 doi:10.1111/j.1479-8425.2010.00450.x

REVIEW ARTICLE

A new perspective on sleepiness

Murray W JOHNS
Optalert Pty Ltd, and Sensory Neuroscience Laboratory, Swinburne University of Technology, Melbourne, Victoria, Australia

Abstract

The aim of this investigation was to examine current methods for measuring sleepiness from a new perspective, highlighting different definitions of sleepiness that are being used, inadequacies of the current conceptual framework for thinking about sleep and wakefulness, and the different timescales used for measuring sleepiness. The currently used methods are briefly described. The results of different tests of sleepiness within the same subjects are then examined, first, using subjectively reported item-scores from the Epworth Sleepiness Scale, and later, using mean sleep latencies in the Multiple Sleep Latency Test and the Maintenance of Wakefulness Test. It is concluded that sleepiness, in the sense of sleep propensity, is not a single characteristic of each person. Measurements of sleepiness involve three sources of variance. First, the person's average sleep propensity in daily life, a hypothetical construct for which we do not have a direct measure. Second, the somnificity of the person's posture, activity and situation at the time. This is a manifestation of how the sensory nervous system, and consequently the wake drive in the central nervous system, is influenced in the majority of people by what they are doing. The third source of variance reflects differences in the way some people respond to particular circumstances. The latter influences are both subject-specific and situation-specific. They are presumably learned, and are not very predictable for individual subjects. It is proposed that, to explain these findings, we need to include a new process (Process-A), reflecting afferent inputs to the central nervous system, in any model of sleep and wakefulness.

Key words: drowsiness, Process-A, sleep propensity, sleep drive, sleepiness, wake drive.

INTRODUCTION

One of the most important concepts in sleep medicine and sleep research is that of sleepiness. It is therefore somewhat surprising that there is no clear consensus at present on what sleepiness is or how it should be measured.[1,2] An early classification of sleep disorders by the American Sleep Disorders Association had "disorders of excessive daytime sleepiness" (DOES) as one of its major categories of disordered sleep.[3] At the time it was believed by most, but not all, clinicians and researchers of sleep medicine that one particular test, the Multiple Sleep Latency Test (MSLT), was the gold standard for measuring sleepiness, a position that was officially adopted by the American Academy of Sleep Medicine in 1992.[4] However, that was subsequently criticized on several grounds. For example, Johns pointed out the absence, for many years, of a clearly established reference range of normal values for mean sleep latencies in the MSLT measured by a generally acceptable method as opposed to a "rule of thumb".[5] When a proper reference range was used, the MSLT was found to be wanting as a gold standard. In recent years, recommendations about the MSLT as a gold standard have been revised and the reference range of normal mean sleep latencies in the MSLT has been clarified,[6] but few people seem to have recognized these changes.

Correspondence: Dr Murray W. Johns, Level 3, Building 5, 658 Church Street, Richmond, Vic. 3121, Australia. Email: mjohns@optalert.com

Received: 15 April 2010/Accepted: 14 June 2010.

The measurement of sleepiness has also been confused by the use of several different terms to describe sleepiness, without adequate definitions. Some researchers have concluded that there must be different "dimensions" of sleepiness based on separate generalized "abilities", such as the "ability to fall asleep" and the "ability to stay awake".[7] The phrase "sleepability" was coined in an attempt to explain an apparent anomaly in which some normal people fall asleep very quickly when they want to, without having any evidence for "excessive daytime sleepiness" in their daily lives.[8] "Subjective sleepiness" has been distinguished from "objective sleepiness" as different dimensions of sleepiness, based on different methods of measurement. The term "hypersomnia" is still used by some people to mean the symptom or a sleep disorder that involves "excessive daytime sleepiness". In some quarters, particularly in relation to drowsy driving, sleepiness has long been confused with fatigue. The whole subject of sleepiness is currently very confused.[9,10]

The purpose of this investigation was to examine the currently used methods for measuring sleepiness from a new perspective. Initially, this involved clarifying some definitions, including consideration of different timescales for measuring sleepiness. The relationships between several different measures of sleepiness in the same subjects are examined, which raises important issues about the nature of sleepiness. It appears that there is an important role for the sensory nervous system in the control of sleep and wakefulness, and therefore in sleepiness, that has been largely neglected. There is no single gold standard method for measuring sleepiness. Discussion about the nature and diagnosis of particular sleep disorders that may involve "excessive daytime sleepiness" is beyond the scope of this investigation.

Definitions of sleepiness

There is no ambiguity about the common English usage of the adjective, sleepy, or of the noun, sleepiness. By dictionary definition, sleepiness means the state of being sleepy, that is, "inclined to sleep, having difficulty in keeping awake, drowsy, somnolent".[11] By this long-held definition, sleepiness is synonymous with the state of drowsiness.

