JCSM
Journal of Clinical Sleep Medicine

SCIENTIFIC INVESTIGATIONS

How Many Polysomnograms Must Sleep Fellows Score Before Becoming Proficient at Scoring Sleep?

Alejandro Chediak, M.D.[1,2]; Belen Esparis, M.D.[1,2]; Richard Isaacson, M.D.[2]; Luis De la Cruz, M.D.[2]; José Ramirez, M.D.[2]; Juan Francisco Rodriguez M.D.[2]; Alexandre Abreu, M.D.[2]

[1]*Miami Sleep Disorders Center, Miami, FL;* [2]*Mount Sinai Medical Center, Miami Beach, FL*

Objectives: In the field of sleep medicine, there is a paucity of evidence-based curriculum development strategies in the literature. We chose to determine the number of polysomnograms (PSG) necessary to be scored by sleep fellows in order to reasonably approximate sleep scoring by a Diplomate of the American Board of Sleep Medicine (DABSM).

Design: The fifth PSG scored by two sleep fellows during the 12 consecutive months of training was chosen for analysis. A DABSM not involved in the training of fellows scored sleep on each of the selected PSG with replication of montage and filter settings. Epoch by epoch comparison of sleep stage scoring is described as the frequency of concordance between fellow and DABSM (f correct).

Measurements And Results: The mean (SD) f correct for all PSG for each fellow was 0.83 (0.06) and 0.83 (0.08) ($p = 0.93$). Concordance between sleep fellow and DABSM approached 0.8 after scoring between 20-30 PSG. This milestone was reached after the fourth month of training. F correct was highest for stage 2 sleep and REM sleep and most variable for slow wave sleep and stage 1 sleep. The variability in f correct for these stages was in part related to the relative paucity of these sleep stages.

Conclusions: Scoring of sleep becomes reasonably proficient after scoring approximately 20-30 PSG and/or four months of dedicated sleep disorders training. A standard measure of concordance that corrects for epoch sample size may be helpful for use in future similar investigations.

Keywords: Sleep scoring, sleep fellowship training, sleep scoring variability, sleep scoring accuracy

Citation: Chediak A; Esparis B; Isaacson R et al. How many polysomnograms must sleep fellows score before becoming proficient at scoring sleep? *J Clin Sleep Med* 2006;2(4):427-430.

In the field of sleep medicine, there is a paucity of evidence-based curriculum development strategies in the literature. In terms of clinical education in graduate medical education, Winkelman describes his approach toward teaching psychiatry residents about sleep and its related disorders.[1] This curriculum assumes that residents have little prior exposure to sleep physiology or the treatment of sleep disorders and was designed to provide a scientific understanding of sleep physiology and its assessment, as well as practical clinical advice for the treatment of patients and for the residents' Board examinations.[1] Sections of his curriculum are based on the National Center on Sleep Disorders Research core competencies for undergraduate medical education.[2] Strohl and colleagues sought to develop a competency-based undergraduate curriculum in sleep medicine for medical students. They utilized a multidisciplinary expert-opinion approach to assess and define education objectives and the potential for implementation.[2]

While one study developed a comprehensive action plan for integrating sleep topics into the medical school curriculum[3], there have been, to date, no published development strategies for a sleep-study curriculum for sleep fellows. Scoring sleep recordings remains a crucial skill typically acquired during the course of a sleep fellowship training program. The optimal mechanism for acquiring the necessary skills to be proficient at scoring sleep in sleep-disorders postgraduate medical education has yet to be defined. It was our goal to determine the number of sleep studies necessary to be scored by a sleep fellow in order to reasonably approximate sleep scoring by a Diplomate of the American Board of Sleep Medicine (DABSM).

STUDY DESIGN AND METHODS

The sleep disorders fellowship training program at Mount Sinai Medical Center consists of a 1-year dedicated training track accredited by the American Academy of Sleep Medicine. Sleep disorders fellowship trainees are assigned to epoch-by-epoch polysomnography scoring 1 day weekly throughout the entire year of the fellowship training experience. Polysomnography scored by a sleep fellow is reviewed with a DABSM the day after scoring. The laboratory technical supervisor instructs sleep fellows on the use of our digital polysomnography system (SomnoStar Pro® version 7.2, Viasys SensorMedics, Yorba Linda, Calif) and observes sleep fellows while they score polysomnography during the first week of their training. Thereafter, the laboratory supervisor serves as a resource when the DABSM is not available to answer questions.

