THE CLINICAL ASSESSMENT OF DAYTIME SLEEPINESS IN PATIENTS WITH OBSTRUCTIVE SLEEP APNEA

M.W. Johns

Epworth Sleep Centre, Epworth Hospital, Melbourne, Victoria, Australia

Abstract

All patients coming to a sleep center should have a quantitative assessment of their daytime sleepiness. This has proved to be more difficult than was thought in the past. Daytime sleepiness is defined here as sleep propensity – how likely the subject is, relative to other people, to doze off while engaged in particular activities. The nature of the subject's activity (*e.g.*, whether standing up or lying down) has a profound influence on sleep propensity at a particular time, so sleepiness can only be measured in relation to specific postures and activities. Different categories are needed to describe sleepiness under different circumstances and time frames. These include the subject's instantaneous sleep propensity (ISP) at a particular time, his situational sleep propensity (SSP) when in the same situation repeatedly, and his average sleep propensity (ASP) across a variety of activities in the course of his daily life. Several objective and subjective methods for measuring a subject's ISP, one or more SSPs, or ASP, are described. Although, in the past, many people believed that the Multiple Sleep Latency Test (MSLT) is the gold standard, recent evidence raises serious doubts about its accuracy as a measure of ASP. The Maintenance of Wakefulness Test (MWT) is the preferred objective method, but it is time-consuming and expensive. The Epworth Sleepiness Scale (ESS) is simple and cheap to use. It has been validated, has very good psychometric properties, and has been translated into many languages. However, it is based on subjective reports that have the potential for falsification and bias. It appears that different SSPs in the same subject are not always closely correlated because there are subject x situation-specific interactions, depending on how each subject responds to each situation in which sleepiness is measured. We do not have a gold standard test for ASP. The severity of obstructive sleep apnea (OSA) is only weakly correlated with any measure of sleepiness used so far. Nevertheless, patients with OSA, as a group, have higher than normal ASPs which can be reduced by successful treatment.

Introduction

Excessive daytime sleepiness (EDS) is a common complaint among patients who present to sleep centers, including those with obstructive sleep apnea (OSA).

Address for correspondence: Murray W. Johns, MD, Epworth Sleep Centre, Epworth Hospital, Melbourne, Victoria 3121, Australia

Surgery for Snoring and Obstructive Sleep Apnea Syndrome, pp. 283-295
edited by M. Fabiani

How best to quantify their EDS is still a vexing question.[1,2] So, too, is the relationship between the severity of OSA and EDS.[3] There appear to be several reasons for this. One is the lack of clear definitions of terms such as sleepiness and fatigue, which are often used interchangeably when they ought not to be. Sleepiness is defined here as sleep propensity – the tendency or likelihood, under a given set of circumstances, of making the transition from wakefulness, via the drowsy state, to sleep. From a practical point of view, the relevant circumstances are usually those encountered in daily life when the subject's intention is to remain awake as, for example, when driving a car. The currently accepted conceptual framework within which to consider sleepiness appears confused and is contributing to our problems.[1] There is little recognition that several factors other than a variety of sleep disorders influence a subject's sleepiness, although not all the mechanisms may yet be clear. These factors are in addition to the time of day (process C) and the duration of prior wakefulness (process S) that are quite well understood.[4] Among other variables are the subject's posture and activity, both physical and mental, and the characteristics of the environment at the time.[5] Johns has called these characteristics of the test situation its somnificity, and has produced an ordinal scale of somnificities for several different situations.[6] We cannot assume that sleepiness is like body height, *i.e.*, similar under whatever circumstances it is measured. On the contrary, it appears that we cannot directly quantify sleepiness without reference to the situation in which it is measured. To complicate matters, there also appear to be subject x situation-specific interactions that modify at least some subjects' responses to a particular situation in ways that are not predictable.[5,6] In addition, there are substantial differences in sleepiness, even between normal subjects, which may be due to a psychophysiological trait, partly inherited.[7]

