J. Sleep Res. (2000) 9, 5–11

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

MURRAY W. JOHNS
Epworth Sleep Centre, Epworth Hospital, Australia

Accepted in revised form 25 August 1999; received 23 April 1999

SUMMARY Excessive daytime sleepiness (EDS) is an important symptom that needs to be quantified, but there is confusion over the best way to do this. Three of the most commonly used tests: the multiple sleep latency test (MSLT), the maintenance of wakefulness test (MWT) and the Epworth sleepiness scale (ESS) give results that are significantly correlated in a statistical sense, but are not closely related. The purpose of this investigation was to help clarify this problem. Previously published data from several investigations were used to calculate the reference range of normal values for each test, defined by the mean $\pm$2 SD or by the 2.5 and 97.5 percentiles. The 'rule of thumb' that many people rely on to interpret MSLT results is shown here to be misleading. Previously published results from each test were also available for narcoleptic patients who were drug-free at the time and who by definition had EDS. This enabled the sensitivity and specificity of the three tests to be compared for the first time, in their ability to distinguish the EDS of narcolepsy from the daytime sleepiness of normal subjects. The receiver operator characteristic curves clearly showed that the ESS is the most discriminating test, the MWT is next best and the MSLT the least discriminating test of daytime sleepiness. The MSLT can no longer be considered the gold standard for such tests.

KEYWORDS daytime sleepiness, sleep propensity

INTRODUCTION

Excessive daytime sleepiness (EDS) is a symptom of increasing interest and importance to sleep physicians and researchers, to those involved with road safety and public health, and to the general public (Dement and Mitler 1993; Lyznicki *et al.* 1998). Unfortunately, there is confusion about what EDS is and how it should be measured (Johns 1998). Part of this confusion arises because the word sleepiness is used by different people to mean different things: from a set of subjective feelings that accompany the drowsy state, to physiological changes (such as in the pupil) that occur during the sleep-onset process, to the propensity to fall asleep under a given set of circumstances (Johns 1998). We are concerned here with normal and abnormal daytime sleepiness, as it refers to the subject's average sleep propensity across a variety of situations in daily life.

Three of the most widely used tests of daytime sleepiness, the MSLT, the MWT, and the ESS give results that are significantly correlated in a statistical sense, but are not closely related (U. S. Modafinil in Narcolepsy Multicenter Study Group 1998). The purpose of this investigation is to help clarify this problem. Here, for the first time, these three tests are compared in terms of their sensitivity and specificity, and hence their ability to distinguish the daytime sleepiness of narcoleptic patients, who by definition have EDS, from normal subjects. This report is concerned with the MSLT only in so far as it measures daytime sleepiness. Other concerns about the sensitivity and specificity of the early onset of REM sleep in the MSLT for diagnosing narcolepsy have been investigated by others (Moscovitch *et al.* 1993; Guilleminault *et al.* 1994).

Correspondence: Murray W. Johns, Epworth Sleep Centre, Epworth Hospital, 187 Hoddle Street, Richmond, Victoria, 3121, Australia. Tel.: [613] 9427 1849; fax: [613] 9427 8522.

METHODS

The data that form the basis of the present investigation have been published previously, particularly as results from a large study of the effects of modafinil in narcoleptics, in which the MSLT, MWT and ESS were used in parallel to quantify their daytime sleepiness (U. S. Modafinil in Narcolepsy Milticenter Study Group 1998; Mitler *et al.* 1998; Sangal *et al.* 1997; Mitler 1997).

This report is in four parts: part A outlines the nature of the three tests of daytime sleepiness and the relationships between them; part B deals with the distribution of results for normal subjects for each test; part C describes the equivalent distributions of results for narcoleptics who were drug-free at the time and part D is concerned with the sensitivity, specificity, and the receiver operator characteristic curves for the three tests, thereby enabling comparisons to be made between them in terms of their ability to distinguish normal from abnormal daytime sleepiness.

RESULTS

The nature of the MSLT

The MSLT measures how quickly the subject falls asleep, when asked to do so, when lying down in a quiet, darkened bedroom while under scrutiny and with electrodes and wires attached so that sleep can be monitored accurately (Richardson *et al.* 1978). The MSLT was first used to measure the sleep latency (SL) repeatedly in normal subjects, but was soon adopted as a means of distinguishing normal from abnormal daytime sleepiness in different subjects, including patients with sleep disorders. What is measured is the mean SL in minutes over 4 (sometimes 5 or 6) naps every 2 h during the day. Many believe the MSLT to be the gold standard for measuring daytime sleepiness, a position backed officially by the American Sleep Disorders Association (Carskadon *et al.* 1986). Certainly the mean MSLT–SL is measured objectively and reasonably accurately; it varies with the time of day, in response to sleep deprivation, and usually with the effects of sedative or stimulant drugs (Thorpy 1992). The mean MSLT–SL's of patients with narcolepsy differ significantly from those of normal subjects (Richardson *et al.* 1978). The mean MSLT–SL is reliable in the test–retest sense ($r=0.65$–0.97) (Zwyghuizen-Doorenbos *et al.* 1988; Van den Hoed *et al.* 1981). It appears therefore that there is substantial evidence to support the contention that the MSLT is valid, reliable and objective – a suitable gold standard.

