

REVIEW ARTICLE

Rethinking the assessment of sleepiness

Murray Johns

Epworth Sleep Centre, Epworth Hospital, Melbourne, 3121, Victoria, Australia

The consequences of dozing off when intending to stay awake, e.g. while driving or at work, are potentially catastrophic. The accurate assessment of this tendency is important, but is currently difficult. Several different methods give disparate results. A way out of this dilemma is suggested that involves modification of existing concepts of sleep and wakefulness to conclude the powerful influence of behaviour on sleep propensity. This propensity at a particular time depends, hypothetically, on a mutually inhibitory interaction between a sleep and a wake drive, not on the magnitude of either drive alone. Measurements of sleep propensity are partly situation-specific, whether measured objectively by laboratory tests or subjectively by a questionnaire such as the Epworth Sleepiness Scale. The latter is believed to measure a general characteristic, the average sleep propensity across a range of specified situations in daily life. Any one situational sleep propensity is not always an accurate predictor of another, even in the same subject. The Multiple Sleep Latency test should not be a gold standard for such measurements. Wider discussion and more research into "sleepiness" is needed.

Key words: somnolence, sleepiness, sleep propensity, Epworth Sleepiness Scale, sleep drive, wake drive.

Introduction

A tendency to doze off during the course of daily life when intending to stay awake is one of the most frequent and important symptoms of patients who present to specialist sleep centres. There is mounting evidence that this tendency represents a major new public health issue because of its consequences, for example in contributing to motor vehicle and industrial accidents [1]. As practitioners of sleep medicine or researchers we should be able to quantify this tendency routinely. This is proving to be more difficult than was previously thought. In this review it is proposed to outline the variety of tests purporting to measure this tendency to doze, which many refer to as "daytime sleepiness". In so doing it will become apparent that there is confusion about the current concept and measurement of "daytime sleepiness" that requires some rethinking and clarification. That confusion is evidenced, for example, by the many different words and phrases that are used, often interchangeably, when discussing "daytime sleepiness", such as drowsiness, hypersomnia, somnolence, sleep propensity, fatigue, tiredness, the ability to fall asleep, sleepability, the ability to stay awake, subjective sleepiness, objective sleepiness and manifest sleepiness.

Somnolence or the somnolent state is assumed here to be equivalent to drowsiness and the drowsy state. It is a transitional state between wakefulness and sleep in which

the "sleep onset process", as it is called by Ogilvie and Wilkinson [2], has already begun, albeit intermittently, and is likely to proceed to sleep. The term "daytime sleepiness" or simply "sleepiness" as used here refers to a sleep propensity (SP), defined as a measure of a subject's tendency at a particular time to doze or fall asleep, at least briefly. This means entering Stage 1 sleep as defined by the EEG, whether or not it progresses to Stage 2 and other stages. This "sleepiness" should be distinguished from the set of feelings and symptoms associated with the drowsy state that can often, but not always, be introspected as "subjective sleepiness". These feelings are sometimes confused with feelings of fatigue or weariness that are not related in any simple or constant way to one's SP. The word "sleepiness" will be used in quotation marks here. It has been used in so many different contexts with different meanings in the past that we cannot define it accurately without qualification.

Some people may see the confusion about "sleepiness" arising mainly from differences between subjective and objective methods of measurement, between the perceived inaccuracy of subjective reports versus the accuracy and objectivity of sleep laboratory measurements. However, it is suggested here that a more important cause of the confusion is the failure of currently accepted concepts of sleep and wakefulness to explain "daytime sleepiness". Some new concepts and a new four-process model of sleep and wakefulness are described that may provide some clarification. It is concluded that currently we do not have a "gold standard" test of "daytime sleepiness" and that the optimum method for any given application depends on the nature and requirements of the test situation. More research with more rigorous conceptual analysis is urgently needed.

Current methods for assessing "sleepiness"

There are several widely used methods purporting to measure "daytime sleepiness" and many others that have been proposed. The list in Table 1 is not exhaustive. They can be divided into two broad categories, objective and subjective tests, with subdivisions in each.

Objective methods

One group of objective tests measures how rapidly the subject falls asleep (the sleep latency) 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. The Multiple Sleep Latency Test (MSLT) is the most widely used test of this type and many people consider it to be the "gold standard" [3], a position with which this reviewer disagrees (see below). In the MSLT the subject is asked to fall asleep when lying down for 20-minute periods, 2 hours apart, four to six times during the day while polysomnographic recordings are made. The time taken to fall asleep, or the sleep latency (SL), after the room lights are switched off can be measured objectively, with reasonable accuracy and with a test-retest reliability of about 0.6 [4]. The mean of the four to six separate SLs (the mean SL), is the measurement of "sleepiness" derived from the MSLT.

