Analysis of Error Profiles Occurring during the OSLER Test

A Sensitive Mean of Detecting Fluctuations in Vigilance in Patients with Obstructive Sleep Apnea Syndrome

Stephanie Mazza, Jean-Louis Pepin, Chrystele Deschaux, Bernadette Naegle, and Patrick Levy

Sleep Laboratory and Neuropsychological Unit, PRETA Laboratory TIMC CNRS 5525, University Hospital, Grenoble, France

The OSLER test represents a simple alternative to the maintenance of wakefulness test. Standard analysis of OSLER test results yields a mean sleep latency (MSL). The aim of this study was to use both MSL and errors (nonresponses to stimulations) to characterize daytime sleepiness in apneic patients. OSLER test results at 9:00 A.M., 11:00 A.M., and 1:30 P.M. were compared in 27 obstructive sleep apnea syndrome patients (30.4 ± 10.4 years; apnea–hypopnea index: 43.05 ± 25.08) and 20 control subjects (C). Not only did obstructive sleep apnea syndrome patients demonstrate earlier sleep onset than control subjects (MSL: 1,788 seconds ± 511 versus 2,335 seconds ± 139, p < 0.001), but they also spent a greater percentage of time making errors than control subjects (5.4% ± 4.7 of total test time versus 0.4% ± 0.4, p < 0.001) with specific error profiles (high prevalence of three to six consecutive errors). When error profile analysis was added to standard sleep latency assessment, up to 40% of patients with normal sleep latency were exhibiting abnormal fluctuation in vigilance. A single 9:00 A.M. OSLER session appeared as sensitive as three consecutive sessions in its use as a means of identifying patients with significant daytime sleepiness. The OSLER test session was least specific to distinguish apneic subjects from normal subjects, suggesting that the OSLER test can identify the afternoon peak in physiologic somnolence.

Keywords: daytime sleepiness; vigilance; OSLER test; obstructive sleep apnea syndrome.

Daytime sleepiness is a major symptom of obstructive sleep apnea syndrome (OSAS). Sleep fragmentation and the resulting excessive daytime sleepiness are considered instrumental in causing impairment of daytime functioning. As a result, patients with OSAS become prone to accidents as well as to marital and occupational problems (1,2).

The quantification of daytime somnolence is not simple. Complaints do not always coincide with an objective impairment of vigilance. Questionnaires used in assessing subjective sleepiness (e.g., the Epworth scale [3] or the Stanford scale [4]) do not strongly correlate with objective quantifications (5). Assessing daytime impairment by using objective quantification tools may provide a better evaluation of daytime function alteration in these patients. The most widely used objective test of sleepiness is the multiple sleep latency test (MSLT) (6). This test measures an individual’s tendency to fall asleep. However, patients with OSAS complain of an inability to stay awake and thus a maintenance of wakefulness test (7), where the instruction is to stay awake rather than fall asleep, may be more clinically relevant in assessing these patients.

These standard tests for objective quantification of sleepiness are labor intensive because they require electroencephalographic (EEG) recordings to determine the exact time of sleep onset. Another drawback is the need for a technician to constantly monitor the EEG trace during the test, which is time consuming. Hence difficulties arise in using such tests in clinical practice and studies in the community.

The OSLER test has been recently proposed as a behavioral test which simplifies the performance of the maintenance of wakefulness test. Indeed, it has been used as a simple alternative to MWT in a study comparing 10 control subjects to 10 severe OSAS patients (8). In this test, the occurrence of sleep is assessed behaviorally rather than by EEG monitoring. The subject is asked to respond by hitting a button each time a dim light flashes. The light-emitting diode flashes regularly for 1 second every 3 seconds. The subject is instructed to remain awake in this soporific situation for a maximum testing time of 40 minutes. When the subject fails to respond for 21 seconds (i.e., seven consecutive illuminations), the test is ended and it is assumed that the patient has fallen asleep. Thus, the OSLER test reproduces many of the MWT characteristics, with the advantage of being a simpler and less expensive tool, which does not require the presence of a trained technician. The simplicity of the test makes it easy to be used outside the sleep laboratory setting.

