substantial agreement in the scoring of the major sleep-wake states with κ = 0.78 (95% CI 0.77-0.78). In Table 3, the κ values are shown with NREM sleep separated into different stages. The scoring of Stages N1 and N2 showed only fair (κ = 0.31) and moderate (κ = 0.60) agreements respectively, although there was still substantial agreement in the scoring of all sleep stages with κ = 0.63 (95% CI 0.62-0.63).

Table 4 shows the agreement of the sleep stage duration (min) and arousal index scoring. There was strong agreement in the scoring of the duration of TST, NREM, and REM sleep, but little to low agreement in the scoring of the duration of Stage N1 and N2. There was moderate agreement in the scoring of the arousal index with an ICC of 0.68 (95% CI: 0.49-0.85).

DISCUSSION

This study investigated the agreement of scoring respiratory events and sleep stages, among experienced scorers using the AASM definitions, in nine international sites (members of SAGIC). The major findings of this study are: (1) there is a very
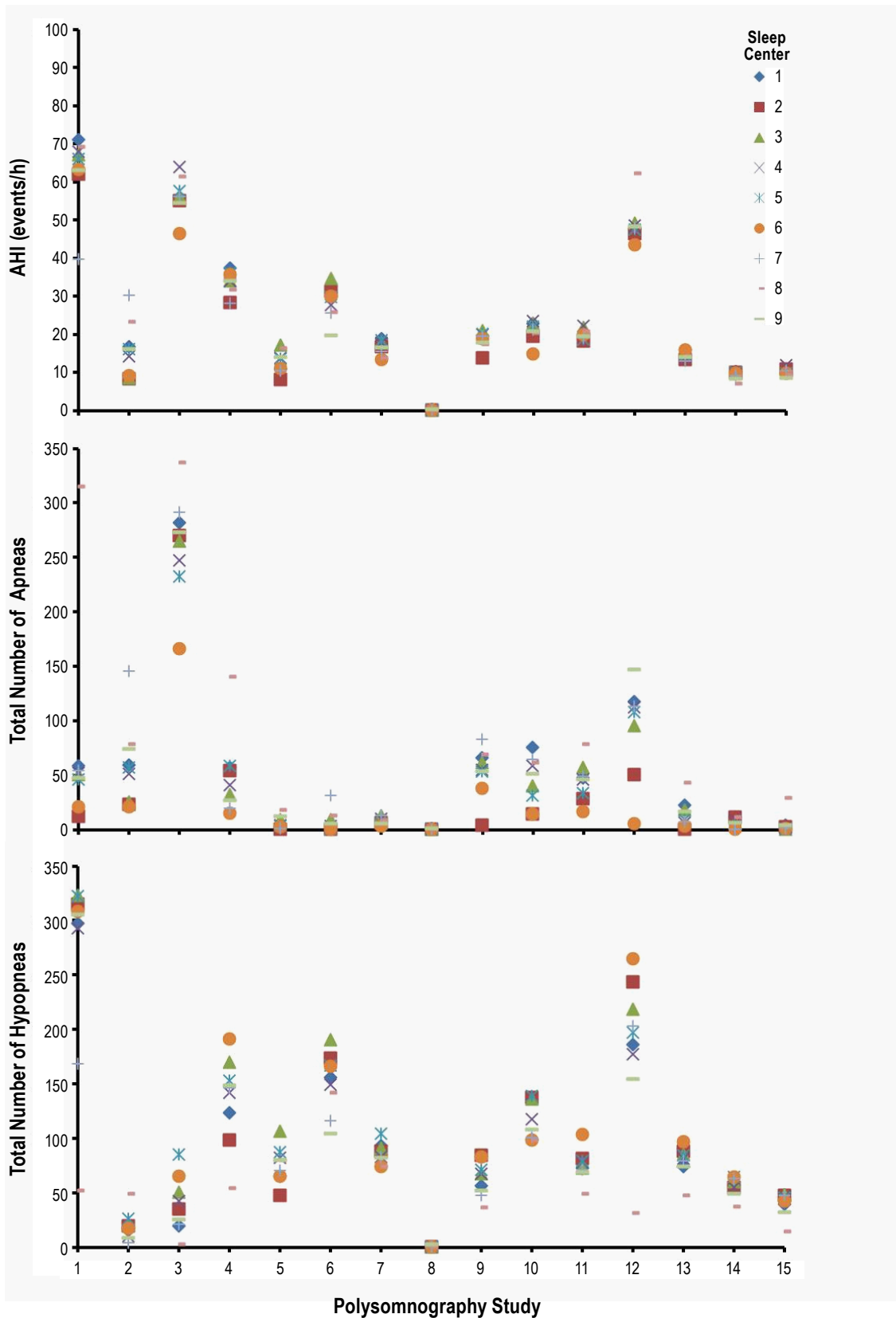
Figure 1—The absolute values of the respiratory event scoring of each of the 15 polysomnographs in the different Sleep Apnea Genetics International Consortium centers are shown. As expected from the intraclass correlation coefficient values, there were more variations in the scoring of the total apneas and hypopneas compared with the apnea-hypopnea index (AHI).

