

Body Mass Index, Gender, and Ethnic Variations Alter the Clinical Implications of the Epworth Sleepiness Scale in Patients with Suspected Obstructive Sleep Apnea[§]



Sean Hesselbacher¹, Shyam Subramanian¹, Jerry Allen², Sara Surani² and Salim Surani^{*,1,2,3,4}

¹Department of Medicine, Section of Pulmonary, Critical Care, and Sleep Medicine, Baylor College of Medicine, Houston, TX, USA

²Torr Sleep Center, Corpus Christi, TX, USA

³Texas A&M University, Corpus Christi, TX, USA

⁴CHRISTUS Spohn Hospital, Corpus Christi, TX, USA

Abstract: *Introduction:* The Epworth Sleepiness Scale (ESS) is often used in the evaluation of obstructive sleep apnea (OSA), though questions remain about the influence gender, ethnicity, and body morphometry have in the responses to this questionnaire. The aim of this study was to examine differences in ESS scores between various demographic groups of patients referred for polysomnography, and the relationship of these score to sleep-disordered breathing

Methods: Nineteen hundred consecutive patients referred for polysomnographic diagnosis of OSA completed questionnaires, including demographic data and ESS. OSA was determined based on a respiratory disturbance index (RDI) ≥ 15 by polysomnography.

Results: In this high risk population for OSA, the ESS was 10.7 ± 5.6 . The highest ESS scores were seen in obese males; non-obese females and non-obese Caucasian males scored the lowest. ESS was weakly correlated with RDI ($r = 0.17$, $P < 0.0001$). The sensitivity of ESS for the diagnosis of OSA was 54% and the specificity was 57%. The positive (PPV) and negative (NPV) predictive values were 64% and 47%, respectively. In obese subjects, the sensitivity and specificity were 55% and 53%, compared with 47% and 63% in non-obese subjects. In obese, Hispanic males, the sensitivity, specificity, and PPV were 59%, 54%, and 84%, respectively. In non-obese, Caucasian females, the sensitivity, specificity, and NPV were 43%, 59%, and 72%.

Conclusions: The ESS appears to be affected by many factors, including gender, ethnicity, and body morphometry. The ability of the ESS to predict OSA is modest, despite a significant correlation with the severity of OSA. The test characteristics improve significantly when applied to select populations, especially those at risk for OSA.

Keywords: Ethnicity, gender, obstructive sleep apnea, epworth sleepiness scale, screening.

INTRODUCTION

Background

Obstructive sleep apnea (OSA) is a common medical disorder with general health and quality-of-life implications [1]. Associations with important medical conditions, including diabetes mellitus, coronary arterial disease, congestive heart failure, hypertension, and cerebrovascular accident, are well-documented, especially in moderate-to-severe OSA [1-6]. Untreated OSA may result in excessive daytime sleepiness, impaired decision-making and auto-

mobile accidents [7-9]. Continuous overnight polysomnography (PSG) performed in a sleep laboratory remains the current gold standard for diagnosis of OSA. Screening strategies for use in the primary care setting have been developed, with the goal of detecting patients at risk for OSA, and subsequent referral for PSG.

One such commonly used screening strategy is the Epworth Sleepiness Scale (ESS). This questionnaire relies on self-reported patient symptoms, asking "How likely are you to doze off or fall asleep?" in a set of 8 hypothetical situations, each scored 0-3, giving a total score 0-24. This test has provided mixed results in the detection of OSA. The scale effectively discriminated between primary snoring and OSA in early studies [10, 11]. An ESS score of 10 is most often considered to be the upper limit of normal, though more recent work has shown that a lower score (8) may be associated with abnormal daytime sleepiness [12, 13].

While primarily designed and developed as a measure of excessive daytime sleepiness, the ESS emerged as an

*Address correspondence to this author at the Department of Medicine, Section of Pulmonary, Critical Care, and Sleep Medicine, Texas A&M University, 613 Elizabeth Street, Suite 813, Corpus Christi, Texas 78413, USA; Tel: 1-361-885-7722; Fax: 361-850-7563; E-mail: srsurani@hotmail.com

[§]Presented in part at the Associated Professional Sleep Societies meeting, Minneapolis, Minnesota, June 11-15, 2011.

important clinical tool in the workup and management of OSA. Some studies have relied on the use of the ESS to screen patients for sleep apnea and clinically this remains the most widespread tool for primary care physicians to triage patients for sleep evaluations [14]. Not only is it part of the screening process in determining referrals to sleep labs for PSG, but it remains part of the clinical decision-making process in determining who should get treatment, especially in cases of mild OSA [15]. The predictive properties of this instrument in the high-risk population of patients referred to the sleep laboratory for a sleep study for evaluation of clinically suspected OSA are not well established. Moreover, the impact of OSA on individual patients is partially affected by their demographic characteristics, i.e. gender and ethnicity [16-18]; body mass index may affect how OSA will impact a patient. The ESS is also subject to variations in self-reported symptoms by these same populations. It has been reported that African American subjects reported higher ESS scores than Caucasians, while gender and age did not influence the average score, in a study of insomnia [19]. Other studies have reported that being Maori in New Zealand is independently associated with elevated ESS score [20]. An analysis of normal patients in the Sleep Heart Health Study showed no association of age, sex, or BMI on the ESS [21], however ESS scores varied despite similar rates of subjective sleepiness [22]. Thus, it is likely that the ESS, which assesses only daytime sleepiness, may vary among different demographic groups; its relationship with OSA may also vary between these same groups.