The assumptions underlying tests purporting to measure sleepiness have seldom been questioned or made explicit. The validity, reliability, and accuracy of many of those tests has never been adequately documented. Even the Multiple Sleep Latency Test (MSLT)[8] which, for the past 20 years, many people believed to be the gold standard, did not have its reference range of normal values published until recently (see below). When it did, the MSLT was found to be wanting as a gold standard, based on the inadequacy of its sensitivity and specificity in distinguishing untreated narcoleptic patients, who by definition have EDS, from normal subjects.[9] At different times, we may want to know how sleepy a person is in daily life, whether he is too sleepy to drive a vehicle at a particular time, or by how much his sleepiness has changed after treatment of his sleep disorder, such as with nasal continuous positive airway pressure (CPAP) treatment of OSA. Whatever tests of sleepiness are used under those different circumstances, they should all have a common frame of reference conceptually, and we should know how one set of results relates to another. Currently, it is assumed by most people that, if a subject is relatively sleepy under one set of circumstances, he should also be relatively sleepy under a different set. This is evidently not always so (see below). Several new categories of sleepiness may help to elucidate this problem, as follows:

Instantaneous sleep propensity (ISP): a subject's sleepiness at a particular time, whatever the circumstances. The ISP varies widely with posture and activity, the time of day, the duration of prior wakefulness, etc. It can change rapidly, decreasing over periods of seconds and increasing somewhat more slowly. It increases progressively with the level of drowsiness.
Situational sleep propensity (SSP): a subject's usual sleepiness in the same situation repeatedly, such as sitting and reading. There would be as many SSPs for each subject as there are different situations in which to measure sleepiness.
Average sleep propensity (ASP): a hypothetical construct based on a subject's average level of sleepiness when engaged in a variety of different activities in situations that he usually meets in daily life. The ASP usually remains constant, but it may change, for example, with the start of a sleep disorder, or with its successful treatment.

Within this context, EDS would occur when a subject's ISP exceeded a critical level that made him temporarily unfit to perform a task that most alert subjects could perform reliably. This might be called occasional EDS. It would be caused by circumstances that were unusual for the subject, such as after missing a night's sleep. In contrast, chronic EDS would be present in a subject whose ASP was perpetually higher than normal and who habitually dozed off in situations in which normal subjects usually do not. The relevance of these categories of sleepiness will be illustrated when considering the tests used for measuring sleepiness (see below).

Tests of sleepiness

There are many methods for measuring sleepiness, some of which have been widely used, and others which have been proposed (Table 1). The list of tests in Table 1 is not exhaustive. They can be divided into two broad categories, objective and subjective, with subdivisions in each.

One group of objective tests measures how rapidly the subject falls asleep in the sleep laboratory during the day. These tests are based on what appears to be a reasonable premise, that the quicker we fall asleep, the sleepier we must be.

Multiple Sleep Latency Test

The MSLT measures how long it takes the subject to fall asleep (the sleep latency in minutes) after instructions to try and sleep after lying down for 20 minutes at two-hour intervals during the day, while polysomnographic record-

Table 1. Different kinds of tests of sleepiness

Objective tests of sleepiness		
	Sleep latency	
		Multiple Sleep Latency Test (MSLT)
		Maintenance of Wakefulness Test (MWT)
	Other parameters	
		EEG frequency, power
		pupillometry
		eye movements (saccades, slow eye movements, tracking)
		eyelid movements (blinks, drooping)
		evoked potentials (visual, auditory)
		performance tests (reaction time, divided attention tasks)
Subjective tests of sleepiness		
	Feelings, symptoms	
		Stanford Sleepiness Scale (SSS)
		Karolinska Sleepiness Scale (KSS)
		Visual Analogue Scale (VAS) of alertness/sleepiness
	Dozing behavior	
		Epworth Sleepiness Scale (ESS)
		Sleep-Wake Activity Inventory (SWAI)