The standard way to analyze the OSLER test is to determine the sleep latency, measured as the delay before the appearance of seven consecutive flashes without response (21 seconds). A 21-second threshold has been chosen because it corresponds to the minimal sleep duration generally used to score one epoch of sleep when using standard scoring rules for night polysomnography (9). However, this threshold could be inappropriate in detecting fluctuation in vigilance occurring before the end of the test (10). Indeed, it is quite possible for a somnolent subject to fall asleep several times during the test for less than 21 seconds without ever missing seven consecutive stimuli, and thus finish the 40 minute test (see example in Figure 1). Alternatively, subjects with low motivation can miss occasional LED flashes without any significant sleepiness. In this case, errors are not expected to be consecutive. Priest and colleagues (10) EEG data recorded during the OSLER test showed that normal subjects missing three consecutive flashes (9 seconds) had a 92% chance of simultaneously manifesting an EEG microsleep. This led the present authors to hypothesize that an increase in the number of consecutive misses would indicate that the subject had fluctuations in vigilance during the intervening period.

Analyzing sleep latencies alone (i.e., seven consecutive misses) does not give a thorough view of fluctuations in vigilance during the period preceding the end of the OSLER test. Therefore, the analysis of all omissions and their distribution...
The OSLER Test

Test procedure. The OSLER test consisted of 40 minute sleep-resistance challenges conducted in a dark room isolated from external noise. The subject, dressed and lying in a semirecumbent position, was asked to remain awake without using specific strategies. By hitting a button placed on a box directly connected to a personal computer, the subject was instructed to respond to a visual stimulus (light-emitting diode flash) which appeared for 1 second every 3 seconds. A total of 800 stimulations per test was emitted. All tests were video recorded to verify that the subjects were following instructions. Each subject underwent the OSLER test three times (at 9:00 a.m., 11:00 a.m. and 1:30 p.m.) allowing the assessment of vigilance at different times of day.

In our clinical experience, the reproducibility of the fourth session of the OSLER test is less satisfactory than that of the three other sessions. Both patients and control subjects were reluctant to accept an additional 40-minute session. As a consequence, we decided to perform three sessions, instead of four as originally described (8).

All patients with OSAS and five of the control subjects underwent polysomnography the night before the OSLER test. Control subjects who did not undergo polysomnography were instructed to follow their usual sleeping habits. They were asked whether sleep quality had been normal during the night before the OSLER test.

The first OSLER session (9:00 a.m.) was started on average 2 hours after subjects had awakened.

Data Analysis from the OSLER Test

Sleep latency. Each session ended automatically after 40 minutes or before if the subject did not give a response to seven consecutive flashes (i.e., 21 seconds), which was taken as an indication of sleep onset (8). Twenty-one seconds has been proposed as a test termination time because it represents approximately the minimal time needed to score a conventional epoch of sleep (9). The computer stored the time to test termination, which is assumed to represent sleep latency. A sleep latency (test duration) was determined for each session (9:00 a.m., 11:00 a.m., and 1:30 p.m.). The mean sleep latency of the three OSLER tests was calculated.

Error Analysis

Errors made by patients and control subjects were compared. The percentage of time corresponding to the appearance of errors during the OSLER test was also taken into account ([3 seconds × number of omissions/sleep latency duration in seconds] × 100).

Lastly, the pattern of consecutive errors during the test was analyzed. The relative proportion of each subtype of consecutive errors (i.e., 2, 3, 4, 5, 6, 7) was used to identify various levels of sleepiness. For each test, consecutive errors were allocated to an EP type, (i.e., two consecutive errors = EP2, three consecutive errors = EP3). The EPs were clustered into type EP1–2 for one and two consecutive errors, indicating lack of attention, type EP3–6 for three to six consecutive errors, indicating a longer period of inattention, and type EP7 for seven consecutive errors, characterizing sleep onset during the OSLER test. Priest and coworkers (10) previously demonstrated in normal subjects that seven consecutive errors (21-second misses) during the OSLER test was almost always associated with the occurrence of an EEG defined episode of microsleep. In addition, the probability of microsleep was 43% and 83% for 3 and 6 second misses, respectively. For 9-second misses, the probability of recording a microsleep on the EEG trace reached 92%.