There is good evidence that comparisons between SLs in the MSLT at different times within the same subject reveal the effects of sleep deprivation, the time of day and

Table 1. Different tests of "sleepiness"

Objective tests of sleepiness

Sleep latency

Multiple Sleep Latency Test (MSLT)

Maintenance of Wakefulness Test (MWT)

Other physiological variables

EEG frequency, power

pupillometry

eye movements (saccades, slow eye movements)

eyelid movements (blinks, drooping)

evoked potentials (visual, auditory)

Performance tests

tracking tasks, 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 behaviour

Epworth Sleepiness Scale (ESS)

Sleep-wake Activity Inventory (SWAI)

the effects of sedative drugs, as expected [3]. That is, within the context of its own test situation which is kept as constant as possible, the MSLT provides a valid measurement of differences in "sleepiness". However, the results of the MSLT are often extrapolated to other situations and that involves assumptions that are seldom made explicit. The first is that the attachment of wires and the scrutiny that the subject is placed under does not influence the results. The second and related assumption is that the "sleepiness" measured by the MSLT is an accurate measurement of the subject's "sleepiness" in other test situations and, more generally, in daily life. There is evidence that the validity of these assumptions cannot be relied upon, as we shall see.

Another widely used test in this category is the maintenance of wakefulness test (MWT) [5]. It is similar to the MSLT except that the subject sits rather than lies down and tries repeatedly to stay awake rather than fall asleep. These two tests measure the same variable, the mean SL, but in different situations. The results of the two tests in the same subjects are significantly correlated (e.g. $r=0.41$, $n=258$, $P<0.001$), but there is discordance between them for about 30% of subjects [4]. Paradoxically, some people fall asleep more quickly when trying to stay awake than when intending to fall asleep. There have been attempts to explain this in terms of different abilities—the ability to fall asleep, or sleepability, versus the ability to stay awake [4]. The assumption has then been made that these tests give accurate measurements of two different but general characteristics of each subject in daily life. This has added to the confusion about "sleepiness" and has provided evidence that we need to do some rethinking about this concept. In addition, the MSLT and the MWT are so time-consuming and expensive that many sleep clinics do not perform either test routinely.

In the transition from alert wakefulness to Stage 2 sleep there are several physiological variables that indicate the presence of the drowsy or somnolent state (Table 1). Changes in the frequency and amplitude of the EEG or its power spectrum, with the blocking of alpha waves and appearance of theta waves, intermittently at first and then continuously for a while, are well known [6]. So, too, are the changes in eye movements, from the rapid, co-ordinated, conjugate movements of each eye during wakefulness to the poorly co-ordinated, slow or rolling eye movements of Stage 1 sleep. Monitoring methods that detect such changes have formed the basis of tests for "sleepiness" in real-life situations, for example in train drivers [7], truck drivers [8], and industrial night-shift workers [9], going about their usual work over periods of hours or days. The frequency, duration and velocity of eyelid movements with blinking may also provide a useful method for detecting the presence of the drowsy state [10]. Changes in the diameter of the pupil and its responses to brief flashes of light have been suggested as yet another test of "sleepiness".

The characteristics of auditory or visual evoked responses have also been suggested as measurements of "sleepiness" [11]. The amplitude of P300 is related to attention whereas the latency to P300 reflects the speed or efficiency of information processing in the central nervous system. Sangal and Sangal [11] studied 283 patients who each had an MSLT, an MWT and evoked response tests on the same day and in the same room. There were some statistically significant correlations between the mean SL in the MSLT and in the MWT with the latency of P300 waves (e.g. $r = -0.15$, $P < 0.05$), but these relationships were not close and it was concluded that evoked responses measure something different again. As yet, no tests based upon such methods have been standardized to the point of being used routinely, either in the sleep laboratory or in real-life situations.

Psychomotor and other performance tests have been used for many years to quantify the effects of sleep deprivation and to measure "daytime sleepiness" [12]. Such tests have varied from simple, continuous tracking tasks and reaction-time tests to tests of higher cognitive function involving decision-making, particularly with divided attention tasks. It appears that decrements in performance of the latter tests depend not only on brief periods of drowsiness and "micro-sleeps", with consequent impairment of attention, short-term memory and volition that interferes even with simple tracking tasks, but also on more subtle impairment of higher cognitive functions, even when the subject is awake and apparently paying attention [13]. Prolonged, non-stimulating performance tests are necessary for demonstrating the former effects of "sleepiness", but evidently not for the latter effects [13].