ings are made.[8] Within the conceptual framework described above, the sleep latency for each nap is a measure of the subject's ISP at the time. When the circumstances of the test are kept as constant as possible, the ISPs of the same subject measured at different times on the same day are moderately correlated, *e.g.*, $r = 0.61$, $n = 258$, $p < 0.001$.[10] However, Bonnet and Arand[11] have shown that the subject's activity during the few minutes before each nap, such as minor exertion at the bedside, can have a marked effect on the subsequent sleep latency. There is a carry-over effect on the subject's ISP for several minutes from one activity to another. The mean of the sleep latencies measured on the same day in the standardized MSLT gives a measure of the subject's SSP for that test situation, the MSLT-SSP. The test-retest reliability of the MSLT-SSP is quite high over periods of days to weeks: the mean correlation coefficient derived from five published series is 0.74 (range, 0.65-0.97). The figure usually quoted ($r = 0.97$) is an extreme value involving only 14 subjects, which has never been replicated.[12] The overall mean sleep latency in the MSLT for normal subjects is 11.5 ± 5.1 (SD) minutes. However, the reference range is more accurately defined by the 2.5 and 97.5 percentiles, and is 3.2-20.0 minutes.[9] The sensitivity and specificity of the MSLT in distinguishing the sleepiness of narcoleptics, who by definition have chronic EDS, from normal subjects, who do not, is the best way to measure the accuracy of the test. Using a cut-off value of 3 minutes for the MSLT, its sensitivity is only 52%, but its specificity is 98.3%. If a cut-off value of < 5 is accepted, the sensitivity would be 80.9% and the specificity 89.8%. Whatever cut-off values are adopted, the MSLT is less sensitive and specific than either the Maintenance of Wakefulness Test (MWT) or the Epworth Sleepiness Scale (ESS) (see below). These results are not compatible with the 'rule of thumb' which, until recently, was the only method available for interpreting MSLT results.[13] It states that normal subjects have mean sleep latencies of between ten and 20 minutes. Those with a sleep latency of less than five minutes have 'pathological sleepiness', and those with between five and ten minutes are in a 'diagnostic gray area'. This 'rule' should be abandoned as it is very misleading. In the light of such evidence, it is difficult to maintain the MSLT as a gold standard, but this is still a matter of contention.[14]

One problem with the MSLT lies in its basic premise that the quicker we fall asleep when lying down in bed, the sleepier we must be. It appears that this is only partly true. Some normal subjects fall asleep in less than three minutes during the MSLT (in the range said to involve a 'pathological sleepiness') without any evidence of EDS in daily life.[15] Therefore, it has been suggested that the MSLT measures 'sleepability' rather than sleepiness.[16] Another explanation for this characteristic of the MSLT is that it only measures one SSP. For each individual subject, this involves a subject x situation-specific interaction, partly learned, which influences how that subject will respond to the particular test-situation, in this case, lying down during the MSLT. Some normal subjects who do not have EDS learn to fall asleep quickly after lying down in circumstances that may keep other subjects awake. This response is not easily predicted. It is simply an unwarranted assumption that, by extrapolation, the mean sleep latency in the MSLT gives an accurate measure of the subject's more general characteristic, his ASP in daily life. It is this latter that we usually want to measure. Johns[5,6] has presented evidence that different SSPs measured in the same subject, but under different circumstances, at about the same time, are only moderately correlated (Spearman's rho = 0.33-0.57). This is true whether the SSPs are measured subjectively or objectively (see below). This problem potentially exists for any test of sleepiness based on one SSP.

Within the conceptual framework proposed by Johns,[1,6] the MSLT-SSP should be seen as a reasonably reliable, but only moderately accurate, measure of the subject's ASP. The MSLT is more suitable for measuring changes in the sleepiness of the same subjects with time, as with the effects of drugs. It has the great advantage of being an objective test, but it is expensive and time-consuming. Despite endorsement by the relevant US authority,[17] it appears that few sleep centers routinely use the MSLT to quantify their patients' sleepiness. Nonetheless, it remains an important diagnostic aid for narcolepsy by demonstrating the occurrence of REM-sleep during daytime naps.[18]

Maintenance of Wakefulness Test

Another widely used test in this category is the MWT.[19] It is similar to the MSLT except that the subject sits in bed propped up on pillows, and tries to stay awake rather than fall asleep. The MWT measures the same variable as the MSLT, the mean of sleep latencies during several naps, but in a different situation. Consequently, the two tests measure different SSPs and, as expected, their results are significantly, but not highly, correlated, *e.g.*, $r = 0.41$, $n = 258$, $p < 0.001$.[10] The mean of normal MWT-sleep latencies is 18.7 ± 2.6 (SD) minutes.[20] The reference range, defined by the 2.5 and 97.5 percentiles, is 12-20 minutes, and the MWT has a sensitivity of 84.3% and a specificity of 98.3%, with a cut-off value of < 12.[9] These results refer to the 20-minute version of the MWT. There is another version in which the subjects are allowed up to 40 minutes to fall asleep. The majority of subjects take longer to fall asleep during the MWT than during the MSLT, *i.e.*, the somnificity of the MSLT situation is higher than that of the MWT. However, some subjects repeatedly fall asleep more quickly when trying to stay awake than when trying to fall asleep.[10] This is an example of the subject x situation-specific interactions that provide an unpredictable source of variance in measurements of sleepiness in individual subjects (see below). The MWT shares with the MSLT its advantage of objectivity and its disadvantage of high cost. The MWT appears to be more useful than the MSLT for distinguishing changes in sleepiness as a result of therapeutic interventions, be that with nasal CPAP for OSA or stimulant drugs for narcolepsy.[10,21] It is the preferred objective test of sleepiness in the sense of one SSP. We simply do not have a gold standard test for the measurement of ASP.