Percent of patients and control subjects demonstrating fluctuation in vigilance during the OSLER test. Based on the findings of Priest and colleagues (10), one can postulate that one period of 9 to 18 second misses (one EP3–6) or several episodes of 3–6 second misses (five EP1–2) are likely to be associated with sleep periods and, therefore, that such errors represent fluctuations in vigilance during the test. For this reason, we extended our analysis to an examination of the percent of patients and control subjects demonstrating such patterns over the three sessions of the OSLER test.

Statistical Analysis

The unpaired t test or Mann-Whitney test was used to compare control and patient groups for quantitative variables. The $\chi^2$ square test was used for qualitative variables. Results are expressed as mean ± SD and statistical significance was accepted for $p < 0.05$.

METHODS

Subjects

A total of 27 patients who had been referred to a tertiary sleep laboratory for clinical suspicion of OSAS that was confirmed by polysomnography (20 males and 7 females, age 50.3 ± 10.4 years) were studied. 20 nonobese, nonsnorer healthy volunteers (14 males and 6 females, age 44.2 ± 10.3 years) served as control subjects. A polysomnography was performed in a random subset of five control subjects (25% of the control group) to confirm the absence of subclinical sleep breathing disorder. None of the participants were remunerated for their participation in the study.

Polysomnography

Continuous recordings of EEG (C3/A2-C4/A1-Cz/O1 from the international 10–20 Electrode Placement System), eye movement measurements, chin electromyogram, and ECG with modified V2 lead were obtained. Airflow was measured with nasal pressure, as well as with the sum of buccal and nasal thermistor signals. Respiratory effort was monitored with uncalibrated inductance respiratory plethysmography. An additional signal of respiratory effort (i.e., pulse transit with the sum of buccal and nasal thermistor signals. Respiratory effort was monitored with uncalibrated inductance respiratory plethysmography. An additional signal of respiratory effort (i.e., pulse transit time or esophageal pressure) was recorded concurrently. Oxygen saturation was measured using a pulse oximeter (Biox-Ohmeda 3700; Ohmeda; Liberty Corner, NJ).

The OSLER test was instructed to respond to a visual stimulus (light-emitting diode flash) which appeared for 1 second every 3 seconds. A total of 800 stimulations per test was emitted. All tests were video recorded to verify that the subjects were following instructions. Each subject underwent the OSLER test three times (at 9:00 a.m., 11:00 a.m. and 1:30 p.m.) allowing the assessment of vigilance at different times of day. In our clinical experience, the reproducibility of the fourth session of the OSLER test was less satisfactory than that of the three other sessions. Both patients and control subjects were reluctant to accept an additional 40-minute session. As a consequence, we decided to perform three sessions, instead of four as originally described (8).

All patients with OSAS and five of the control subjects underwent polysomnography the night before the OSLER test. Control subjects who did not undergo polysomnography were instructed to follow their usual sleeping habits. They were asked whether sleep quality had been normal during the night before the OSLER test.

The first OSLER session (9:00 a.m.) was started on average 2 hours after subjects had awakened.

Data Analysis from the OSLER Test

Sleep latency. Each session ended automatically after 40 minutes or before if the subject did not give a response to seven consecutive flashes (i.e., 21 seconds), which was taken as an indication of sleep onset (8). Twenty-one seconds has been proposed as a test termination time because it represents approximately the minimal time needed to score a conventional epoch of sleep (9). The computer stored the time to test termination, which is assumed to represent sleep latency. A sleep latency (test duration) was determined for each session (9:00 a.m., 11:00 a.m., and 1:30 p.m.). The mean sleep latency of the three OSLER tests was calculated.

Error Analysis

Errors made by patients and control subjects were compared. The percentage of time corresponding to the appearance of errors during the OSLER test was also taken into account ([3 seconds × number of omissions/sleep latency duration in seconds] × 100).