Disclosure Statement

Drs. Chediak, Esparis, Isaacson, De la Cruz, Ramirez, Rodriguez, and Abreu have indicated no financial conflicts of interest.

Submitted for publication March 15, 2006
Accepted for publication April 7, 2006

Address correspondence to: Alejandro Chediak, MD, Miami Sleep Disorders Center, 7029 SW 61 Avenue, Miami, FL 33143; Tel: (305) 666-2224; Fax: (305) 666-2297; E-mail: MiamiSleep@BellSouth.net

The sleep-disorders fellows who participated in this investigation had just completed a 3-year pulmonary-critical care fellowship at the University of Miami that included a 3-week rotation in our sleep disorders laboratory. The rotation was clinically oriented, and pulmonary fellows were assigned 3 to 4 new and 3 to 4 follow-up patient encounters weekly. The encounters were attended by the sleep-disorders fellowship program director. The technical experience was limited to completely scoring 2 polysomnograms by the end of the rotation. These fellows rotated through to the sleep laboratory at a time that the facility was using analog paper polysomnography recordings, whereas when the sleep disorders fellows were in training, the laboratory had transitioned into full digital polysomnography. Hence, the exposure to scoring polysomnography prior to the sleep disorders fellowship was nominal.

The sleep-disorders laboratory uses identical equipment and montages for recording and reviewing sleep-related physiologic waveforms. Sleep-stage scoring is performed using identically configured Dell computers and 20-inch high-resolution monitors. Scoring PSG is conducted systematically and repeatedly using the following protocol. Firstly, sleep and arousal are scored using a montage that limits the visible waveforms to only those necessary to stage sleep (C3-A2, C4-A1, O1-A2, O2-A1, right and left electrooculogram, and digastric electromyogram). Filter settings and sensitivities are preset within the montage and are not routinely changed during the scoring of polysomnography.

To minimize selection bias, we chose the fifth polysomnogram scored by each fellow during each month of training. We chose the fifth study of each month to allow the fellows to become familiar with the SomnoStar Pro® software used for scoring studies prior to the first data point of analysis. Using 1 of the sleep laboratory computer systems, a DABSM, who was not involved in the training of the fellows, scored sleep on each of the selected polysomnograms, again limiting the screen view to those waveforms necessary for sleep-stage scoring and without modifying filter settings and sensitivity. It was deemed that this retrospective selection with prospective scoring would allow for the most representative real-world evaluation of the effect of our training on our fellows' proficiency at sleep-stage scoring. The number of polysomnograms scored by fellows between sampled studies was collected to allow for comparison across both time and number of studies scored.

SomnoStar Pro® allows for epoch-by-epoch comparison of sleep-stage scoring, comparison of latency to sleep stages, and more global quantitative measures of sleep. Epoch-by-epoch comparison of sleep-stage scoring is described as the frequency of concordance between fellow and DABSM (f correct). This index is calculated by the quotient of the number of epochs in which the fellow and DABSM were in agreement and the total number of epochs recorded in the polysomnogram. We recognize that, with sleep stages of relatively lesser occurrence, f correct overestimates the lack of concordance, but since frequency of concordance is the customary measure of accuracy used to proficiency of scoring sleep by sleep technologists in American Academy of Sleep Medicine-accredited sleep disorders centers, we chose to not introduce a correction.

Unless specified otherwise, data are expressed as mean ± SD. Statistical analysis on absolute measures of sleep time was accomplished by the 2-tailed variant of the Student paired t test.

RESULTS

Polysomnography consisted of 24 samples collected over 12 months. All polysomnograms were done in the course of investigation for sleep apnea. Mean (SD) for age for patients scored by fellows A and B was 50 (15) and 53 (12) years, respectively. Two patients had concomitant complaints of insomnia (1 scored by each fellow) and 1 other patient was also diagnosed with periodic limb movement disorder. Selected demographic and clinical features of the subjects whose polysomnograms were used for the analysis by each fellow can be found in the Table.