However, there have been dissenting opinions for several years (Jahnke and Aldrich 1990; Johnson *et al.* 1990; Lammers and van Dijk 1992; Chervin *et al.* 1995; Steinberg *et al.* 1996; Geisler *et al.* 1998). Even after widespread use for 20 years, the sensitivity and specificity of the mean MSLT–SL as a measure of EDS have not been published. There is confusion about the cut-off point that best distinguishes normal from abnormal daytime sleepiness. Most people have relied, uncritically, on a 'rule of thumb' (Carskadon *et al.* 1986). This rule states that a mean MSLT–SL < 5 min represents 'pathological sleepiness', 10–20 min is normal, and 5–10 min is in a 'diagnostic grey area' (Carskadon *et al.* 1986). Others have adapted this rule, more or less to suit their own requirements, so that the lower cut-off point is $\leq 5.5, <6, <8$ or ≤ 8 min (Guilleminault *et al.* 1994; Van den Hoed *et al.* 1981). For reasons that are not clear, the usual methods for determining a reference range of normal values for such a test (e.g. the mean ± 2 SD) (Nicoll *et al.* 1999) have been avoided with the MSLT (Carskadon *et al.* 1986; Thorpy 1992; Roehrs and Roth 1992). In fact, the 'rule of thumb' was based on little more than a clinical impression, as was stated at the time (Richardson *et al.* 1978).

There is another problem with the MSLT in relation to the sleepiness that it measures. It is assumed that MSLT-sleepiness accurately reflects the more general characteristic of sleepiness in other situations in daily life, which is what we want to know about in the present context. There is evidence that this assumption is not generally valid (Johns 1994). Johns (1998) has pointed out that measurements of sleepiness in one test situation, no matter how accurately and objectively they are made, cannot be relied upon as accurate predictors of sleepiness in a different situation. This problem is highlighted by comparison of MSLT–SL's with the results of a different, but equally objective test, the MWT (see below).

The nature of the MWT

In this test the subject sits in bed, resting against pillows, in a quiet dimly lit room, attempting to stay awake for 20 (or sometimes 40) min, with the same sense of scrutiny and with the same electrodes attached as in the MSLT (Mitler *et al.* 1982). The mean MWT–SL in minutes, determined from 4 or more naps, is the measure of sleepiness. The main differences between the MSLT and MWT test–situations are the subject's posture (lying vs. sitting in bed), the degree of support for the head, having eyes open vs. closed, the room being dimly lit rather than dark, and the intention to stay awake vs. to sleep.

The mean MSLT–SL's are significantly correlated with MWT–SL's in the same subjects [$r=0.41$, $n=258$, $P<0.001$ (Sangal *et al.* 1992); $r=0.52$, $n=522$, $P<0.001$ (Sangal *et al.* 1997)], but only about 20–25% of variance in one test is shared by the other. This has led some researchers to conclude that the two tests measure different abilities or kinds of sleepiness (Sangal *et al.* 1992). However, Johns (1998) has argued that there are as many different kinds of sleepiness, which he describes as situational sleep propensities, as there are different situations in which to measure it. All measurements of sleep propensity are always partly situation-specific as well as subject-specific. Within a new conceptual framework proposed by Johns (1998), the subject's posture and activity, both physical and mental, become one of several major determinants of sleep propensity in any situation and at any particular time. This makes intuitive sense when we realise that we are much more likely to fall asleep when lying down than when standing up, even when all environmental stimuli and psychological factors are kept constant. It is postulated that standing up decreases

our sleep propensity because feedback from the activity of postural muscles, particularly of the neck, back and legs, enhances a putative 'wake drive' rather than directly reducing a putative 'sleep drive'. Sleep propensity at any particular time depends on the relative strengths, not the absolute strengths, of the wake drive' and the sleep drive' that interact continuously by mutually inhibiting each other. It was within this new conceptual framework that a very different test of sleepiness was developed – the ESS (Johns 1991).