Other objective tests of sleepiness

Several other objective tests of sleepiness rely on detection of the drowsy state at a particular time. Drowsiness is the fluctuating transitional state between wakefulness and sleep. Some physiological concomitants of the drowsy state, such as slow rolling eye movements recorded from the electro-oculogram,[22] spontaneous fluctuations in size of the pupil,[23] or changes in the frequency and amplitude of the EEG,[24] are objective and specific. If the requirement is to monitor a subject's ISP accurately over limited periods of time, from several hours to a few days (*e.g.*, while driving a truck or a train), then the continuous monitoring of some such physiological variable may be essential. One promising method uses video camera images of the subject's face and eyes (*e.g.*, of a truck or bus driver) to indicate the presence of the drowsy state from the pattern of eyelid movements and eye closures.[25,26] None of these methods has yet been standardized to form the basis of a clinically useful objective test of sleepiness, whether in the sense of continuous measurements of ISP, with indications of when a critical ISP has been reached, or a more limited test of a SSP in the laboratory.

Bennett *et al.*[27] have described a behavioral test of sleepiness, the Osler test, in which subjects indicate whenever they see a small light come on in front of them, as it does automatically for one second every three seconds. The subjects are said to be asleep after they have failed to respond to seven consecutive lights, when the test ends. The advantage of the Osler test over the MWT is said to be the objectivity of its automatic scoring and the need for less equipment. Nevertheless, the Osler test takes all day, as do the MSLT and the MWT. It does not yet have a reference range of normal values, and its role is yet to be determined.

A variety of different driving simulator tests has been devised to address the problem of drowsy driving. All such tests are based on the premise that we can accurately predict a driver's sleepiness while driving, on the basis of measurements of his sleepiness at other times and in different circumstances. This premise is of doubtful validity (see below). Some tests such as Steer Clear are little more than simple reaction time tests.[28] Other tests use driving simulators that incorporate tracking and divided attention tasks that simulate actual driving more exactly. Such a test can readily distinguish the higher levels of sleepiness in groups of patients with OSA compared to normals. However, differences between individual subjects are much less clear-cut.[29] Many patients with OSA and EDS, by other criteria, perform normally in such tests. As yet, there is no standardized test of sleepiness that, by itself, can be used to withhold a driving license. However, if a subject's clinical history and safety record suggested a problem of EDS while driving, the diagnosis of a sleep disorder should be sought by overnight polysomnography, and some laboratory testing of sleepiness would be advantageous.

Subjective tests of sleepiness

Some other tests of sleepiness rely on introspection and self-reports of feelings and symptoms that appear and change with drowsiness. Such subjective reports have formed the basis of several scales for measuring a subject's ISP. They are said to measure 'subjective sleepiness' as opposed to 'objective sleepiness' or sleep propensity. The Stanford Sleepiness Scale (SSS) is one such scale that is commonly used.[30] However, the SSS is not a unitary scale. Factor analysis of its item-scores reveals two factors, one apparently reflecting ISP, the other related more to fatigue, which is a disadvantage.[31] Scores on the SSS are not closely related to objective measurements of ISP made a few minutes later during the MSLT.[32] This should not surprise us because someone's ISP can change in a matter of seconds and because self-reports of feelings may not be very reliable, particularly for comparisons between subjects. The KSS is another method for measuring ISP.[22] So, too, is a visual analogue scale on which the respondent places a mark between two extremes of alertness and drowsiness to represent his ISP at the time. Scores on these tests have not been standardized.

