In: Sleep Deprivation: Causes, Effects and Treatment
Editors: Pedr Fulke and Sior Vaughan
ISBN: 978-1-60741-974-7

Chapter 2

What is Excessive Daytime Sleepiness?

Murray W Johns*

Optalert Pty Ltd, Melbourne, Australia, Epworth Sleep Centre, Melbourne, Australia, Sensory Neuroscience Laboratory, Swinburne University of Technology, Melbourne, Australia

Introduction

It is commonly recognized that sleep deprivation causes an increased tendency to fall asleep; that is, it causes an increase in the sleep-deprived person's sleep propensity. Their increased sleep propensity may reach a level that causes excessive daytime sleepiness (EDS). According to one current definition, EDS is "sleepiness in a situation when an individual would be expected to be awake and alert" [1] . EDS is not a disorder. It is a symptom that can have many different causes, not only sleep deprivation. EDS is a very common symptom among patients who present to sleep clinics around the world. Unfortunately, the concepts of "sleepiness" and EDS have not been well developed and the methods used for their measurement, including both subjective and objective methods, remain a matter of contention [2]. Before we can define what EDS is, we must be able to say what "sleepiness" is and how we can measure it.

What follows in this Chapter is an attempt to answer the question "what is excessive daytime sleepiness?" This requires an examination of the basic assumptions behind the current view of EDS, many of which are implicit rather than explicit. There is evidently a need for some new definitions and a broader frame of reference within which to consider "sleepiness" before we can define EDS. A new conceptual model of sleep and wakefulness is outlined that recognizes major influences on "sleepiness" derived from the combined effects of sensory inputs to the central nervous system [3]. They have not been included in other models. The methods that are commonly used, and several others that have been proposed for measuring "sleepiness" are then described within this new conceptual framework As others

* Contact address: Dr M.W.Johns, Optalert Pty Ltd, Suite 9, 150 Chestnut Street, Richmond, Melbourne, VIC, 3121 Australia, Phone: +61 3 9425 5000, Fax : +61 3 9425 5001, Email: mjohns@optalert.com

have commented, it is very difficult to discuss EDS meaningfully without a clear understanding of what it is and how it can be measured [2]. Our first task therefore is to clarify those issues. We shall not be concerned here with the epidemiology of EDS or the disorders that cause it. The contents of this Chapter are based largely, but not exclusively, on the published work of the present author. This involves several terms and concepts that are new to sleep medicine, some of which may still be considered hypothetical.

Let us begin by considering the common English language definitions of words used in this context, and then consider how some of them have taken on different meanings in recent times. By a long-standing dictionary definition, the adjective *sleepy* means "inclined to sleep, having difficulty in keeping awake, drowsy, somnolent" [4]. The noun, *sleepiness*, therefore means the state of being *sleepy*. The definition of the adjective *drowsy* is virtually the same as for *sleepy* – "inclined to sleep, heavy with sleepiness, half asleep, dozing". By contrast, *fatigue* is defined as "weariness resulting from bodily or mental exertion", and *tiredness* is defined as being "fatigued, weary". That is, *sleepiness* is synonymous with *drowsiness*, and *fatigue* is synonymous with *tiredness* according to the common usage of the words.

In the 1980's, a new meaning of the word *sleepiness* was adopted by some sleep researchers - a "physiological need state that leads to an increased tendency to sleep" [5] or as "a physiological drive usually resulting from sleep deprivation" [6]. This loosely equated *sleepiness* with what might now be called the *sleep drive*, a measure of the strength of activity in the neuronal system within the central nervous system that directly promotes sleep as opposed to wakefulness. However, this approach made it necessary for another concept to be introduced, that of *masking* [5]. It is self-evident that we can usually avoid falling asleep by remaining active and, in particularly, by not lying down. This led to the idea that "physiological sleepiness may not necessarily be manifest" [5]. It was stated that "heavy meals, warm rooms, boring lectures, and the monotony of long-distance driving unmask the presence of physiological sleepiness but do not cause it" [5]. That is, *masking* would not affect *physiological sleepiness,* but it would prevent sleep onset as an expression of *manifest sleepiness*. These concepts were never sufficiently developed to be of practical use, but their on-going influence is pervasive.

In recent years, sleep researchers and clinicians have also used the word *sleepiness* to mean *sleep propensity,* which is quite different from the common language meaning of the word *sleepiness* as a state. The current concept of EDS appears to have arisen from this use of the word *sleepiness* to mean *sleep propensity*. The common meaning of the word *excessive* is clear enough, "being greater in amount or degree than is usual, necessary or right", and *daytime* clearly means "in the hours of daylight". What appears to be meant by the term *excessive daytime sleepiness* or EDS might be defined as follows: **EDS is a symptom arising at any time from an excessive propensity to become drowsy or to fall asleep, when the intention and expectation is to remain awake and alert at the time.** Several implications and corollaries arise from this and other similar definitions:

- The symptom of EDS usually involves the subject involuntarily entering the drowsy state, contrary to intentions, whether or not drowsiness progresses to sleep.

- The symptom of EDS is most obvious in situations of low somnificity (see below), but sleep propensity may be relatively increased in all situations, even without a high level of drowsiness being experienced (see below).
- Use of the word *drowsiness* avoids confusion surrounding the different meanings of the word *sleepiness*. However, the term *sleep propensity* is used as a dimension of *sleepiness*.
- EDS is not synonymous with high "sleepability", a term which refers to the ability or speed of falling asleep voluntarily [7],
- EDS is generally considered to be undesirable or inappropriate, but it is also associated with increased health and safety risks because of "performance failures" e.g. the risk of crashing while driving.
- For people who remain active at night (e.g. working on night-shift), EDS can be more of a problem at night than during the daytime, in which case the term *excessive daytime sleepiness* is somewhat misleading.
- EDS does not by itself cause "hypersomnia" or unusually prolonged sleep duration
- EDS can be a long-term and continuing problem in daily life, or it may arise only occasionally and be short-lived. It should be considered on different time-scales.

Difficulties with Current Models of Sleep and Wakefulness

Johns [3, 8] has previously drawn attention to inadequacies of the current models of sleep and wakefulness, and how they limit our understanding of "sleepiness" and EDS. The currently accepted models are based mainly on Process-C (a function of the time of day and the phase of a circadian rhythm) and Process-S (a function of the duration of prior wakefulness) as described by Borbély et al. in 1982 [9]. A third process (Process-W) has been proposed by Folkard and Åkerstedt (1987) [10] to explain "sleep inertia" during the first 20 or 30 minutes after waking up from a period of sleep.

These processes are undoubtedly important, especially Processes C and S, but they are not sufficient to explain much of what obviously happens in real life. It is self-evident that we are much more likely to doze off when lying down with eyes closed than when standing up with our yes open, no matter what time of day it is or how long we have been awake. Nor do the currently accepted models explain how interaction with other people can affect *sleep propensity*. We are more likely to doze off when sitting alone than when sitting and talking to someone, even in the same physical environment with its stimulation due to hot/cold, light intensity and visual input [11-13]. None of the currently accepted models explains these observations. Some previous discussions of EDS have mentioned that "sleepiness" is affected by the level of environmental stimulation at the time, providing a context within which all such measurements must be considered [14]. In 1990 the International Classification of Sleep Disorders defined "mild", "moderate" and "severe sleepiness" in terms of the different activities in which the subject might fall asleep inadvertently [15]. Those activities were thought to differ in the levels of physical exertion and the attention required to engage in them. The subject's posture was not mentioned explicitly as a characteristic of those activities. The effects on "sleepiness" of a subject's posture, cognitive state and the environmental sources of sensory stimulation at the time have only occasionally been measured directly in the laboratory [16-18]. However, that does not justify their omission from a model of sleep and wakefulness, and their absence limits the usefulness of the currently accepted models for understanding "sleepiness" and EDS.