Lastly, the pattern of consecutive errors during the test was analyzed. The relative proportion of each subtype of consecutive errors (i.e., 2, 3, 4, 5, 6, 7) was used to identify various levels of sleepiness. For each test, consecutive errors were allocated to an EP type, (i.e., two consecutive errors = EP2, three consecutive errors = EP3). The EPs were clustered into type EP1–2 for one and two consecutive errors, indicating lack of attention, type EP3–6 for three to six consecutive errors, indicating a longer period of inattention, and type EP7 for seven consecutive errors, characterizing sleep onset during the OSLER test. Priest and coworkers (10) previously demonstrated in normal subjects that seven consecutive errors (21-second misses) during the OSLER test was almost always associated with the occurrence of an EEG defined episode of microsleep. In addition, the probability of microsleep was 43% and 83% for 3 and 6 second misses, respectively. For 9-second misses, the probability of recording a microsleep on the EEG trace reached 92%.

Percent of patients and control subjects demonstrating fluctuation in vigilance during the OSLER test. Based on the findings of Priest and colleagues (10), one can postulate that one period of 9 to 18 second misses (one EP3–6) or several episodes of 3–6 second misses (five EP1–2) are likely to be associated with sleep periods and, therefore, that such errors represent fluctuations in vigilance during the test. For this reason, we extended our analysis to an examination of the percent of patients and control subjects demonstrating such patterns over the three sessions of the OSLER test.

Statistical Analysis

The unpaired t test or Mann-Whitney test was used to compare control and patient groups for quantitative variables. The $\chi^2$ square test was used for qualitative variables. Results are expressed as mean ± SD and statistical significance was accepted for $p < 0.05$. 

Figure 1. OSLER test results: misses scale on y-axis (0–7), time scale on x-axis (2,400 seconds).
RESULTS

Subjects
Table 1 summarizes control subjects and patients’ anthropometric and sleep data. Patients were middle-aged and had moderate to severe OSAS associated with nocturnal desaturations. Control subjects did not significantly differ from patients with OSAS for sex ratio and age but were leaner.

Assessment of Vigilance: OSLER Test Analysis
None of the subjects was excluded on the basis of monitoring by video recording. All subjects followed instructions correctly.

Sleep latencies and mean sleep latency. Sleep latencies for each test throughout the day are shown in Figure 2. For all tests, patients demonstrated significantly shorter sleep latencies than control subjects. The mean sleep latency in the control and the patient groups was 2.335 ± 139 seconds (38.9 minutes) and 1.788 ± 511 seconds (29.8 minutes), respectively (p < 0.001) (Figure 2).

Error analysis. The mean percent of subjects making errors during the OSLER test was significantly higher in the OSAS group than in the control group (91.3% of patients with OSAS versus 51.7% of control subjects).

For all three sessions, patients with OSAS had more errors in each category of EP (1–2, p < 0.0001; 3–6, p < 0.00001; 7, p < 0.0001) than control subjects. Moreover, the difference in the number of errors between the two groups was probably underestimated, the recording time being shorter in the OSAS group (i.e., shorter sleep latency).

Figure 3 shows the percentage of OSLER test duration spent making errors ([3 seconds × number of omissions/sleep latency duration in seconds] × 100). For the three sessions, OSAS patients spent more time in omissions than the control subjects. On average, 5.4% of the time of each session represented missed responses in the group of patients with OSAS, whereas only 0.4% of the control subjects’ test time was spent in omissions (p < 0.001).

An analysis of the distribution of errors shows that both groups did not demonstrate the same error profile (Table 2). The control group demonstrated a significantly larger proportion of type 1–2 errors (EP1–2) (3 or 6 seconds of omission) at 9:00 A.M. and 11:00 A.M. compared with patients (9:00 A.M.: p < 0.0009, 11:00 A.M.: p < 0.001). Throughout the three sessions, the patients with OSAS had a significantly larger proportion of type 3–6 errors (EP3–6) than control subjects (9:00 A.M., p < 0.01; 11:00 A.M., p = 0.006; 1:30 P.M., p < 0.01). Similar results were found in EP7, the patients group demonstrating more EP7 (seven consecutive omissions) than the control subjects at 9:00 A.M. and 11:00 A.M. (Table 2).

