

Correlating Subjective and Objective Sleepiness: Revisiting the Association Using Survival Analysis

R. Nisha Aurora, MD¹; Brian Caffo, PhD²; Ciprian Crainiceanu, PhD²; Naresh M. Punjabi, MD, PhD^{1,3}

¹Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; Departments of ²Biostatistics and ³Epidemiology, Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD

Study Objectives: The Epworth Sleepiness Scale (ESS) and multiple sleep latency test (MSLT) are the most commonly used measures of subjective and objective sleepiness, respectively. The strength of the association between these measures as well as the optimal ESS threshold that indicates objective sleepiness remains a topic of significant interest in the clinical and research arenas. The current investigation sought to: (a) examine the association between the ESS and the average sleep latency from the MSLT using the techniques of survival analysis; (b) determine whether specific patient factors influence the association; (c) examine the utility of each ESS question; and (d) identify the optimal ESS threshold that indicates objective sleepiness.

Design: Cross-sectional study.

Patients and Settings: Patients (N = 675) referred for polysomnography and MSLT.

Measurements and Results: Using techniques of survival analysis, a significant association was noted between the ESS score and the average sleep latency. The adjusted hazard ratios for sleep onset during the MSLT for the ESS quartiles were 1.00 (ESS < 9), 1.32 (ESS: 10–13), 1.85 (ESS: 14–17), and 2.53 (ESS ≥ 18), respectively. The association was independent of several patient factors and was distinct for the 4 naps. Furthermore, most of the ESS questions were individually predictive of the average sleep latency except the tendency to doze off when lying down to rest in the afternoon, which was only predictive in patients with less than a college education. Finally, an ESS score ≥ 13 optimally predicted an average sleep latency < 8 minutes.

Conclusions: In contrast to previous reports, the association between the ESS and the average sleep latency is clearly apparent when the data are analyzed by survival analysis, and most of the ESS questions are predictive of objective sleepiness. An ESS score ≥ 13 most effectively predicts objective sleepiness, which is higher than what has typically been used in clinical practice. Given the ease of administering the ESS, it represents a relatively simple and cost-effective method for identifying individuals at risk for daytime sleepiness.

Keywords: Epworth Sleepiness Scale, Multiple sleep latency test, subjective sleepiness

Citation: Aurora RN; Caffo B; Crainiceanu C; Punjabi NM. Correlating subjective and objective sleepiness: revisiting the association using survival analysis. *SLEEP* 2011;34(12):1707-1714.

INTRODUCTION

Sleepiness is a universal phenomenon in humans. While it is welcome at bedtime, the consequences of intrusive sleepiness at times when wakefulness is required can range from mild nuisance to devastating impairment. Excessive daytime sleepiness may be due to habitual short sleep duration or indicate the presence of an underlying medical or sleep disorder (e.g., sleep-disordered breathing). The generally accepted “gold standard” assessment of sleepiness is the multiple sleep latency test (MSLT), a series of 5 nap opportunities in the sleep laboratory scheduled at 2-hour intervals. The MSLT has been employed extensively in both clinical and research settings. It is reasonably sensitive to manipulations of sleep duration and has been shown to detect sleepiness in a variety of pathological states such as obstructive sleep apnea and narcolepsy.^{1–3} However, because the MSLT is labor intensive and quantifies sleepiness in only one situation, its utility as a screening tool for sleepiness is limited.

Several different subjective scales have been utilized to assess self-reported sleepiness in both clinical and research settings. Perhaps the most widely used is the Epworth Sleepiness Scale (ESS), which requires that patients estimate their “likelihood of dozing” in 8 different situations.⁴ The degree to which the ESS scores are correlated with objective measures of sleepiness (i.e., MSLT) has generated much debate, and the strength of the association in the published literature is, at best, modest.⁵ The lack of a strong association between the ESS and the MSLT may be due to several factors. First, it is clear that the ESS and the MSLT characterize different aspects of daytime sleepiness,^{6,7} and thus some degree of disagreement is to be expected. Second, the ability to detect an association may have been hampered by the limited sample sizes included in many of the previous studies.^{8–14} Third, because the MSLT quantifies the latency to the occurrence of an event (i.e., sleep onset), classical correlation and regression methods are not well suited for the analysis of such data and perhaps may not detect the association of interest even with more sizeable samples.^{7,15–17} The technique of survival analysis, which has been used to model sleep latency, has consistently uncovered associations that have previously not been appreciated.^{18–20} In addition to these issues, it remains to be determined whether specific patient factors modify the association between subjective and objective measures of sleepiness. For example, if characteristics such as age or sex influence the degree of association, then targeted use of the ESS may improve its utility. Furthermore, alternative approaches

Submitted for publication November, 2010

Submitted in final revised form June, 2011

Accepted for publication June, 2011

Address correspondence to: Naresh M. Punjabi, MD, PhD, Division of Pulmonary and Critical Care Medicine, Johns Hopkins University, 5501 Hopkins Bayview Circle, Baltimore, MD 21224; Tel: (410) 550-5405; Fax: (410) 550-2612; E-mail: npunjabi@jhmi.edu

could also then be tailored for those patient groups in which the ESS may not accurately identify daytime sleepiness. Another major gap in defining the value of the ESS is whether there are differences in the association between the ESS score and sleep latency across the different daytime naps and whether specific ESS questions are more predictive of objective sleepiness than others. Finally, there is a considerable need to empirically define thresholds in the ESS that predict the presence of objective sleepiness. Based on the original description by Johns,⁴ a threshold value of 10 is commonly used in clinical practice but is based on limited evidence. Thus, the objectives of the present study were to: (a) assess whether ESS scores correlate with the degree of MSLT-defined sleepiness using the techniques of survival analysis; (b) establish whether patient factors such as age, sex, race, and educational status modify the association; (c) define the utility of each ESS question in predicting objective sleepiness; (d) evaluate whether the association between ESS scores and average sleep latency varies across the daytime; and (e) determine the optimal threshold of the ESS score that indicates objective sleepiness.