This retrospective study was designed to examine the differences in ESS scores between demographic groups (and subgroups) of a cohort of patients referred for PSG. Because the ESS is commonly used to screen patients for OSA, we sought to determine if the predictive value of the ESS for OSA varies between these same groups and subgroups.

MATERIALS AND METHODOLOGY

Study Participants

Data were collected from 2112 consecutive eligible patients referred to the Torr Sleep Center (Corpus Christi, TX) for PSG evaluation of suspected OSA between February 5, 2007 and June 26, 2009. All subjects were ≥ 12 years of age without a prior diagnosis of OSA. Subjects were excluded from the final analysis if the ESS was incomplete or if they failed to undergo PSG and recording of the respiratory disturbance index (RDI). The protocol was approved by the CHRISTUS Spohn Hospital Corpus Christi Institutional Review Board (#08 08013).

Baseline Evaluation

Prior to undergoing PSG, the subjects completed questionnaires. A general health questionnaire obtained information about demographics and general sleep health, including objective daytime sleepiness, as measured by the Epworth Sleepiness Scale [10]. Gender and ethnicity were self-reported by the participants. Physical examination was performed by trained technologists. The general examination included height and weight measurement.

Polysomnographic Evaluation

Overnight comprehensive PSG was performed in the sleep laboratory, with multichannel recordings monitoring electroencephalogram, electrocardiogram, electrooculograms, submental is electromyogram, airflow, respiratory effort, oxygen saturation, and anterior tibialis electromyogram. Data were scored by a technologist manually, according to the American Academy of Sleep Medicine Scoring Guidelines [23]. An apnea was scored if there was cessation of airflow for ≥ 10 seconds; a hypopnea was scored if there was $\geq 30\%$ reduction in airflow for ≥ 10 seconds, associated with a drop in $\text{SaO}_2 \geq 4\%$; a respiratory effort related arousal (RERA) was scored if there was a sequence of breaths, not qualifying as apnea or hypopnea, lasting ≥ 10 seconds with increasing respiratory effort or flattening of the nasal pressure waveform leading to an arousal from sleep. The RDI was calculated by summing the number of obstructive apneas, hypopneas, and RERAs per hour. Technologists were chosen with minimum experience of scoring 500 PSG, and intra- and inter-scorer variability were standardized by means of a point system in place at the sleep center [24].

Statistics

Comparisons between the means of 2 normally distributed groups were performed with the unpaired t-test, between 2 non-normally distributed groups with the Mann-Whitney U test, and between 3 or more non-normally distributed groups with the Kruskal-Wallis test. Sensitivity and specificity of the ESS in the detection of OSA were calculated based using ESS ≥ 10 as a positive test result and RDI ≥ 15 events/hour as the diagnostic standard for OSA. Correlation analysis was performed by calculating the Spearman rank correlation coefficient for nonparametric samples. A p-value of < 0.05 was considered statistically significant.

RESULTS

Clinical Characteristics and Demographics

A total of 2112 patient records were evaluated for inclusion into the study, of which 1900 had complete data collected and were included in the final analysis. Of the excluded patients, the vast majority were because of missing ESS questionnaires. Excluded patients did not differ significantly from those included in terms of age, height, body mass index (BMI), RDI, or lowest O_2 saturation during the PSG; the excluded patients did have a higher mean weight than included patients (71 vs 68 kg, $p=0.047$) and were more likely to be male (65% vs 58%, $P=0.037$). The baseline clinical characteristics and demographics of the included participants, separated by gender are shown in Table 1. In general, men were larger by height and weight, though the average body mass index (BMI) was higher in females. Indices of OSA, such as RDI and lowest SaO_2 were more severe in males, and the mean ESS was higher in males (11 vs 10). The subjects predominantly identified themselves as Hispanic or Caucasian in both groups, and the mean BMI fell within the range of obesity; subject characteristics were

Table 1. Demographic and Clinical Characteristics of the Participants

Group	Males	Females	P-Value
Number	1092	805	
Age, years (\pm s.d.)	53 \pm 15	55 \pm 14	0.0020
Height, cm (\pm s.d.)	177 \pm 8	161 \pm 9	<0.0001
Weight, kg (\pm s.d.)	73 \pm 18	63 \pm 17	<0.0001
Body mass index, kg/m ² (\pm s.d.)	35 \pm 8	36 \pm 9	0.0003
Neck circumference, cm (\pm s.d.)	17 \pm 2	15 \pm 3	<0.0001
Ethnicity, no. (%)			
Caucasian	593 (54%)	401 (50%)	ns
Hispanic	471 (43%)	385 (48%)	0.045
Other [†]	28 (3%)	18 (2%)	ns
TST, minutes (\pm s.d.)	343 \pm 79	346 \pm 76	ns
Sleep efficiency, % (\pm s.d.)	77 \pm 16	78 \pm 17	ns
RDI, events/hour (\pm s.d.)	37 \pm 30	23 \pm 27	<0.0001
Lowest SaO ₂ , % (\pm s.d.)	77 \pm 12	81 \pm 10	<0.0001

*s.d. = standard deviation; cm = centimeters; kg = kilograms; m = meters; no. = number; ESS = Epworth Sleepiness Scale score; AHI = apnea-hypopnea index, SaO₂ = oxyhemoglobin percentage, ns = not significant.