Percent of patients and control subjects demonstrating fluctuation in vigilance during the OSLER procedures is shown in Figure 3. From normal subjects, indicating that the OSLER test may indeed detect mild fluctuations in vigilance. Our findings suggest that the OSLER test is a sensitive tool in the detection of mild fluctuations in vigilance.

DISCUSSION
This is the first study using both sleep latencies and error profiles during the OSLER test as a means of characterizing daytime sleepiness in patients with OSAS. Patients with OSAS not only fell asleep earlier compared with control subjects (mean sleep latency: 1.788 seconds ± 511 versus 2.335 seconds ± 139, respectively; p < 0.001) but they also demonstrated specific patterns of error profiles (a high prevalence of three to six consecutive errors). When an error profile analysis was added to the classical sleep latency assessment, a loss of vigilance could be suspected in up to 40% of patients who presented normal sleep latencies. A single 9:00 A.M. OSLER session appeared as sensitive as three consecutive sessions were in identifying patients with significant daytime sleepiness. Indeed, only two patients with OSAS out of 27 demonstrated errors on the 11:00 A.M. session, whereas the 9:00 A.M. session was normal. On the other hand, the 1:30 P.M. OSLER test session was less specific in distinguishing patients with OSAS from normal subjects, indicating that the OSLER test may identify the afternoon peak in physiologic somnolence. Our findings suggest that the OSLER test is a sensitive tool in the detection of mild fluctuations in vigilance.

The OSLER test has been proposed as a simplified MWT which does not require EEG recordings or technician supervision. Its ability to distinguish normal subjects from severe OSAS patients (8) in terms of hypsomnolence and to detect vigilance fluctuations after sleep deprivation in normal subjects (10) has been demonstrated. Moreover, the OSLER test appears sensitive enough to identify improvement in sleepiness after continuous positive airway pressure treatment (12).

TABLE 1. ANTHROPOMETRIC AND SLEEP DATA ON PATIENTS AND CONTROL SUBJECTS

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>Age, yr (mean, SD)</th>
<th>BMI, kg/m² (mean, SD)</th>
<th>ESS (mean, SD)</th>
<th>AH1 (mean, SD)</th>
<th>Mean Nocturnal</th>
<th>Minimal Nocturnal</th>
<th>Time Spent at S90% (mean, SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control subjects</td>
<td>20, 14/6</td>
<td>44.17, 13.32</td>
<td>22.99, 3.33</td>
<td>5.55, 3.38</td>
<td>11.49, 6.18</td>
<td>96.4, 1.12</td>
<td>92.25, 2.37</td>
<td>0.025, 0.05</td>
</tr>
<tr>
<td>Patients</td>
<td>27, 20/7</td>
<td>50.31, 10.37</td>
<td>28.5, 5.12*</td>
<td>13.3, 5.13*</td>
<td>43.05, 25.08*</td>
<td>92.90, 2.68*</td>
<td>74.27, 17.57*</td>
<td>11.15, 18.7*</td>
</tr>
</tbody>
</table>

Definition of abbreviations: AH1 = apnea-hypopnea index; BMI = body mass index; ESS = Epworth sleepiness scale; NS = not significant; S90% = oxygen saturation.

* p < 0.05.
However, normative data are available only from two small groups of 12 and 10 subjects, respectively (8, 10), and a complete analysis of error profiles has never been reported in unselected groups of OSAS and control subjects.

In our study, all the normal subjects were able to stay awake for 40 minutes during the OSLER tests performed at 9 and 11:00 A.M. In the reports by Bennett and coworkers (8) and Priest and coworkers (10), the majority of control subjects were able to finish both 40 minute sessions without 21 second misses. On the other hand, the significant reduction in sleep latencies observed in OSAS patients in our study confirms the OSLER test's capacity to distinguish between control subjects and OSAS patients. The mean sleep latency in our subjects tended to be longer than in the initial validation study (29 minutes versus 10 minutes, respectively). This initial study was performed in a group of severe, highly symptomatic patients with OSAS (Epworth sleepiness scale = 17, O₂ saturation dip rate = 32.7/hour [9.7–65.6]) (8). Mean sleep latencies later reported by the Oxford group in a more representative OSAS population (12) were in fact close to the 30 minutes reported in the present study.