METHODS

Study Sample

A cross-sectional analysis was conducted of all new adult patients (age ≥ 18 years) referred for overnight polysomnography and a daytime MSLT over a 2-year period (August 1998 and July 2000). Prior to the overnight study in the sleep laboratory, each patient completed a detailed questionnaire which included demographic information such as age, sex, race, educational level, and marital status (i.e., single, married, widowed, separated, or divorced). Subjective ratings of sleep propensity on the ESS were also acquired on the clinical questionnaire. The MSLT was conducted following the overnight study if requested by the evaluating physician as part of the assessment for daytime sleepiness. The study sample was heterogeneous with regards to the final clinical diagnoses and included disorders such as obstructive sleep apnea (48.1%), restless legs syndrome (12.1%), psychophysiological insomnia (6.1%), delayed sleep phase syndrome (4.7%), insufficient sleep syndrome (4.6%), sleep disorder due to a mood disorder (4.3%), narcolepsy (3.7%), central sleep apnea (3.3%), idiopathic hypersomnia (2.4%), and other sleep disorders (10.7%) such as various parasomnias and chronic fatigue syndrome. Weight and height were obtained on the night of the sleep study. The study protocol was approved by the local institutional review board on human research.

Polysomnography and Multiple Sleep Latency Test

The overnight polysomnogram consisted of continuous recordings of a modified electrocardiographic (V_6) lead, right and left electrooculographic leads, submental and bilateral anterior tibialis surface electromyograms, and the electroencephalogram (C_3-A_2 , C_3-O_1). Respiration was monitored throughout the night with a nasal pressure transducer, thermocouples at the nose and mouth, and thoracic and abdominal strain gauges. Continuous recording of the oxyhemoglobin saturation was obtained with an oximeter. Physiologic signals were digitized for off-line analysis of sleep and breathing patterns. Sleep-stage scoring was performed on 30-sec epochs according to the cri-

teria of Rechtschaffen and Kales.²¹ Apneas were identified if airflow was absent in the thermistor and nasal cannula channels for ≥ 10 sec. Hypopneas were identified if there was $\geq 30\%$ reduction in airflow which occurred for ≥ 10 sec and was associated with a 4% oxyhemoglobin desaturation or an EEG arousal. The apnea-hypopnea index (AHI) was defined as the number of apneas and hypopneas.

The MSLT, which consisted of a series of four 20-min nap trials at 2-h intervals, was performed in accordance with the 1992 guidelines for the objective assessment of sleepiness.³ Patients were instructed at 09:00, 11:00, 13:00, and 15:00 to lie down on a bed in a quiet and darkened bedroom and were allowed to sleep. The recording montage during the MSLT was similar to overnight study, except that patients did not wear the nasal cannula, thermistor, or the thoracic and abdominal strain gauges. Each nap trial lasted 20 min if sleep did not occur. If the patient fell asleep within 20 min, the trial was terminated 15 min after sleep onset. If no sleep occurred during a trial, a value of 20 min was recorded. Between naps, patients were instructed not to sleep and were monitored by trained technicians. The sleep latency for each nap trial was defined as the time to the first 30-sec epoch composed of stage 1 or any other stage of sleep.

Statistical Analysis

For examining the association between the ESS score and the average sleep latency, techniques of survival analysis were used as previously described.¹⁸⁻²⁰ Descriptive methods of survival analysis include examining the survivorship function as proposed by Kaplan and Meier.²² An important feature of the MSLT is that all individuals are observed for the entire period of observation (20 min). Therefore, there is no interim censoring, and the Kaplan-Meier survivorship function is merely the proportion of individuals that remain awake at each time t for $t \leq 20$ min. For the current analyses, the Kaplan-Meier survival curve was determined for the entire 20-min duration of the MSLT.

To determine whether the ESS scores were associated with the average sleep latency, Kaplan-Meier curves were constructed for categories based on quartiles of the ESS score, and the log-rank test was used to assess differences across the 4 curves. The magnitude of association between the ESS score and MSLT-defined sleepiness was then quantified using proportional hazards regression.²³ In the development of the multivariable proportional hazards regression model, the 20-min time point (i.e., end of the nap trial) was used for censoring those individuals that did not experience sleep onset. Covariates such as age, sex, race, and body mass index (BMI) were included in a multivariable proportional hazards model to determine the adjusted hazard ratio for falling asleep during the MSLT for the quartiles of the ESS score. Other covariates considered in model construction were educational level (< 12 th grade, high school graduate or equivalent, some college, and college degree or higher), marital status, total sleep time, time in bed, and the apnea-hypopnea index (AHI). Interaction terms between the ESS score, age, sex, race, marital status, educational level, BMI, and AHI were examined to determine whether any of these factors modified the association between the ESS score and the average sleep latency. Subsequently, sleep latency from each MSLT nap was modeled separately and heterogeneity of effects for a particular covariate across the naps was examined.

Finally, receiver operating curves were constructed using an average sleep latency < 8 min to define excessive daytime sleepiness. Statistical significance of all hazard ratios and odds ratios was determined by the 2-sided test of the regression coefficient. All statistical analyses were conducted using the SAS 9.1 statistical software.