[†]Other reported ethnicities: Black (n = 36), Asian (n = 4), American Indian (n = 2), Filipino (n = 1), Indian (n = 1), Lebanese (n = 1), and Portuguese (n = 1).

Table 2. Demographic and Clinical Characteristics of the Participants by Ethnicity

Group	Caucasian	Hispanic	Other	P-Value
Number	998	856	46	
Age, years (\pm s.d.)	56 \pm 14	52 \pm 14	52 \pm 16	<0.0001
Height, cm (\pm s.d.)	173 \pm 10	166 \pm 11	174 \pm 11	<0.0001
Weight, kg (\pm s.d.)	68 \pm 17	69 \pm 18	75 \pm 22	ns
Body mass index, kg/m ² (\pm s.d.)	34 \pm 8	37 \pm 9	37 \pm 11	<0.0001
Neck circumference, cm (\pm s.d.)	16 \pm 3	17 \pm 3	17 \pm 2	ns
Male, no. (%) [†]	593 (60%)	471 (55%)	28 (61%)	ns
TST, minutes (\pm s.d.)	340 \pm 78	348 \pm 79	346 \pm 66	ns
Sleep efficiency, % (\pm s.d.)	77 \pm 16	79 \pm 17	78 \pm 13	0.015
RDI, events/hour (\pm s.d.)	27 \pm 27	36 \pm 31	30 \pm 30	<0.0001
Lowest SaO ₂ , % (\pm s.d.)	80 \pm 10	77 \pm 12	77 \pm 15	0.0001

*s.d. = standard deviation; cm = centimeters; kg = kilograms; m = meters; no. = number; ESS = Epworth Sleepiness Scale score; AHI = apnea-hypopnea index; SaO₂ = oxyhemoglobin percentage; ns = not significant.

[†]Gender was not reported by 3 participants (all Caucasian).

otherwise unremarkable. Three subjects (all Caucasian) did not report their gender on the pre-PSG questionnaires. Table 2 shows the demographic and clinical characteristics separated by race.

Epworth Sleepiness Scale Scores Differ by Demographic and Clinical Characteristics

The mean \pm s.d. ESS score for all patients was 10.7 \pm 5.6. The distributions of ESS within each of the major groups evaluated (separated by gender, ethnicity, and obesity status) and subgroups (combinations of demographic and clinical factors such that only one subgroup is appropriate for each

patient) are shown in Fig. (1A, B), respectively. ESS scores varied significantly by group and subgroup (P<0.0001 by 1-way ANOVA for both sets). Within the major groups, the non-obese patients had significantly lower ESS scores than obese patients and the cohort as a whole. Males reported significantly higher ESS scores than females. The highest scores were noted in males (11.1 \pm 5.7), Hispanics (11.1 \pm 6.0), and obese patients (11.1 \pm 5.6). The only group with a mean score <10 was non-obese patients (9.7 \pm 5.5).

Only subgroups involving either Caucasian or Hispanic patients were analyzed, because a) the "other ethnicities" group was too small to effectively subdivide and b) clinically