By analyzing the errors made before the end of the test (i.e., seven consecutive omissions), we found that errors systematically occurred during the test in most OSAS patients (91.3% of patients with OSAS versus 51.7% of control subjects). Overall, control subjects made fewer errors than patients with OSAS. Moreover, the number of errors would have presumably been higher in patients with OSAS if the test had been prolonged beyond the sleep latency cut off (EP 7). In the OSAS group, error time (time spent making errors) represented 5.4% of their sleep latency, whereas it represented only 0.4% in the control group. This observation suggests that patients with OSAS have difficulties not only maintaining wakefulness during a soporific task, but also in maintaining a constant level of vigilance during the test. This could explain the results obtained from driving simulation in patients with OSAS (13). Not only do patients with OSAS get into accidents such as crashes (during sleep episodes) but also show a greater tendency to deviate from the ideal lane position (short fluctua-
tions in vigilance) than control subjects. In our study, type 1–2 errors were the most prevalent, both in control subjects and patients. However, type 3–6 errors accounted for more than 12% of patients total errors, in contrast with the control group (only 4% of their total errors). Seven type errors never occurred in the control group during morning OSLER sessions. The different error profiles are likely to be related to different aspects of vigilance fluctuations.

In a recent study, Priest and colleagues (10) hypothesized that the probability that a subject was asleep would increase as the number of consecutive misses increased. Indeed, these authors were able to show a 92% probability of microsleep onset in subjects who presented three consecutive errors. Microsleep was defined as the occurrence of a period of at least 3 seconds where θ rhythm replaces α rhythm or appears on a background of desynchronized EEG without eye-blinding artifacts. Another intermediate state of vigilance termed ‘attention lapse’ (13, 14) has been defined as the appearance of α rather than θ rhythm for less than 5 seconds on EEG tracings. These different definitions underscore the difficulty in characterizing attention and vigilance fluctuations. Different error subtypes, defined by the number of consecutive misses, and their frequency during the OSLER test are likely to represent a range of situations, from physiologic variations in attention to sleepiness that could potentially lead to increased accident risk. Based on findings in normal subjects (10), it is difficult to attribute the 3–6 EP in the OSAS group to a simple lack of attention. The number of 3–6 errors gives an indication of overall microsleep time before the onset of sleep (seven consecutive errors). Further studies are needed to determine whether an increase in the number of errors with normal sleep latencies is correlated to modifications of outcome measures such as subjective sleepiness, abnormal cognitive functioning, and automobile crashes.

By considering sleep latency as the only measurement of vigilance in the OSLER test, one risks neglecting interesting indicators, such as error analyses, as a useful means of identifying patients with hypersomnolence. For example, some patients with OSAS were able to complete all 40 minutes of testing time although their level of vigilance was not constant throughout the test. Analyzing error profiles may provide useful clinical insight and eventually contribute to a more sensitive diagnosis of daytime sleepiness. In other words, using the sleep onset criterion alone, a number of patients were categorized as “not hypersomnolent” when they should have been diagnosed as “hypersomnolent.” Using a more sensitive assessment of sleepiness, one can expect a better correlation between the number of microarousals during the night and daytime functioning. Moreover, OSAS treatments could be aimed not only at normalizing sleep latencies but also at suppressing fluctuations in vigilance.

Conclusion
The results of this study confirm that the OSLER test is a strong indicator of daytime sleepiness in patients with OSAS. However, the standard way of analyzing the OSLER test (determination of sleep latency) does not identify abnormal fluctuations of vigilance levels during the test. Profile analysis, in conjunction with sleep latency, shows that OSAS patients have difficulty both in maintaining wakefulness and maintaining a constant level of vigilance over short periods. In light of these results, one might argue for a multifactorial approach to vigilance deficits in OSAS. Whether OSAS treatment is accompanied by improvement in sleep latencies and the capacity to maintain a constant level of vigilance remains to be determined.

Acknowledgment: The authors are indebted to C. Grunwald, S. Launois, and J. Plante for their helpful editorial suggestions, and C. Loiodice and M. Selek for their technical assistance.

References