A different kind of subjective test of sleepiness has been developed, the first and most commonly used of which is the ESS.[33] This depends on the subject's retrospective reports of dozing behavior in different situations. The ESS is based on the common experience that, if we doze off while sitting with our head unsupported, the neck muscles that hold our head erect when we are awake relax, and this allows our head to drop forward. This nodding movement often rouses us briefly and makes us aware of having just dozed off and of our eyes having been closed, without necessarily being aware of the preceding drowsy state. The ESS is a simple self-administered questionnaire that asks the subject to rate on a scale of 0 to 3 his usual chances of dozing off in eight different situations that are commonly met in daily life. These situations are graded according to their somnificity. The item-scores give estimates of eight different SSPs. Each SSP is reasonably reliable in a test-retest sense over a period of months, *e.g.*, $r = 0.56$, $n = 87$, $p < 0.001$.[34] The ESS is the sum of the eight item-scores and can vary from zero to 24. It represents the subject's ASP across the eight situations in daily life. It does not to measure 'subjective sleepiness' because it does not assess feelings. That the ESS refers to observable behavior rather than subjective feelings is supported by the high correlation between the patients' and their partners' independent reports of the patients' dozing behavior (*e.g.*, rho = 0.74, $n = 50$, $p < 0.001$).[5] Rather, it measures the subject's sleep propensity subjectively in relation to a variety of particular situations. It is the only method that recognizes and addresses the requirement for measuring ASP. It is not a subjective equivalent of the MSLT or the MWT, each of which only measures one particular SSP.

Normal subjects in Australia have a mean ESS score of 4.6 ± 2.8 (SD),[35] and in the UK, 4.5 ± 3.3.[36] Based on the Australian sample ($n = 72$) and the 2.5 and 97.5 percentiles, the reference range of normal values is 0-10. The ESS has a high sensitivity (93.5%) and specificity (100%) in distinguishing narcoleptics from normal subjects. This has been independently confirmed in the UK, where the sensitivity was 97% and the specificity 100%.[36] Despite its reliance on retrospective subjective reports, the ESS has a high test-retest reliability, as high as the MSLT ($r = 0.81$ versus 0.74), and has very good evidence for its validity, accuracy, internal consistency, and unitary structure.[9,34] The ESS has been translated into many languages besides English, some translations being standardized, others not. To answer the ESS, the subject must have had recent experience of most of or all the situations described in its items. This has proved to be impossible in the case of some patients suffering from severe medical illnesses.[37] ESS scores do not vary consistently with age or gender.[35] However, evidence from Brazil suggests that there may be some ethnic or cultural differences.[38] The ESS costs very little and is easy to administer compared to the MSLT or MWT. Nevertheless, the fact that the ESS is based on subjective reports, with at least the potential for falsification and error, means that it cannot be a gold standard. For that we need an objective test to make the same measurements as the ESS makes subjectively. There is no such test at present. Any test of sleepiness should be used with a clear understanding of its limitations.

Relationships between the results of different tests of sleepiness

The relationship between the results of the MSLT and the ESS in the same subjects is a matter of contention, particularly for those researchers who have reported the lack of a significant relationship.[14] However, of all the 12 published series with correlations available at the time of writing, ten reported significant correlations ($p < 0.05$-0.001) with a mean $r = -0.36$ (range, -0.23 to -0.61). In five of these ten series, the number of subjects exceeded 100 and the largest series involved 522 subjects ($r = -0.29$, $p < 0.001$).[39] Thus, it cannot be said that we have insufficient data. It is clear that the relationship between the MSLT and the ESS is usually statistically significant, but it is not a close one. Nor is the relationship between the MWT and the ESS much closer; of three correlations published so far, the mean r was -0.42 (range, -0.29 to -0.48), all being statistically significant (*e.g.*, Sangal *et al.*[10]). This is consistent with other evidence about the limited relationships between different SSPs, based on item-scores in the ESS.[5,6] In 987 Australian subjects (patients with sleep disorders as well as normal subjects), the mean of 28 Spearman correlation coefficients between their eight ESS item-scores was rho = 0.49 (range, 0.33-0.57). All these were statistically significant ($p < 0.0001$), but none was a very close relationship.[6] Even when different SSPs are measured objectively by the MSLT and the MWT, they are not more closely correlated than the SSPs measured by ESS item-scores. The results of recent experiments by Bonnet and Arand[11] are very relevant to this. They measured SSPs objectively in five different test situations, including the MSLT. The mean of four Pearson correlation coefficients between the results of the MSLT and the other tests was 0.53. With $n = 14$, only two of those correlations were statistically significant.