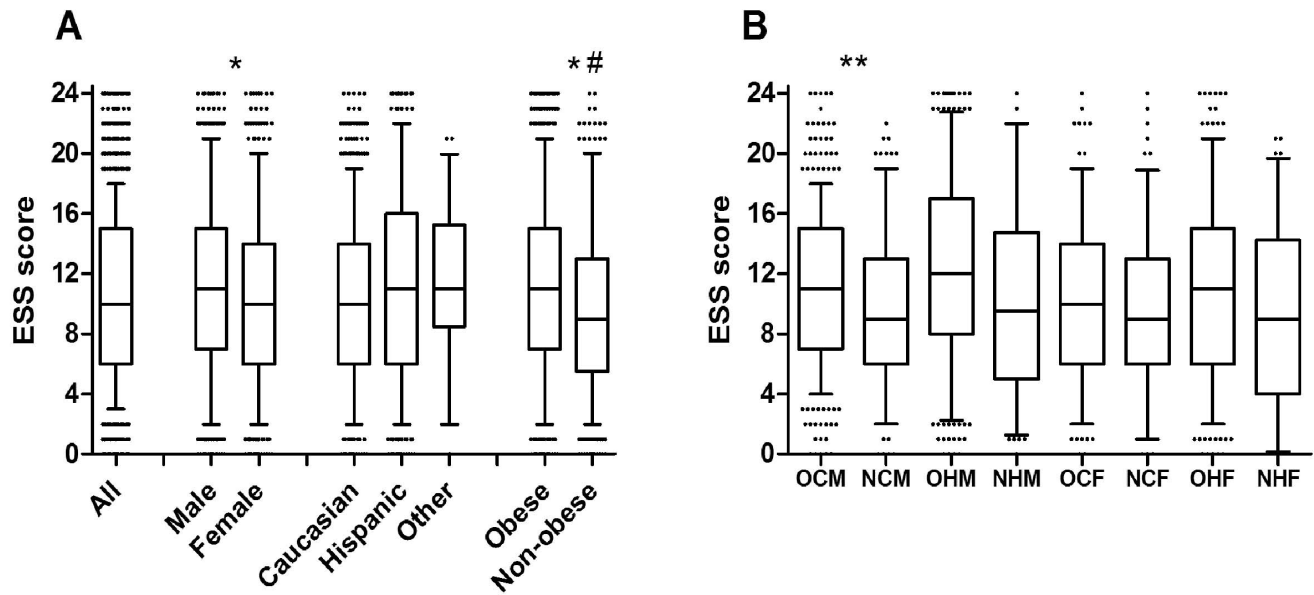


Fig. (1). Epworth Sleepiness Scale (ESS) scores of patients referred for polysomnography. (A) Box (25-75 percentile) and whisker (10-90 percentile) plots are shown depicting the ESS scores of All (N = 1900), Male (N = 1092), Female (N = 805), Caucasian (N = 998), Hispanic (856), Other ethnicities (Other, N = 46), Obese (N = 1370), and Non-obese (N = 529) patients. (B) Box and whisker plots are shown for subgroups of patients: obese Caucasian males (OCM, N = 397), non-obese Caucasian males (NCM, N = 198), obese Hispanic males (OHM, N = 364), non-obese Hispanic males (NHM, N = 104), obese Caucasian females (OCF, N = 269), non-obese Caucasian females (NCF, N = 130), obese Hispanic females (OHF, N = 303), and non-obese Hispanic females (NHF, N = 82). # Non-obese patients differed significantly in ESS scores than all patients. * ESS scores differed significantly between males and females, and between obese and non-obese patients. ** ESS scores differed significantly between obese and non-obese Caucasian males.

important conclusions could not be drawn from such heterogeneous subgroups. Most of the subgroups showed similar ESS scores to other subgroups that differed by a single factor (i.e., obese Caucasian males were similar to obese Caucasian females); the exception was obese Caucasian males who reported higher ESS scores than non-obese Caucasian males. Otherwise, obese Hispanic males scored higher than non-obese Caucasian males, obese Caucasian females, non-obese Caucasian females, and non-obese Hispanic females. The highest ESS scores were from obese Hispanic males (12.0 ± 6.0) and obese Caucasian males (11.2 ± 5.4). Non-obese Caucasian males (9.5 ± 5.1), non-obese Caucasian females (9.5 ± 5.2), non-obese Hispanic females (9.5 ± 5.8), and obese Caucasian females (9.97 ± 5.08) scored the lowest.

Relationship of the Epworth Sleepiness Scale Score to Obstructive Sleep Apnea

In the overall cohort of patients, those with OSA (RDI ≥ 15) had significantly higher ESS scores than those without OSA ($P < 0.0001$). The mean \pm s.d. ESS scores are shown in Table 3 for patients with and without OSA in the groups and subgroups mentioned above. The major groups for which the ESS was significantly higher in those with OSA included males, Caucasians, Hispanics, obese, and non-obese patients. For subgroups, the same held true for obese Caucasian males, obese Hispanic males, and non-obese Hispanic males. In non-obese Caucasian females, the ESS was actually higher in patients without OSA, though not significantly. The 90th percentile ESS scores for patients without OSA was calculated for each group, as this would serve as a potential

upper limit of normal (ULN) when the ESS is applied as a screening tool for OSA. Each group produced a calculated ULN too high to be clinically useful for detection of OSA. The 10th percentile of ESS in patients with OSA (i.e., lower limit of abnormal) was consistently in the 3-4 range for all groups and subgroups.

The correlations of between ESS and indices of OSA severity (RDI and lowest SaO₂) also varied between groups. The correlations for the overall study group, as well as each group and subgroup are listed in Table 4. In the overall study cohort, the ESS correlated significantly with RDI ($r = 0.17$, $P < 0.0001$) and lowest SaO₂ ($r = -0.20$, $P < 0.0001$). Of the groups that demonstrated a significant correlation between ESS and the markers of OSA, that association was almost universally stronger with the lowest SaO₂ than with RDI, though if there was a significant correlation with one of the markers, there was most likely also a significant correlation with the other within the same group. The correlation between ESS and RDI was most pronounced in males ($r = 0.21$), Hispanics ($r = 0.19$), and those that had OSA on the PSG ($r = 0.20$); significant correlations were seen with lowest SaO₂ were in the overall group ($r = -0.20$), males ($r = -0.22$), Caucasians ($r = -0.22$), Hispanics ($r = -0.24$), obese ($r = -0.20$), and those with OSA ($r = -0.25$). In the subgroups, consistently significant associations between the ESS and severity of OSA were seen in obese Caucasian males ($r = 0.21$ for RDI; $r = -0.21$ for lowest SaO₂), obese Hispanic males ($r = 0.18$ for RDI; $r = -0.22$ for lowest SaO₂), and non-obese Hispanic males ($r = 0.29$ for RDI; $r = -0.27$ for lowest SaO₂).