Nature of the ESS

The ESS is a self-administered questionnaire with eight items, described in detail elsewhere (Johns 1991). It asks the subject to rate on a four point scale his/her chances of dozing in each of eight different situations that are often encountered in daily life. The situations were chosen on a priori grounds to differ in their soporific nature – their ability to induce dozing. That they do differ in this regard has been confirmed by very similar ranking of the eight item-scores from different groups of patients and normal subjects (Johns 1994). Each of the eight item-scores of the ESS provides a measure of one situational sleep propensity; the habitual tendency to doze or stay awake in that situation as part of daily life. These item-scores are about as reliable in the test–retest sense as are MSLT–SL's or MWT–SL's in consecutive naps on the same day (Johns 1994). The ESS score is the sum of the eight item-scores and can vary from 0 to 24. It provides a measurement of the subject's average sleep propensity in daily life with a reliability similar to that of the mean MSLT–SL ($r=0.81$, $n=87$, $P<0.001$) (Johns 1994).

Because it relies on subjective reports, some people equate the sleepiness measured by the ESS with 'subjective sleepiness', as measured by the Stanford sleepiness scale, the Karolinska sleepiness scale, or a visual analogue scale of alertness (Johns 1998). The latter are based on the presence or intensity of a set of feelings that signal the onset of sleep at the particular time. The ESS, by contrast, relies on retrospective reports of dozing behaviour, such as a nodding head and lapses of attention, that we are better able to recognise immediately after the event, when we rouse, than before or during it. Thus, the ESS relies on subjective reports that equate with sleep propensity, not with 'subjective sleepiness'. Having a high sleep propensity (EDS) means having a history of dozing in situations that have a relatively low soporific nature, in which normal subjects seldom doze.

The ESS has a major advantage over the MSLT or MWT in its very low cost and ease of administration. An MSLT costs about 1000 times more than an ESS. It can be argued also that the average sleep propensity measured by the ESS is a more appropriately generalised test than one based on only one particular situational sleep propensity measured either by the MSLT or MWT. The main disadvantage of the ESS is its reliance on subjective reports rather than objective measurements. There are many published reports about the correlation between ESS scores and mean MSLT–SL's, and in none of them is it a high correlation (e.g. $r=-0.27$, $n=522$, $P<0.001$) (Mitler *et al.* 1998). The mean of nine such correlation coefficients published so far is –0.3. Most, but not all, were statistically significant. In a more detailed analysis of this relationship, Chua *et al.* (1998) have shown a higher correlation ($r=-0.57$, $P<0.001$) between ESS scores and mean MSLT–SL's for tests in which the subject fell asleep at each opportunity than for tests in which the subject sometimes did not fall asleep ($r=0.009$, $P>0.1$). There have been fewer reports of the correlation between ESS scores and mean MWT–SL's, but they are of similar magnitude to those with the MSLT [e.g. $r=-0.48$, $n=41$, $P<0.001$ (Sangal *et al.* 1997); $r=-0.29$, $n=522$, $P<0.001$ (Sangal *et al.* 1997)].

Reference range for the MSLT in normal subjects

As noted above, a reference range for normal MSLT–SL's, calculated in any of the usual ways, has never been used. However, we can estimate it from published data. For this we need a histogram, or its equivalent, of normal MSLT–SL's. The largest such series yet published is for two groups of male and female normal sleepers, assessed by history and overnight polysomnography; 129 aged 18–29 years, and 47 aged 30–80 years (Levine *et al.* 1988). The mean of their mean MSLT–SL's was 11.1 ± 5.2 SD for the first group and 12.5 ± 4.7 SD for the second. The SD's were not quoted in the original report, but have been calculated from the published data by the present author. It was reported by the original authors that these two series differed significantly (using a 2-factor ANOVA, $P<0.001$), but when recalculated by the present author they did not differ significantly either by *t*-test ($P=0.12$) or, more appropriately, by the Mann–Whitney *U*-test ($P=0.098$) because the distribution for 129 subjects differed significantly from normal (χ^2, $P=0.006$). Consequently, the two groups have been combined here to form a larger one ($n=176$) with a broader age range (18–80 years) for which the overall mean SL was 11.5 ± 5.1 SD (95% CI = 10.7–12.2). There are several other reports of mean MSLT–SL's from smaller groups of normal subjects that give comparable mean's and SD's, which lends credence to these figures [e.g. 11.1 ± 4.6, $n=31$ (Steinberg *et al.* 1996); 11.4 ± 6.1 $n=35$ (Bliwise *et al.* 1991)].

If the reference range for MSLT–SL's is defined simply as the mean ±2 SD it would be 1.3–20 min for the combined series of 176 subjects. This should theoretically include 95% of normal MSLT–SL's, with another 2.5% expected to be < 1.3 min and 2.5% = 20 min. In fact, the whole range of values was 2–20 min. Because MSLT–SL's are seldom normally distributed, a more accurate definition of the reference range would be the 2.5 and 97.5 percentiles, in this case 3.2 and 20 min. This reference range is entirely compatible with several other published reports that some normal subjects, of all ages, fall asleep in < 5 min in the MSLT without any evidence of EDS (Johnson *et al.* 1990; Levine *et al.* 1988). This is not compatible with the 'rule of thumb'. There have been attempts to explain this discrepancy by saying that such subjects are not normal, and that they suffer from chronic sleep deprivation (Roehrs *et*

al. 1990). This has not been substantiated by later investigations (Roehrs *et al.* 1996).