However, over the past 30 years or so, new meanings of the word sleepiness have been introduced, mainly in the field of sleep medicine. For many people, sleepiness now means "sleep propensity" – the speed, ease or likelihood of falling asleep as opposed to remaining awake.[9] The term sleepiness has also been used sometimes to mean "sleep drive", that is, a "physiological drive usually resulting from sleep deprivation".[12] This use of the word sleepiness implies that sleepiness is a function solely of a putative "sleep drive", and not a joint function of interactions between a "sleep drive" and a "wake drive", each involving multiple neuronal centres in the central nervous system. By this meaning, any measurement of sleepiness is a direct measure of the "sleep drive". It is probably no longer appropriate to use the word sleepiness in this sense.

Methods for measuring sleepiness

The different ways of thinking about sleepiness have formed the basis of different methods for measuring it, either as sleep propensity or as a state of drowsiness.

For example, the MSLT objectively measures a person's sleep propensity in terms of sleep latency – how many minutes it takes them to fall asleep in that particular test-situation.[13] The Maintenance of Wakefulness Test (MWT) also measures sleep latency, using the same criteria as the MSLT, but in a different test situation.[14] In fact, both tests measure sleep latency several times on one day (usually four or five times) and then provide a mean sleep latency for each subject. Both tests are based on objective measurements, not subjective reports. However, an important point is that neither of these tests measure sleepiness as a state, in the sense of the presence of drowsiness or as a distinction between different levels of drowsiness. The relationship between results of the MSLT and the MWT in the same subjects will be discussed below.

Several other methods have been used to indicate the presence of the drowsy state or to distinguish different levels of drowsiness at some particular time.[10] For example, the Karolinska Sleepiness Scale (KSS) asks respondents to choose which of seven statements most accurately reflects their subjective state of alertness/drowsiness over the past few minutes.[15] Those statements form a Likert scale, ranging from "extremely alert" (=1) to "extremely sleepy, fighting sleep" (=9). The Stanford Sleepiness Scale (SSS)[16] does something similar with a 7-point Likert scale, but has the disadvantage of using some poorly defined words such as "vital", "foggy" and "woozy". A visual analogue scale of "alertness–drowsiness" has also been used. The respondent makes a mark on a horizontal line 100 mm long with "very alert" at one end and "very drowsy" at the other. The position of their mark along that line is the measure of their sleepiness. These methods provide subjective measurements of sleepiness at some particular time, whatever the person is doing.

Other proposed methods for measuring sleepiness have used a variety of physiological variables as objective indicators of the drowsy state.[10] They include changes in the frequency, amplitude, and power of the EEG (electroencephalogram), often divided into frequency ranges corresponding to, for example, alpha or theta waves.[17] Another method measures the brain's evoked responses to a visual or an auditory stimulus repeated frequently over a short time. These EEG changes have not formed the basis of a standardized test of sleepiness at a particular time, but they have been used to demonstrate changes across time. Other physiological variables have been derived from pupillometry – the pattern of spontaneous changes in the diameter of the pupil under some conditions, the rate of change in the diameter of the pupil in response to a flash of light, etc.[18] As with the EEG-based methods, pupillometric methods do not provide a standardized measure of drowsiness, but can be used repeatedly to track changes over time.

Various performance tests have been proposed as indicators of drowsiness. One of the most commonly used is the Psychomotor Vigilance Test (PVT), a simple visual reaction-time test that takes about 10 min to complete.[19] With drowsiness, reaction-times become longer and more variable, and there are more frequent "lapses" in performance.[20] The PVT shows the effects of one night's sleep deprivation, but not as sensitively as measurements of sleep latency in the MSLT.[21] Several other performance tests have been used that relate more specifically to higher cognitive functions and memory. Some people believe that sleepiness can be measured more accurately by performance tests that involve higher cognitive functions than by simple reaction-time tests, but the evidence for this is equivocal.[21]

In recent years, new methods have been specifically developed for objectively and continuously measuring drowsiness while driving a vehicle. The need for this arose from considerations of safety and operator performance rather than the diagnosis of sleep disorders. These methods have been based mainly on eyelid movements and eye-closures. One such method uses video camera images of the person's eyes. The proportion of time that the eyelids are at least 80% closed (PERCLOS) increases with drowsiness.[22] However, there have been practical issues with the implementation of video-camera methods for monitoring the drowsiness of drivers because of difficulties with the quality of the video images in sunlight or when the person wears glasses.

An alternative method for continuously measuring sleepiness, in the sense of drowsiness, has recently been introduced. It uses infrared reflectance oculography to measure the relative velocity and duration of eyelid movements during blinks, and especially the short-term variability of those characteristics (Optalert™, manufactured by Optalert Pty Ltd, Melbourne, Australia).[23,24] This method is currently being used for monitoring the drowsiness of drivers. Its role in clinical sleep medicine is yet to be established.

One of the most commonly used methods for measuring sleepiness, the Epworth Sleepiness Scale (ESS),[25] is quite different from the other methods described above. It measures sleepiness as sleep propensity, but not in terms of sleep latency, rather as retrospective subjective reports about the likelihood of dozing off in particular situations as part of daily life. The ESS does not refer to subjective feelings of drowsiness or sleepiness, and should not be confused with other methods that do, such as the KSS or SSS. The different situations in the eight items of the ESS were chosen on *a priori* grounds to differ in what Johns has called their somnificity, that is, their capacity to facilitate sleep onset in the majority of people.[9,26] People with high levels of sleepiness are likely to doze off in situations associated with low somnificity. The ESS is the only method available at present that provides a direct estimate of a person's average level of sleepiness in daily life. The total ESS score (0–24) is the sum of eight item-scores that will be the subject of further discussion below.