Table 3. Epworth Sleepiness Scale (ESS) Scores in Groups of Patients Referred for Polysomnography

Group	No OSA (Mean ± s.d.)	OSA (Mean ± s.d.)	P-Value	90 th pct for No OSA
All participants	9.8 ± 5.4	11.3 ± 5.7	<0.0001	17
Male	9.8 ± 5.5	11.7 ± 5.7	<0.0001	18
Female	9.8 ± 5.3	10.6 ± 5.5	ns	17
Caucasian	9.8 ± 5.2	10.7 ± 5.2	0.004	17
Hispanic	9.8 ± 5.6	11.9 ± 6.1	<0.0001	17
Other Ethnicities	10.0 ± 5.7	11.5 ± 4.8	ns	16
Obese	10.2 ± 5.3	11.6 ± 5.7	<0.0001	17
Non-obese	9.2 ± 5.4	10.3 ± 5.4	0.015	17
Subgroups				
Obese Caucasian Males	10.4 ± 5.5	11.5 ± 5.3	0.044	18
Non-obese Caucasian Males	9.1 ± 5.2	10.0 ± 5.0	ns	18
Obese Hispanic Males	10.3 ± 5.7	12.4 ± 6.1	0.018	18
Non-obese Hispanic Males	8.3 ± 5.5	11.7 ± 6.3	0.005	18
Obese Caucasian Females	9.8 ± 4.9	10.1 ± 5.3	ns	16
Non-obese Caucasian Females	9.7 ± 5.2	9.2 ± 4.5	ns	17
Obese Hispanic Females	10.2 ± 5.5	11.3 ± 6.1	ns	17
Non-obese Hispanic Females	9.2 ± 5.8	10.3 ± 6.0	ns	17

*s.d. = standard deviation; OSA = obstructive sleep apnea (respiratory disturbance index \geq 15 events/hour); pct = percentile; ns = not significant.

Performance of the Epworth Sleepiness Scale in Screening for Obstructive Sleep Apnea

In all participants, the sensitivity of ESS for predicting OSA was 54% and specificity was 57%. The positive predictive value (PPV) was 64% and negative predictive value (NPV) was 47. In males, the sensitivity was 56%, specificity was 58%, PPV was 74%, and NPV was 38%; in females, sensitivity was 49%, specificity was 56%, PPV was 48%, and NPV was 57%. Among the self-reported ethnicity groups, the ESS displayed the best sensitivity in Hispanics (56%) and best specificity in Caucasians (57%). The PPV was substantially better in Hispanics (69%) than either Caucasian (59%) or other races (50%), while the NPV was <50% for all 3 ethnicity groups. In obese subjects, sensitivity was 55%, specificity was 53%, PPV was 68%, and NPV was 40%, compared with sensitivity of 47%, specificity of 63%, PPV of 49%, and NPV of 61% for non-obese subjects. Sensitivities and specificities for each group are shown in Fig. (2A).

The sensitivities and specificities in clinical and demographic subgroups are shown in Fig. (2B). The highest sensitivity was seen in obese Hispanic males (59%); the corresponding specificity was 54%. The highest specificity was seen in non-obese Hispanic males (76%), which also had a sensitivity of 55%, making this subgroup the one in which the ESS is most accurate in predicting OSA. As above, sensitivity and specificity analyses were not performed for males or females of other ethnicities due to small group sizes. Calculations of the test characteristics were repeated using other ESS ULN to determine if what ULN would result in a sensitivity >90% in each subgroup. These results were similar to those from the overall group: in

each group, only an ESS ULN in the 3-4 range resulted in sufficient sensitivity to serve as a screening test, though also provided very poor specificity (7-22%).

DISCUSSION

In this cohort of patients referred for PSG evaluation of suspected OSA, we found that ESS scores differed by ethnicity, gender, and body morphometry. The ESS was highest in obese males (Hispanic and Caucasian); scores were lowest in non-obese females and non-obese Caucasian males. Some of these differences can be explained by the presence or absence of OSA; there were weak, but statistically significant correlations between ESS score and PSG indices of OSA severity (RDI and lowest SaO₂). However, the enormous overlap of ESS scores between those that have OSA and those that don't suggests additional unmeasured factors influencing ESS score.

It is unclear from these data why an obese male and non-obese female with similar severities of OSA may report different degrees of sleepiness using the ESS. Reported symptoms of sleep-disordered breathing are known to vary across ethnicities and genders [25, 26]. This is likely to have a strong influence on PSG referral patterns from primary care. Any cultural or language barriers that inhibit medical history-taking could also influence the likelihood of referral. This is an important factor in interpretation of our data, since our cohort was comprised of patients referred for PSG. Comparison with the at-large community may help elucidate factors behind these findings.

Demographic and clinical characteristics help stratify risk for a range of sleep disorders, not limited to OSA, many of

Table 4. Correlation Coefficients of the Epworth Sleepiness Scale with Severity of Sleep-Disordered Breathing in Groups and Subgroups of the Study Population

Group	Correlation with RDI	Correlation with Lowest SaO ₂
Overall	0.17**	-0.20**
Males	0.21**	-0.22**
Females	0.09*	-0.14**
Caucasian	0.14**	-0.22**
Hispanic	0.19**	-0.24**
Other ethnicities	0.15	-0.16
Obese	0.17**	-0.20**
Non-obese	0.10*	-0.09
OSA	0.20**	-0.25**
No OSA	0.01	-0.01
Subgroup combinations		
Obese Caucasian males	0.21**	-0.21**
Non-obese Caucasian males	0.07	-0.08
Obese Hispanic males	0.18**	-0.22**
Non-obese Hispanic males	0.29**	-0.27**
Obese males – other ethnicities	0.14	-0.19
Non-obese males – other ethnicities	-0.36	-0.05
Obese Caucasian females	0.03	-0.04
Non-obese Caucasian females	0.03	-0.03
Obese Hispanic females	0.10	-0.21*
Non-obese Hispanic females	0.08	-0.03
Obese females – other ethnicities	0.22	-0.02
Non-obese females – other ethnicities	0.80	-0.80

* P < 0.05.