Table 4—Interscorer reliability of sleep stage duration and arousal index

Variable	Sleep Center (mean)									ICC	95%CI
	1	2	3	4	5	6	7	8	9		
TST	342.2 ± 51.6	355.7 ± 46.7	365.2 ± 43.8	336.3 ± 54.5	362.3 ± 44.7	347.4 ± 51.9	370.8 ± 48.7	367.6 ± 45.3	352.2 ± 48.9	0.88	0.76-0.95
NREM	298.3 ± 43.7	301.7 ± 36.6	313.8 ± 35.3	289.7 ± 45.8	318.9 ± 37.7	300.2 ± 42.8	328.7 ± 41.5	297.1 ± 33.6	288.0 ± 42.2	0.77	0.59-0.90
Stage N1	59.2 ± 26.1	35.2 ± 20.4	31.8 ± 17.8	38.8 ± 14.1	43.8 ± 24.7	110.9 ± 43.6	44.4 ± 19.1	55.6 ± 29.4	61.3 ± 25.1	0.21	0.08-0.45
Stage N2	200.7 ± 44.5	238.3 ± 41.1	231.2 ± 26.3	192.9 ± 36.2	210.6 ± 31.5	164.0 ± 58.3	211.7 ± 41.1	180.9 ± 42.3	191.2 ± 39.3	0.41	0.22-0.66
Stage N3	38.4 ± 30.8	28.2 ± 29.8	50.8 ± 37.6	58.0 ± 40.2	64.5 ± 36.3	25.4 ± 18.5	72.6 ± 61.2	60.5 ± 32.2	35.4 ± 26.8	0.62	0.40-0.82
REM	43.9 ± 22.3	54.0 ± 26.8	51.5 ± 29.3	46.6 ± 22.8	43.4 ± 22.8	47.9 ± 26.9	42.0 ± 25.9	70.5 ± 31.2	64.3 ± 31.2	0.78	0.59-0.91
Arousal Index (events/h)	38.0 ± 16.4	30.2 ± 19.4	25.9 ± 17.7	35.1 ±18.0	19.9 ± 11.0	25.4 ± 16.3	22.5 ± 16.1	30.9 ± 18.0	26.3 ± 16.6	0.68	0.49-0.85

CI, confidence interval; ICC, intraclass correlation coefficient; NREM, nonrapid eye movement; REM, rapid eye movement; TST, total sleep time.

strong agreement in the scoring of the AHI among experienced sleep technologists using the AASM recommended definitions of respiratory events; (2) there is strong agreement in the scoring of the total number of apneas and hypopneas; and (3) there is substantial epoch-by-epoch agreement in scoring wake, NREM, and REM sleep. The computer-assisted scoring of the ODI also showed very strong agreement despite using various scoring software to derive this index. To our knowledge, this is the first study examining the interscorer agreement of PSG scoring of respiratory events and sleep among international sleep centers based on the 2007 AASM guidelines. The issue of agreement is not only important to researchers who are planning multicenter collaborative efforts but also to clinicians because such information is essential for evaluation of the findings and interpretation. The use of different scoring software across the nine centers is a strength of this study as it provides a greater challenge to good agreement and is a real-world application.