The disparity between the results of the MSLT and the MWT has led some researchers to believe that the MSLT measures the 'ability to fall asleep', whereas the MWT measures the 'ability to stay awake'.[10] While at face value this is so, it implies that the two 'abilities' represent two different generalized characteristics of each subject. Johns[5,6] argues differently, claiming that the MSLT and the MWT measure two different SSPs that, in part, relate specifically to each test situation, and are not generalized 'abilities'. If they were, there would be as many generalized 'abilities' as there are different situations in which to measure SSPs. He explains the relationship between any two SSPs in the same subject in terms of three different factors or sources of variance. The first relates to the subject's general level of sleepiness, his ASP. The second relates to differences in the somnificity of different postures, activities, and situations that are highly predictable for groups of subjects, less so for individuals. The third relates to subject x situation-specific interactions due to differences in the usual reaction of each subject to each test situation. It is such interactions that prevent patients suffering from psychophysiological insomnia from falling asleep in bed at night when, in a different cognitive setting, they can readily fall asleep in a chair, watching TV in the evening. Similarly, such patients can sleep better than average on the first night in the sleep laboratory, when many others sleep worse than at home because of the first night effect. In 'normal' circumstances, most people fall asleep more quickly after lying in bed and preparing for sleep at night than they do when sitting and watching TV, *i.e.*, the somnificity of the former situation is usually higher than the latter. However, the 'unusual' responses of some subjects can increase, decrease, or reverse the effects of somnificity in some situations, but not in others.

The evidence to date suggests a tentative conclusion: we cannot rely on one particular SSP as an accurate predictor of a different SSP in the same subject, whether the SSPs are measured objectively by the MSLT or the MWT, subjectively by the ESS, or by any other means. Two different SSPs in the same subjects will often be moderately correlated, but sometimes not. The same will be true for predictions of a subject's ISP at a particular time. If this conclusion is accepted, it has important ramifications; for example, with the assessment of a driver's ISP at a particular time, based on a measurement of his ISP or an SSP when he is not driving. Much more research is needed on this topic. However, a corollary of this conclusion is that a measurement of a subject's general level of sleepiness in his daily life (his ASP) is likely to be more accurate if it is based on a variety of different SSPs rather than on one. This could explain the documented accuracy of ESS as a measure of ASP, despite its perceived inaccuracy because it is based on subjective reports.

Sleepiness in patients with obstructive sleep apnea/snoring

The assessment of sleepiness in individual patients is complicated. So, too, is the assessment of the severity of OSA. However, one thing is clear – groups of patients with OSA have higher levels of sleepiness than do groups of normal subjects.[21,27,40] It has generally been assumed that the repeated fragmentation of sleep by respiratory arousals, combined with repeated episodes of arterial oxygen desaturation, causes the EDS. Certainly, when the OSA is successfully treated by nasal CPAP, the levels of sleepiness in groups of patients are invariably reduced when assessed by the ESS,[41,42] less reliably so by the MWT, and least reliably by the MSLT.[10] However, the levels of sleepiness in treated patients often remain in the upper half of the normal range for reasons that are not clear. The frequency of apneas and hypopneas per hour of sleep (the apnea-hypopnea index, or AHI, otherwise known as the respiratory disturbance index, or RDI) is commonly used as the main measure of the severity of OSA. Another measure is the lowest level of arterial oxygen saturation reached during apneas and hypopneas overnight (minimum SaO_2). There have been many attempts to correlate the RDI and minimum SaO_2 with patients' sleepiness, whether assessed by the MSLT, MWT, or ESS. None of the correlations has been high. Some have been statistically significant,[21,40] but many have not.[43] For some time, it was thought that a much better index of the severity of OSA would be the respiratory arousal index, separating those apneas and hypopneas that caused arousal. This has not been found to be the case, even when subcortical arousals and a sensitive index of respiratory effort were included.[3] The mechanisms by which EDS is caused in OSA remain uncertain.

Some patients who snore persistently without having OSA (*i.e.*, RDI < 5/hour) have levels of sleepiness significantly higher than normal subjects, but lower than many patients with OSA.[35,40] The mechanism for this is also uncertain. There is no measure of EDS that can be used alone to quantify the severity of OSA in individual patients. Sleepiness should be used as one measure among many others, such as body mass index, that describe a patient's clinical condition. There are occasional patients with severe OSA (RDI > 50/hour) with frequent episodes of arterial oxygen desaturation who do not have EDS, clinically or by ESS scores, etc. There is no evidence that this arises simply from the inaccuracy of our measurements of sleepiness. There may be several reasons why EDS and OSA are not closely related in individual subjects, as others have described.[3,43] One reason is that much of the variance in measurements of sleepiness between subjects is due to factors other than sleep disorders, as is now recognized. Another reason may be that there are sleep disorders other than OSA, such as restless legs syndrome and periodic limb movement disorder, which are common disorders that often occur with OSA and contribute to EDS, but which are under-diagnosed.