Different timescales for measurements of sleepiness

It is important to recognize that different methods for measuring sleepiness are based on different timescales.[9,10] For example, each nap opportunity in the MSLT measures the subject's sleep latency as a measure of sleep propensity in that test-situation at one particular time. Johns has called this the subject's MSLT-instantaneous sleep propensity (MSLT-ISP), although the reality is that it also reflects their activities and their sleep propensity over the preceding few minutes.

By contrast, the mean sleep latency in the MSLT, measured from four or five naps in one day, gives a different measure – that of the subject's usual sleepiness in that test situation when repeated several times. Johns calls this the subject's MSLT-situational sleep propensity (MSLT-SSP). However, for many practical purposes, what we really want to know about is the person's general level of sleepiness in daily life, which Johns calls their average sleep propensity (ASP). Most people assume that sleepiness is a general characteristic of each subject that can be measured independently of the test-situation. A corollary of that is to assume that a person's MSLT-SSP gives an accurate measure of their ASP. That is, if a person has excessive daytime sleepiness (EDS) in one test-situation and at a particular time, it is assumed they will have EDS in other situations, at other times, and more generally in daily life. This is a source of much confusion, as we shall see below.

The relationships between different measures of sleepiness in the same subject

With a variety of different methods for measuring sleepiness, either as sleep propensity or as drowsiness, and with measurements made on different timescales, using either subjective or objective methods, the question arises: what is the relationship between the results of these different tests? Much of the available evidence has recently been summarized.[9,10] However, as part of that ongoing discussion, we shall now consider how a person's sleep propensity depends on what they are doing at the time.

Analysis of ESS scores from four groups of subjects with and without sleep disorders

The ESS asks each respondent to give an estimate (rated 0–3) of his/her usual chances of dozing off (i.e. their usual sleep propensity) while engaged in each of eight different activities that involve differences in posture and environmental stimulation. Let us first consider total ESS scores, and then ESS item-scores, from the following four groups of subjects involving a total of 990 people, aged between 17 and 78 years. All have been the subjects of previous reports.

Group 1: patients with sleep disorders who had overnight polysomnography at Epworth Sleep Centre to diagnose their sleep disorders that included sleep-disordered breathing (ranging from simple snoring to severe obstructive sleep apnea), narcolepsy, idiopathic hypersomnia, etc. They answered the ESS, unaided, at their first sleep clinic presentation ($n = 240$, male/female = 176/74).

Group 2: workers from one regional industry in Victoria selected without regard to their sleep habits or sleep disorders. They anonymously answered a detailed sleep questionnaire and the ESS ($n = 324$, M/F = 271/67).

Group 3: undergraduate and graduate students at Swinburne University, Melbourne, mainly in the School of Life and Social Sciences, who anonymously answered a questionnaire about sleep habits and driving safety, including the ESS ($n = 322$, M/F = 157/155).

Group 4: fourth-year medical students at Monash University, Melbourne, who were all present at one particular teaching session and who were asked to answer the ESS anonymously, without regard to their sleep habits or sleep disorders ($n = 104$, M/F = 55/49).

One-way ANOVA demonstrated statistically significant differences in the total ESS scores between these groups ($F = 39.7$, d.f. = 986, $P < 0.001$) (Table 1). Post-hoc Scheffé tests showed that the total ESS scores for Group 1 (sleep clinic patients) were significantly higher than for Group 2 (industrial workers; $P = 0.01$), and those in turn were higher than for either Group 3 (university students; $P < 0.001$) or Group 4 (medical students; $P < 0.001$), which did not differ significantly ($P > 0.99$). These group differences were as expected, with the highest total ESS scores in patients diagnosed with sleep disorders known to cause "excessive daytime sleepiness". Many subjects among the industrial workers and the two groups of students would presumably have had normal sleep habits, but some would probably have had undiagnosed sleep disorders, including chronically insufficient sleep. The groups also differed in their ages and their gender distributions. The reasons for such group differences are not a concern for this analysis.

Within each group, item 5 ("lying down to rest in the afternoon when circumstances permit") had the highest scores. Scores for item 2 ("watching TV") were the next highest, followed by those for item 1 ("sitting and reading"). Item 6 ("sitting and talking to someone") and item 8 ("in a car, while stopped for a few minutes in the traffic") had the lowest scores within each group. Scores for item 4 were higher than for item 3 in all groups except Group 1. Scores for item 7 were intermediate in all groups. Thus, there was a fairly consistent pattern of differences between different SSPs, reported as ESS item-scores, but with minor differences for some people in particular situations.