Johns [3, 8] has proposed a new conceptual model of sleep and wakefulness in which there is a major influence on *sleep propensity* from the integrated effects of all sensory inputs to the central nervous system. That includes extero-ceptive inputs from the environment, (visual input, the direct effect of light, noise, hot and cold sensation, etc) as well as enter-oceptive sensory inputs from within the subject's body and brain. The latter arise from many sources, such as the afferent nerves associated with stretch receptors and spindles in muscles and joint capsules that are activated by the tonic activity of postural muscles and the phasic activity of muscles and joints during body movements, as well as from vestibular inputs from the ears, baroreceptor inputs from stretch receptors in the aortic arch, and central inputs related to ongoing cognitive and affective mental activity. This new conceptual model invokes separate neuronal systems within the central nervous system that provide two drives, a wake drive and a sleep drive, that interact with each other by mutual inhibition. Other researchers have described some neuroanatomical and neurophysiological characteristics of those systems [19-23]. However, in the Johns model, the wake drive is divided into two components – a primary wake drive, which is the equivalent to Process-C of Borbély et al., [9] derived from spontaneous activity in the suprachiasmatic nucleus with its endogenous circadian rhythm, and a secondary wake drive derived from the integration of all entero- and extero-ceptive sensory inputs to the central nervous system, especially in the thalamus with its widespread projections to the cerebral cortex, hypothalamus and ascending reticular activating system. It is suggested that changes in this secondary wake-drive are a major factor, indeed a controlling factor, in determining whether we are awake or asleep at any particular time. There is no need for Process-W in this model, in which "sleep inertia" is a function of the inhibitory interaction between the sleep-drive and the wake-drive. After wakening, it takes some time for the wake-drive to inhibit the sleep-drive adequately. The Johns model is therefore unique in several ways – it recognizes the secondary wake-drive as a separate functional entity to be distinguished from the primary-wake drive, and it emphasizes the role of all entero-ceptive as well as extero-ceptive sensory inputs as a major influence that activates the secondary wake-drive.

Within this new conceptual framework, we will become drowsy and perhaps fall asleep whenever our primary and secondary wake-drives, combined, are not strong enough to counter and inhibit our sleep-drive sufficiently so as to maintain alert wakefulness, as they usually do for about 16 or 17 hours of the day for most people. We can choose to fall asleep purposely at almost any time by reducing our secondary wake-drive voluntarily by a process Johns calls *sleepening* [12]. This involves the active choice of lying down in a comfortable place, typically on a bed in a suitably warm, dark and quiet bedroom, and then actively closing our eyes to further reduce the effect of light and of visual attention, relaxing our postural muscles, and ceasing voluntary movements. As part of our usual sleep habits, we choose to begin *sleepening* at about the same time each night. By then we may have been awake for about 17 or 18 hours, and our primary wake-drive would have begun to decrease as a result of its circadian rhythm, having reached its highest level at about 7 pm. However, it is largely by controlling our secondary wake-drive, by voluntary *sleepening,* that we fall asleep when we choose to. The word *sleepening* was originally coined in the late 19th century by William Gowers, a prominent neurologist in the UK, to mean the process of falling asleep, but it was never widely adopted. Johns has narrowed the meaning of the word to the process that is under voluntary control, at least initially, by which we reduce our secondary wake-drive to facilitate sleep onset.

As a practical example of the importance of these influences on *sleep propensity*, consider what might happen when we choose voluntarily to stay awake all night, which most of us can do without much difficulty. By 5 am, the sleep-drive (Process-S) would be relatively high because we would have been awake for about 22 hours. The primary wake-drive (Process-C) would be relatively low because of the nadir in its circadian rhythm. Both of these changes would increase our sleep propensity. However, we can overcome this voluntarily and remain awake by increasing our secondary wake drive, especially by increasing the activity of postural muscles by standing up (increasing entero-ceptive sensory inputs).This does not entail any increase in the levels of environmental stimulation (extero-cetive sensory inputs) at the time. Simply to sit down and close our eyes under those circumstances greatly reduces our secondary wake-drive and increases our sleep propensity to the point that we may well fall asleep involuntarily. However, there is a psychophysiological cost in maintaining our secondary wake-drive to remain awake all night, which is manifested as *fatigue* – muscular fatigue because of the prolonged activity of postural and other muscles, and mental fatigue because focused attention and vigilance have been maintained for longer than usual. With sleep deprivation, the subjective feelings of fatigue and dysphoria may be more obvious than the awareness of drowsiness.

According to this model, sleep-onset insomnia would arise whenever *sleepening* was not successful, for whatever reason [24]. Conversely, we would fall asleep inadvertently whenever we failed to maintain our secondary wake-drive at high enough levels, in the absence of *sleepening*. Thus, EDS may be caused by several different mechanisms. By measuring *sleep propensity* alone we cannot distinguish which mechanisms are involved in particular cases:

- a relatively high sleep-drive,
- a relatively low primary wake- drive,
- a relatively low secondary wake-drive,
- reduced inhibition of the sleep drive by the wake drive,
- some combination of the above.

Johns [12, 13] has combined the effects of extero- and entero-ceptive sensory influences on *sleep propensity* in a new variable, the *somnificity* of each activity, with its typical posture, levels of physical and mental activity and of environmental stimulation at the time. *Somnificity* is a measure of the relative capacity of an activity and situation to induce drowsiness in the majority of people. *Somnificity* is not a characteristic of individual subjects, or of their sleep disorders or levels of "sleepiness". It is a measure of the effect that particular activities have on the secondary wake-drive in the majority of people. There are potentially as many different *somnificities* as there are different combinations of posture and activity for people to engage in. An ordinal scale of several different *somnificities* has been described for some postures and activities that are commonly met in daily life, such as those described briefly in the Epworth Sleepiness Scale (ESS) [12]. There is widespread consistency in the relative *somnificities* of those activities, regardless of differences in the levels of *sleep propensity* between subjects [13]. The symptom of EDS is most likely to occur when we are engaged in activities that have a high *somnificity*, i.e. activities that have an innate tendency to reduce our secondary wake-drive and consequently to increase our *sleep propensity* involuntarily at the time. By contrast, when we choose to fall asleep, our intentional *sleepening* activities put us in a situation with a high *somnificity*.

Measuring Sleep Propensity on Different Time-Scales

To further increase our understanding of the nature of "sleepiness", EDS, and the various methods for measuring it, Johns [25, 8] has proposed several different categories of *sleep propensity* based on different time-scales. These categories do not refer to the severity of *sleep propensity*.

- ***Instantaneous sleep propensity (ISP):*** a subject's *sleep propensity* at some particular time, whatever the circumstances. The ISP can vary rapidly over a few seconds with a change of posture or physical activity, and more slowly with the time of day and the duration of prior wakefulness. The ISP reflects the subject's posture and activity during the preceding few minutes as well as at the actual time, presumably because of the nature of the integrator of sensory inputs. Mild exertion, such as walking for 5 minutes instead of sitting and watching TV, reduces the ISP for the next few minutes [26]. A subject's ISP can be measured either by how long it takes them to fall asleep, or their probability of falling asleep at the time. The latter can be estimated indirectly by measuring their level of *drowsiness* at the time - the drowsier the subject, the greater the probability of them falling asleep. This requires a scale upon which to measure different levels of *drowsiness* (see below).

- ***Situational sleep propensity (SSP):*** a subject's usual *sleep propensity* when measured in the same situation repeatedly. For example, one's usual *sleep propensity* when sitting and watching television in the evening, intending to stay awake. There are as many different SSPs as there are different situations in which to measure *sleep propensity*. Different SSPs within the same subject are only moderately correlated (r = approx 0.4), whether measured objectively or subjectively (see below).

- ***Average sleep propensity(ASP):*** this is a hypothetical construct based on a subjects' general level of *sleep propensity* across the whole range of different activities they engage in as part of their daily lives [11]. This is a reflection of many different SSPs and ISPs combined, depending on the nature of the person's activities engaged in during an average day. The person's ASP increases with the onset of a sleep disorder such as obstructive sleep apnea or narcolepsy, and decreases again after successful treatment of such a disorder [27].