The mean sleep-onset latency for fellows A and B versus the DABSM reference was 13.13 (10.70) versus 14.38 (13.80) minutes and 11.88 (11.22) versus 14.04 (12.13) minutes, respectively ($p > .05$ for both fellows). For all polysomnograms combined, total sleep time scored by fellow A was 343.42 (78.24) minutes, whereas the DABSM recorded 331.38 (73.53) minutes, $p = .08$. Fellow B scored 328.25 (71.51) minutes of sleep, compared with 317.13 (66.87) minutes of total sleep time recorded by the DABSM, $p = .01$. The mean f correct for all polysomnograms for fellow A was 0.83 (0.06), whereas fellow B achieved overall concordance of 0.83 (0.08); values were not statistically significant from each other when compared by unpaired 2- tailed t-test. ($p = .93$).

Concordance (f correct) of behavioral stages as a function of the number of polysomnography scored by a fellow can be found in Figures 1 to 4. Concordance between fellows and DABSM approaches 0.8 after scoring between 20 and 30 polysomnograms, a number that was achieved by approximately the fourth month of training. Thereafter, there is continued gradual improvement, and at the end of the 1-year fellowship training program, mean concordance for the year was 0.83 for both fellows. Significant variability in concordance is noted for scoring non-rapid eye movement (NREM) stage 1 sleep (Figure 2) and NREM stage 3 + 4 (Figure 3). Longitudinal analysis of concordance in rapid eye movement (REM) sleep (Figure 4) exhibited slightly more variability than that observed for NREM stage 2 sleep (Figure 2). The variability in scoring sleep for these stages is in part related

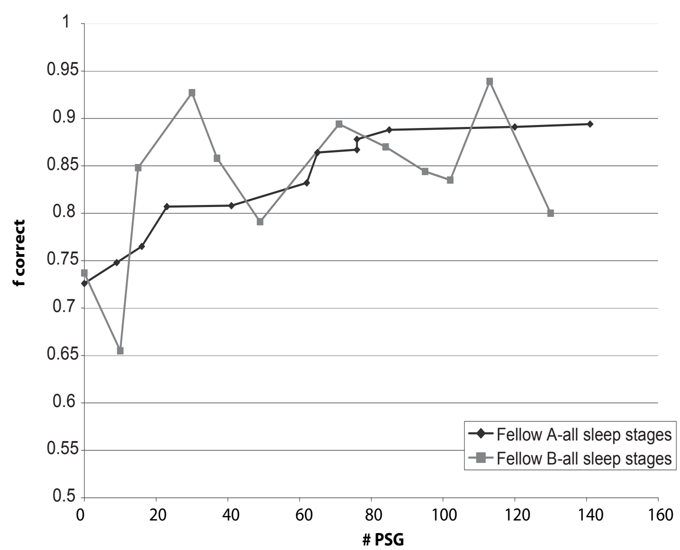

Figure 1—Concordance determined as f correct versus number of polysomnograms (PSG) scored for all stages of sleep. Each point depicts sequential months of sleep-disorders fellowship training.

Table 1—Selected Features of Subjects and Polysomnography Scored by Fellows

Study Date	Sex	Age, y	Diagnosis	# in series of PSG scored	TRT	TST_f	$TSTD_{ABSM}$
Fellow A							
7/6/2004	M	77	OSDB[a]	5	456.5	133.5	132.5
8/11/2004	M	45	OSDB	7	453.0	410.5	380.5
9/15/2004	M	40	OSDB	16	470.0	374.0	304.0
10/11/2004	F	36	OSDB	25	462.5	389.0	381.5
11/9/2004	M	39	OSDB	36	425.0	296.5	291.0
12/30/2004	F	46	OSDB	54	422.0	365.5	351.0
1/3/2005	M	67	OSDB[b]	57	476.0	342.0	334.5
2/9/2005	M	30	OSDB	78	424.0	364.5	373.5
3/7/2005	M	38	OSDB	85	383.5	316.0	326.0
4/13/2005	M	53	OSDB	106	487.5	419.0	404.5
5/3/2005	M	65	OSDB	126	433.0	301.0	301.5
6/3/2005	M	60	OSDB	161	478.5	409.5	396.0
Fellow B							
8/12/2004	M	33	OSDB	5	455.0	378.5	348.5
9/28/2004	F	55	OSDB	10	486.0	454.0	429.5
10/7/2004	F	49	OSDB	15	458.0	385.0	368.5
11/23/2004	F	38	OSDB	30	399.0	338.5	332.0
12/14/2004	M	68	OSDB	37	386.5	288.5	265.5
1/4/2005	F	65	OSDB	49	485.0	352.0	325.0
2/15/2005	M	50	OSDB[a]	71	453.5	262.0	256.0
3/14/2005	M	44	OSDB	84	369.5	312.0	309.0
4/12/2005	F	47	OSDB	95	482.0	421.0	425.0
5/19/2005	F	61	OSDB	102	371.0	242.0	240.0
6/7/2005	F	69	OSDB	113	431.5	239.5	238.5
7/8/2005	M	51	OSDB	130	392.5	266.0	268.0