RESULTS

The study sample consisted of 675 consecutive patients (415 men and 260 women) with an average age of 48.8 years (SD 12.2). The median ESS score of the sample was 13 with an interquartile range (25th–75th percentile) from 9 through 17. The distribution of race in the study sample was as follows: White (74.7%), African American (20.9%), and Other (4.4%). Approximately two-thirds (65.2%) of the sample reported being married, with another 18.3% being single and the remaining 16.5% being widowed, separated, or divorced. Many patients (44.6%) had a college or advanced degree, with the rest having some college experience (27.5%), a high school diploma or equivalent (19.0%), or less than high school education (8.8%). The average of all sleep onset latencies for the study sample was 6.67 min (SD 4.6). Objective sleepiness, defined as an average sleep latency < 8 min, was present in 461 patients (68.3%), and severe daytime sleepiness, defined as an average sleep latency ≤ 5 min, was present in 312 patients (46.2%). The Spearman rank correlation between the ESS score and average sleep latency was -0.30 (95% CI: $-0.24, -0.38$; $P < 0.0001$) suggesting a weak but statistically significant association.

Kaplan-Meier survival curves were initially used to characterize the association between the ESS score and the average sleep latency. The study sample was categorized into quartiles based on the ESS score distribution as follows: ≤ 9 (quartile 1), 10–13 (quartile 2), 14–17 (quartile 3), and ≥ 18 (quartile 4). The unadjusted Kaplan-Meier survivor functions across the ESS score quartiles were distinct and revealed that being in the second through fourth quartiles was progressively associated with a greater tendency for falling asleep (Figure 1; $P < 0.0001$ by log-rank test). Proportional hazards regression models were used to derive the unadjusted and adjusted hazards ratios relating the ESS score to the average sleep latency. Compared to the patients in the first quartile, the unadjusted hazards ratio for sleep onset for patients in the second through fourth quartiles were 1.28 (95% CI: 1.04–1.58), 1.75 (95% CI: 1.41–2.16), and 2.62 (95% CI: 2.09–3.28), respectively. Thus, an ESS score between 10 and 13 was associated with a 28% increase in risk of falling asleep, whereas ESS scores of 14–17 and ≥ 18 were associated with 75% and 162% increase in risk of falling asleep. Multivariable proportional hazards models that included adjustments for covariates such as age, sex, race, education level, marital status, BMI, AHI, total sleep time, and time in bed were also constructed and showed that the parameter estimates relating the ESS score to the average sleep latency were materially unchanged. Thus, the most parsimonious multivariable model that adjusted for demographic data was constructed and included the following variables: age, sex, race, education level, marital status, and BMI. Using the first ESS quartile as the reference, the multivariable adjusted hazard ratios for the second, third, and fourth quartile were 1.32 (95% CI: 1.06–1.64), 1.85 (95% CI: 1.48–2.31), and 2.54 (95% CI: 2.01–3.19), respectively. In-

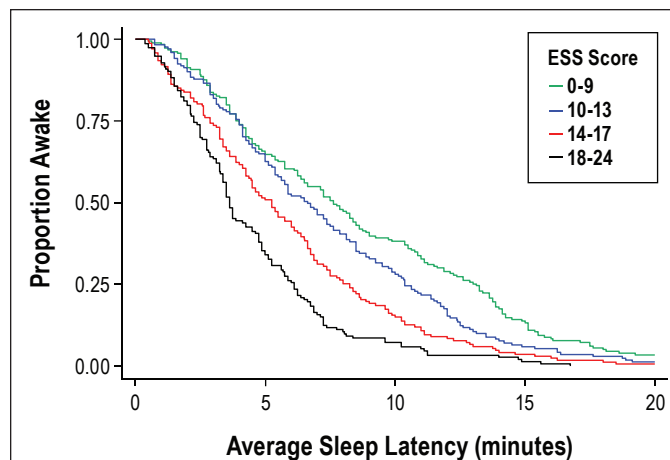


Figure 1—Kaplan-Meier survival curves for quartiles of ESS scores.

clusion of interaction terms between ESS score and categories of age, sex, race, educational level, marital status, BMI, and AHI showed no evidence of heterogeneity in the association across different strata of these variables.

To determine whether there was heterogeneity in the association between the ESS score and the sleep latency across the different nap trials, Kaplan-Meier curves and multivariable proportional hazards models corresponding to each nap trial were constructed separately, with partial and full covariate adjustments as before. Across each of the nap trials, the Kaplan-Meier survivor functions for the ESS quartiles were distinct, with the smallest and the largest separation being in the first and fourth nap trial, respectively (Figure 2). Proportional hazards models confirmed the association between the ESS score and sleep latency for each nap and showed that while there were group differences in hazard ratios for different ESS scores in each nap, the differences were most prominent in the fourth nap. A significant interaction between nap trial number and ESS score was observed (Table 1), suggesting that there is heterogeneity in the association across the different naps.

Additional analyses were subsequently conducted to characterize the association between each of the ESS questions and the average sleep latency. In unadjusted and adjusted analyses, there was consistency in the hazard ratios for objective daytime sleepiness across most of the ESS questions (Table 2). Assessment of dozing tendency in response to the question “How likely are you to doze off when lying down to rest in the afternoon when circumstances permit” was the only question not associated with objective sleepiness compared to the other 7 questions. Interestingly, when the study sample was restricted to those without a college or advanced degree, this question became a significant predictor of objective daytime sleepiness. Using the response of “never doze” as the reference, the adjusted hazards ratios for “slight chance,” “moderate chance,” and “high chance” were 1.10 (95% CI: 0.57–2.10), 1.46 (95% CI: 0.80–2.67), and 1.74 (95% CI: 0.99–3.06), respectively, with a corresponding P-value of 0.002 for a linear trend. Inferences regarding the association between the individual ESS questions and sleep latency were similar regardless of whether the overall average or the individual sleep latencies from the different naps were used in the proportional hazards models (Appendix).