Methods for Measuring Sleepiness

Much of the current confusion about the best method(s) for measuring "sleepiness" seems to arise from the lack of a clear understanding about what "sleepiness" is and what each test is measuring. That includes a lack of understanding of how drowsiness and sleep propensity are measured by different tests and on different time-scales, the difficulty in extrapolating from one time-frame to another or from one test-situation to another, and the inaccuracies that are thereby introduced. To illustrate this, we shall now consider several different methods for measuring "sleepiness", some based on variables measured objectively, others subjectively.

Multiple Sleep Latency Test (MSLT)

This was the first standardized test of its kind [28]. It is performed in a sleep laboratory, in a warm, darkened, and quiet bedroom in which the subject is asked to lie down and try to fall asleep while their EEG/EOG/EMG are monitored by electrodes attached to the head and neck [29]. It has been assumed by some that this test-situation does not involve any alerting stimuli [28], but that may not be so with the attachment of electrodes and the sense of having to pass a test by falling asleep quickly. The subject's ISP at the time is measured as the sleep latency (SL), how long it takes them to fall asleep after they have settled own and the lights have been switched off in the room. If the subject does not fall asleep within 20 minutes the attempt is interrupted and a SL of 20 min is assumed. Each subject has 4 or 5 such nap opportunities two hours apart during the day, starting at 10 am. The mean SL is calculated for the whole test, for which the reference range of normal values is 1.8 – 19.0 minutes (mean = 10.4 +/- 4.3 SD minutes). The test takes all day to perform, requires the presence of a sleep technician, and is expensive. The MSLT has been externally validated and is reliable in the test-retest sense [29].

Within the conceptual framework suggested here, the SL for each nap in the MSLT gives an objective measure of the subject's ISP at the time. The mean SL for the whole test measures a slightly different characteristic - one particular SSP, the MSLT-SSP. The SLs measured at two-hourly intervals within the same subject and on the same day vary somewhat, randomly more than consistently, but they are moderately correlated with one another ($r = 0.61$, $n = 258$, $p<0.001$) [30]. Many people believe that the mean SL in the MSLT is the gold standard for measuring a person's "sleepiness". This belief was promoted for many years by the American Academy of Sleep Medicine [31], but that has now changed [1]. However, it is still assumed by many that the MSLT can provide an accurate measure of a subject's sleep propensity in situations other than the MSLT and, more generally, their sleep propensity in daily life (the ASP as we have called it). Nonetheless, the MSLT-SSP does appear to be moderately and significantly correlated with other SSPs, so it can provide one estimate of a subject's ASP, to be used as one piece of evidence when assessing a person's ASP and EDS [1]. Currently there is no gold standard objective test of ASP. Such a test might reasonably be called the holy grail of sleep medicine. The MSLT remains very important for other reasons – for making the diagnosis of narcolepsy, based on the early appearance of REM-sleep in one or more naps, and for comparing the effects of drugs on sleepiness, or comparing multiple measurements of sleepiness for some other reason in the same situation at different times [1].

Maintenance of Wakefulness Test (MWT)

This test is also performed in a sleep laboratory, in a similar environment and with the same electrodes attached as in the MSLT, and with similar high cost [32]. The main difference is that the subject sits up in bed with their back and head partially supported by pillows, and they are asked to stay awake rather than fall asleep during four periods of 40 minutes, two hours apart during the day. The MWT-SL is measured objectively for each of those recording periods and, if the subject does not fall asleep, a SL of 40 minutes is assumed. The mean SL is calculated for the whole test, for which the normal range is 8-40 minutes (mean = 30.4 +/- 11.2 minutes) [29]. The MWT-SL has been externally validated and shown to be reliable [1].

However, it is claimed that the MWT is a test "to evaluate alertness" [29], and herein lies another source of confusion. Both the MSLT and the MWT measure the same variable, the time taken to fall asleep (SL), but under different circumstances. It does not seem appropriate for the same variable to be interpreted as a measure of *sleep propensity* in one test and as a measure of *alertness* in another. Within the context of the terminology used here, both tests are measuring different SSPs. The longer SL in the MWT (mean = 30.4 minutes) than in the MSLT (mean = 10.4 minutes) is a reflection of the different *somnificities* of the two test-situations. In the MSLT, the subject's head and neck are completely supported by one or more pillows, the trunk and limbs are supported by a mattress, and the subject is asked to close his/her eyes so that visual input is blocked. That overall posture is maintained without the need for any postural muscles activity. Consequently, the level of entero-ceptive sensory inputs would usually be low in that situation. In the quite, dark and comfortably warm bedroom in which the MSLT is performed, extero-ceptive inputs would also be low but not negligible because of the electrode attachments etc. By contrast, in the MWT, the subject's head and neck are less supported than in the MSLT, and that posture requires tonic postural muscle activity in the neck and back muscles to be maintained, albeit unconsciously. In addition, the subject's eyes are mostly open, except when they close involuntarily with sleep onset, so visual inputs would provide further sensory stimulation during the MWT. Thus, the combined effects of extero-ceptive and entero-ceptive sensory inputs would usually be higher during the MWT than the MSLT. That would explain the difference in their SSPs, as reflected in their respective mean SLs.

However, there is another complicating factor. Some subjects, paradoxically, fall asleep more quickly in the MWT than they do in the MSLT [30], contrary to expectations based on the somnificities of the test-situations. This may be because of "paradoxical intention", in which the symbolic meaning of the test for some subjects, and the perceived threat of failing the test influences their secondary wake-drive at the time. For this to happen, the secondary wake-drive must be influenced by learned associations and their emotional concomitants, in addition to the physical effects of the sensory inputs at the time. In an entirely different context, learned associations may increase the secondary wake-drive in patients suffering from psychophysiological insomnia, who can readily fall asleep unintentionally in a chair while watching TV, but who have much more trouble when they go to bed and try to fall asleep [15].

Epworth Sleepiness Scale

The Epworth Sleepiness Scale (ESS) was introduced in 1991 [33] and is now the most commonly used methodor assessing a person's ASP subjectively. It is based on retrospective reports of *dozing behaviour* when engaged in a variety of activities and environmental situations that are commonly experienced in daily life. The term *dozing behaviour* requires some explanation. *Drowsiness* is characterized, among other things, by intermittent lack of awareness of the here-and-now [34, 35]. There is also inhibition of tonic and phasic muscle activity [36]. This often becomes apparent first in the muscles that keep the eyelids open during wakefulness (levator palpebrae superioris), and other muscles that contract to close the eyelids during blinks and longer eyelid closures (orbicularis oculi). As an indication of the reduction in the force of contraction of those muscles in the drowsy state, the velocity of eyelid movements during blinks is reduced, the eyelids take longer to close and reopen during blinks, they tend to remain closed for longer, and the upper eyelids droop [37-39]. If we doze off while sitting with our head unsupported, the muscles that hold our head erect when we are awake begin to have their tonic activity inhibited, and this eventually allows the head to drop forward. That movement often rouses us briefly, and we then become aware of having just dozed off, without having being aware of the beginning of the nodding movement or the level of drowsiness at that time.

The ESS is a simple self-administered questionnaire that asks subjects to rate on a scale of 0 to 3 their usual chances of dozing off in each of eight different situations "in recent times" (0 = "would never doze " and 3 = "a high chance of dozing") (Fig 1).

Epworth Sleepiness Scale

Name: ______________________________ Today's date: __________

Your age (Yrs): ____________ Your sex (Male = M, Female = F): _______

How likely are you to doze off or fall asleep in the following situations, in contrast to feeling just tired?

This refers to your usual way of life in recent times.

Even if you haven't done some of these things recently try to work out how they would have affected you.

Use the following scale to choose the **most appropriate number** for each situation:

0 = would **never** doze
1 = **slight chance** of dozing
2 = **moderate chance** of dozing
3 = **high chance** of dozing

It is important that you answer each question as best you can.