Reference range for the MWT in normal subjects

Mitler *et al.* (1998) have published the distribution of mean MWT–SL's for 64 normal subjects, men and women aged 30–70 years, who had polysomnography followed by the MWT next day. There are variations of the MWT, but that of special interest here involves a 20-min protocol, defining sleep onset as the start of the first three consecutive 30-sec epochs of stage 1 sleep or the first epoch of any other stage of sleep (Doghramji *et al.* 1997). The mean of normal MWT–SL's was 18.7+2.6 SD, so the reference range defined by the mean±2 SD would be 13.5–20 min (Doghramji *et al.* 1997). The distribution of these SL's was skewed to the right, so the reference range, defined by the 2.5 and 97.5 percentiles, would be approximately 12–20 min.

Reference range for ESS scores in normal subjects

Johns and Hocking (1997) have published the distribution of ESS scores for 72 normal subjects. Their mean score was 4.6±2.8 SD, which gives a reference range of 0–10. This also coincides with the 2.5 and 97.5 percentiles. These 72 subjects were chosen from a group of 331 ostensibly healthy male and female workers, aged 22–59 years, in Australia. They did not have polysomnography, but were selected by strict criteria derived from a detailed sleep questionnaire that screened out those suffering from most sleep disorders including insomnia and snoring or sleep-disordered breathing (Johns and Hocking 1997). A larger sample of 188 normal subjects from the UK had a mean ESS score of 4.5±3.3, which suggests that the reference range may be 0–11 (Parkes *et al.* 1998). However, their three criteria for normality may not have screened adequately for snoring and sleep-disordered breathing. As with any reference range, that for the ESS may change slightly when more data are collected.

MSLT, MWT and ESS in narcoleptics

The US modafinil in narcolepsy multicenter study group have published the distributions of mean MSLT–SL's, mean MWT–SL's and ESS scores in a large group of drug-free narcoleptics before they took modafinil and while taking it (Mitler *et al.* 1998; Sangal *et al.* 1997; Mitler 1997). The subjects were 530 men and women aged 17–68 years who had previously been diagnosed as having narcolepsy according to the International Classification of Sleep Disorders, with some modifications. They must all have had the complaint of EDS and either cataplexy with a mean MSLT–SL ≤8 min and >1 min sleep onset REM episode, or a mean MSLT–SL ≤5 min and >1 min sleep onset REM episode without cataplexy (Sangal *et al.* 1997). In fact, there were several narcoleptics' with mean MSLT–SL's >10 min, so the authors did not adhere strictly to their own diagnostic criteria. The MSLT results formed an important part of those criteria, which makes it difficult then to test the MSLT's ability to distinguish normal from abnormal sleepiness without using a circular argument. However, let us assume initially that all diagnoses of narcolepsy were accurate and that all subjects had EDS.

There is no question that the mean MSLT–SL's, the mean MWT–SL's and ESS scores of these narcoleptics differed significantly from normals. The mean MSLT–SL in the modafinil study before treatment was 2.8±3.8 SD min, the mean MWT–SL was 6±4.8 min, and the mean ESS score was 17.7±3.8 ($n=522$). Similar results have been reported from other investigations of narcoleptics (Parkes *et al.* 1998; Aldrich *et al.* 1997). However, such figures do not enable us to compare the discriminating power of the three tests of daytime sleepiness. For that we must calculate the sensitivity and specificity of each test at different cut-off values, and then construct the ROC curves (Nicoll *et al.* 1999). The accuracy of such calculations depends not only on the assumption that all the narcoleptics had EDS, but also that none of the normal subjects had EDS.

Sensitivity, specificity and ROC curves

The frequency histograms used to derive the mean MSLT–SL's, mean MWT–SL's and ESS scores for narcoleptics were those described by Mitler *et al.* (1998), Sangal *et al.* (1997) and Harsh *et al.* (1998). The mean MSLT–SL had a sensitivity of only 80.9% and a specificity of 89.8% when the cut-off of <5 min was used. The sensitivity increased to a more acceptable 94.5% when the cut-off was <8 min, but the specificity fell then to 73.3%. The sensitivity fell to only 52% when the cut-off was <3 min, but the specificity was 98.3%.

The MWT–SL had a sensitivity of 84.3% when the cut-off was at <12 min, at which point the specificity was 98.4%.

The ESS had a high sensitivity (93.5%) and high specificity (100%) with a cut-off score >10.

The ROC curves (Fig. 1) clearly demonstrate that the MWT–SL is more effective than the MSLT–SL in distinguishing EDS from normal daytime sleepiness, but that the ESS is better than either the MSLT or MWT. Thus, the MSLT fails as a gold standard.