[a]Insomnia
[b]Periodic limb movement disorder
TRT refers to total recording time in minutes; TST_f, total sleep time in minutes scored by fellow; $TSTD_{ABSM}$, total sleep time in minutes scored by Diplomate American Board of Sleep Medicine; OSDB = obstructive sleep disordered breathing.

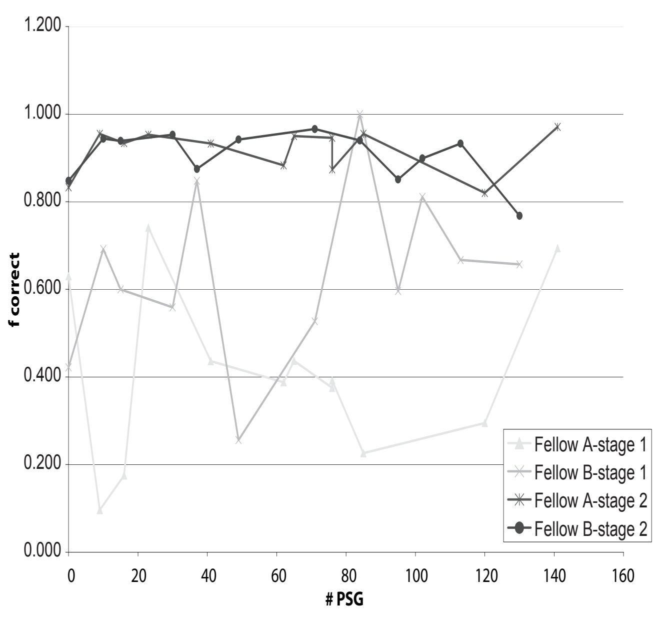

Figure 2—Concordance determined as f correct versus number of polysomnograms (PSG) scored for stage 1 and stage 2 sleep. Each point depicts sequential months of sleep-disorders fellowship training.

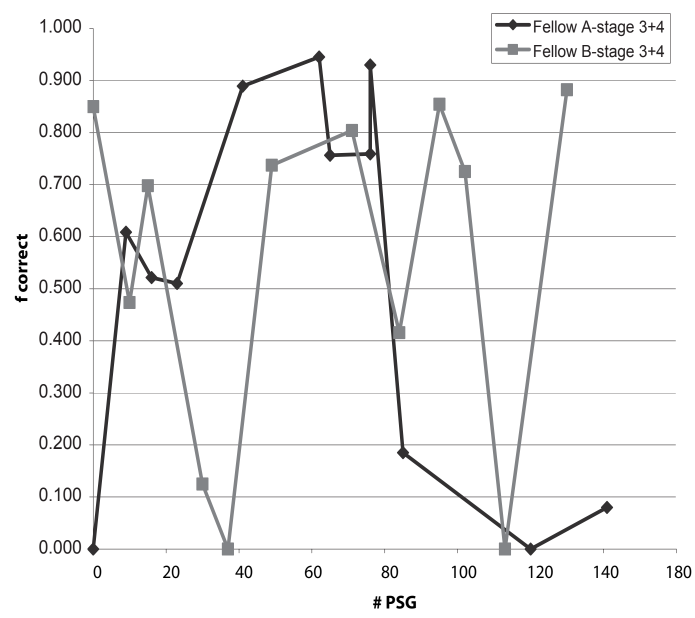

Figure 3—Concordance determined as f correct versus number of polysomnograms (PSG) scored for slow-wave sleep (combined stage 3 and 4). Each point depicts sequential months of sleep disorders fellowship training.

to the relative paucity of these sleep stages. For example, Fellow B scored 1 polysomnogram without any REM sleep (Figure 4), whereas the same study contained 21 epochs of REM sleep when scored by the DABSM. Hence, f correct (number of correct epochs divided by the number of total epochs = 0/21) without adjustment for sample size amplifies the error in the concordance