Situation	Chance of Dozing (0-3)
Sitting and reading	___
Watching TV	___
Sitting, inactive in a public place (e.g. a theatre or a meeting)	___
As a passenger in a car for an hour without a break	___
Lying down to rest in the afternoon when circumstances permit	___
Sitting and talking to someone	___
Sitting quietly after a lunch without alcohol	___
In a car, while stopped for a few minutes in the traffic	___

THANK YOU FOR YOUR COOPERATION

Figure 1: Epworth Sleepiness Scale (modified from Johns [33])

The total ESS score, ranging from 0 to 24, is the sum of 8 item-scores, each between 0 and 3. The ESS does not measure *subjective sleepiness* as some people have thought [40] because it does not ask about subjective feelings associated with different levels of *drowsiness*. It asks only for the subject's retrospective recall of dozing behaviour, from which is inferred their usual *sleep propensity* (SSP) in each specified situation. That the ESS refers to observable behaviour rather than subjective feelings is supported by the close relationship ($r = 0.74$, $n = 50$, $p<0.001$) between each subject's ESS score and that reported independently about the subject by his/her partner [11]. That is comparable to the test-retest reliability of ESS scores when repeated by the same subject after a few months ($r = 0.81$, $n = 87$, $p<0.001$) [41].

The situations were chosen on a priori grounds to vary in what Johns initially called their soporific nature [33], and later their *somnificity* [12, 13]. The higher the ESS score, the more likely the subject is to doze off in situations of low *somnificity*, i.e. the subject has a higher sleep propensity in those situations than other people have. A large epidemiological study of identical and non-identical twins who answered the ESS independently found that about 40% of the variance in total ESS scores between subjects represented a genetically determined trait, reflected in long-term differences in ASP that are partially inherited [42].

When choosing the questions to be included in the ESS, it was necessary to refer only those activities that most people would encounter in their daily lives, not necessarily very often, but often enough for them to form an estimate of their *dozing behaviour* in each. This precluded asking about *dozing behaviour* in some situations that might otherwise be of special interest to investigators, such as dozing while driving a car, but which some subjects do not engage in at all. That is why the descriptor for question 8 is "in a car, while stopped for a few minutes in the traffic". It does not specify as a driver or a passenger because it must allow for those who do not drive. If one such question remained unanswered, the total ESS score can not be calculated. Interpolation of scores is not possible. The ESS has been translated into many languages, some published but many more unpublished. The 1991 version of the ESS was modified in 1997 by the addition of an extra sentence of instructions- "It is important that you answer each question as best you can". This greatly decreased the proportion of missing ESS scores, to the order of 1%. Unfortunately, some people have apparently used the ESS without the standardized instructions to respondents [29], which invalidates the questionnaire.

The external validity of the ESS has been demonstrated in several ways, e.g. by its high sensitivity (93.5 to 97%) and specificity (100 %) for distinguishing narcoleptics, who have EDS, from normal subjects who do not [43, 44]; by the change in ESS scores after CPAP treatment for obstructive sleep apnea syndrome, typically reducing scores by 5 or more units [41, 45]; and by the change in ESS scores after the treatment of narcoleptics with Modafinil as compared with placebo [46]. Total ESS scores have a high test-retest reliability over a period of several months, $r = 0.81$ ($n = 87$, $p<0.001$) [41]. The distinctive nature of what the ESS measures is reflected in the fact that total ESS scores are only weakly correlated with subjective measures of *drowsiness/fatigue* at a particular time derived from the SSS, KSS or a VAS (see below). ESS scores are not highly correlated with scores on other scales measuring the longer-term state of *fatigue,* such as the Fatigue Severity Scale (FSS) e.g. $r = 0.33$, $n = 489$, $p<0.001$) [47].

There is good evidence for the internal consistency, unitary structure and consistent hierarchical item structure of the ESS as a sum-scale. That evidence has come from several quarters, including factor analysis that has usually, but not always, found only one factor that includes all ESS items, but with somewhat variable weightings in different groups of subjects [11, 48, 49]. Other evidence has come from item analysis and Cronbach's alpha statistic, alpha varying between 0.73 and 0.88 in different groups of subjects [41, 49]. Item Response Theory has also been applied to ESS item-scores [49, 50], which has provided the most detailed evidence yet about the internal structure of the ESS.

The mean of ESS scores in normal people, who do not have a sleep disorder including frequent snoring, is 4.6 +/- 2.8 (SD) [51]. The reference range of normal ESS scores is therefore defined as 0 to 10, which is the mean +/- 2SD. This also coincides with the 2.5th and 97.5th percentiles. Similar results have been reported from Italy (4.4 +/- 2.8) [52] and England (4.5 +/- 3.3) [43]. The ESS has been used in many surveys, often in languages other than English, to estimate the proportions of various groups that have EDS [53, 54]. Between about 7% and 18% of the adult population in different countries, selected without reference to their age, gender or sleep habits, have ESS scores >10. However, the ESS is not a gold standard. Like all methods based on subjective reports, it can sometimes be inaccurate because some subjects may not be able to distinguish their own behavioural states retrospectively or may give biased reports, particularly when major decisions might be made on the basis of the results (e.g. about granting or withholding a driver's license).

Karolinska Sleepiness Scale

The Karolinska Sleepiness Scale (KSS) is a 9-point Likert scale based on a self-reported, subjective assessment of the subject's level of *drowsiness* at the time [55]. In its original format the KSS had word descriptors only for scores of 1, 3, 5, 7 and 9. Those descriptors varied from 1= "very alert" to 9="very sleepy, fighting sleep, an effort to keep awake". However, additional descriptors were later added for all scores, as shown in Fig 2 [56].

The KSS is assumed to be an ordinal scale with a unitary structure, although that has not been confirmed. The KSS has been used widely, particularly for describing changes over time within subjects [57-59]. KSS scores may require standardization to control for differences between subjects [57, 58]. The changes observed in the EEG/EOG with *drowsiness* do not usually appear until KSS scores reach 7 and higher [57]. Lower KSS scores (<5) may reflect differences in the subjective awareness of *fatigue* as much, or more than, levels of *drowsiness*. Higher KSS scores (7+) may refer more specifically to the state of *drowsiness* because the subject may then have experienced involuntary dozing behaviour, with "lapsing" episodes and brief loss of awareness of the here-and-now, followed by arousal and the return of awareness, including some awareness of recently having dozed off.

Karolinska Sleepiness Scale
Here are some descriptors about how alert or sleepy you might be feeling right now. Please read them carefully and CIRCLE the number that best corresponds to the statement describing how you feel at the moment.
1. Extremely alert 2. Very alert 3. Alert 4. Rather alert 5. Neither alert nor sleepy 6. Some sighs of sleepiness 7. Sleepy, but no difficulty remaining awake 8. Sleepy, some effort to keep alert 9. Extremely sleepy, fighting sleep

Figure 2. A modified version of the KSS (after Reyner and Horne [56])

Stanford Sleepiness Scale

The Stanford Sleepiness Scale (SSS) is also a Likert scale based on a series of statements, numbered 1 to 7, that range from "feeling active, vital, alert, wide awake" to "almost in reverie, cannot stay awake, sleep onset appears imminent" (Fig 3) [60]. The different statements are assumed to represent an ordinal scale that reflects different positions along a continuum of states between alert wakefulness, through progressively deeper levels of *drowsiness*. Respondents are asked to choose which statement most accurately describes how they feel at the time

The SSS has been widely used, particularly for studying the effects of sleep deprivation [61], sleep fragmentation [62], and circadian rhythms [63]. However, scores on the SSS are not closely related to SLs measured a few minutes later in MSLTs [64]. Another problem with the SSS is that factor analysis suggests it is not a unitary scale [65]. It seems to measure "sleepiness" in a way that confounds *drowsiness* and *fatigue*. There are several poorly defined words in the SSS, such as "responsive", "foggy", "vital", and "woozy", that are not consistently associated with the drowsy state. The SSS can be used to measure changes in "sleepiness" within subjects over time, particularly over periods of hours and days, but the KSS or a VAS are preferred. Scores on the SSS often require standardization (e.g. to z-scores) to remove differences between subjects. The SSS cannot provide an overall measure of a subject's ASP or EDS in daily life.

Stanford Sleepiness Scale	
Circle the ONE number that best describes your level of alertness or sleepiness right now.	
1.	Feeling active, vital, alert, wide awake.
2.	Functioning at a high level but not at peak, able to concentrate.
3.	Relaxed, awake but not fully alert, responsive
4.	A little foggy, let down.
5.	Foggy, beginning to lose track, difficulty staying awake.
6.	Sleepy, prefer to lie down, woozy.
7.	Almost in reverie, cannot stay awake, sleep onset appears imminent.