** P < 0.001.

RDI = respiratory disturbance index; SaO₂ = oxyhemoglobin percentage; OSA = obstructive sleep apnea (RDI ≥ 15).

which may result in the symptom of sleepiness. The most obvious example of this in our study is obesity, which is a known predisposing factor for OSA. Other, more subtle, factors are also likely to play a role. Females, especially premenopausal women, are at a somewhat reduced risk of OSA compared with males; therefore the likelihood that another sleep disorder is causing the increased sleepiness is relatively higher. The role of ethnicity in the risk of OSA is less clear and bears further study. One study has shown that African American patients with OSA often present younger and with more severe disease [27]. While this is not likely to affect our results due to the small proportion of African American participants, the Hispanic group in our current study was also younger and had higher overall RDI than the Caucasian group. Another factor that must be considered is the possibility that an RDI ≥ 15 holds different meanings for different populations. For example, while the Sleep Heart Health Study demonstrated increased risk of cardiovascular diseases and mortality in men with moderate-to-severe OSA, the same relationships did not always hold up in women [1, 2, 4]. Current categorization of OSA is the same across all

adult populations as large, population-based studies, generating normative data based on ethnicity and gender are lacking to this point.

The ESS performed modestly overall in predicting significant OSA in our cohort. The test characteristics varied based on the population to which it was applied. Overall, the ESS demonstrated the highest sensitivity for OSA in obese males of all ethnicities, and the specificity was highest in non-obese Caucasian and Hispanic males. In all analyses, the test performed best in non-obese Hispanic men. When applied to women, especially non-obese women, the sensitivity suffered. Thus, while the ESS can be a useful tool for the evaluation of OSA, the results must be taken in context of the overall clinical picture; the clinical and demographic attributes of the patient can be somewhat helpful in interpreting the ESS.

When comparing the present study with an earlier study evaluating the utility of ESS in patients with OSA [11], there are some significant differences in the results. The prior study reported significant correlations between the ESS and

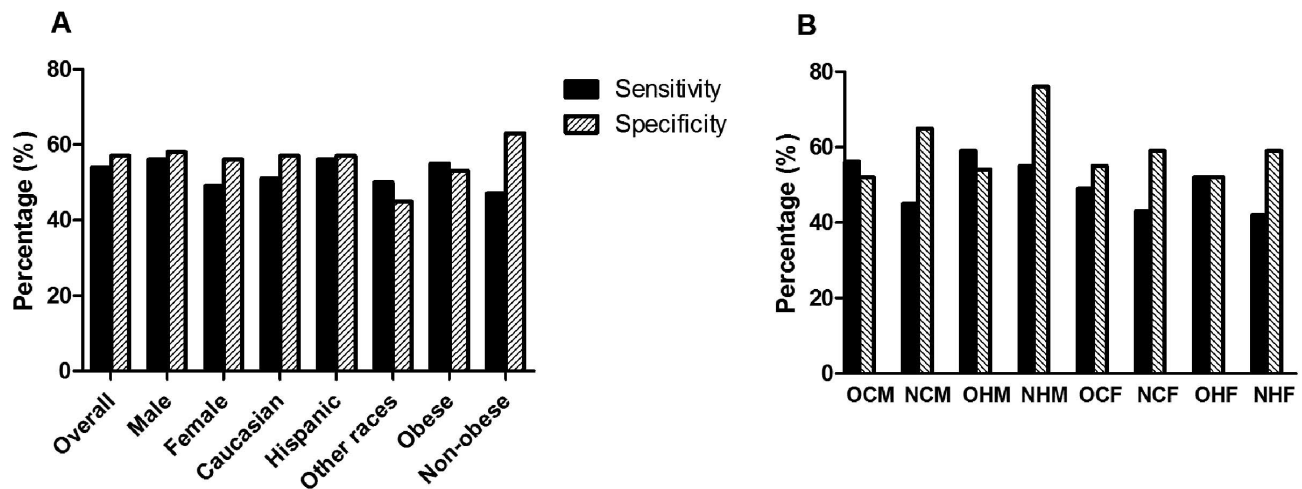


Fig. (2). Test characteristics of the Epworth Sleepiness Scale (ESS) in the diagnosis of obstructive sleep apnea (OSA), related to gender, ethnicity, and body mass index. (A) The sensitivity and specificity of the ESS for the diagnosis of OSA are plotted for the overall study population (N = 1900), males (N = 1092), females (N = 805), Caucasians (N = 998), Hispanics (N = 856), other races (N = 46), obese (N = 1370), and non-obese (N = 529). (B) The sensitivity and specificity of the ESS for the diagnosis of OSA are plotted for obese Caucasian males (OCM, N = 397), non-obese Caucasian males (NCM, N = 198), obese Hispanic males (OHM, N = 364), non-obese Hispanic males (NHM, N = 104), obese Caucasian females (OCF, N = 269), non-obese Caucasian females (NCF, N = 130), obese Hispanic females (OHF, N = 303), and non-obese Hispanic females (NHF, N = 82).

both RDI ($r = 0.44$) and minimum SaO_2 ($r = -0.40$). In the current study, while the ESS is significantly correlated with both PSG measures of OSA, the correlations are not nearly as strong. Some technical details of the analysis may contribute to this discrepancy. The scoring criteria have changed substantially since 1993; therefore while RDI represents similar results in both studies, the absolute and relative values may be markedly different. Most likely, the divergent results are due to dissimilar study populations. Though both studies made use of patients referred to a sleep lab for suspicion of OSA, 40% of the subjects in the previous study did not have any significant OSA ($\text{RDI} < 5$), compared with 17% in the current study. In addition, the present study population tended to be older and more obese than the earlier study group. It is unclear if gender or ethnic differences play a role in the discrepancies.