Factorial ANOVA of the ESS item-scores within and between the four groups of subjects allowed us to investigate differences in sleep propensity when the same subjects were engaged in a variety of different activities, using the same method of measurement for each. This analysis enabled effect-sizes to be calculated for different

Table 1 The mean, with standard deviation in brackets, of Epworth Sleepiness Scale (ESS) item-scores and total ESS scores for four groups of subjects. The sleep-clinic patients had significantly higher total ESS scores than the industrial workers, who had higher total scores than either the university or medical students ($P < 0.01$)

		Group of subjects			
ESS item no.	Activity	1. Sleep clinic patients $n = 240$	2. Industrial workers $n = 324$	3. University students $n = 322$	4. Medical students $n = 104$
1.	Sitting and reading	1.6 (1.1)	1.2 (1.0)	1.04 (0.9)	0.9 (0.9)
2.	Watching TV	1.8 (1.1)	1.5(1.1)	1.1 (0.9)	1.2 (0.9)
3.	Sitting, public place	1.2 (1.1)	0.9 (1.0)	0.6 (0.8)	0.7 (0.8)
4.	Car passenger	1.2 (1.2)	1.0 (1.1)	0.9 (0.9)	0.8 (0.9)
5.	Lying down to rest	2.2 (1.0)	1.8 (1.2)	1.7 (1.0)	1.6 (1.0)
6.	Sitting and talking	0.7 (0.9)	0.3 (0.7)	0.1 (0.4)	0.1 (0.4)
7.	Sitting quietly after lunch	1.3 (1.1)	0.8 (1.0)	0.6 (0.8)	0.7 (0.8)
8.	In a car, stopped in traffic	0.6 (0.9)	0.4 (0.8)	0.1 (0.4)	0.1 (0.3)
Total ESS scores		10.5 (6.4)	7.8 (5.4)	6.1 (3.9)	5.9 (4.2)

Table 2 Factorial ANOVA of Epworth Sleepiness Sca item-scores in four groups of subjects

Effect	SS	D of F	MS	F	P	Partial eta-squared
ESS item	1468.2	7	209.74	248.20	<0.0001	0.180
Group of Ss	379.2	3	126.41	149.58	<0.0001	0.054
ESS item*Group	29.9	21	1.42	1.69	=0.026	0.004
Intercept	5799.7	1	5799.70	6863.02	<0.0001	0.465
Error	6679.4	7904	0.85			

12 SS, sums of squares. D of F, degrees of freedom. MS, mean square. F, variance ratio.

sources of variance among the ESS item-scores, using partial eta-squared (Table 2).

There were highly significant differences between the groups, across all ESS-items, consistent with the differences in their total ESS scores described above. These differences presumably reflected the presence or absence of sleep disorders in some subjects but not others. However, there were also significant differences between the sleep propensities in different situations, represented by different ESS item-scores within subjects. That is, there were consistent differences in sleepiness depending on what the subjects were doing at the time. Within the context of our earlier discussions, their SSPs differed significantly and generally in a consistent way. The effect-size for this source of variance was much larger than for the differences between groups (partial eta-squared = 0.180 *vs* 0.05). However, there was another small but statistically significant source of variance in ESS item-scores, represented by a significant group-situation interaction. The effects that some activities had on some subjects' SSPs were different from the fairly consistent effects seen in the majority of subjects.

In summary, factorial ANOVA enabled us to distinguish several statistically significant sources of variance in ESS item-scores, as follows:

1 A general effect across all activities/situations that presumably reflected the presence or absence of sleep disorders within subjects and which influenced their total ESS scores and, by inference, their ASPs.
2 A general effect of what subjects were doing at the time (ESS activities).
3 A group/situation specific effect of some particular activities in some subjects.

The concept of somnificity

Johns coined the term somnificity to explain the effects that different activities and situations have on sleep propensity.[26] Somnificity is defined as the general characteristic of a posture, activity and situation that reflects its capacity to facilitate sleep-onset in the majority of subjects. It is not a characteristic of people or their sleep disorders. This can be explained by considering further analysis of the above ESS item-scores for all 990 subjects together. In this analysis, the ESS item-scores were ranked within subjects, eliminating the differences in ASP between them. One-way Friedman (non-parametric) ANOVA showed highly significant overall differences between the ranks scored for each item ($P < 0.001$) (Table 3). Post-hoc analysis by Wilcoxon's *T*-tests for pairs of item-ranks showed that they formed an ordinal scale with six significantly different levels of somnificity (each $P < 0.01$). However, items 3 and 7 did not differ significantly, nor did items 6 and 8 ($P > 0.5$). This scale of somnificities for Australian subjects was very similar to that for a group of 614 subjects in the USA,[27] suggesting that the concept of somnificity has wide application. The important point is that there were consistencies in the way that sleep propensity was influenced by what people were doing at the time. Some such effects can probably be ascribed to differences of posture, such as "lying down" in item 5 versus "sitting" in items 1, 3, 6 and 7 and, by implication, also in items 2, 4 and 8. However, more subtle effects need to be explained, including the significantly higher somnificity for item 1 ("sitting and reading") than for item 7 ("sitting quietly after a lunch without alcohol").