Figure 3: Stanford Sleepiness Scale (after Hoddes et al. [60])

Visual analogue scales

A Visual Analogue Scale (VAS) is typically a horizontal line 100 mm long across a page, with a word at each end representing two extremes, for example, "very sleepy" and "very alert" [66]. The subject is asked to place a mark at that point on the line that represents his current state along that continuum. The VAS score is the distance (measured in mm) between the subject's mark and one or other end of the line. Scores on such a VAS respond to time-of-day effects and to sleep deprivation at least as well as and perhaps better than the SSS [66]. Several different VASs can be used in parallel if required, with different pairs of words at the extremes for each scale [67-69]. Clearly, the choice of words is critical so that each VAS represents a single dimension of variation between two extremes. VAS scores often have to be standardised (e.g. to z-scores) to allow for the differences between subjects. VAS scores have not been used to diagnose or quantify EDS in daily life, but presumably could be used, for example, to quantify the degree of "difficulty" that EDS causes.

Sleep-Wake Activity Inventory

The Sleep-Wake Activity Inventory (SWAI) has 35 items with several subscales enquiring about sleep habits and sleep disorders, of which one subscale is called "daytime sleepiness" [70]. It asks questions about how frequently the subject dozes off in different situations, and how frequently they have "difficulty staying alert throughout the day". These questions are not the same as those in the ESS that specifically do not ask about the frequency with which the subject dozes off in different situations. Answers to the SWAI questions would dependent on how often the subject was in those situations, which may have nothing to do with their "sleepiness". That is a disadvantage with the SWAI, but it can nonetheless assess several different aspects of a person's sleep-wake habits at the same time. It has not been widely used as a measure of "sleepiness".

Adjective Check-lists about Mood and Feelings

The Profile of Mood States (POMS) is a questionnaire that has occasionally been used to investigate "sleepiness" [69]. It comprises a list of 65 adjectives, and the subjects are asked to select or not select each according to how they are feeling at the time. One dimension of the POMS relates to "vigor–activity" and another to "fatigue–inertia", scores on which subscales change with sleep deprivation [69]. In general, these POMS subscales are of limited use because other scales, such as a VAS of "alertness–drowsiness", appear to be better (see below).

Responses to a Single Question About "Sleepiness"

When enquiring about "sleepiness" within a particular context, it may be advisable to assess the subject's specific SSP of relevance, such as the propensity to doze off while driving, rather than their more general ASP. For example, one question may ask "how many times have you dozed off at the wheel while driving during the last year?". Additional questions may be required, such as "how many times in the last year has the vehicle you were driving been involved in a crash caused at least in part by your drowsiness at the time?". Answers to such questions are probably the simplest and best subjective indicator we have of a driver's SSP while driving. Some have advocated more widespread use of single but specific questions in sleep medicine [71].

Osler Test

The Osler test was proposed in 1997 as a simpler and cheaper alternative to the MWT [72]. The subject sits "semi-reclining" in a chair, with instructions to try and stay awake during a 40-minute recording period, and to push a button every time an LED light appears in front of them at eye height. The light appears regularly every 3 seconds and stays on for 1 second. The test does not measure reaction-times but assesses whether the subject has responded to each stimulus. After failing to respond to 7 consecutive stimuli the subject is said to be asleep. As proposed initially, the test provides a measurement of the SL in each of 3 or 4 recording sessions and a mean SL for the whole test. The Osler-SL is reported to be comparable to the MWT-SL, but without the need for EEG, EOG and EMG recording equipment and the attachment of electrodes [72]. More recently it has been proposed that the pattern of errors of omission, when there are less than 7 consecutive missed responses, should also be taken into account. In some cases the test may involve only one 40-minute recording session at 9 am [73]. The role of the Osler test for measuring a subject's ISP, or their SSP after multiple recording sessions, is not yet clear, but some see it as a screening test.

Changes in the EEG and EOG – Microsleeps and Slow Eye Movements

Changes in the EEG with sleep onset have been known for many years, typically described by the appearance and disappearance of waves in particular frequency bands, assessed visually. More detailed analysis of those changes by power spectrum analysis using the Fast Fourier Transform has confirmed the earlier descriptions, particularly with an increase of spectral power in the theta-range of frequencies [55]. Power in the alpha-frequency range, which is normally recorded predominantly from occipital sites, especially when the eyes are closed, moves to the front of the head in the drowsy state. Indeed, it has been suggested that the ratio of EEG power in the alpha-frequency range measured at occipital vs. frontal sights can be used as a measure of the subject's drowsiness at the time [74].

Intermittently in the drowsy state there are burst of theta-waves lasting only a few seconds, called microsleeps. Williams et al., [75] reported that such events were associated with lapses in performance (errors of omission) during a vigilance task. However, Thomas et al. [76] studied 66 subjects who were either partially or completely sleep deprivation while they drove a car simulator repeatedly, with their EEG being monitored. There were 619 "accidents", such as driving off the road. In only 1% of those "accidents" was there a microsleep at the time, and in only 14% was there a microsleep during the preceding minute. This suggests that impaired performance in the drowsy state is not simply a function of the occurrence of microsleeps or of frank sleep onset defined by the EEG. A similar conclusion was arrived at after a major investigation of drowsiness in Canadian and US truck drivers who had their EEG and EOG recorded, along with many other variables, while they drove their usual commercial routes at work [77]. Some drivers had clear signs of episodic drowsiness while driving, although there were no crashes. There were far more episodes of drowsiness detected by long eyelid closures in video camera images of the driver's eyes than were detected as microsleeps in the EEG.

A series of slow, pendular eye movements (SEMs), each lasting about 1-4 seconds, begins in the drowsy state before sleep onset. They are mainly horizontal movements with poor binocular coordination, usually beginning after the eyes close involuntarily with drowsiness [78]. The subject is not usually aware of those SEMs, presumably because awareness of the here-and-now is inhibited at the time. SEMs recorded by the EOG have been used to monitor drowsiness in active people [79]. However, SEMs can be difficult to distinguish from some other eye movements that are common during wakefulness in active people, such as smooth pursuit movements [55]. Most drowsy episodes in the US-Canadian truck-driver investigation referred to above were not accompanied by SEMs detected in the EOG [77]. It is clear that some lapses in psychomotor performance in the drowsy state occur with the eyes open and without SEMs being present [80]. Monitoring the EEG/EOG is probably not a gold standard method for measuring "sleepiness".

Psychomotor Vigilance Test - PVT

Reaction-time tests have long been used to investigate the effects of sleep deprivation. Several different types of test have been used, the commonest being a simple reaction- time test, in which a visual stimulus is presented intermittently and the subject is asked to respond as soon as they detect the stimulus, e.g. by pushing a button. Those responses are affected in more than one way by sleep deprivation:

- the mean RT increases, even when the subject is able to respond to all stimuli
- the frequency of errors of omission increases, in which the subject fails to respond within a given time-frame to a particular stimulus
- the effect of "time-on-task" is increased so that RTs get progressively longer during a test that lasts more than a few minutes
- the frequency of errors of commission increases, in which the subject makes a response before the stimulus is presented, or too soon after its presentation to represent a physiological responses (e.g. <120 ms).

Wilkinson and Houghton (1982) [81] pioneered the development of an automated system for recording RTs in the investigation of sleep deprivation. Dinges and Powell (1985) [82] later developed a similar system which they called the Psychomotor Vigilance Test (PVT), and extended Wilkinson's findings. At random intervals of a few seconds, an LED digital counter begins to display elapsed time in milliseconds until the subject pushes a button, when the counter stops and the RT is displayed digitally for a period of 1 second. A "lapse" in performance is said to occur either when there was no push-button response within 5 seconds, or when there was a response but the RT was greater than 500 ms. That is, moderately delayed responses, with RTs typically in the range of 501– 1000 ms, are not distinguished from errors of omission. Dorian et al., (2005) [83] have described the sensitivity the PVT to the effects of sleep deprivation, with RTs and the frequency of "lapses" increasing. There are persistent differences in RTs between subjects, even when alert, but comparisons within-subjects overcome these differences. So far, there is no generally accepted and standardized measure of "sleepiness" or of EDS based on RTs, whether measured by the PVT or some other method.