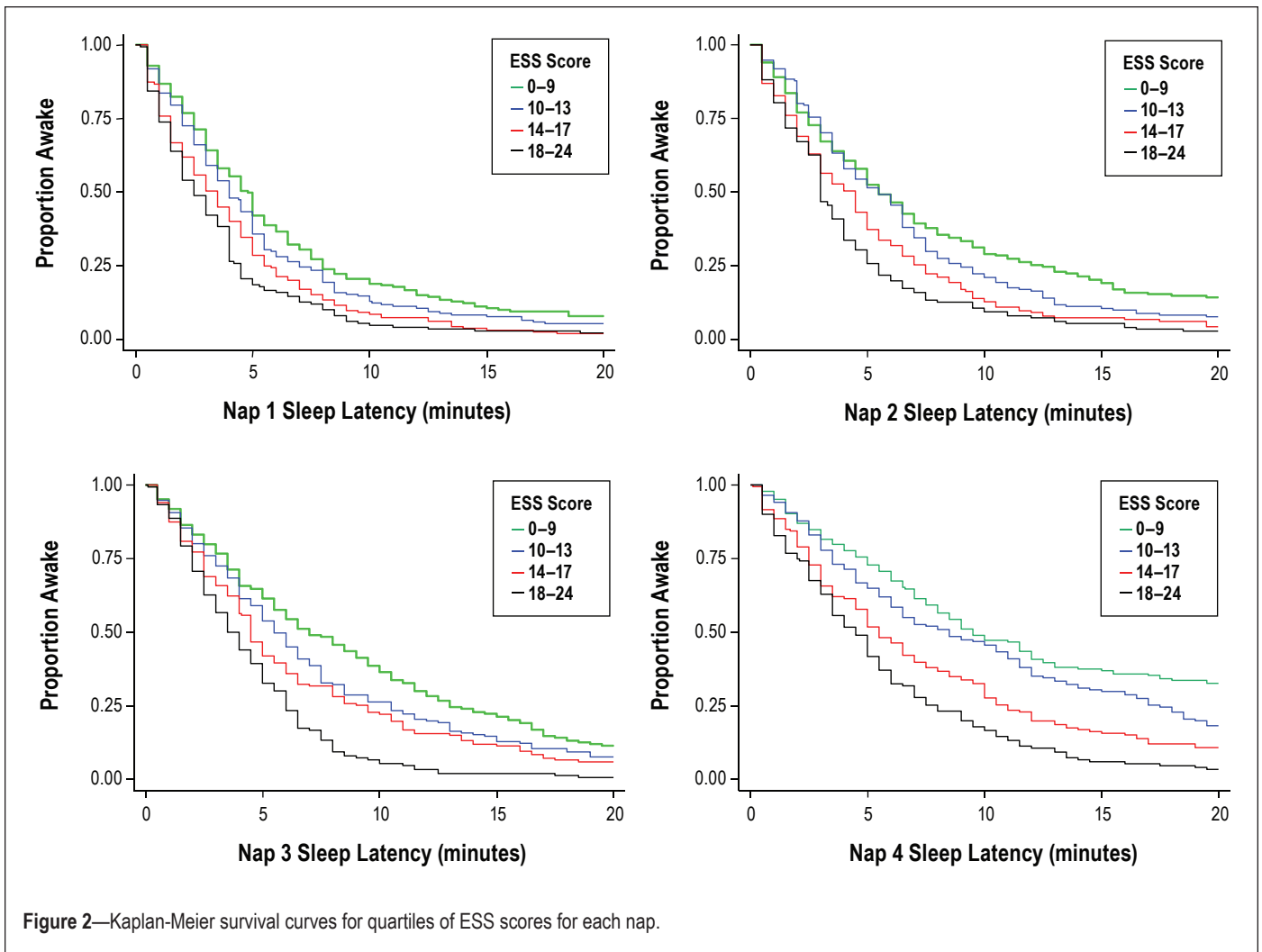


Figure 2—Kaplan-Meier survival curves for quartiles of ESS scores for each nap.

Table 1—Multivariable hazard ratios* (95% CI) derived from proportional hazards regression modeling of the average and individual nap latencies and the Epworth Sleepiness Scale (ESS)

ESS Quartile	Average MSLT	Individual Naps by Start Time				P-value†
		09:00	11:00	13:00	15:00	
I (0–9)	1.00	1.00	1.00	1.00	1.00	—
II (10–13)	1.32 (1.06–1.64)	1.15 (0.93–1.44)	1.18 (0.94–1.47)	1.28 (1.03–1.61)	1.35 (1.06–1.73)	0.24
III (14–17)	1.85 (1.48–2.31)	1.43 (1.15–1.79)	1.60 (1.28–2.01)	1.45 (1.16–1.82)	1.99 (1.56–2.55)	< 0.08
IV (≥ 18)	2.54 (2.01–3.19)	1.69 (1.34–2.12)	1.81 (1.44–2.29)	2.26 (1.79–2.85)	2.56 (1.99–3.29)	< 0.001

*Hazard ratios are adjusted for age, sex, race, marital status, education, and BMI. †P-value for the test for trend for the hazard ratio across naps for each ESS quartile.

To determine the threshold in the ESS score that was most predictive of objective daytime sleepiness (defined as an average sleep latency < 8 min), logistic regression analyses were used. The odds ratios relating the ESS score quartiles to objective daytime sleepiness were as follows: 1.00 (Reference; ESS ≤ 9), 1.37 (95% CI: 0.88–2.14; ESS: 10–13), 2.54 (95% CI: 1.58–4.11; ESS: 14–17), and 7.14 (95% CI: 3.87–13.18). Receiver operating curve analysis showed an area under the curve of 0.68 (95% CI: 0.64–0.72), indicating the ESS can predict objective daytime sleepiness to a moderate extent. As expected, increasing the cut-point in the ESS score to define excessive sleepiness was

associated with decreased sensitivity and increased specificity (Figure 3). Furthermore, the percentage of individuals correctly classified as having an average sleep latency < 8 min decreased with an increase in the ESS cut-point (Figure 3).