References

1. Johns MW: Rethinking the assessment of sleepiness. Sleep Med Rev 2:3-15, 1998
2. Krieger J: Clinical approach to excessive daytime sleepiness. Sleep 23(Suppl 4):S95-S98, 2000
3. Stradling JR, Barbour C, Glennon J, Langford BA, Crosby JH: Prevalence of sleepiness and its relation to autonomic evidence of arousals and increased inspiratory effort in a community based population of men and women. J Sleep Res 9:381-388, 2000
4. Borbély AA, Achermann P, Trachsel L, Tobler I et al: Sleep initiation and sleep intensity: interaction of homeostatic and circadian mechanisms. J Biol Rhythms 4:149-160, 1989
5. Johns MW: Sleepiness in different situations measured by the Epworth Sleepiness Scale. Sleep 17:703-710, 1994
6. Johns MW: Sleep propensity varies with behaviour and the situation in which it is measured: the concept of somnificity. J Sleep Res 11:61-67, 2002
7. Carmelli D, Bliwise DL, Swan GE, Reed T: A genetic study of the Epworth Sleepiness Scale in 1608 male-male twin pairs from the NAS-NRC World War II twin registry. Sleep (Abstract Suppl)2:A92, 2000
8. Carskadon MA, Dement WC, Mitler MM, Roth T, Westbrook RR, Keenan S: Guidelines for the Multiple Sleep Latency Test (MSLT): a standard measure of sleepiness. Sleep 9:519-524, 1986
9. Johns MW. Sensitivity and specificity of the multiple sleep latency test (MSLT), the maintenance of wakefulness test and the Epworth sleepiness scale: failure of the MSLT as a gold standard. J Sleep Res 9:5-11, 2000
10. Sangal RB, Thomas L, Mitler MM: Maintenance of wakefulness test and Multiple Sleep

Latency Test: measurement of different abilities in patients with sleep disorders. Chest 101:898-902, 1992