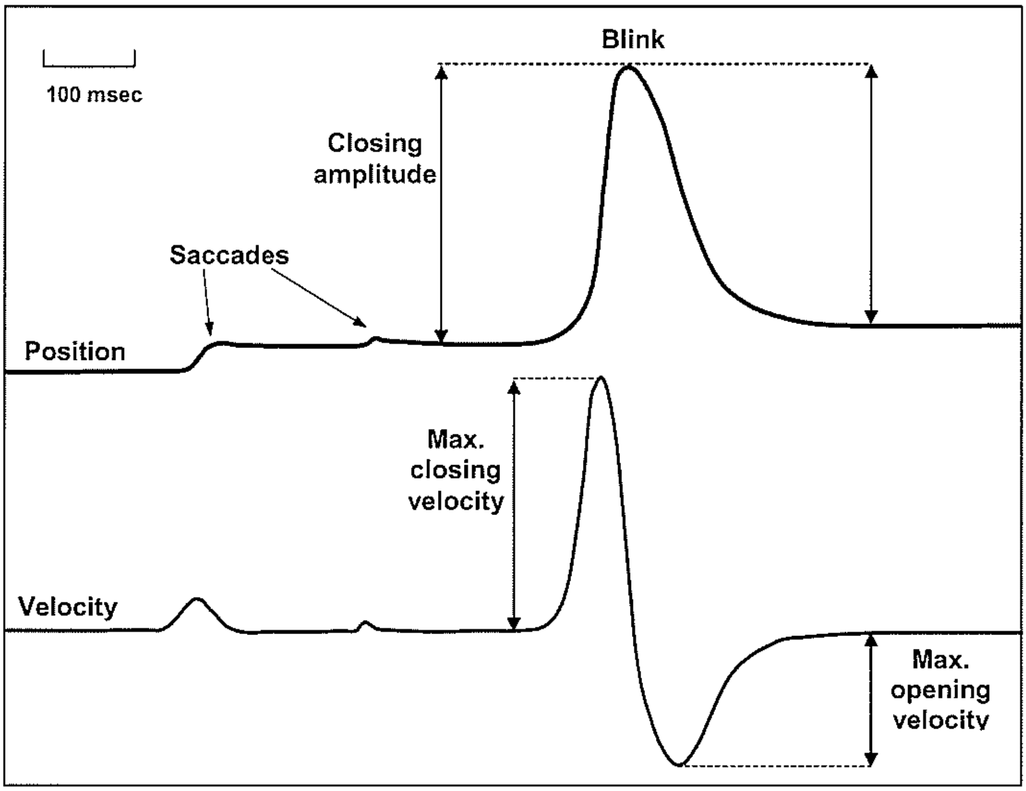

Figure 4. The position and velocity signals recorded by the infrared reflectance oculography system, showinh two saccades and a blink

Infrared Reflectance Oculography During a Vigilance Test - JTV

This test is relatively new and not yet widely known. It was made possible by the development of a new system of infrared reflectance oculography (Optalert Pty Ltd, Melbourne, Australia) [84]. It uses a special glasses frame that subjects wear in the same way that other spectacle frames are worn, and which can also hold prescription lenses and sunglasses if required. The frame incorporates tiny infrared (IR) transducers (an LED and a phototransistor) and a microprocessor. Brief pulses of IR light, just outside the visible range, are emitted in a 30-degree cone from below the eye pointing up at the upper eyelid 500 times per second. The output from this system, shown in Fig 4, enables the position of the eyes and eyelids and the "velocity" of eye movements to be monitored.

The system calculates the amplitude-velocity ratio for each eyelid movement and saccadic eye movement as a measure of their relative velocities, without the need for calibration of those movements in absolute terms [84]. The system also calculates the duration of those movements and of eyelid closures. Means and standard deviations are calculated for each variable per minute. The relative velocity of those movements decreases with drowsiness (amplitude-velocity ratios increase) [85] and their duration increases [86]. A new scale of drowsiness, the JDS with scores from 0 to 10, is based on a weighted combination of variables derived by multiple regression comparing results from subjects in two conditions, first when they are alert and performing very well in a vigilance test, and second when sleep-deprived and performing poorly in the vigilance test, with frequent errors of omission and delayed responses [84].

The JTV is a computer-based test that lasts 10-15 minutes. The subject performs a reaction-time test, at the same time as drowsiness is measured by a JDS score each minute. Three circles are displayed across a monitor screen and, at intervals that vary randomly between 5 and 15 seconds, they change shape to become either squares or diamonds of the same size for 400 milliseconds, before reverting to circles. The JTV task is for the subject to push a button held in the dominant hand as soon as possible after seeing the change of shapes, which happens 80-85 on average. The mean RT in the JTV for alert subjects is 390 ± 103 milliseconds [87], which is longer than in the PVT. There is evidently some extra visual processing to be done in the JTV to distinguish that a shape has changed to one or other of two shapes, rather than simply detecting a red LED coming on, as in the PVT. Otherwise the changes in RT with drowsiness are very similar in the two tests.

One advantage of this system is that it records and analyzes the results of a vigilance test at the same time as it measures *drowsiness* objectively by IR reflectance oculograhy. The whole recording can be replayed to identify particular events visually, such as individual eye movements or lapses in performance. The system automatically produces means and standard deviations for JDS scores and RTs, and counts errors of omission and commission. The JTV system also records video images of the subject's eyes on the same computer during the test, but they are not used to measure *drowsiness.*

The JDS scores measured during JTV tests have been validated as a measure of different levels of *drowsiness* in several different ways. Mean RTs were significantly correlated with mean JDS scores during JTVs in alert and sleep deprived subjects ($r = 0.53$, $n = 140$, $p<0.001$) [85]. Mean JDS scores during the JTV repeated every three hours during prolonged wakefulness showed the expected circadian rhythm [85]. There was a high test-retest reliability for mean JDS scores during JDVs repeated within 2 hours under similar circumstances (intra-class $r= 0.80$, $n = 42$, $p<0.001$) [88]. In a double-blind study involving well-rested young adults, a single 200 mg dose of caffeine reduced their JDS scores significantly for 3 to 4 hours, even though their JDS scores were mostly within the normal range [89]. The EEG power of theta waves was shown to increase at the time of errors of omission in the JTV and to be significantly correlated with JDS scores [90]. JDS scores have been used to measure the risk of "performance failure", assessed each minute in two different scenarios after sleep deprivation: first, as a measure of the risk of making an error of omission in the JTV, and second, the risk of driving with all four wheels out of the lane in a high fidelity car simulator [91, 92]. JDS scores greater than 5, particularly greater than 7, are associated with a 5 to 10-fold increase in risk of such "performance failures".