DISCUSSION

The results of the current study provide several novel and important clinical insights regarding the utility of the ESS in predicting daytime sleep tendency. First, ratings of sleep propensity on the ESS were observed to be associated with objective sleep onset during the MSLT particularly when subjected

Table 2—Multivariable hazard ratios* (95% CI) derived from proportional regression modeling of the average sleep latency and each question of the Epworth Sleepiness Scale (ESS)

ESS Question	Chance of Dozing			
	Never	Slight	Moderate	High
Sitting and reading	1.00	1.44 (1.02–2.02)	1.83 (1.32–2.53)	2.23 (1.63–3.05)
Watching television	1.00	1.74 (1.18–2.56)	1.82 (1.25–2.63)	2.12 (1.47–3.06)
Sitting, inactive in a public place	1.00	1.58 (1.25–1.99)	1.90 (1.50–2.40)	2.43 (1.91–3.10)
As a passenger in a car for an hour without a break	1.00	1.39 (1.07–1.81)	1.79 (1.38–2.32)	2.35 (1.83–3.01)
Lying down to rest in the afternoon when circumstances permit	1.00	1.00 (0.64–1.58)	0.96 (0.63–1.47)	1.39 (0.94–2.05)
Sitting and talking to someone	1.00	1.45 (1.21–1.73)	1.85 (1.41–2.43)	2.66 (1.84–3.84)
Sitting quietly after a lunch without alcohol	1.00	1.28 (1.02–1.60)	1.51 (1.21–1.90)	1.86 (1.47–2.34)
In a car, while stopped for a few minutes in the traffic	1.00	1.45 (1.20–1.76)	1.70 (1.34–2.15)	3.15 (2.19–4.53)

*Hazard ratios are adjusted for age, sex, race, marital status, education, and BMI.

ESS Cutpoint	Sensitivity	Specificity	Correctly Classified
0	100.0	0.0	68.3%
1	99.6	1.4	68.4%
2	99.1	2.8	68.6%
3	98.1	4.2	68.3%
4	96.3	10.3	69.0%
5	95.2	14.0	69.5%
6	93.3	17.8	69.3%
7	89.8	22.4	68.4%
8	87.2	26.2	67.9%
9	83.3	34.1	67.7%
10	79.0	40.7	66.8%
11	74.6	48.1	66.2%
12	68.8	56.5	64.9%
13	63.6	66.8	64.6%
14	56.8	72.9	61.9%
15	48.8	77.1	57.8%
16	41.2	80.4	53.6%
17	34.1	87.9	51.1%
18	29.7	92.5	49.6%
19	23.4	93.9	45.8%
20	16.9	95.8	41.9%
21	11.7	97.2	38.8%
22	6.5	99.1	35.9%
23	3.0	100.0	33.8%
24	1.5	100.0	32.7%
24	0.0	100.0	31.7%

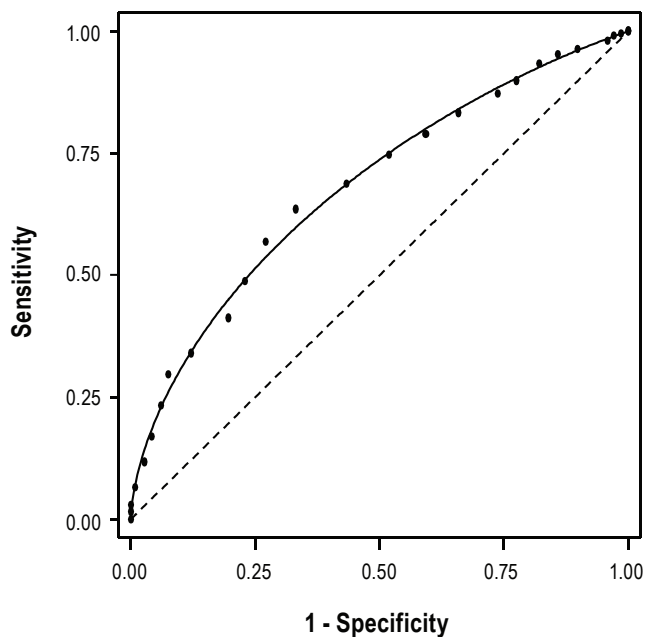


Figure 3—Receiver Operating Curve Analysis for the Epworth Sleepiness Scale. An ESS cutpoint of 13 optimized both sensitivity and specificity for an average latency < 8 minutes.

to survival analysis. Second, the association between the ESS score and sleep onset latency displayed heterogeneity across the four daytime naps with the association being strongest with the afternoon nap. Third, the association was independent of age, sex, race, educational level, marital status, and BMI and no significant effect modification was observed due to these factors. Fourth, sleep-related factors such as total sleep time, total time in bed, and AHI did not appear to confound or modify the association. Fifth, while almost all of the questions on the ESS were associated with objective sleepiness, the question of dozing tendency when lying down to rest in the afternoon was not predictive of sleep latency except in those patients with less than a college education. Finally, an ESS score threshold of 13 provided the optimal value indica-

tive of objective sleepiness, defined as a mean sleep latency of less than 8 minutes.