Table 3 The mean and standard deviation of somnificities for each Epworth Sleepiness Scale item, based on the ranks of item-scores in 990 subjects. The somnificities differed significantly from each other, apart from item 3 which did not differ significantly from item 7, and item 6 which did not differ significantly from item 8. Thus, there were six significantly different levels of somnificity, shown here from highest to lowest levels on an ordinal scale

ESS Item No.	Activity	Somnificity Mean	SD
5	Lying down to rest in the afternoon when circumstances permit	6.5	1.7
2	Watching TV	5.7	1.7
1	Sitting & reading	5.5	1.6
4	As a passenger in a car for an hour without a break	4.6	1.8
3	Sitting, inactive in a public place	4.3	1.5
7	Sitting quietly after a lunch without alcohol	4.1	1.5
6	Sitting & talking to someone	2.8	1.2
8	In a car, while stopped for a few minutes in the traffic	2.8	1.2

Correlations between different SSPs within subjects

Considering the raw ESS item-scores (not their ranks) for all 990 subjects, we can calculate 28 Spearman correlation coefficients between pairs of item-scores within subjects. All 28 coefficients were statistically significant ($P < 0.001$), consistent with a general effect on all SSPs within the same subject, and hence on their ASPs. The mean the correlation coefficients was 0.45 and their range was 0.31–0.57, accounting for only about 20% of the variance involved (range, 10–33% for different items). For comparison, the mean test–retest reliability of individual ESS item-scores, repeated after a period of 5 months in the same subjects, has previously been reported as 0.82 ($n = 87$, $P < 0.001$), which explains about 65% of the variance involved.[28] It appears therefore that not only is a person's SSP in one situation often quite different from that in a different situation, reflecting differences in somnificity, but the different SSPs are likely to be only moderately correlated with one another. We cannot expect to make an accurate assessment of a person's sleepiness in one situation from a measurement of their sleepiness in a different situation. More generally, we cannot accurately measure a person's sleepiness without reference to their posture, activity and situation at the time.

The argument above was based on the analysis of ESS item-scores. These are subjective reports that are potentially inaccurate because of faulty recall and biased reporting. However, we can also examine the results of objective tests for two different SSPs, estimated from the mean sleep latencies in the MSLT and the MWT performed on the same subjects on the same days. Sangal *et al.* reported such results from 258 subjects[7] (Fig. 1).

The sleep latency data from Sangal *et al.* are not strictly comparable to ESS item-scores, but the two investigations share some features. They both refer to different SSPs measured within the same subjects, albeit in eight different situations in the ESS *versus* two in Sangal *et al.* The details of the Sangal *et al.* data were not available and his subjects were not separated into groups with different ASPs, which would enable factorial ANOVA to be performed, comparable to that with ESS item-scores. However, both investigations provided correlation coefficients between pairs of different SSPs, measured on their respective scales.

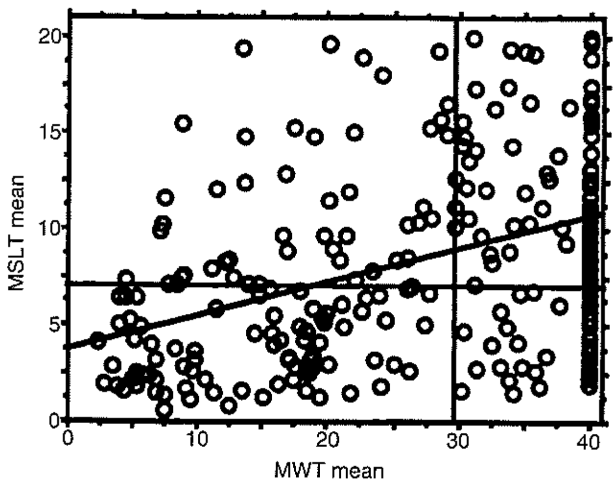

Figure 1 The mean sleep latencies in the Multiple Sleep Latency Test and Maintenance of Wakefulness Test (min) measured on the same day in 258 subjects with a variety of sleep disorders. The data points are divided into four quadrants according to the median score on each scale, and the overall regression line is shown ($r = 0.41$, $P < 0.001$) (after Sangal *et al.* 1992[7]).

The mean sleep latencies in the MWT were much longer than in the MSLT[7] (mean, 18 *vs* 11 min, respectively) which reflects the different somnificity of the test-situations. In the MSLT, the subject is lying down with eyes closed and the head, trunk and limbs fully supported by the bed and pillow(s).[6] By contrast, in the MWT the subject is "propped up", semi-reclining in bed, presumably with little support for the head, and with eyes open, intending to remain awake.[6] It should not be surprising that these differences in the MSLT and MWT test-situations can have a marked effect on the subject's sleep propensity at the time. In Sangal *et al.*, the sleep latencies in the two test-situations within the same subjects were only moderately correlated ($r = 0.41$, $n = 258$, $P < 0.001$),[7] and explained approximately 20% of the relevant variance. This is comparable to the mean of 28 correlation coefficients between 8 ESS item-scores in 990 subjects, which was 0.45 ($P < 0.001$). Nevertheless, for the majority (70%) of subjects in Sangal *et al.*, a relatively low or high sleep propensity in the MSLT was associated with a correspondingly low or high sleep propensity in the MWT (lower-left and upper-right quadrants of Fig. 1). The remaining 30% of subjects had discordant results, with a relatively high sleep propensity in one test but not in the other (top-left and bottom-right quadrants of Fig. 1). Some subjects fell asleep more quickly in the MWT, when asked to remain awake, than in the MSLT, when asked to fall asleep.