The effects of overnight sleep deprivation on JDS score during JTV tests have been investigated by an experiment referred to in Fig 5. A series of 146 JTVs were performed by 51 subjects at different times, with and without sleep deprivation for about 30 hours. Without sleep deprivation, and presumably therefore when alert, most subjects made no errors of

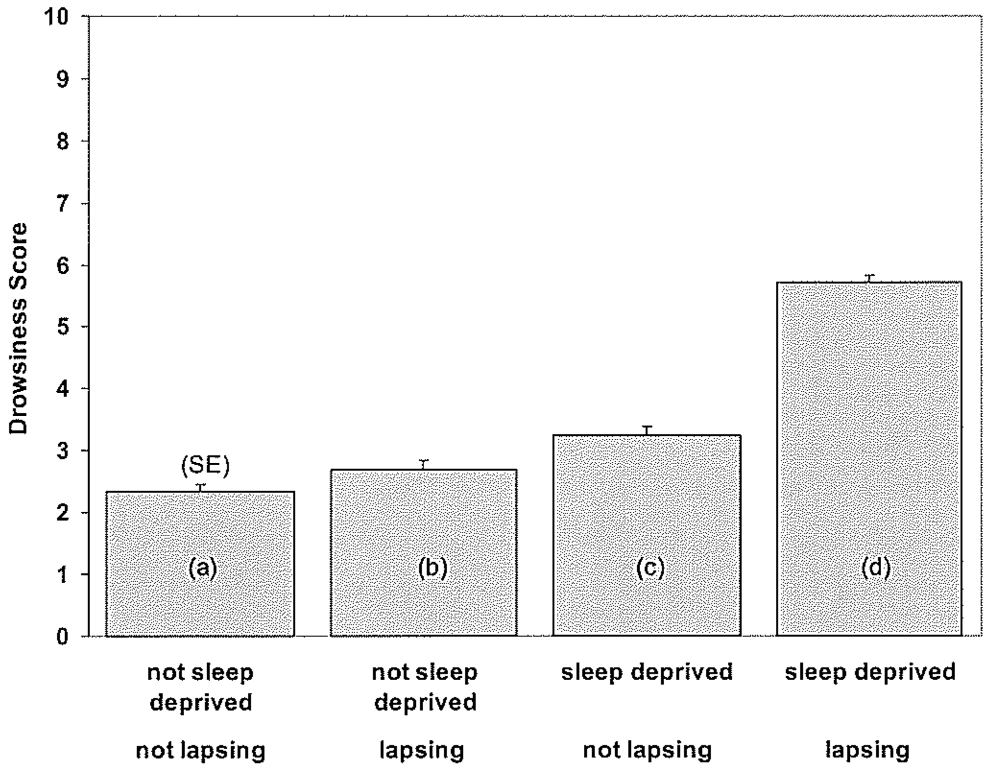

Figure 5 Mean JDS scores (and standard errors) for subjects who lapsed or did not lapse during JTV tests, with and without sleep deprivation

omission, but some made occasional errors (<2%). After sleep deprivation, many subjects made frequent errors of omission, but some still did not make any. The means of JDS scores in those 4 groups are shown in Fig 5.

There was a significant overall difference in JDS scores between groups, (ANOVA, df = 3,872; F = 193.1; p<0.001). Post-hoc Scheffé tests showed that JDS scores for subjects who were not sleep-deprived, but who made occasional errors of omission (<2%), were not significantly different from those who made no such errors (p>0.3). Sleep deprivation significantly increased JDS scores (p<0.001), but that effect was much greater in subjects who repeatedly made of errors of omission than in those who continued to make no such errors, despite their sleep deprivation. JDS scores could evidently distinguish lower levels of drowsiness that were still associated with reasonably good performance, albeit with somewhat slower responses, from higher levels of drowsiness associated with much more consequential "performance failures".

So far, the JTV has not been widely used by clinicians, so it is too early to assess its broader role in sleep medicine. Suffice it to say that it appears to be promising. Its main role so far has been with the development and validation of the Optalert ™ system for monitoring drowsiness continuously in drivers [84, 92].

Video Camera Methods for Detecting Eyelid Closure - PERCLOS

Video camera methods have been developed for monitoring a subject's eyes and eyelids, detecting their eyelid closures, both as longer-than-average blinks and as more prolonged eyelid closures [93]. Sophisticated software has been developed to detect the position of the eyelids and pupil in the video images. These methods have been proposed mainly for monitoring "sleepiness", in the sense of *drowsiness,* in drivers. The variable that has most commonly been measured is PERCLOS, the percentage of time (over an interval that might be a few minutes) that the subject's eyelids cover the pupil by at least 80% for periods in excess of 500 milliseconds at a time [94]. Those events occur when the subject is at a high level of drowsiness, when their eyelid movements become slower so that blinks last longer, and longer eyelid closures begin, at which point their psychomotor performance is likely to be impaired. Dinges et al., (1998) [94] demonstrated a close relationship between PERCLOS measurements and "lapses" in the PVT during a period of sleep deprivation.

Video camera methods have the advantage that there is no physical contact with the subject, either from the attachment of electrodes for the EEG/EOG, or by wearing glasses for the JTV and Optalert ™. Video camera methods appeared to be very promising at first, but several problems have arisen with their practical implementation. The video images may be compromised when the subject wears prescription lenses or sunglasses, or when the test is performed in sunlight. The camera can lose track of the subject's eyes during major head movements made when changing gaze, especially when looking down. In the drowsy state, many episodes of "performance failure" occur with the eyes open at the time, so video cameras can not detect those episodes [80]. Most video cameras have a frame-rate of only about 30 frames per second and are unable to measure the velocity of eyelid movements and of saccades with sufficient accuracy to detect the earliest stages of drowsiness, as the JTV test does.

Other Oculometric Tests

The pupil increases in size in darkness and with relaxation, and it fluctuates spontaneously in the drowsy state. A Pupillary Unrest Index has been described that varies with the time of day and increases with sleep deprivation [95]. This variable is correlated moderately and statistically significantly with the SL in the MSLT [96]. Sleep deprivation has also been shown to reduce the velocity of saccadic eye movements and to increase the latency to pupil constriction in response to a flash of light [97]. So far these variables have not been widely used as the basis for a standardized test of "sleepiness" that can diagnose EDS reliably [96].

External Validity of Different Methods for Measuring "Sleepiness"

The external validity of some commonly used methods for measuring "sleepiness" has already been noted above i.e. their ability to measure what they are intended to measure. Many of those methods and several others have been compared by Balkin and his colleagues [98] who tested the same subjects by 26 different methods, with and without sleep deprivation, to see how relatively effective each test was in distinguishing the two conditions. In any group of subjects, individuals will be affected differently by the same duration of sleep deprivation. However, it was not a requirement of Balkin's experiment that all subjects had the same level of "sleepiness", either initially or after sleep deprivation. What was important was the difference between each subject's "sleepiness" measured by different methods under two conditions, non-sleep-deprived and sleep-deprived. That difference was then assessed in terms of an effect-size for each test, calculated as the difference between the mean results for each test in the two conditions, divided by the standard deviation in the non-sleep-deprived condition. The largest effect- size (0.447) was for measurements of SL in a modified MSLT (with 2 rather than 4 naps), and the next largest (0.208) was for the speed of responses in the PVT. The smallest effect-size (0.001) was for the speed of logical reasoning. In only 9 of the 26 tests was the effect-size statistically significant. This investigation did not include measurements based on infrared reflectance oculography combined with a psychomotor vigilance test (the JTV test). The results provided no support for the idea that higher cortical functioning is preferentially impaired by sleep deprivation.

The results indicated reasonable efficacy for some of the tests that have been used or advocated for measuring "sleepiness", but many others were found wanting. However, those authors did not report correlations between the different test results within the same subject. There are several other sources of information about such correlations that we shall now consider.

Comparing the Results of Different Measures of "Sleepiness" in the Same Subjects

Here we shall examine correlations between the results of different measures of "sleepiness" in the same subject and at about the same time, but measured by different methods. We shall first compare one objective method with another objective method, then subjective vs. subjective methods, and finally subjective vs. objective methods.

Comparing Different Objective Measures of "Sleepiness"- The MSLT and MWT

If the mean SLs measured objectively in the MSLT and the MWT in the same subjects at about the same time simply reflected each subject's ASP and the different somnificities of the two test-situations, the results from those two tests would be highly correlated within subjects. In fact, only moderate correlations have been reported from two large series, with $r = 0.41$, $n = 258$, $p<0.001$ [30] and $r = 0.52$, $n = 522$, $p<0.001$ [99]. Those correlations coefficients are lower than those between MSLT-SLs repeated in the same subjects on different days (r studies = 0.65 – 0.97) [44]. Thus, two different SSPs measured objectively within the same subject on the same day, their MSLT-SSP and the MWT-SSP, are only moderately correlated, but significantly so in a statistical sense. Johns [44] has suggested that this is because each SSP is partly influenced by each subject's individual response to each test-situation. Those responses are therefore partially situation-specific as well as subject-specific. A response to one particular situation may not always be predictable from the same subject's response to another situation, even after allowing for the different somnificities of those situations. That makes accurate measurements of a subject's "sleepiness" even more difficult than previously thought.