The observation that the ESS score is associated with MSLT-defined sleepiness independent of several factors is in contrast to the findings noted in a majority of the earlier studies.^{5-17,24-26} In clinical and nonclinical samples, several investigators have reported either no correlation or a low correlation between the subjective ratings of sleep propensity on the ESS and the average sleep latency (Table 3). One possible explanation for the disparity between the current study and previous reports is our use of survival analysis to characterize the association. Survival analysis confers a specific advantage over correlation and regression methods by more appropriately handling censored observations. During the MSLT, sleep onset may not occur during

Table 3—Studies on the association between the Epworth Sleepiness Scale and the MSLT-defined sleepiness

Degree of Association*	First Author	Year	N	Patient Sample	Reported Correlation (ρ)	Statistical Methods Used**
Poor to None	Chervin ¹⁷	1999	237	Sleep clinic patients	0.03	Correlation/Regression
	Benbadis ¹⁶	1999	102	Sleep clinic patients	0.17	Correlation/Regression
	Furuta ⁸	1999	10	Sleep clinic patients	0.23	Correlation/Regression
	Masel ³⁴	2001	71	Patient with traumatic brain injury	0.16	Correlation/Regression
	Parker ¹⁰	2003	46	Patients on hemodialysis	0.14	Correlation/Regression
	Razmy ³⁵	2004	78	Patients with Parkinson disease	0.12	Correlation/Regression
	Fong ⁷	2005	296	Sleep clinic patients	0.15	Correlation/Regression
	Blaivas ³⁶	2007	54	Sleep clinic patients	0.06	Correlation/Regression
	Castriotta ³⁷	2007	87	Patients with posttraumatic brain injury	0.10	Correlation/Regression
	Beiske ¹²	2009	37	Sleep clinic patients	0.09	Correlation/Regression
Laberge ¹⁴	2009	43	Patients with myotonic dystrophy	0.14	Correlation/Regression	
Fair	Briones ²⁸	1996	129	Clinic + community samples	0.27	Correlation/Regression
	Chervin ²⁴	1997	60	Sleep clinic patients	0.37	Correlation/Regression
	Olson ¹⁵	1998	225	Sleep clinic patients	0.30	Correlation/Regression
	Vignatelli ²⁹	2003	91	Sleep clinic patients	0.31	Correlation/Regression
	Watson ¹¹	2004	40	Patients with chronic fatigue syndrome	0.40	Correlation/Survival
	Leng ³¹	2005	72	Sleep clinic patients	0.36	Correlation/Regression
	Weaver ³⁸	2007	149	Sleep clinic patients	0.31	Correlation/Regression
	Jimenez-Correa ¹³	2009	23	Sleep clinic patients	0.40	Correlation/Regression
Moderate	Johns ⁴	1991	27	Sleep clinic patients	0.51	Correlation/Regression
	Johns ²⁷	1994	44	Sleep clinic patients	0.42	Correlation/Regression
	Chung ²⁵	2000	161	Clinic + community samples	0.42	Correlation/Regression
	Barnes ⁹	2002	42	Sleep clinic patients	0.50	Correlation/Regression
Strong	Komada ³⁹	2005	131	Clinic sample and healthy controls	0.86	Correlation/Regression
	Bravo ⁴⁰	2007	70	Clinic sample and healthy controls	0.74	Correlation/Regression
	Poryazova ⁴¹	2010	30	Patients with Parkinson disease	0.65	Correlation/Regression

*Degree of association is defined as the following: $\rho < 0.20$: poor or no association; $\rho = 0.21-0.40$: fair; $\rho = 0.41-0.60$: moderate; $0.61-0.80$: good; $0.81-1.00$: Strong. **Statistical methods used to characterize the association between the Epworth Sleepiness Scale and the MSLT-defined sleepiness.

each nap trial. A value of 20 minutes is used for such nap trials, and this value is included in the overall average of four or five naps despite the fact that sleep onset did not occur. Classical regression and correlation methods, if employed to model bounded or censored data, can lead to biased estimates given their inability to handle such data. Survival analysis, on the other hand, can more accurately model the sleep latency data because of its capacity to handle non-censored (i.e., occurrence of sleep onset) and censored (i.e., no occurrence of sleep) observations.

Additional explanations for the disparity in the findings between the current study and previous reports include the potential impact of limited sample size and the lack of consideration for potential confounding from patient factors such as age, sex, race, and educational status. Studies with small patient samples are at risk for not being able to capture an association when, in fact, one truly exists. The current study obviates such limitations by not only including a relatively large and heterogeneous clinic-based sample but also carefully considering the effects of several patient factors. Interestingly, while others have found that the ESS is influenced by factors such as sex,¹⁷ the current study showed that the association is not confounded or modified by several patient factors. The only exception was the predic-

tive utility of the question regarding the tendency to doze in the mid-afternoon when circumstances permit, which was associated with objective sleepiness in patients with less than a college education, but not in those with a college or advanced degree.

The fact that the ESS and MSLT differ somewhat in their classification of sleepiness is not surprising. The ESS was designed to gauge a subject's self-reported level of sleep tendency in particular situations. Specifically, the scale requires that the responder estimate the probability of falling asleep in eight situations but does not explicitly request the actual frequency of having fallen asleep in these situations. The ESS also requires that the assessment take into consideration "recent times" which likely reflects a period of week or more. Thus, the ESS rating embodies a subject's overall or average propensity for falling asleep (i.e., a trait) without an emphasis on an exact time or situation (i.e., the state). Furthermore, the ESS does not appraise the potential of any trait-state interactions, which are crucial given that a patient's risk for falling asleep is most likely determined by not only the surrounding circumstances but also his or her response to those circumstances. The MSLT, on the other hand, uses physiological data to assess the rate of falling asleep (sleep latency) in an environment that is most conducive for

sleep (a quite darkened room). Although an objective measure of daytime sleepiness, the MSLT is not without its limitations, given that it can be certainly influenced by various factors such as duration of prior sleep and subject motivation. Given that the ESS and MSLT assess distinct aspects of sleepiness, some level of disagreement between the two measures is to be expected.