A more recent study has re-evaluated ESS cutoff points for the identification of OSA [13]. That study used a receiver-operator curve (ROC) to evaluate the study performance, finding an area under ROC = 0.60 for detection of $\text{AHI} \geq 5$ and the optimal cutoff score was $\text{ESS} > 8$. The calculated sensitivity in our present study is lower than reported in that study, even when using $\text{RDI} \geq 5$ and $\text{ESS} > 8$ as cutoff values (76% in the previous study vs 63% now). This again may be due to a change in respiratory event scoring (scoring rules changed in the interim), as the previously reported specificity was 31% compared with 44% in our data set. A reduction in sensitivity, combined with increase in specificity, suggests a consistent shift in the data; a similar result is seen when moving the ESS cutoff to 6 in our data, but keep the $\text{RDI} \geq 15$ cutoff (sensitivity = 75%; specificity = 31%). Because of these differences in scoring techniques or other reasons, our data do not support the use of either $\text{ESS} > 10$ or $\text{ESS} > 8$ alone as screening tools for OSA, given the low sensitivity.

A limitation of the present study that may hinder generalization of the results is the population under

investigation. The study was performed at a single sleep center, which maintains consistency in scoring and implementation of the tests, but the population in this area is heavily skewed toward Caucasians and Hispanics. Given such few participants from other ethnicities, few if any, conclusions should be drawn about those subgroups. Additionally, our population had a high prevalence of OSA. The cutoff of $\text{RDI} \geq 15$ was chosen in part to minimize bias introduced by the high prevalence of OSA in the population, which can alter the PPV and NPV results, but may also impact the sensitivity and specificity calculations. The results were similar when we analyzed our data using an $\text{RDI} \geq 5$ as the cutoff for OSA.

Another major limitation is that all patients were referred to the sleep center for a sleep test and thus the results may not be generalizable to a more heterogeneous primary care practice which includes patients with varied symptom and clinical profiles. At the same time, the ESS marks the single most commonly used modality to trigger a decision point with regards to referral for a sleep evaluation. It is our view that a more realistic appraisal of the predictive characteristics of this measure will enable primary care physicians and other specialty physicians, who are trying to ascertain sleep apnea risk in their heterogeneous population, to avoid an exclusive reliance on this one metric in influencing their referral decisions. We suspect that the modest associations seen in a population that was referred to a sleep center will only be rendered less specific in a more heterogeneous population which may include patients with varied other that may be associated with excessive daytime sleepiness.

While our study analyzed the effects of gender, ethnicity, and obesity on the ability of the ESS to detect OSA, other clinical and demographic characteristics should also be considered for analysis. Chief among the potential candidates are age and mood disorders. A study reported that these affect self-reported daytime sleepiness independently of sleep apnea [28], these also may alter the way in which sleep-disordered

breathing is perceived by the patient. Data on the effects of age on self-reported daytime sleepiness have been inconsistent [19-21]. If the proper clinical characteristics can be identified, normative data for the ESS could then be determined for those subgroups, optimizing its utility as a screening tool for OSA.

CONCLUSIONS

While the ESS is well-validated to detect sleepiness, including patients with OSA, it is probably influenced by other factors, including gender, ethnicity, and body morphometry. Its sensitivity to detect clinically important OSA is insufficient to be used as a screening tool in the absence of other clinical data. Therefore, patients in whom there is clinical suspicion for OSA should undergo diagnostic testing, even in cases with normal ESS scores. The test characteristics of the ESS improve significantly when applied to select populations with increased risk for OSA, such as obese males.

ACKNOWLEDGEMENTS

Salim Surani: Design, collection of data, and preparation and review of manuscript.

Sean Hesselbacher: Data analysis and preparation of manuscript.

Jerry Allen: Data collection and entry.

Sara Surani: Data collection and entry.

Shyam Subramanian: Study design, review of data and review of manuscript.

CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest.