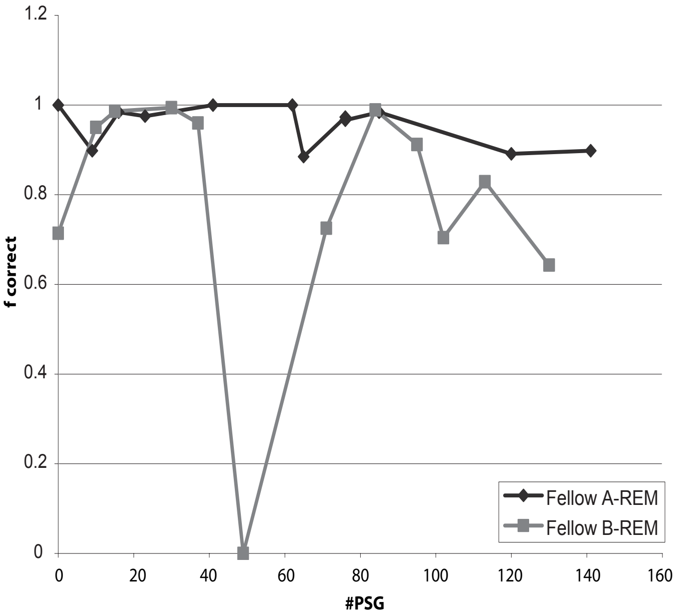

Figure 4—Concordance determined as f correct versus number of polysomnograms (PSG) scored for rapid eye movement (REM) sleep. Each point depicts sequential months of sleep-disorders fellowship training

of REM-sleep scoring on this polysomnogram.

DISCUSSION

Other investigators have published their findings using epoch-by-epoch analysis of sleep-stage scoring to ascertain interobserver agreement. Most of these have compared sleep scoring by sleep technologists or visual scoring of sleep by experienced personnel.[4-8] One of these investigations utilized recordings obtained by unattended home monitoring,[8] whereas another compared the reliability of sleep-stage scoring using the Multiple Sleep Latency Test.[7] In general, these investigations have supported the finding that concordance is greatest for NREM stage 2 sleep and REM sleep.

To our knowledge this is the first investigation that probes the rate at which physician trainees involved in a sleep fellowship training program acquire proficiency at scoring sleep. Using our methodology to teach sleep-stage scoring, it appears that sleep fellows become reasonably proficient after scoring approximately 20 to 30 polysomnograms and 4 months of dedicated sleep-disorders training. The variability in concordance observed for NREM sleep stages 1 and 3+4 sleep can be influenced by the number of epochs and therefore may not be suitable for standard concordance analysis as described in this manuscript. A standard measure of concordance that corrects for epoch sample size should be developed for use in future similar investigations.

REFERENCES

1. Winkelman JW. Designing a sleep disorders curriculum for psychiatry residents. Harv Rev Psychiatry 2005;12:54-6.
2. Strohl KP, Veasey S, Harding S et al. Competency-based goals for sleep and chronobiology in undergraduate medical education. Sleep 2003;26:333-6.
3. Harding SM, Berner ES. Developing an Action Plan for Integrating Sleep Topics into the Medical School Curriculum. Sleep Breath 2002;6:155-60.
4. Collop NA. Scoring variability between polysomnography technologists in different sleep laboratories. Sleep Med 2002;3:43-7.
5. Norman RG, Pal I, Stewart C, Walsleben JA and Rapoport DM. Interobserver agreement among sleep scorers from different centers in a large database. Sleep 2000;23:901-8.
6. Kim Y, Kurachi M, Horita M, Matsurra K, Kamikawa Y. agreement of visual scoring of sleep stages among many laboratories in Japan: effect of a supplementary definition of slow wave on scoring of slow wave sleep. Jpn J Psychiatry Neurol. 1993;47:91-7.
7. Drake CL, Rice MF, Roehrs TA, Rosenthal L, Guido P, Roth T. Scoring reliability of the multiple sleep latency test in a clinical population. Sleep 2000;23:911-3.
8. Whitney CW, Gottlieb DJ, Redline S, et al. Reliability of scoring respiratory disturbance indices and sleep staging. Sleep 1998 1;21:749-57.