In the current study, seven of the eight ESS questions were independently associated with objective sleepiness which is in contrast with other reports where only a few, if any, of the questions were of value.^{12,17,24,25,27} The consistency of the association across most of the ESS questions perhaps echoes one of the major assertions of the current study that the use of survival analysis to model sleep latency data can uncover associations that may not be otherwise revealed with classical regression or correlation analyses. Interestingly, the question “How likely are you to doze off when lying down to rest in the afternoon when circumstances permit” was not associated with the average sleep latency, a finding that has also been noted by others. The lack of association is surprising given that this question approximates the conditions of the afternoon nap trial, a time that coincides with the circadian trough seen as part of normal diurnal variability. It is likely that because many patients typically do not have the opportunity to lie down during the day, their ability to estimate the likelihood of falling asleep in the afternoon may not be accurate. Additional research is needed to further delineate the underlying factors that can explicate the lack of association between the tendency for “lying down in the rest in the afternoon” and objective sleep latency.

In clinical practice, an ESS score greater than 10 has been commonly used to define excessive daytime sleepiness.^{12,28-31} While such dichotomization is appealing for both clinical screening and research purposes, supporting evidence for any particular cut point is limited. The results provided herein indicate that a threshold value of 13 has the optimal sensitivity and specificity in correctly classifying individuals with a mean MSLT of less than 8 minutes. Not surprisingly, as the definition of objective sleepiness is varied (i.e., different cut points in the MSLT), the corresponding value on the ESS that optimally predicts objective sleepiness will also vary.

There are several strengths and limitations of the present study that merit discussion. Strengths include the large number of patients with the full range of sleepiness on both the subjective and objective scales, the use of survival analysis to model the overall average and individual nap data, and the careful consideration of confounding and effect modification due to several factors. Limitations include the use of a clinic-based sample which restricts the generalizability of the results, the relatively small proportion of minority patients, the use of four naps instead of the currently recommended five naps for the MSLT, and the lack of data on socioeconomic and employment status. Availability of such data would have permitted a more rigorous assessment as to whether these factors can alter the performance characteristics of the ESS as it relates to objective daytime sleepiness. Finally, it is important to recognize that because the proportional hazards model used to characterize the association between the ESS score and the average sleep latency is semi-parametric and does not require the specification of the underlying hazard function, it is not possible to predict the degree of objective sleepiness given a patient’s sub-

jective rating on the ESS. However, parametric survival functions (e.g., exponential, Weibull, log-normal) can be used to develop a prediction model which can estimate the expected sleep latency for a given ESS score and represents a logical extension of the current work.

The potential implications of the objective-subjective sleepiness association are broad in scope. Given that sleepiness is a ubiquitous phenomenon that can have deleterious consequences when present at inappropriate times, having a simple screening tool to identify sleepy people is appealing. The ability to assess subjective sleepiness in a simple and convenient manner that predicts one’s actual sleep propensity is critical as interventions can then be implemented to perhaps prevent mistakes and accidents. However, it is important to recognize that even though some studies have found that a high ESS score is associated with accident risk,^{32,33} the scale is inherently limited secondary to its subjective nature and therefore cannot accurately identify an individual’s tendency for lapses in performance particularly during periods of vulnerability (e.g., 5 AM).

ACKNOWLEDGMENT

This study was supported by National Institutes of Health Grants HL075078 and HL086862.

DISCLOSURE STATEMENT

This was not an industry supported study. Dr. Punjabi has received research support from ResMed Inc. The other authors have indicated no financial conflicts of interest.

REFERENCES

1. Carskadon MA, Dement WC, Mitler MM, Roth T, Westbrook PR, Keenan S. Guidelines for the multiple sleep latency test (MSLT): a standard measure of sleepiness. *Sleep* 1986;9:519-24.
2. Roehrs T, Roth T. Multiple sleep latency test: technical aspects and normal values. *J Clin Neurophysiol* 1992;9:63-7.
3. Thorpy MJ, Westbrook P, Ferber R, et al. The clinical use of the Multiple Sleep Latency Test. Standards of Practice Committee of the American Sleep Disorders Association. *Sleep* 1992;15:268-76.
4. Johns MW. A new method for measuring daytime sleepiness: the Epworth sleepiness scale. *Sleep* 1991;14:540-5.
5. Chervin RD. The multiple sleep latency test and Epworth sleepiness scale in the assessment of daytime sleepiness. *J Sleep Res* 2000;9:399-401.
6. Miletin MS, Hanly PJ. Measurement properties of the Epworth sleepiness scale. *Sleep Med* 2003;4:195-9.
7. Fong SY, Ho CK, Wing YK. Comparing MSLT and ESS in the measurement of excessive daytime sleepiness in obstructive sleep apnoea syndrome. *J Psychosom Res* 2005;58:55-60.
8. Furuta H, Kaneda R, Kosaka K, Arai H, Sano J, Koshino Y. Epworth Sleepiness Scale and sleep studies in patients with obstructive sleep apnea syndrome. *Psychiatry Clin Neurosci* 1999;53:301-2.
9. Barnes M, Houston D, Worsnop CJ, et al. A randomized controlled trial of continuous positive airway pressure in mild obstructive sleep apnea. *Am J Respir Crit Care Med* 2002;165:773-80.
10. Parker KP, Bliwise DL, Bailey JL, Rye DB. Daytime sleepiness in stable hemodialysis patients. *Am J Kidney Dis* 2003;41:394-402.
11. Watson NF, Jacobsen C, Goldberg J, Kapur V, Buchwald D. Subjective and objective sleepiness in monozygotic twins discordant for chronic fatigue syndrome. *Sleep* 2004;27:973-7.
12. Beiske KK, Kjelsberg FN, Ruud EA, Stavem K. Reliability and validity of a Norwegian version of the Epworth sleepiness scale. *Sleep Breath* 2009;13:65-72.
13. Jimenez-Correa U, Haro R, Obdulia GR, Velazquez-Moctezuma J. Correlations between subjective and objective features of nocturnal sleep and excessive diurnal sleepiness in patients with narcolepsy. *Arq Neuropsiquiatr* 2009;67:995-1000.