REFERENCES

- [1] Punjabi NM, Caffo BS, Goodwin JL, Gottlieb DJ. Sleep-disordered breathing and mortality: a prospective cohort study. *PLoS Med* 2009; 6: e1000132.
- [2] Nieto FJ, Young TB, Lind BK, *et al.* Pickering TG for the Sleep Heart Health Study. Association of sleep-disordered breathing, sleep apnea, and hypertension in a large community-based study. *JAMA* 2000; 283: 1829-36.
- [3] Punjabi NM, Shahar E, Redline S, Gottlieb DJ, Givelber R, Resnick HE. Sleep Heart Health Study Investigators. Sleep-disordered breathing, glucose intolerance, and insulin resistance. *Am J Epidemiol* 2004; 160: 521-30.
- [4] Chami HA, Devereux RB, Gottdiener JS, *et al.* Left ventricular morphology and systolic function in sleep-disordered breathing: the Sleep Heart Health Study. *Circulation* 2008; 117: 2599-607.
- [5] O'Connor GT, Caffo B, Newman AB, *et al.* Prospective study of sleep-disordered breathing and hypertension. *Am J Respir Crit Care Med* 2009; 179: 1159-64.
- [6] Redline S, Yenokyan G, Gottlieb DJ, *et al.* Obstructive sleep apnea-hypopnea and incident stroke: the Sleep Heart Health Study. *Am J Respir Crit Care Med* 2010; 182: 269-77.
- [7] George CF, Boudreau AC, Smiley A. Effects of nasal CPAP on simulated driving performance in patients with obstructive sleep apnoea. *Thorax* 1997; 52: 648-53.
- [8] Turkington PM, Sircar M, Allgar V, Elliott MW. Relationship between obstructive sleep apnoea, simulated driving performance, and risk of road traffic accidents. *Thorax* 2001; 56: 800-5.
- [9] Pichel F, Zamarrón C, Magán F, Rodríguez JR. Sustained attention measurements in obstructive sleep apnea and risk of traffic accidents. *Respir Med* 2006; 100: 1020-7.
- [10] Johns MW. A new method for measuring daytime sleepiness: the Epworth Sleepiness Scale. *Sleep* 1991; 14: 540-5.
- [11] Johns MW. Daytime sleepiness, snoring, and obstructive sleep apnea: the epworth sleepiness scale. *Chest* 1993; 103: 30-6.
- [12] Hirshkowitz M, Bibbs M, Smith JM. Epworth Sleepiness Scale normative values [abstract]. *Sleep Med* 2003; 4(Suppl 1): S18.
- [13] Rosenthal LD, Dolan DC. The epworth sleepiness scale in the identification of obstructive sleep apnea. *J Nerv Ment Dis* 2008; 196: 429-31.
- [14] Doghramji PP. Recognition of obstructive sleep apnea and associated excessive daytime sleepiness in primary care. *J Fam Pract* 2008; 57 (Suppl 8): S17-23.
- [15] NCD for continuous positive airway pressure (CPAP) therapy for obstructive sleep apnea. 15 Oct 2008. Available at: <http://www.cms.gov/transmittals/downloads/R96NCD.pdf> [Accessed on 20 Mar 2011].
- [16] Subramanian S, Guntupalli B, Murugan T, *et al.* Gender and ethnic differences in prevalence of self-reported insomnia among patients with obstructive sleep apnea. *Sleep Breath* 2011; 15: 711-5.
- [17] Ye L, Pien GW, Ratcliffe SJ, Weaver TE. Gender differences in obstructive sleep apnea and treatment response to continuous positive airway pressure. *J Clin Sleep Med* 2009; 5: 512-8.
- [18] Baldwin CM, Ervin AM, Mays MZ, *et al.* Sleep disturbances, quality of life, and ethnicity: the Sleep Heart Health Study. *J Clin Sleep Med* 2010; 6: 176-83.
- [19] Sanford SD, Lichstein KL, Durrence HH, Riedel BW, Taylor DJ, Bush AJ. The influence of age, gender, ethnicity, and insomnia on Epworth sleepiness scores: a normative US population. *Sleep Med* 2006; 7: 319-26.
- [20] Gander PH, Marshall NS, Harris R, Reid P. The Epworth Sleepiness Scale: influence of age, ethnicity, and socioeconomic deprivation. Epworth sleepiness scores of adults in New Zealand. *Sleep* 2005; 28: 249-53.
- [21] Walsleben JA, Kapur VK, Newman AB, *et al.* Sleep and reported daytime sleepiness in normal subjects: the Sleep Heart Health Study. *Sleep* 2004; 27: 293-8.
- [22] Baldwin CM, Kapur VK, Holberg CJ, Rosen C, Nieto FJ. Sleep Heart Health Study Group. Associations between gender and measures of daytime somnolence in the Sleep Heart Health Study. *Sleep* 2004; 27: 305-11.
- [23] Iber C, Ancoli-Israel S, Chesson A, Quan S. 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.
- [24] Surani S, Aguillar R, Aguillar R, Subramanian S. Standardization of quality assurance for technologist: a model. *Sleep Breath* 2010; 14: 3-12.
- [25] O'Connor GT, Lind BK, Lee ET, *et al.* Variation in symptoms of sleep-disordered breathing with race and ethnicity: The Sleep Heart Health Study. *Sleep* 2003; 26: 74-9.
- [26] Valipour A, Lothaller H, Rauscher H, Zwick H, Burghuber OC, Lavie P. Gender-related differences in symptoms of patients with suspected breathing disorders in sleep: a clinical population study using the sleep disorders questionnaire. *Sleep* 2007; 30: 312-9.
- [27] Scharf SM, Seiden L, DeMore J, Carter-Pokras O. Racial differences in clinical presentation of patients with sleep-disordered breathing. *Sleep Breath* 2004; 8: 173-83.
- [28] Bixler EO, Vgontzas AN, Lin HM, Calhoun SL, Vela-Bueno A, Kales A. Excessive daytime sleepiness in a general population sample: the role of sleep apnea, age, obesity, diabetes, and depression. *J Clin Endocrinol Metab* 2005; 90: 4510-5.