DISCUSSION

The results throw some much needed light on the problem of measuring daytime sleepiness. Far from being the gold standard, the MSLT is the least accurate of the three tests used here. The 'rule of thumb' should be abandoned as misleading. The reference range for the mean MSLT–SL appears from the present results to be 3.2–20 min. It is certainly not 10–20 min as suggested by the 'rule of thumb'. That the ESS is a more discriminating test of sleepiness in daily life than either the MSLT or the MWT is supported by the observation of the US modafinil study group that the effects of modafinil in reducing the EDS of narcoleptics were more clearly shown by the ESS than by the MWT, and least by the MSLT (U. S. Modafinil in Narcolepsy Multicenter Study Group 1998).

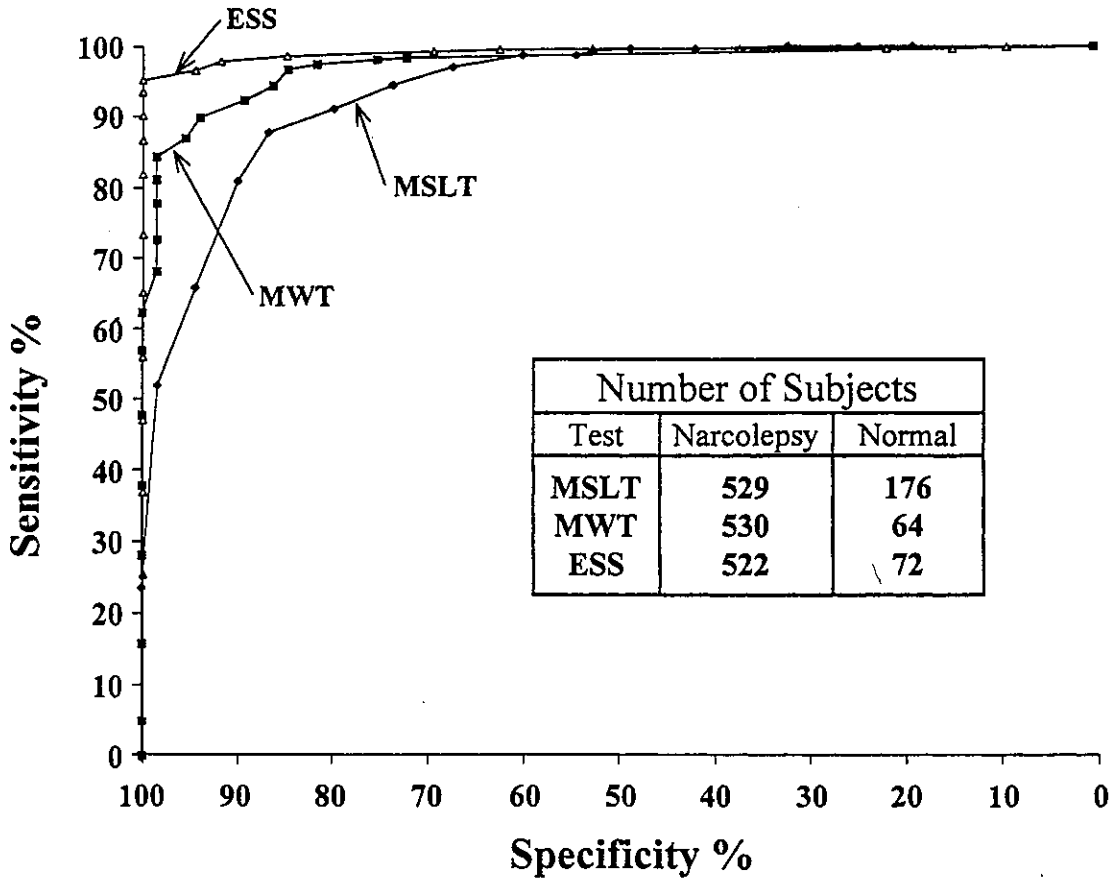

Figure 1. The receiver operator characteristic curves for the multiple sleep latency test (MSLT), the maintenance of wakefulness test (MWT) and the Epworth sleepiness scale (ESS). Note that specificity percentage is plotted from right to left, rather than as (100 – specificity) from left to right.

Some people may have difficulty understanding how the ESS, which is based on retrospective subjective reports, could be more accurate than the objective measurements made of the mean SL in the MSLT or MWT. One reason is that the MSLT and MWT can each measure only one situational sleep propensity, each of which is correlated in an uncertain way with what is intended to be measured – the average of many different situational sleep propensities that reflect the activities of daily life. The ESS provides an estimate of the latter, the average sleep propensity in eight specified situations (Johns 1998; Johns 1994).