14. Laberge L, Begin P, Dauvilliers Y, et al. A polysomnographic study of daytime sleepiness in myotonic dystrophy type 1. *J Neurol Neurosurg Psychiatry* 2009;80:642-6.
15. Olson LG, Cole MF, Ambrogetti A. Correlations among Epworth Sleepiness Scale scores, multiple sleep latency tests and psychological symptoms. *J Sleep Res* 1998;7:248-53.
16. Benbadis SR, Mascha E, Perry MC, Wolgamuth BR, Smolley LA, Dinner DS. Association between the Epworth sleepiness scale and the multiple sleep latency test in a clinical population. *Ann Intern Med* 1999;130:289-92.
17. Chervin RD, Aldrich MS. The Epworth Sleepiness Scale may not reflect objective measures of sleepiness or sleep apnea. *Neurology* 1999;52:125-31.
18. Punjabi NM, O'hearn DJ, Neubauer DN, et al. Modeling hypersomnolence in sleep-disordered breathing. A novel approach using survival analysis. *Am J Respir Crit Care Med* 1999;159:1703-9.
19. Punjabi NM, Bandeen-Roche K, Marx JJ, Neubauer DN, Smith PL, Schwartz AR. The association between daytime sleepiness and sleep-disordered breathing in NREM and REM sleep. *Sleep* 2002;25:307-14.
20. Punjabi NM, Bandeen-Roche K, Young T. Predictors of objective sleep tendency in the general population. *Sleep* 2003;26:678-83.
21. Rechtschaffen A, Kales A. Manual of standardized terminology, techniques and scoring system for sleep stages of human subjects. Washington, DC: US Government Printing Office; 1968.
22. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc* 1958;53:457-81.
23. Cox DR. Regression models and life tables. *J R Stat Soc B* 1972;34:187-220.
24. Chervin RD, Aldrich MS, Pickett R, Guilleminault C. Comparison of the results of the Epworth Sleepiness Scale and the Multiple Sleep Latency Test. *J Psychosom Res* 1997;42:145-55.
25. Chung KF. Use of the Epworth Sleepiness Scale in Chinese patients with obstructive sleep apnea and normal hospital employees. *J Psychosom Res* 2000;49:367-72.
26. Weaver TE. Outcome measurement in sleep medicine practice and research. Part 1: assessment of symptoms, subjective and objective daytime sleepiness, health-related quality of life and functional status. *Sleep Med Rev* 2001;5:103-28.
27. Johns MW. Sleepiness in different situations measured by the Epworth Sleepiness Scale. *Sleep* 1994;17:703-10.
28. Briones B, Adams N, Strauss M, et al. Relationship between sleepiness and general health status. *Sleep* 1996;19:583-8.
29. Vignatelli L, Plazzi G, Barbato A, et al. Italian version of the Epworth sleepiness scale: external validity. *Neurol Sci* 2003;23:295-300.
30. Zallek SN, Redenius R, Fisk H, Murphy C, O'Neill E. A single question as a sleepiness screening tool. *J Clin Sleep Med* 2008;4:143-8.
31. Leng PH, Low SY, Hsu A, Chong SF. The clinical predictors of sleepiness correlated with the multiple sleep latency test in an Asian Singapore population. *Sleep* 2003;26:878-81.
32. Maycock G. Sleepiness and driving: the experience of heavy goods vehicle drivers in the UK. *J Sleep Res* 1997;6:238-44.
33. Howard ME, Desai AV, Grunstein RR, et al. Sleepiness, sleep-disordered breathing, and accident risk factors in commercial vehicle drivers. *Am J Respir Crit Care Med* 2004;170:1014-21.
34. Masel BE, Scheibel RS, Kimbark T, Kuna ST. Excessive daytime sleepiness in adults with brain injuries. *Arch Phys Med Rehabil* 2001;82:1526-32.
35. Razmy A, Lang AE, Shapiro CM. Predictors of impaired daytime sleep and wakefulness in patients with Parkinson disease treated with older (ergot) vs newer (nonergot) dopamine agonists. *Arch Neurol* 2004;61:97-102.
36. Blaivas AJ, Patel R, Hom D, Antigua K, Ashtyani H. Quantifying micro-sleep to help assess subjective sleepiness. *Sleep Med* 2007;8:156-9.
37. Castriotta RJ, Atanasov S, Wilde MC, Masel BE, Lai JM, Kuna ST. Treatment of sleep disorders after traumatic brain injury. *J Clin Sleep Med* 2009;5:137-44.
38. Weaver TE, Maislin G, Dinges DF et al. Relationship between hours of CPAP use and achieving normal levels of sleepiness and daily functioning. *Sleep* 2007;30:711-9.
39. Komada Y, Inoue Y, Mukai J, Shirakawa S, Takahashi K, Honda Y. Difference in the characteristics of subjective and objective sleepiness between narcolepsy and essential hypersomnia. *Psychiatry Clin Neurosci* 2005;59:194-9.
40. Bravo ML, Serpero LD, Barcelo A, Barbe F, Agustí A, Gozal D. Inflammatory proteins in patients with obstructive sleep apnea with and without daytime sleepiness. *Sleep Breath* 2007;11:177-85.
41. Poryazova R, Benninger D, Waldvogel D, Bassetti CL. Excessive daytime sleepiness in Parkinson's disease: characteristics and determinants. *Eur Neurol* 2010;63:129-35.