11. Bonnet MH, Arand DL: Arousal components which differentiate the MWT from the MSLT. Sleep 24:441-447, 2001
12. Zwyghuizen-Doorenbos A, Roehrs T, Schaeffer M, Roth T: Test-retest reliability of the MSLT. Sleep 11:562-565, 1988
13. Richardson GS, Carskadon MA, Flagg W, Van den Hoad J, Dement WC, Mitler MM: Excessive sleepiness in man: sleep latency measurement in narcoleptic and control subjects. Electroencephalogr Clin Neurophysiol 45:621-627, 1978
14. Chervin RD, Aldrich MS: The Epworth Sleepiness Scale may not reflect objective measures of sleepiness or sleep apnea. Neurology 52:125-131, 1999
15. Johnson LC, Spinweber CL, Gomez SA, Matteson LT: Daytime sleepiness, performance, mood, nocturnal sleep: the effect of benzodiazepine and caffeine on their relationship. Sleep 13:121-135, 1990
16. Harrison Y, Horne JA: 'High sleepability without sleepiness': the ability to fall asleep rapidly without other signs of sleepiness. Neurophysiol Clin 26:15-20, 1996
17. Thorpy MT: The clinical use of the Multiple Sleep Latency Test. Sleep 15:268-276, 1992
18. Moscovitch A, Partinen M, Guilleminault C: The positive diagnosis of narcolepsy and narcolepsy's borderland. Neurology 43:55-60, 1993
19. Mitler MM, Gujavarty KS, Browman CP: Maintenance of wakefulness test: a polysomnographic technique for evaluating treatment in patients with excessive somnolence. Electroencephalogr Clin Neurophysiol 53:658-661, 1982
20. Doghramji K, Mitler MM, Sangal RB, Shapiro C, Taylor S, Walsleben J et al: A normative study of the maintenance of wakefulness test (MWT). Electroencephogr Clin Neurophysiol 103:554-562, 1996
21. Poceta JS, Ho S-L, Jeong D-U, Mitler MM: The maintenance of wakefulness test in obstructive sleep apnea. Sleep Res 19:268, 1990
22. Åkerstedt T, Gillberg M: Subjective and objective sleepiness in the active individual. Int J Neurosci 52:29-37, 1990
23. Wilhelm B, Wilhelm H, Lüdkte H, Streicher P, Adler M: Pupillographic assessment of sleepiness in sleep-deprived healthy subjects. Sleep 258-265, 1998
24. Makeig S, Inlow M: Lapses in alertness: coherence of fluctuation in performance and EEG spectrum. Electroencephalogr Clin Neurophysiol 86:22-35, 1993
25. Mallis MM, Maislin G, Powell JW, Konowal NM, Dinges DF: Perclos predicts both PVT lapse frequency and cumulative lapse duration. Sleep 22:S148-S149, 1999
26. Häkkäanan H, Summala H, Partinen M, Tiihonen M, Silvo J: Blink duration as an indication of driver sleepiness in professional bus drivers. Sleep 22:798-802, 1999
27. Bennett LS, Stradling JR, Davies JO: A behavioural test to assess daytime sleepiness in obstructive sleep apnea. J Sleep Res 6:142-145, 1997
28. Findley L, Unverzagt M, Guchu R, Fabrizio M, Buckner J, Suratt P: Vigilance and automobile accidents in patients with sleep apnea or narcolepsy. Chest 108:619-624, 1995
29. George CFP: Vigilance impairment: assessment of driving simulators. Sleep 23(Suppl 4):S115-S118, 2000
30. Hoddes E, Zarcone V, Smythe H, Phillips R, Dement W et al: Quantification of sleepiness: a new approach. Psychophysiology 10:431-436, 1973
31. Maclean AW, Fekkon GC, Saskin P, Knowles JB: Psychometric evaluation of the Stanford sleepiness scale. J Sleep Res 1:35-39, 1992
32. Pressman MR, Fry JM: Relationship of autonomic nervous system activity to daytime sleepiness and prior sleep. Sleep 12:239-245, 1989
33. Johns MW: A new method for measuring daytime sleepiness: the Epworth Sleepiness Scale. Sleep 14:540-545, 1991
34. Johns MW: Reliability and factor analysis of the Epworth Sleepiness Scale. Sleep 15:376-381, 1992

35. Johns MW, Hocking B: Daytime sleepiness and sleep habits of Australian workers. Sleep 20:844-849, 1997
36. Parkes JD, Chen SY, Clift SJ, Dahlitz MT, Dunn G: The clinical diagnosis of the narcoleptic syndrome. J Sleep Res 7:41-52, 1998
37. Sancisi E, Mostacci B, Mondini S, Buzzi G, Cirignotta F: Evaluation of sleepiness in patients with chronic medical disease: unreliability of the Epworth sleepiness scale. Sleep (Abstract Suppl)24:A388-A389, 2001
38. Pedroso A, Nosek A, Armaganijan L, Levites M, Bismark M, Aloe F et al: Epworth Sleepiness Scale outcome in 2893 Brazilian students. Sleep (Suppl):275, 1998
39. Sangal RB, Mitler MM, Sangal JAM: US modafinil in narcolepsy multicenter study group: MSLT, MWT and ESS indices of sleepiness in 522 drug-free patients with narcolepsy. Sleep Res 26:492, 1997
40. Johns MW: Daytime sleepiness, snoring and obstructive sleep apnea: the Epworth Sleepiness Scale. Chest 103:30-36, 1993
41. Hardinge FM, Pitson DJ, Stradling JR: Use of the Epworth sleepiness scale to demonstrate response to treatment and nasal continuous positive airway pressure in patients with obstructive sleep apnea. Respir Med 89:617-620, 1995
42. Engleman HM, Kingshott RN, Wraith PK, Mackay TW, Deary IJ, Douglas NJ: Randomized placebo-controlled crossover trial of continuous positive airway pressure for mild sleep apnea/hypopnea syndrome. Am J Respir Crit Care Med 159:461-467, 1999
43. Sauter C, Asenbaum S, Popovic R, Bauer H, Lamm C, Klösch G, Zeitlhofer J: Excessive daytime sleepiness in patients suffering from different levels of obstructive sleep apnea syndrome. J Sleep Res 9:293-301, 2000