Whatever cut-off point is used with the mean MSLT–SL it will tend either to misdiagnose a significant number of normal subjects as having EDS (false positives) or will fail to diagnose some narcoleptics or patients with other sleep disorders as having EDS (false negatives). That such misdiagnosis is happening now is suggested by recent results reported by Olson *et al.* (1998). All patients who were referred to their sleep centre with a suspicion of having EDS had overnight polysomnography and then an MSLT done routinely in an attempt to quantify their daytime sleepiness by what those authors believed to be the gold standard test. Of 225 'near-consecutive' patients 41 (18.2%) were diagnosed as having narcolepsy or idiopathic hypersomnia with a mean ESS score of 11 and with a range from 4 to 22, i.e. many had ESS scores in the normal range, of < 11. By contrast, at Epworth Sleep Centre, which serves a similar population in Australia, several surveys of 200 consecutive patients having polysomnography have shown that narcolepsy is diagnosed in 0–2.5% and idiopathic hypersomnia in another 0–3% of patients. Their mean ESS scores at Epworth Sleep Centre are 17.5 ± 3.5 and 17.9 ± 3.1, respectively (Johns 1991). So far, none has had an ESS score < 12. The percentage of all patients having polysomnography at Epworth Sleep Centre who have EDS, with ESS scores > 10, varies between 48 and 54%, similar to that shown graphically by Olson *et al.* (1998) and calculated as 52% by the present author. Thus, the 'narcolepsy' and 'idiopathic hypersomnia' diagnosed by Olson *et al.* (1998) are much more prevalent than at another similar sleep centre in Australia. In addition, the daytime sleepiness of those patients, measured by ESS scores, is much lower than in any of the 14 other published series of narcoleptics currently available from many centres in several countries in which mean ESS scores varied between 16 and 20.5 and in which the overall mean score was 18.6 (e.g. Parkes *et al.* 1998; Ezpeleta et al. 1998; Anic-Labat *et al.* 1999).

This raises an important question, not only about the diagnoses made by Olson *et al.* (1998) but also about some of the narcoleptics in the US modafinil study who had ESS scores 10, one as low as 1.0 (Sangal *et al.* 1997). It is difficult, intuitively, to understand how someone could have sufficient EDS (to be distinguished from fatigue) to enable a diagnosis of narcolepsy or idiopathic hypersomnia to be made and for treatment with a stimulant drug to be considered when their tendency to doze, even in relatively soporific situations as assessed by ESS scores, is less than that of an average normal subject. If the diagnosis is correct, the subjects have either not comprehended the ESS questions or have not answered them truthfully, which is possible. However, it must also be asked whether they were wrongly diagnosed as having narcolepsy or idiopathic hypersomnia on the basis of MSLT results, particularly a mean MSLT–SL < 5 min. It is also well known that EOREM's can occur in normal subjects and in patients suffering from sleep disorders other than narcolepsy (Bishop *et al.* 1996). Many people also have cataplexy-like symptoms without having narcolepsy, so misdiagnosis is likely (Parkes *et al.* 1998; Anic-Labat *et al.* 1999).

It is suggested therefore that instead of a mean MSLT–SL < 5 or < 8 min, an ESS score > 10 or perhaps > 11 should be a diagnostic feature of narcolepsy and of idiopathic hypersomnia. By contrast, Parkes *et al.* (1998) chose a higher cut-off point of > 14 for the ESS, which gave a sensitivity of 97% and a specificity of 100% in distinguishing their narcoleptics from normals. Their results provide independent confirmation of the accuracy of the ESS.

To do an MSLT or MWT on all patients suspected of having EDS is hard to justify when the ESS is a more discriminating test and is far cheaper and easier to administer. However, there is still good reason to do an MSLT on someone suspected of having narcolepsy, using EOREM's as a diagnostic aid (Moscovitch *et al.* 1993; Guilleminault *et al.* 1994). The ESS cannot measure rapid changes in sleepiness over periods of hours, but can over periods of a few weeks, as with the benefits of nasal continuous positive airway pressure treatment for obstructive sleep apnea (Hardinge *et al.* 1995). The MSLT or MWT are better suited for more rapidly repeated measurements of one situational sleep propensity. None of these tests has been validated under circumstances in which the results may prove incriminating in a legal sense. Nor are any of these tests of sleepiness much use for monitoring the state of alertness, say, of a truck driver while driving his truck. For that, we may require continuous methods based, for example, on the EEG or eye and eyelid movements (Mitler *et al.* 1997).

All methods for measuring daytime sleepiness should be used only with a clear understanding of their respective strengths and weaknesses. The ESS appears to be the most discriminating test of average sleep propensity so far but, ultimately, we need an objective test that is at least as good as the ESS. Such a test may have to measure sleep propensity objectively in a variety of different situations in which the subject's posture and activity is also monitored. Then we may truly have a gold standard, but it will be complicated and expensive.