These results, from two well-established objective tests of sleep propensity each measuring a different SSP in terms of the same variable (the mean sleep latency in the particular test-situation), are entirely consistent with those derived from subjectively reported SSPs, measured by ESS item-scores. This suggests that measurements of particular SSPs, whether based on objective or subjective methods, reflect more than one source of variance. One source reflects a general characteristic in each subject, their ASP, which is presumably what most people think they are measuring as a single general characteristic of each subject when they perform those tests, for example, as a measure of the severity of a sleep disorder. However, another source of variance will reflect a general characteristic of what the subjects are doing at the time, its somnificity, the effects of which are largely predictable. A third source of variance is much less predictable. It reflects the responses of some people to particular activities and circumstances. These responses are both subject- and situation-specific and are presumably learned.

This third source of variance can explain why, in Sangal *et al.*, a minority of subjects fell asleep more quickly in the MWT than in the MSLT. It can also explain why some people, without any complaint of excessive daytime sleepiness, fall asleep in less than 5 min in the MSLT. It also explains why there are some minor differences in ESS item-scores between some groups of subjects that do reflect the somnificity of those activities, as described above. This third source of variance in SSPs is presumably intrinsic to the nature of sleepiness, and not due to the failure of particular methods of measurement. Its presence emphasizes the fact that sleepiness is not a single characteristic that can be measured for each person.

These conclusions are supported by the results of other investigations that have compared sleep latencies in the MSLT, and MWT and ESS scores in narcoleptic patients.[29] In addition, Bonnet *et al.* have extended such comparisons to sleep latencies in four different situations.[30] They found that the sleep latencies differed significantly between the various situations, but were only moderately intercorrelated within subjects. In summary, it appears that sleepiness cannot be measured accurately without reference to what the subject is doing at the time. The accuracy of all currently used objective and subjective methods for measuring sleepiness is limited by subject- and situation-specific interaction effects, which are largely unpredictable.

Need for a revised model of sleep and wakefulness

These conclusions provide an entirely new perspective on the measurement of sleepiness, and raise important issues about the currently accepted models for the control of sleep and wakefulness. For example, what is it about "sitting and reading" that gives it a significantly higher somnificity than either "sitting, inactive in a public place" or "sitting quietly after a lunch without alcohol"? Then again, why does "watching TV" have a significantly higher somnificity than "sitting and talking to someone"? Perhaps those activities tend to occur at different times of the day, involving different levels of Process-C (varying with the time of day) and Process-S (varying with the duration of prior wakefulness) that influence sleep propensity at those times. Johns has proposed an another explanation, by which most differences of somnificity are mediated by effects on the sensory nervous system, with all exterosensory and somatosensory inputs continually influencing the wake drive.[9,10,26] This hypothesis places particular emphasis on somatosensory inputs, especially from postural muscles, in addition to exteroceptive inputs from vision and non-visual light detection, etc., with additive effects on the central nervous system promoting wakefulness. Activities with a high somnificity would presumably induce less wake drive, particularly in that which Johns calls the secondary wake drive, by reducing sensory inputs and thereby facilitating sleep onset.

This suggests the need for revision of the currently accepted models of sleep and wakefulness. We need a new process, to be called Process-A, along with Process-C and Process-S [31] to explain sleepiness. The letter-A in Process-A stands for afferent, to emphasize the afferent nervous system as the basis of this process. Measurements of sensory inputs to the central nervous system under different circumstances, how they are integrated over periods of a few minutes, and how they influence the wake drive all require more research and more detailed elaboration than is possible here. Currently, there is no single, direct and objective measure of the overall Process-A. However, we can postulate the need for Process-A to explain what happens when we choose to lie down in a comfortable environment, close our eyes, stop talking, and cease making other voluntary movements. In that situation, the mean of normal sleep latencies is about 11 min, even at 10 AM after a night's sleep, as in the MSLT. By contrast, most people would be very unlikely to fall asleep if they were walking around at that time. In a different context, it is presumably because of the need to control Process-A that a constant routine is necessary for the laboratory investigation of Process-C and -S, with reduced light stimulation and controls over posture and movement. However, our understanding of sleepiness in real-life situations will remain limited until Process-A, or some other mechanism that allows for the influence of the sensory nervous system on sleepiness, becomes incorporated into our models of sleep and wakefulness.

Which test of sleepiness to use?

With so many different methods currently in use or having been proposed for measuring sleepiness, we might be forgiven for being confused about which test to use, and when. We can classify those methods according to whether sleepiness is measured as sleep propensity or the drowsy state, using either objective or subjective methods, and on different time-scales.