Subjective methods

One of the most widely used subjective tests of "sleepiness" is the Stanford Sleepiness Scale (SSS) [14]. The subject identifies his level of "sleepiness" at a particular time on a 7-point scale according to the presence or absence of particular feelings and symptoms that can be introspected and which are thought to reflect the progression from alert wakefulness to the drowsy state. Unfortunately, there is not a close correlation between SSS scores and SP measured objectively by SL [15]. The symptoms reported in the SSS have two components on principal components analysis, one reflecting "sleepiness", the other perhaps "fatigue". Nevertheless, repeated responses to the SSS by the same subjects do change with sleep deprivation and with the time of day, and the highest

scores do seem to reflect imminent sleep onset [2]. The same may also be said of scores from a similar test, the 9-point Karolinska Sleepiness Scale [16]. However, a single visual analogue scale (VAS) distinguishing degrees of introspected alertness-sleepiness on a 10-cm analogue scale may be the simplest and best way of quantifying "subjective sleepiness" at a particular time [16].

It has been pointed out in the past that subjects with a high level of "sleepiness", for whatever reason, tend to doze off during the day when alert subjects do not [17]. Johns developed the Epworth Sleepiness Scale (ESS) to standardize assessments of dozing behaviour in different situations [18]. It is a matter of common experience that if we doze off while sitting with our head unsupported, the postural muscles that hold the head erect when we are awake relax, and this allows the head to drop forward. That nodding movement often rouses us briefly and makes us aware of having just dozed, 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 in eight different situations that are commonly met in daily life. Those situations were chosen on a priori grounds to differ in their soporific nature. The eight item-scores each represent a different situational sleep propensity, that is the subjects' usual tendency to doze in the same situation at different times. Many different groups of subjects have been shown to rate the relative soporific nature of the eight situations of the ESS in the same way [19]. Lying down to rest in the afternoon when circumstances permit is the most soporific; sitting and talking to someone usually the least soporific.

The Epworth score is the sum of the eight item-scores and can vary from zero to 24. It represents the subject's average sleep propensity (ASP) across the eight different situations in daily life. The higher the Epworth score the more likely it is that the subject will doze in situations that are not soporific for normal people. Epworth scores are usually highly reliable in a test-retest sense over periods of months, (e.g. $r = 0.81$, $n = 87$, $P < 0.001$) [20]. That the ESS refers to observable behaviour rather than subjective feelings is supported by the high correlation between the patients' and their partners' independent reports of the patients' dozing behaviour e.g. ($\rho = 0.74$, $n = 50$, $P < 0.001$) [19]. There is mounting evidence about the validity of the ESS and its usefulness in quantifying changes in ASP after the treatment of obstructive sleep apnea by nasal continuous positive airway pressure (CPAP) [20], or after treatment of narcolepsy by stimulant drugs [21]. To date, the ESS is the only method for measuring "daytime sleepiness" that clearly distinguishes patients with narcolepsy or idiopathic hypersomnia (Epworth scores > 12) from normals (Epworth scores 0-10) [22]. The ESS is very cheap and easy to use repeatedly. The cost of the MSLT or MWT is three orders of magnitude greater than that of the ESS.

A 35-item Sleep Wake Activity Inventory (SWAI) has also been developed with one sub-scale that measures "sleepiness" based on reports of the frequency of dozing in different situations [23]. This sub-scale appears to be similar to the ESS.

The need for new concepts of sleepiness

We have already seen with the MSLT and MWT, which measure the same variable (mean sleep latency) in different situations, that the results of one test are not accurately predictable from the other. Only about 20% of variance in the results is common to the two tests. ESS item-scores provided us with a new opportunity to assess the

relationships between eight different situational sleep propensities in the same subject, using the same method in each [19]. Many of the 28 paired comparisons between different situational SPs in the same subject were correlated significantly in a statistical sense, but were never very closely correlated; not as closely as were assessments of the same situational SP repeated at different times in the same subject [19]. These results led Johns to the conclusion that all measurements of sleep propensity have several components of variation as follows:

- (1) a general component reflecting the average SP (ASP) of the subject across a variety of specified situations in daily life.
- (2) a situation-specific component reflecting the soporific nature of each particular situation and activity, and its effect on the SP of an "average" adult.
- (3) a situation \times subject interaction for each subject, reflecting their habits and their usual behavioural and psychological responses to particular situations.
- (4) variation in situational SPs with the time of day according to the combined effects of circadian and ultradian rhythms, and the duration of prior sleep and wakefulness (see below).

If these conclusions are accepted they have far-reaching implications for measurements of "sleepiness". They do not pose much of a problem for measuring changes in one particular situational SP over time, as with investigations of circadian and ultradian rhythms, or the effect of a drug on the same subject and in the same situation repeatedly. The MSLT and MWT measure such situational SPs every 2 hours [4]. Lavie's ultra-short sleep-wake paradigm measures a situational SP every 20 min, with 7 min available for sleep and 13 min for wakefulness [24]. Such techniques clearly show circadian and ultradian variations in their particular situational SPs. A problem arises if we attempt to extrapolate such results from one test situation to another and assume that any one situational SP accurately reflects a different situational SP or the subject's ASP in daily life. None of these findings can be explained by the currently accepted concepts or models of sleep and wakefulness—hence the need for some rethinking.