REFERENCES

Aldrich, M. S., Chervin, R. D. and Malow, B. A. Value of the multiple sleep latency test (MSLT) for the diagnosis of narcolepsy. *Sleep*, 1997, 20: 620–629.

Anic-Labat, S., Guilleminault, C., Kraemer, H. C., Meehan, J., Arrigoni, J. and Mignot, E. Validation of a cataplexy questionnaire in 983 sleep-disorders patients. *Sleep*, 1999, 22: 77–87.

Bishop, C., Rosenthal, L., Helmus, T., Roehrs, T. and Roth, T. The frequency of multiple sleep onset REM periods among subjects with no excessive daytime sleepiness. *Sleep*, 1996, 19: 727–730.

Bliwise, D. L., Carskadon, M. A., Seidel, W. F., Nekich, J. C. and Dement, W. C. MSLT-defined sleepiness and neuropsychological test performance do not correlate in the elderly. *Neurobiol. Aging*, 1991, 12: 463–468.

Carskadon, M. A., Dement, W. C., Mitler, M. M., Roth, T., Westbrook, P. R. and Keenan, S. Guidelines for the Multiple Sleep Latency Test (MSLT): a standard measure of sleepiness. *Sleep*, 1986, 9: 519–524.

Chervin, R. D., Kraemer, H. C. and Guilleminault, C. Correlates of sleep latency on the multiple sleep latency test in a clinical population. *Electroencephalogr. Clin. Neurophysiol.*, 1995, 95: 147–153.

Chua, L. W. Y., Yu, N. C., Golish, J. A., Nelson, D. R., Perry, M. C., Foldvary, N. and Dinner, D. S. Epworth sleepiness scale and the multiple sleep latency test: Dilemma of the elusive link. *Sleep*, 1998, 21 (Suppl.): 184.

Dement, W. C. and Mitler, M. M. It's time to wake up to the importance of sleep disorders. *JAMA*, 1993, 269: 1548–1549.

Doghramji, K., Mitler, M. M., Sangal, R. B., Shapiro, C., Taylor, S., Walsleben, J., Belisle, C., Erman, M. K., Hayduk, R., Hosn, R., O'Malley, E. B., Sangal, J. M., Schutte, S. L. and Youakim, J. M. A normative study of the maintenance of wakefulness test (MWT). *Electroencephalogr. Clin. Neurophysiol.*, 1997, 103: 554–562.

Ezpeleta, D., Vicario, JL., Marin, P., Balas, A. and Peraita-Adrados, R. Clinical, polysomnographic and genetic characteristics of narcolepsy-cataplexy syndrome in a Spanish sample. *J. Sleep Res.*, 1998, 7 (Suppl. 2): 81.

Geisler, P., Crönlein, T., Tracik, F. and Zulley, J. M. S. L. T. Sleep latency in normals is age and sex related. *J. Sleep Res.*, 1998, 7 (Suppl. 2): 99.

Guilleminault, C., Mignot, E. and Partinen, M. Controversies in the diagnosis of narcolepsy. *Sleep*, 1994, 17: S1–S6.

Hardinge, F. M., Pitson, D. J. and Stradling, J. R. Use of the Epworth Sleepiness Scale to demonstrate response to treatment with nasal continuous airway pressure in patients with sleep apnoea. *Respir Med. J.*, 1995, 89: 617–620.

Harsh, J. R., Peszka, J., Hartwig, G. and Mitler, M. Night time sleep and daytime sleepiness in narcolepsy. *Sleep*, 1998, 21 (Suppl.): 198.

Jahnke, B. and Aldrich, M. S. The multiple sleep latency test (MSLT) is not infallible. *Sleep Res.*, 1990, 19: 240.

Johns, M. W. A new method for measuring daytime sleepiness: the Epworth sleepiness scale. *Sleep*, 1991, 14: 540–545.

Johns, M. W. and Hocking, B. Daytime sleepiness and sleep habits of Australian workers. *Sleep*, 1997, 20: 844–849.

Johns, M. W. Rethinking the assessment of sleepiness. *Sleep Med. Rev.*, 1998, 2: 3–15.

Johns, M. W. Sleepiness in different situations measured by the Epworth sleepiness scale. *Sleep*, 1994, 17: 703–710.

Johnson, L. C., Spinweber, C. L., Gomez, S. A. and Matteson, L. T. Daytime sleepiness, performance, mood, nocturnal sleep: The effect of benzodiazepine and caffeine on their relationship. *Sleep*, 1990, 13: 121–135.

Lammers, G. J. and van Dijk, J. G. The Multiple Sleep Latency Test: a paradoxical test? *Clin. Neurol. Neurosurg*, 1992, 94: S108–S110.

Levine, B., Roehrs, T., Zorick, F. and Roth, T. Daytime sleepiness in young adults. *Sleep*, 1988, 11: 39–46.