1. Methods based on measurements of sleep propensity
   - 1.1 Methods for measuring ISP (instantaneous sleep propensity)
     - 1.1.1 Objective methods (e.g. sleep latency for each nap in the MSLT or MWT)
     - 1.1.2 Subjective methods (none at present)
   - 1.2 Methods for measuring particular SSPs (situational sleep propensities)
     - 1.2.1 Objective methods (e.g. mean sleep latency in the MSLT or MWT)
     - 1.2.2 Subjective methods (e.g. ESS item-scores, answers to other questions, for example, about drowsy driving)
   - 1.3 Methods for measuring ASP (average sleep propensity)
     - 1.3.1 Objective methods (none at present)
     - 1.3.2 Subjective methods (e.g. total ESS scores)
2. Methods based on indicators of drowsiness
   - 2.1 Measurements made at particular times, perhaps repeatedly
     - 2.1.1 Objective methods (e.g. psychomotor performance tests such as PVT, EEG characteristics [alpha and theta waves])
     - 2.1.2 Subjective methods (e.g. KSS, SSS, visual analogue scales [VAS])
   - 2.2 Measurements of drowsiness made continuously
     - 2.2.1 Objective methods (e.g. monitoring EEG characteristics, eyelid movements by infrared reflectance oculography [Optalert]), video-camera images of the eyes

[PERCLOS], EOG (electrooculogram) recordings)

2.2.2 Subjective methods (none at present)

The choice of which of the currently available methods for measuring sleepiness to use will depend on the circumstances and purpose of those measurements. It is beyond the scope of this investigation to discuss the advantages and disadvantages of each method in a wide variety of different circumstances. However, some methods are cheap and easy to perform on many subjects (e.g. ESS), others are expensive and time-consuming (e.g. MSLT and MWT). If the intention is to measure a person's average level of daytime sleepiness in daily life (their ASP), then there is simply not a direct objective measure of that at present. It may be best to use some combination of methods and not rely on any one measure, whether based on objective or subjective methods.

It must be remembered that the MSLT fulfils two separate diagnostic functions, only one of which is to measure sleepiness. The other is to demonstrate the presence or absence of the early onset of REM-sleep, important for the diagnosis of narcolepsy. There is evidence that the MWT is a more sensitive and specific objective test of sleepiness than the MSLT.[5] This raises the possibility that both tests could be used, perhaps in combination on the same day, for diagnosing narcolepsy and related conditions. This would presumably provide a more accurate measure of sleepiness, based on two objectively measured SSPs rather than one, as well as information about REM-sleep latency.

CONCLUSIONS

The measurement of sleepiness is evidently much more complicated than previously thought. Different definitions and ways of thinking about sleepiness, and the use of different timescales and different methods for its measurement have created much confusion. The discussion here raises important new issues about sleepiness, the need for a new perspective, and the need to expand our models of sleep and wakefulness. Sleepiness, in the sense of sleep propensity, is very much influenced by what the subject is doing at the time, that is, the somnificity of that activity. In addition, a person's sleep propensity in one set of circumstances is usually only moderately correlated with their sleep propensity in a different set of circumstances, whether measured objectively or subjectively. These effects are presumably independent, at least to some extent, of Process-C (time of day) and Process-S (duration of prior wakefulness). To explain this, we require an entirely different process, Process-A, to be incorporated into our models of sleep and wakefulness.

Mean sleep latencies in the MSLT or the MWT, or total ESS scores, are all reasonably accurate for distinguishing normal subjects from narcoleptic patients, who by definition have abnormally high ASPs.[5] However, the ESS appears to be the most accurate of those three tests and the MSLT the least accurate. We now have a plausible explanation for that. The MSLT and the MWT each measures one SSP, and they are different SSPs. By contrast, total ESS scores are based on eight different SSPs, which give a better estimate of a person's ASP than any one SSP can. However, the activities described in the ESS were never intended to represent average daily routines. As with any subjective reports, ESS scores can potentially suffer from subjective bias, particularly if the subject's responses are likely to influence future events in a way that is perceived as adverse. For example, we should not expect unbiased responses to the ESS from someone who could lose their driver's license on the basis of those responses. Thus, the ESS cannot be viewed as a gold standard method for measuring sleepiness in the sense that the MSLT was previously thought to be. We need more research based on the new perspective to investigate objective, rather than subjective, measures that relate directly to Process-A. This will help explain the physiological basis for the scale of somnificity.

Potential conflict of interest statement

This investigation was not industry financed. However, Dr Johns is Director, Chief Scientist and share-holder of Optalert Pty Ltd, a research and development company that manufactures technology (Optalert™) for monitoring drowsiness in active people such as drivers.

REFERENCES

1. Johns M. Rethinking the assessment of sleepiness. *Sleep Med. Rev.* 1998; 2: 3–15.
2. Ohayon MM. From wakefulness to excessive sleepiness: what we know and still need to know. *Sleep Med. Rev.* 2008; 12: 129–41.
3. Diagnostic Classification Steering Committee. *ICSD-International Classification of of Sleep Disorders: Diagnostic and Coding Manual. Diagnostic Classification Steering Committee, Thorpy MJ, Chairman.* American Sleep Disorders Association: Rochester, MN, 1990.