Current models of sleep and wakefulness

Currently there are several conceptual and mathematical models that describe some features of sleep and wakefulness, such as the occurrence of NREM-REM sleep cycles, and some of the changes in delta-wave sleep during a night's sleep [25]. There is good evidence, for example, that a subject's SP in a given situation increases progressively with the duration of prior wakefulness and decreases during sleep [26]. Borbély has called this process S. A different process (process C of Borbély) varies with a circadian rhythm that interacts with process S. A third process, W, has been postulated to explain a period of "sleep inertia" after waking.

Most current thinking about the "sleepiness" that is measured by the MSLT emphasizes the role of environmental stimulation. Carskadon and Dement have written of a "physiological sleep drive" being unmasked to reveal "manifest sleepiness" when we enter a "low-stimulus" situation, which the MSLT is said to involve [27]. The nature of this "unmasking" process, whether it is partial or complete, and how it interacts with the "physiological sleep drive" was unexplained. Dinges pointed out that it is

not simply the intensity of environmental stimulation, but also its meaning and significance for the subject that modifies the "expression of sleepiness" which is therefore a "contextually-dependent phenomenon" [28]. However, what has been largely ignored is the fact that many people can fall asleep within 15 min at any time of day simply because they lie down, close their eyes, stop talking and moving, and allow sleep to begin—a process I call sleepening. Alternatively, they can remain awake simply by standing up, even after prolonged sleep deprivation. Any clinically useful model of sleep and wakefulness must include some mechanism for these postural and behavioural changes to have a controlling influence. The currently accepted models do not.

A new conceptual model of sleep and wakefulness

Johns [29] has previously postulated a four-process conceptual model of sleep and wakefulness that takes into account the powerful behavioural influences on SP. A basic tenet of this model is that whether we are awake or asleep at any particular time depends on the relative strengths of two mutually inhibiting drives, the wake drive and the sleep drive, not on the absolute strength of either drive alone (Fig. 1). It is further postulated that each of these drives has a primary and a secondary component, the primary components being derived mainly from intrinsic activity in different neuronal groups within the central nervous system. The secondary components are homeostatic and behaviourally influenced.

The study of some other physiological variables has also revealed two main components of variation, comparable to that postulated for the sleep and wake drives, and manifested, first, as a circadian rhythm which is largely intrinsic and independent of behavioural changes and, second, a behaviourally responsive or homeostatic variation. This is true, for example, for core body temperature. Sleep researchers use a

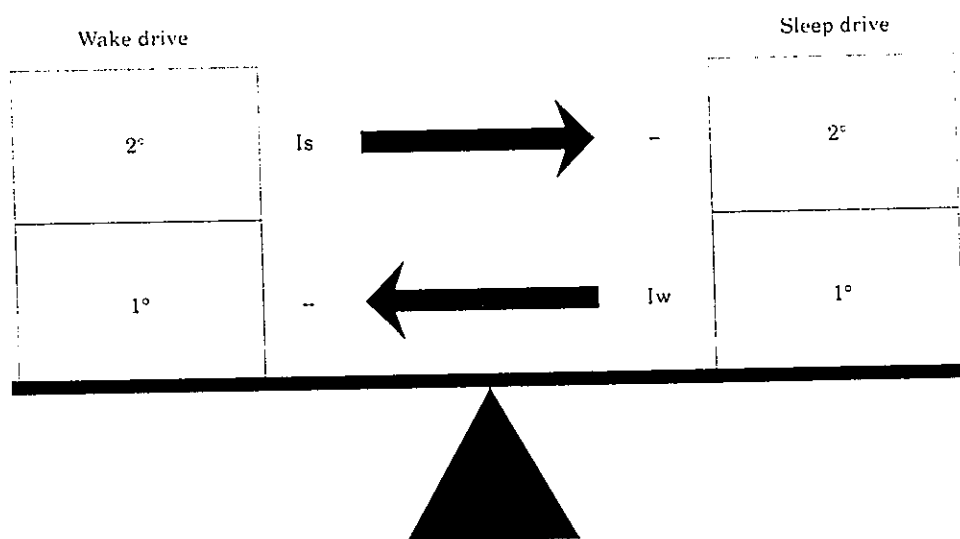


Figure 1. The sleep and wake drives, each with primary and secondary components, and each inhibiting the other. They function as a "flip-flop" oscillator, analogous to a see-saw.

constant routine in an attempt to minimize the behavioural effects on core temperature and thus reveal its intrinsic circadian and ultradian rhythms [30]. Cortisol secretion from the adrenal cortex also shows evidence of intrinsically generated as well as homeostatic changes with time, even though its secretion is episodic with an ultradian and a circadian rhythm [31]. It was with these variables in mind that the imputed sleep and wake drives were postulated.