A previous study reported by Whitney and colleagues[5] showed that the interrater agreement of scoring respiratory indices by three sleep technologists at one centralized scoring center was excellent when respiratory indices were defined in association with an oxyhemoglobin desaturation rather than the presence of associated EEG arousals. Our study extends this finding to international sleep centers and showed that using the 2007 AASM recommended scoring guidelines (which bases the hypopnea definition on the occurrence of a ≥ 4% oxyhemoglobin desaturation), there is very strong agreement in AHI scoring (ICC = 0.95; 95% CI 0.91-0.98). This is similar to the ICC of 0.99 (95% CI not reported) in the prior study of three scorers using the same hypopnea definition.[5] In our current study, the agreement in the scoring of central and mixed apneas was lower than obstructive apneas with an ICC of 0.46 and 0.42, respectively. The reasons for these findings are unclear, but suggest that it is an area that needs to be emphasized in ongoing quality improvement of PSG scoring among the SAGIC centers.

We also found similar levels of agreement in regard to sleep stage scoring with a previously reported κ of 0.81-0.83 among two scorers in the study by Whitney et al.[5] compared with the κ of 0.78 observed among the nine scorers in our study. It should be pointed out that this previous study used the Rechtshaffen and Kales (R & K) criteria in scoring sleep stages because it was conducted prior to the publication of the 2007 AASM scoring guidelines. With regard to sleep staging, our results indicated that the lowest agreement is in the scoring of Stage N1, a

finding also reported in the study by Whitney et al. The interrater agreement of sleep stage scoring was also previously examined by Danker-Hopfe and colleagues[4] in 12 PSGs among six scorers from three sleep centers located in Austria and Germany. These investigators found a substantial agreement in sleep stage scoring using the AASM guidelines (0.61 < κ < 0.80) and their study suggested that the interrater agreement was higher when using the AASM guidelines compared with the R & K criteria. In addition, they also found the lowest agreement in Stage N1 scoring. Unlike our study, however, these investigators also found a low interrater agreement in Stage N3 scoring.

Our study has some limitations. First, we only included one experienced scorer at each site. Additional research would be needed to determine PSG scoring agreement if multiple scorers from each center would be involved in SAGIC research projects. Second, we did not determine the intrascorer reliability of respiratory event and PSG scoring. Nonetheless, the previous study by Whitney and colleagues[5] suggests that intrascorer agreement of respiratory indices and sleep staging was also excellent. Third, the 15 PSGs used in this study were attended, in-laboratory sleep studies performed in one member center. However, data acquisition in these PSGs used standard techniques that would not be expected to be different given the equipment used at the nine SAGIC sites. It is not clear whether our findings can be generalized to unattended PSGs. Fourth, we did not examine the scoring of periodic limb movements because our aim was primarily to assess the interrater agreement of scoring of respiratory indices and sleep stages. Finally, our current findings should be viewed within the context of new AASM respiratory scoring rules that will soon be published that call for a change in hypopnea definition and stipulate the following: a drop of peak signal excursion by ≥ 30% of pre-event baseline using the nasal pressure signal for a duration of ≥ 10 sec with a ≥ 3% oxygen desaturation from pre-event baseline or an associated arousal. The effect of the upcoming change in hypopnea definition (which incorporates the presence of arousals) on the interrater reliability of scoring respiratory events among international centers is yet to be examined.

There are several key implications in research collaboration continuing from the results of this important study. Although there was very strong agreement in the scoring of the AHI, one site (sleep center number six) tended to score the respiratory events as hypopneas, whereas another (sleep center number eight) tended to score the respiratory events as apneas (Table 1).

We also found lower levels of agreement in the scoring of specific NREM sleep stages, particularly stage N1 (Table 3). Findings from this initial project will help in the ongoing quality improvement of PSG scoring among the SAGIC centers because they suggest that efforts in scoring education among the sleep technologists should be centered on the differentiation between apneas and hypopneas, identification of central apneas, and improving recognition of sleep stage N1.