4. Thorpy MJ. The clinical use of the Multiple Sleep Latency Test. The Standards of Practice Committee of the American Sleep Disorders Association. *Sleep* 1992; **15**: 268–76.
5. 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.* 2000; **9**: 5–11.
6. Arand D, Bonnet M, Hurwitz T, Mitler M, Rosa R, Sangal RB. The clinical use of the MSLT and MWT. *Sleep* 2005; **28**: 123–44.
7. 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.
8. Harrison Y, Horne JA. High sleepability without sleepiness. The ability to fall asleep rapidly without other signs of sleepiness. *Neurophysiol. Clin.* 1996; **26**: 15–20.
9. Johns MW. The subjective measurement of excessive daytime sleepiness. In: Pandi-Perumal SR, Verster JC, Monti JM, Lader M, Langer SZ, eds. *Sleep Disorders: Diagnosis and Therapeutics*. Informa Healthcare: London, 2008; 643–57.
10. Johns MW. What is excessive daytime sleepiness? In: Fulke P, Vaughan S, eds. *Sleep Deprivation: Causes, Effects and Treatment*. Nova Science Publishers: New York, 2009; 59–94.
11. Onions CT. *The Shorter Oxford English Dictionary*. Oxford University Press: Oxford, UK, 1959.
12. Roth T, Roehrs T, Carskadon MA, Dement WC. Daytime sleepiness and alertness. In: Kryger MH, Roth T, Dement WC, eds. *Principles Ad Practice of Sleep Medicine*, 1st edn. W.B. Saunders: Philadelphia, PA, 1989; 14–23.
13. Carskadon MA, Dement WC. The multiple sleep latency test: what does it measure? *Sleep* 1982; **5** (Suppl 2): S67–72.
14. Mitler MM, Gujavarty KS, Browman CP. Maintenance of wakefulness test: a polysomnographic technique for evaluation treatment efficacy in patients with excessive somnolence. *Electroencephalogr. Clin. Neurophysiol.* 1982; **53**: 658–61.
15. Åkerstedt T, Gillberg M. Subjective and objective sleepiness in the active individual. *Int. J. Neurosci.* 1990; **52**: 29–37.
16. Hoddes E, Zarcone V, Smythe H, Phillips R, Dement WC. Quantification of sleepiness: a new approach. *Psychophysiology* 1973; **10**: 431–6.
17. Williams HL, Granda AM, Jones RC, Lubin A, Armington JC. EEC-frequency and finger pulse volume as predictors of reaction time during sleep loss. *Electroencephalogr. Clin. Neurophysiol.* 1962; **14**: 64–70.
18. Wilhelm B, Giedke H, Ludtke H, Bittner E, Hofmann A, Wilhelm H. Daytime variations in central nervous system activation measured by a pupillographic sleepiness test. *J. Sleep Res.* 2001; **10**: 1–7.
19. Dinges DF, Powell JW. Microcomputer analysis of performance on a portable, simple visual RT task during sustained operations. *Behav. Res. Meth. Instr. Comp.* 1985; **17**: 652–5.
20. Dorrian J, Rogers NL, Dinges DF. Psychomotor vigilance performance: neurocognitive assay sensitive to sleep loss. In: Kushida CA, ed. *Sleep Deprivation: Clinical Issues, Pharmacology, and Sleep Loss Effects*. Marcel Dekker: New York, 2005; 39–70.
21. Balkin TJ, Bliese PD, Belenky G *et al.* Comparative utility of instruments for monitoring sleepiness-related performance decrements in the operational environment. *J. Sleep Res.* 2004; **13**: 219–27.
22. Dinges DF, Mallis MM, Maslin G, Powell JW. *Evaluation of Techniques for Ocular Measurement as an Index of Fatigue and the Basis for Alertness Management, in DOT HS 808 762*. U.S. Department of Transportation National Highway Traffic Safety Administration: Washington DC, 1998.
23. Johns MW, Tucker A, Chapman R, Crowley K, Michael NJ. Monitoring eye and eyelid movements by infrared reflectance oculography to measure drowsiness in drivers. *Somnologie* 2007; **11**: 234–42.
24. Johns MW, Chapman R, Crowley K, Tucker A. A new method for assessing the risks of drowsiness while driving. *Somnologie* 2008; **12**: 66–74.
25. Johns MW. A new method for measuring daytime sleepiness: the Epworth Sleepiness Scale. *Sleep* 1991; **14**: 540–5.
26. Johns MW. Sleep propensity varies with behaviour and the situation in which it is measured: the concept of somnificity. *J. Sleep Res.* 2002; **11**: 61–7.
27. Johns MW, Knutson KL. The somnificity of different activities described in the Epworth Sleepiness Scale. *Sleep* 2006; **29**: (Abstract Supplement), A30.
28. Johns MW. Reliability and factor analysis of the Epworth Sleepiness Scale. *Sleep* 1992; **15**: 376–81.
29. Sangal RB, Mitler MM, Sangal JAM, US Modafinil in Narcolepsy Multicenter Study, G. r. MSLT MWT and ESS: indices of sleepiness in 522 drug-free patients with narcolepsy. *Sleep Res.* 1997; **26**: 492.
30. Bonnet MH, Arand DL. Arousal components which differentiate the MWT from the MSLT. *Sleep* 2001; **24**: 441–7.
31. Borbély AA. A two process model of sleep regulation. *Hum. Neurobiol.* 1982; **1**: 195–204.