The functions of these drives can be outlined hypothetically as follows:

Primary wake drive: arises from intrinsic activity in some or all of several neuronal centres in the central nervous system whose action promotes wakefulness [32]. This activity varies with a circadian rhythm, usually entrained to have its peak at 7 to 9 p.m. and its trough at 4 to 5 a.m. Behavioural and environmental influences would set the phase of such a rhythm, but would not generate it. This is reflected in core temperature, cortisol secretion, melatonin secretion, and REM sleep, although each of these variables also have other control mechanisms. The primary wake drive is the equivalent of Borbély's process C [26].

Secondary wake drive: due to the additive effects of inputs to the central nervous system from postural muscles, joints and other enteroceptive nerve tracts, as well as from visual and other exteroceptive inputs, with collaterals to the thalamic projection system and the limbic system. This secondary wake drive is partly under voluntary control.

Primary sleep drive: arises from intrinsic activity in various neuronal centres whose co-ordinated action promotes non-REM sleep [32]. It may have low-amplitude circadian and ultradian rhythms. It may be related to the secretion of thyroid stimulating hormone. Its peak of activity would usually be between 10 p.m. and midnight. This is the most conjectural part of the four-process model. However, it could explain the reappearance of delta-wave in the latter part of a night's sleep. It may also explain the substantial differences in the amount of delta-wave obtained by different people of the same age that are otherwise unexplained.

Secondary sleep drive: this is the equivalent of Borbély's process S [26]. It increases progressively during wakefulness and is discharged during non-REM sleep. Its build-up is prevented by frequent naps during the day.

It is postulated that the primary and secondary wake drives have additive effects which, together, would constitute the total wake drive. Similarly with the total sleep drive. These sleep and wake drives would mutually inhibit each other (I_s and I_w in Fig. 1) to produce an oscillator of the "flip-flop" type. This would be analogous to an electronic astable multivibrator. Another analogy could be drawn with a see-saw as in Figure 1. We would be awake whenever the total wake drive exceeded the total sleep drive and asleep when the sleep drive exceeded the wake drive. Once asleep, progression from stage 1 to stage 2 probably requires further active inhibition of the wake drive by the sleep drive (I_w in Fig. 1). The sleep onset process takes some time and is not instantaneous, as others have pointed out [2]. Other processes during sleep would control the interaction between non-REM and REM sleep. This model is not unique in postulating opposing sleep and wake drives [33], but is unique in separating primary and secondary components of those drives. The model needs further detailed elaboration in terms of neuroanatomy and neurophysiology.

Under most circumstances it is the magnitude of the secondary wake drive that is the single most important determinant of whether we fall asleep or stay awake at any

particular time. Changes in the secondary wake drive are partly under voluntary control, for example, when we choose to lie down in bed, relax and close our eyes, i.e. to begin the process of sleepening. Other changes in SP with the time of day (due to changes in the primary wake and sleep drives) and with the duration of prior wakefulness (due to changes in the secondary sleep drive) are not under much voluntary control, but would influence us in choosing when to go to bed each night as part of our sleep habits. Different situational SPs in the same subject would arise because of differences in the secondary wake drive which would be greater, for example, when sitting and talking to someone than when lying down to rest. These changes would be largely independent of the time of day or of the duration of prior sleep or wakefulness.

Corollaries of the four-process model of sleep and wakefulness

This model can explain many differences in the patterns of sleep and wakefulness between normal subjects, and between normal and abnormal. A relatively high ASP could be due either to a high sleep drive, a low wake drive, or both; or perhaps by unusually efficient inhibition of the wake drive by the sleep drive. Such differences may be partly inherited and partly learned as psychophysiological traits. Other differences in ASP between subjects could be due to sleep disorders that increase the sleep drive or decrease the wake drive, or change the interaction between those drives.

A very "sleepy" subject (e.g. with untreated narcolepsy or severe obstructive sleep apnoea) could spend much of the day in a state close to drowsiness while still remaining technically awake most of the time. It would require only a minimal decrease in the secondary wake drive, by engaging in a more soporific activity, to produce dozing behaviour in the drowsy state. If such a "sleepy" person attempted to stay awake all night, his secondary sleep drive would increase progressively throughout that time. His primary wake drive would decrease after about 9 p.m. and would be minimal by about 4 a.m. It would then be very difficult for him to remain awake, for example while driving a car, for his wakefulness would depend almost entirely on his secondary wake drive. There is evidence that performance failure under such circumstances is due not only to brief lapses into sleep, but also to a general decrement in central nervous system function when technically awake but with a high SP [28,34]. This has little to do with the concept of fatigue that is often invoked in similar discussions elsewhere, inappropriately in this reviewer's opinion.