OSA is a highly prevalent condition not only in Western countries but also worldwide.[10-12] The condition is associated with a variety of adverse consequences including hypertension, stroke, cardiovascular disease, and motor vehicle crashes.[13-16] There is a growing interest in performing multicenter studies involving patients with OSA at the international level.[17] Centralized scoring of PSGs in multicenter studies involving different countries in various continents will require a greater magnitude of resources that may not always be available to investigators. An alternative approach would be scoring of the sleep studies locally in the sleep centers where the PSGs are acquired. Our results suggest that this is feasible in future research collaboration among national and international groups such as SAGIC, particularly in studies where the AHI, ODI, and duration of sleep are the primary variables of interest. Additional steps required to ensure standardization of data collection would include ensuring similar procedures (such as sampling rates and filters) and comparable pulse oximeters[18] are used at each center, and technically adequate signals are obtained during the data acquisition of PSGs.

In summary, our findings show a substantial agreement in the scoring of the respiratory events as well as sleep stages among international sleep centers working in the SAGIC consortium using the 2007 AASM guidelines where experienced sleep technologists are involved. Our results suggest an alternative approach to centralized scoring of PSGs in research collaboration among national and international centers, particularly if the AHI, TST, and ODI are the primary variables of interest. This approach could be used in multicenter studies providing a similar scoring validation is conducted.

ACKNOWLEDGMENTS

DISCLOSURE STATEMENT

REFERENCES

1. Kushida CA, Littner MR, Morgenthaler T, et al. Practice parameters for the indications for polysomnography and related procedures: an update for 2005. Sleep 2005;28:499-521.
2. Sleep-related breathing disorders in adults: recommendations for syndrome definition and measurement techniques in clinical research. The Report of an American Academy of Sleep Medicine Task Force. Sleep 1999;22:667-89.
3. Iber C, Ancoli-Israel S, Chesson A, Quan SF, eds. 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.
4. Danker-Hopfe H, Anderer P, Zeitlhofer J, et al. Interrater reliability for sleep scoring according to the Rechtschaffen & Kales and the new AASM standard. J Sleep Res 2009;18:74-84.
5. Whitney CW, Gottlieb DJ, Redline S, et al. Reliability of scoring respiratory disturbance indices and sleep staging. Sleep 1998;21:749-57.
6. Munro BH. Statistical methods for health care research. 5th ed. Philadelphia: Lippincott Williams Wilkins, 2005:248-9.
7. Cheng JW, Tsai WC, Yu TY, Huang KY. Reproducibility of sonographic measurement of thickness and echogenicity of the plantar fascia. J Clin Ultrasound 2012;40:14-9.
8. Portney LG, Watkins MP. Foundations of clinical research. Applications to practice. New Jersey: Prentice Hall Inc., 2000:560-7.
9. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159-74.
10. Young T, Palta M, Dempsey J, Skatrud J, Weber S, Badr S. The occurrence of sleep-disordered breathing among middle-aged adults. N Engl J Med 1993;328:1230-5.
11. Davies RJ, Stradling JR. The epidemiology of sleep apnoea. Thorax 1996;51:S65-70.
12. Lam B, Lam DC, Ip MS. Obstructive sleep apnoea in Asia. Int J Tuberc Lung Dis 2007;11:2-11.
13. Young T, Peppard P, Palta M, et al. Population-based study of sleep-disordered breathing as a risk factor for hypertension. Arch Intern Med 1997;157:1746-52.
14. Yaggi HK, Concato J, Kernan WN, Lichtman JH, Brass LM, Mohsenin V. Obstructive sleep apnea as a risk factor for stroke and death. N Engl J Med 2005;353:2034-41.
15. Somers VK. Sleep--a new cardiovascular frontier. N Engl J Med 2005;353:2070-3.
16. George CF. Reduction in motor vehicle collisions following treatment of sleep apnoea with nasal CPAP. Thorax 2001;56:508-12.
17. McEvoy RD, Anderson CS, Antic NA, et al. The sleep apnea cardiovascular endpoints (SAVE) trial: Rationale and start-up phase. J Thorac Dis 2010;2:138-43.
18. Zafar S, Ayappa I, Norman RG, Krieger AC, Walsleben JA, Rapoport DM. Choice of oximeter affects apnea-hypopnea index. Chest 2005;127:80-8.