Lyznicki, J. M., Doege, T. C., Davis, R. M. and Williams, M. A. Sleepiness, driving and motor vehicle crashes. *JAMA*, 1998, 279: 1908–1913.

Mitler, M. M., Gujavarty, K. S. and Browman, C. P. Maintenance of wakefulness test: a polysomnographic technique for evaluating treatment in patients with excessive somnolence. *Electroencephagr Clin. Neurophysiol.*, 1982, 153: 658–661.

Mitler, M. M., Miller, J. C., Lipsitz, J. J., Walsh, J. K. and Wylie, C. D. The sleep of long-haul truck drivers. *N Engl. J. Med.*, 1997, 337: 755–761.

Mitler, M. M. US Modafinil in narcolepsy multicentre study group. Modafinil for the treatment of pathological somnolence in patients with narcolepsy. *Sleep Res.*, 1997, 26: 527.

Mitler, M. M., Walsleben, J. A., Sangal, R. B. and Hirshkowitz, M. Sleep latency on the maintenance of wakefulness test (MWT) for 530 patients with narcolepsy while free of psychoactive drugs. *Electroencephalog Clin. Neurophysiol.*, 1998, 107: 33–38.

U. S. Modafinil in Narcolepsy Multicenter Study Group. Randomized trial of modafinil for the treatment of pathological somnolence in narcolepsy. *Ann. Neurol.*, 1998, 43: 88–97.

Moscovitch, A., Partinen, M. and Guilleminault, C. The positive diagnosis of narcolepsy and narcolepsy's borderland. *Neurology*, 1993, 43: 55–60.

Nicoll, D. and Detmer, W. M. Diagnostic testing and medical decision making. In: L. M. Tierney, S. J. McPhee and M. A. Papadakis. (Eds) *Current Medical Diagnosis and Treatment.* Appleton and Lange, Stanford, 1999:1550–1560.

Olson, L. G., Cole, M. F. and Ambrogetti, A. Correlations among Epworth sleepiness scale scores, multiple sleep latency tests and psychological symptoms. *J. Sleep Res.*, 1998, 7: 248–253.

Parkes, J. D., Chen, S. Y., Clift, S. J., Dahlitz, M. T. and Dunn, G. The clinical diagnosis of the narcoleptic syndrome. *J. Sleep Res.*, 1998, 7: 41–52.

Richardson, G. S., Carskadon, M. A., Flagg, W., van den Hoed, J., Dement, W. C. and Mitler, M. M. Excessive daytime sleepiness in man: multiple sleep latency measurement in narcoleptic and control subjects. *Electroencephalogr. Clin. Neurophysiol.*, 1978, 45: 621–627.

Roehrs, T. and Roth, T. Multiple sleep latency test: technical aspects and normal values. *J. Clin. Neurophysiol.*, 1992, 9: 63–67.

Roehrs, T., Shore, E., Papineau, K., Rosenthal, L. and Roth, T. A two-week sleep extension in sleepy normals. *Sleep*, 1996, 19: 576–582.

Roehrs, T., Timms, V., Zwyghuizen-Doorenbos, A., Buzenski, R. and Roth, T. Polysomnographic, performance, and personality differences of sleepy and alert normals. *Sleep*, 1990, 13: 395–402.

Sangal, R. B., Mitler, M. M. and Sangal, J. A. M. US modafinil in narcolepsy multicenter study group. MSLT, MWT and ESS. indices of sleepiness in 522 drug-free patients with narcolepsy. *Sleep Res.*, 1997, 26: 492.

Sangal, R. B., Sangal, J. M. and Belisle, C. MWT and ESS measure different abilities in 41 patients with snoring and daytime sleepiness. *Sleep Res.*, 1997, 26: 493.

Sangal, R. B., Thomas, L. and Mitler, M. M. Maintenance of wakefulness test and multiple sleep latency test: measurement of different abilities in patients with sleep disorders. *Chest*, 1992, 101: 898–902.

Steinberg, R., Schönberg, C., Weeß, H. G., Schneider, C. and Pritzel, M. The validity of the multiple sleep latency test. *J. Sleep Res.*, 1996, 5 (Suppl.1): 220.

Thorpy, M. J. The clinical use of the Multiple Sleep Latency Test. *Sleep*, 1992, 15: 268–276.

Van den Hoed, J., Kraemer, H., Guilleminault, C., Zarcone, V. P., Miles, L. E. and Dement, W. C. Disorders of excessive somnolence: polygraphic and clinical date for 100 patients. *Sleep*, 1981, 4: 23–37.

Zwyghuizen-Doorenbos, A., Roehrs, T., Schaefer, M. and Roth, T. Test-retest reliability of the MSLT. *Sleep*, 1988, 16: 562–565.