So-called "sleep inertia" could be explained as part of the process of waking which, as with the sleep-onset process, takes some time. Wakening involves active inhibition of the sleep drive by the wake drive (Is in Fig. 1), even after the point is reached when the magnitude of the wake drive first exceeds that of the sleep drive. Different subjects will vary in the efficiency and speed of that active inhibition. Thus, in some people the wakening process can take longer than in others, which could be interpreted as "inertia".

The four-process model described here can explain the existence of a "sleep gate" at night after a period of very low SP (a "forbidden zone") in the evening when an ultra-short sleep-wake paradigm is imposed (7 min for sleep, 13 awake) [24]. This would prevent the build up of the secondary sleep drive that normally proceeds

during hours of wakefulness, and would keep the secondary wake drive relatively low and constant during the repeated measurements of the same situational SP. The SP would then vary only in relation to the interaction between the primary sleep and wake drives. When the sleep drive exceeded the wake drive in magnitude, the "sleep gate" would be suddenly opened.

Is the MSLT a gold standard for measurements of "sleepiness"?

Many people believe that measurements of the mean SL in the MSLT provide the gold standard against which other methods for assessing "sleepiness" should be compared [3]. There are several problems with that, some of which have been pointed out by others [35,36]. One problem is with the assumption that the attachment of wires and the scrutiny involved in the test situation usually have a negligible effect on the results, i.e. the MSLT is perceived by the subject to be a "low-stimulus" situation. This may be true for some people, but surely cannot be assumed for all subjects. This may explain why some patients who complain of "daytime sleepiness" that affects their daily lives do not fall asleep in <10 min in the MSLT.

According to the conceptual model of sleep and wakefulness outlined above, the MSLT measures one particular situational SP, the MWT another, and the 8-item scores of the ESS other situational SPs again. They share some commonality which reflects the ASP, but are not always closely related. Thus, we ought not to refer to the results of the MSLT as reflecting a general "ability to fall asleep", nor of the MWT reflecting a general "ability to stay awake".

Of even greater concern is the fact that despite widespread use of the MSLT for almost two decades its proponents have not given an accurate description of the distribution of mean SLs in normal subjects. Most users of the MSLT have relied uncritically on a "rule of thumb" suggested by Richardson *et al.* in 1978 [37], or on some minor variation of that rule. That rule is that a mean SL ≤ 5 min in the MSLT indicates "pathological sleepiness", SLs between 5 and 10 min are in a "diagnostic grey area", and normal subjects have SLs in the range 10–20 min. The limited published data about normal subjects in the MSLT suggest that this "rule of thumb" cannot be relied upon. The mean SL for normal adults is approximately 11 min with an SD that has varied in different series from about 3 to 6 min (not always reported). The range of SLs in normal subjects is 2 to 20 min [38]. While 80–90% of patients with narcolepsy have a SL ≤ 5 min, as do about 15% of normal adults, the latter are without complaints of "excessive daytime sleepiness" in their daily lives [38]. Almost as many normal subjects have a SL ≤ 10 min (i.e. in the "pathological" and "grey" areas) as have an SL >10 min (the so-called "normal"). These exact proportions have not been published, but can be inferred from a study of 193 normal subjects aged 18–55 years, of whom 44.6% had a mean SL ≤ 9 min and 55.4% >9 min [38]. Thus the MSLT is not a particularly useful test for discriminating between normal and abnormal "sleepiness". In this author's opinion, the MSLT cannot be a gold standard for measuring ASP. However, it is likely that no other test of one situational SP can be a gold standard either. We simply do not have a gold standard at present. The ESS cannot assume such a role because of potential difficulties with it and all questionnaires in being able to be falsified, whether consciously or not.

Optimum methods for assessing "sleepiness"

The optimum method for assessing "sleepiness" depends to some extent on the circumstances of that assessment. This could vary from the need to diagnose a patient's sleep disorder such as narcolepsy, or to determine a driver's fitness to drive a truck at a particular time, or to determine the level of ASP in a large group of subjects by epidemiological surveys. It may be that different methods of assessment should be used under these different circumstances when costs and ease of administration are taken into account. Nevertheless, the results from the different tests should all be compatible with one another. Each of the many psychophysiological tests and variables measured in the laboratory that can be shown to reflect somnolence and different levels of "daytime sleepiness" may have a role to play in particular circumstances. It is important to point out, however, that any test in one particular situation, no matter how accurately the measurements are made, will provide only one situational SP and will face the same problems inherent in the MSLT so far as extrapolation of the results to other situations is concerned.

When deciding who is fit or unfit to drive a truck or fly an aeroplane from the point of view of "sleepiness", there may be more than one criterion to be applied. Certainly, an argument can be made that someone whose ASP is very high (e.g. with untreated narcolepsy) would not be fit to do so at all. ESS scores can detect such people, within the limits of their subjective reports. However, other people may be fit to drive most of the time, but unfit at particular times because of temporary sleep deprivation, use of drugs, etc. compounded with time-of-day effects. What we really need under such circumstances is an ongoing indicator of their impending performance failure [39]. It may not be sufficient to detect sleep onset and their actual performance failure, as for example from EEG recordings [8], if the potential consequences of that failure are catastrophic. We must attempt to predict imminent drowsiness and, hopefully, avert performance failures before they happen. Drivers' education about the nature of this problem may increase the likelihood that introspection of their own behavioural state will convince some drivers not to continue driving when drowsy. This important public health issue requires much more research for which, it is hoped, this review may provide a stimulus.

Practice Points

1. Somnolence and drowsiness are part of a transitional state between wakefulness and sleep during which time the sleep onset process has begun, at least intermittently.
2. All measurements of sleep propensity (SP) are partly situation-specific and can be made only in relation to the subject's activity, posture and situation at the time.
3. A high situational sleep propensity in one situation is not always an accurate predictor of the situational sleep propensity in a different situation. This may be true even for a simulated activity, as in a driving simulator, as compared with the real-life situation.
4. To measure sleep propensity at a particular time and place is not to measure the sleep drive, but rather the relative magnitude of, and interaction between, the sleep drive and the wake drive.

5. A subject's sleep propensity at a particular time and place can be seen as a measure of how far the wake drive, particularly the secondary wake drive, must decrease to reach the point where the sleep drive exceeds the wake drive and sleep onset begins.
6. The MSLT should no longer be seen as the gold standard for measuring a subject's average sleep propensity. We do not currently have such a gold standard.
7. The ESS is a cheap and simple method for measuring average SP in large numbers of people, but it shares the limitations inherent in all self-administered questionnaires.
8. All methods for assessing "sleepiness" should be used with a clear understanding of their respective limitations.

Research Agenda

More research is needed into the drowsy state and its predictors that may help prevent the public health problems with which performance failures in the drowsy state are often associated.

References

- 1 Dement WC, Mitler MM. It's time to wake up to the importance of sleep disorders. *JAMA* 1993; 269:1548-1549.
- 2 Ogilvie RD, Wilkinson RT, Allison S. The detection of sleep onset: behavioural, physiological, and subjective convergence. *Sleep* 1989; 12:458-474.
- *3 Thorpy MJ. The clinical use of the Multiple Sleep Latency Test. *Sleep* 1992; 15:268-276.
- *4 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* 1992; 101:898-902.
- 5 Mitler MM, Gujavarty KS, Browman CP. Maintenance of wakefulness test: a polysomnographic technique for evaluating treatment in patients with excessive somnolence. *Electroencephalogr Clin Neurophysiol* 1982; 153:658-661.
- *6 Broughton R, Hasan J. Quantitative topographic electroencephalographic mapping during drowsiness and sleep onset. *J Clin Neurophysiol* 1995; 12:372-386.
- 7 Torsvall L, Åkerstedt T. Sleepiness on the job: continuously measured EEG changes in train drivers. *Electroencephalogr Clin Neurophysiol* 1987; 66:502-511.
- 8 Mitler MM, Miller JC, Lipsitz JJ, Walsh JK, Wylie CD. The sleep of long-haul truck drivers. *N Engl J Med* 1997; 337:755-761.
- 9 Torsvall L, Åkerstedt T, Gillander K, et al. Sleep on the night shift: 24-hour EEG monitoring of spontaneous sleep/wake behaviour. *Psychophysiology* 1989; 26:352-358.
- 10 De Gennaro L, Casegrande M, Andropoli D, Pellegrino C, Violani C, Bertini M. Oculomotor activity as predictor of sleep onset at the Multiple Sleep Latency Test (MSLT). *J Sleep Res* 1996; 5(Suppl. 1):45.
- 11 Sangal RB, Sangal JM. Measurement of P300 and sleep characteristics in patients with hypersomnia: do P300 latencies, P300 amplitudes, and Multiple Sleep Latency and Maintenance of Wakefulness tests measure different factors? *Clin Electroenceph* 1997; 28:179-181.
- 12 Wilkinson RT. Sleep deprivation: performance tests for partial and selective sleep deprivation.

The most important references are denoted by an asterisk.

- In: Abt LE, Riess BF (eds). *Progress in Clinical Psychology: Volume 8. Dreams and Dreaming*. New York: Grune & Stratton 1968, pp. 28-43.
- 13 Monk TH, Carrier J. Speed of mental processing in the middle of the night. *Sleep* 1997; 20: 399-401.
 - 14 Hoddes E, Zarcone V, Smythe H, Phillips R, Dement W, et al. Quantification of sleepiness: a new approach. *Psychophysiology* 1973; 10:431-436.
 - 15 Johnson LC, Spinweber CL, Gomez SA et al. Daytime sleepiness, performance, mood, nocturnal sleep: the effect of benzodiazepine and caffeine on their relationship. *Sleep* 1990; 13:121-135.
 - 16 Åkerstedt T, Gillberg M. Subjective and objective sleepiness in the active individual. *Intern J Neuroscience* 1990; 52:29-37.
 - 17 Roth T, Roehrs T, Carskadon M, Dement W. Daytime sleepiness and alertness. In: Kryger MH, Roth T, Dement WC (eds). *Principles and Practice of Sleep Medicine*. Philadelphia: W.B. Saunders, 1989, pp. 14-23.
 - *18 Johns MW. A new method for measuring daytime sleepiness: the Epworth sleepiness scale. *Sleep* 1991; 14:540-545.
 - *19 Johns MW. Sleepiness in different situations measured by the Epworth sleepiness scale. *Sleep* 1994; 17:703-710.
 - 20 Engleman HM, Asgari-Jirhandeh N, McLeod AL, Ramsay CF, Deary IJ, Douglas NJ. Self-reported use of CPAP and benefits of CPAP therapy. *Chest* 1996; 109:1470-1476.
 - 21 Parkes D. Introduction to the mechanism of action of different treatments of narcolepsy. *Sleep* 1994; 17:S93-S96.
 - 22 Johns M, Hocking B. Daytime sleepiness and sleep habits of Australian workers. *Sleep* 1997; 20:844-849.
 - 23 Rosenthal LD, Rosen A, Wittig RM, Zorick FJ, Roehrs TA, Roth T. A sleep-wake activity inventory to measure daytime sleepiness. *Sleep Res* 1991; 20:130.
 - 24 Lavie P. Ultrashort sleep-waking schedule-111: "Gates" and "forbidden" zones. *Electroencephalogr Clin Neurophysiol* 1986; 63:414-425.
 - *25 Borbély AA, Achermann P. Concepts and models of sleep regulation: an overview. *J Sleep Res* 1992; 1:63-79.
 - *26 Borbély AA, Achermann P, Trachsel L, Tobler I, et al. Sleep initiation and sleep intensity: interaction of homeostatic and circadian mechanisms. *J Biol Rhythms* 1989; 4:149-160.
 - 27 Carskadon MA, Dement WC. The multiple sleep latency test: what does it measure? *Sleep* 1982; 5(Suppl. 2):S67-S72.
 - 28 Dinges DF. The nature of sleepiness: causes, contexts, and consequences. In: Stunkard AJ, Baum A (eds). *Perspectives in Behavioral Medicine: Eating, Sleeping, and Sex*. Hillsdale, NJ: Lawrence Erlbaum Assoc. 1989, pp. 147-179.
 - *29 Johns MW. Daytime sleepiness, snoring, and obstructive sleep apnea. The Epworth sleepiness scale. *Chest* 1993; 103:30-36.
 - 30 Lack LC, Lushington K. The rhythms of human sleep propensity and core body temperature. *J Sleep Res* 1996; 5:1-11.
 - 31 van Canter E, Refetoff S. Multifactorial control of the 24-hour secretory profiles of pituitary hormones. *J Endocrinol Invest* 1985; 8:381-391.
 - 32 Jones BE. Basic mechanisms of sleep-wake states. In: Kryger MH, Roth T, Dement WC (eds). *Principles and Practice of Sleep Medicine*. Philadelphia: W.B. Saunders, 1989, pp. 121-138.
 - 33 Edgar DM, Dement WC, Fuller CA. Effect of SCN lesions on sleep in squirrel monkeys: evidence for opponent processes in sleep-wake regulation. *J Neurosci* 1993; 13:1065-1079.
 - *34 Valley V, Broughton R. The physiological (EEG) nature of drowsiness and its relation to performance deficits in narcoleptics. *Electroencephalogr Clin Neurophysiol* 1983; 55:243-251.
 - *35 Chervin RD, Kraemer HC, Guilleminault C. Correlates of sleep latency on the multiple sleep latency test in a clinical population. *Electroencephalogr Clin Neurophysiol* 1995; 95:147-153.
 - 36 Nicholson AN, Stone BM, Wright NA, Belyavin AJ. Daytime sleep latencies: relationships with the electroencephalogram and with performance. *J Psychophysiol* 1989; 3:387-395.
 - 37 Richardson GS, Carskadon MA, Flagg W, van den Hoed J, Dement WC, Mitler MM. Excessive daytime sleepiness in man: multiple sleep latency measurement in narcoleptic and control subjects. *Electroencephalogr Clin Neurophysiol* 1978; 45:621-627.
 - *38 Roth T, Roehrs TA, Livesay D, et al. Nocturnal sleep and daytime sleepiness in normal volunteers for research studies. *Sleep Res* 1993; 22:101.
 - 39 Horne JA, Reyner LA. Driver sleepiness. *J Sleep Res* 1995; 4(Suppl. 2):23-29.