Comparing Different Subjective Measures of "Sleepiness" in the Same Subjects

We can make two kinds of comparison between the results of different subjective tests of "sleepiness" in the same subjects at about the same time. First, by comparing the results of different tests in the same set of circumstances, i.e. measuring the same SSP bybdifferent subjective methods. Second, by comparing the results from the same test performed under different circumstances, i.e., measuring different SSPs by the same subjective method.

Comparing Different Subjective Tests of "Sleepiness" in the Same Circumstances

Several comparisons of this kind have been reported. For example, scores reported every two hours between 7 am and 11 pm by 40 healthy airmen have been used to compare results from the SSS, KSS and two kinds of VAS, first as a single scale of "sleepiness", and second as the mean score from 10 VASs asking about "tired eyes", "heavy eyelids", etc. [58]. After conversion of scores to Z-scores, variations on all four scales during this period of wakefulness were similar and were sensitive to the time-of-day. However, the effect-size was larger for the VASs than for either the SSS or KSS. This suggests that a VAS of "alertness-sleepiness" may be the preferred method among those tested here.

In a different experiment, Pilcher and her colleagues (2003) [69] reported on the relationships between seven different subjective measures of "sleepiness", all completed on the same day by psychology students. The tests included the SSS, three different VAS scales (*sleepy – awake*, *active – lethargic*, and *alert – drowsy*), two sub-scales from the POMS adjective check-list (*fatigue* and *vigor*), and the ESS. The ESS was used in two modes, first in a modified mode in which the respondents were asked to rate their chances of dozing off in each situation "now", and second, in the original mode of dozing off "in recent times". Raw scores from each pair of scales were inter-correlated significantly, but the highest correlations were between those scales other than the ESS (mean r = 0.59; range of r = 0.32 to 0.79). By contrast, the mean of correlations between the ESS and all other scales was r = 0.26 with the range from r = 0.17 to 0.32. Factor analysis clearly showed that ESS scores, whether reported in the modified or the original mode, formed a different factor from all the other scales which together formed a single factor.

Likert scales such as the SSS, KSS, VASs and adjective check-lists such as POMS ask about feelings and symptoms related to "sleepiness", in the sense of *drowsiness*, at a particular time. For many people, the state of *drowsiness*, particularly in its early stages, may not be readily distinguished subjectively from the state of *fatigue*. The subjective reports used in these scales may reflect a combination of feelings and symptoms of *drowsiness* and *fatigue*. All provide similar measures that are quite highly inter-correlated, but they are evidently not measuring what the ESS measures. We might conclude that scores on the SSS, KSS, a VAS of "alertness-sleepiness", and adjective check-lists all provide a subjective estimate of the subject's level of *drowsiness* confounded with *fatigue* at the time. That may also be confounded with changes of mood and feelings of dysphoria. Some people refer to what these scales measure as "subjective sleepiness". The important point is that "subjective sleepiness" is not closely related to the objectively measured ISP or level of drowsiness at the time, but there is often a weak correlation between these variables. This would explain the uncertain relationship between the SLs measured in individual naps in the MSLT and SSS scores reported before each nap.

Comparing Subjective Tests of "Sleepiness" under Different Circumstances

Kim and Young [100] described an investigation in which they performed factor analysis of subjective measures of "sleepiness" using responses to individual questions which, in the light of the categories described here, included measures of different SSPs. They used the eight ESS item-scores as well as responses to six other questions asking about the frequency of occurrence of particular feelings and problems, such as "not feeling rested during the day", the "need for coffee or other stimulants to stay awake during the day", and finding it "very difficult to get up in the morning". The latter responses were given as estimates of the frequency of occurrence, with a range of times per month on a 5-point Likert scale. Factor analysis was performed on raw scores for 13 variables, followed by oblique rotation so that highly correlated factors could be retained.

They reported three factors, the first of which they called "perceived daytime sleepiness", based on reports of frequently having a "feeling of excessive daytime sleepiness", and a "need for coffee". This factor could be interpreted as the frequency of occurrence of feelings of *drowsiness* during the day. The authors called their second factor "subjective sleep propensity in active situations" and their third factor "subjective sleep propensity in passive situations". Unfortunately, their distinction between "active" and "passive situations" was poorly based. They described "sitting quietly after lunch without alcohol" and "sitting inactive in a public place (e.g. a theatre or a meeting)" as "active situations", without explanation to justify that distinction. Their second and third factors were highly correlated oblique factors ($r = 0.59$, $p<0.05$), so a clinically significant distinction between them may be difficult to establish.

It seems reasonable to conclude that Kim and Young's analysis provided further evidence that the frequency of feelings of *drowsiness* during the day (which they called "perceived daytime sleepiness") represents a different aspect of "sleepiness" from that which each ESS item-score measures. Different levels of *drowsiness* and of *sleep propensity*, in its various time-categories, could be considered to be different dimensions of "sleepiness".

Frey and her colleagues (2004) [101] compared the effects of sleep deprivation for two nights on 25 healthy young adults using a battery of 22 different measures of "neuro-behavioral" function, including the MWT-SL, SSS scores, PVT lapses, speed and accuracy of mental arithmetic, etc. Seventeen of the 22 tests showed significant effects of sleep deprivation. Exploratory factor analysis of the results revealed seven different factors, with PVT results forming a different factor from SSS scores, and different again from MWT-SLs. Subjects who showed the greatest impairment on some tests after sleep deprivation often showed little impairment on others. The authors concluded that each subject's responses to sleep deprivation were "task" specific.

Another investigation compared scores on different items of the ESS within the same subjects [11]. Total ESS scores in a composite group of 990 subjects, including patients with a variety of sleep disorders and normal subjects, varied between 0 and 24. Because the ESS item-scores were not always normally distributed, Spearman's non-parametric correlations were calculated for each pair of item-scores for the 8 items in each subject. All 28 correlation coefficients were statistically significant ($p<0.0001$). The mean of those correlation coefficients was 0.45 and their range 0.31 to 0.57. Principal components analysis revealed a single factor that included all items, which is consistent with previous results [11]. This is good evidence that ESS item-scores are each tapping a common source of variance that represents the subject's overall level of "sleepiness" or ASP. However, each of those 28 correlation coefficients accounted for only about 20% of the variance between different SSPs within the same subject (range 10-33% for different items). Allowing for the relative inaccuracy of ESS item-scores, this still suggests that, as noted above, each different SSP has a situation-specific component that is not always predictable from other SSPs or other measurements of "sleepiness" in the same subject.

Comparing Subjective and Objective Measures of "Sleepiness" in the Same Subjects

Objective and subjective measurements of "sleepiness" cannot usually be made at precisely the same time. A subject's ISP can change very quickly, in a matter of a few seconds, and subjective reports of *drowsiness* in the KSS and other such scales may not reflect that accurately. This may be another reason why an SSS score reported immediately before an MSLT nap does not always closely reflect the SL recorded objectively a few minutes later [64, 102].

Many researchers have investigated the relationship between total ESS scores and mean SLs measured in the MSLT [33, 103]. The mean of Pearson correlation coefficients reported from 9 separate studies, some with very large groups involving 522 subjects, was -.030 [44]. Most, but not all, were statistically significant, but none indicated a very close relationship.

What Does all This Mean for the Measurement of "Sleepiness"?

Currently the different methods for assessing a subject's "sleepiness" can be classified as follows:

1. Methods based on measurements of *sleep propensity*

1.1: methods for measuring ISP

1.1.1 objective methods

e.g. SL for each nap in the MSLT or MWT

SL in the Osler test

1.1.2 subjective methods

there are none at present

: **1.2. methods for measuring SSP**

1.2.1 objective methods

e.g. mean SL in the MSLT or MWT

mean SL in the Osler test

1.2.2 subjective methods

e.g. ESS item-scores

1.3 methods for measuring ASP

1.3.1 objective methods

there are none at present

1.3.2 subjective methods

total ESS scores

2. Methods based on measurements of drowsiness

2.1.1 objective methods

e.g. PVT

JTV with JDS scores

EEG/EOG monitoring (microsleeps and SEMs)

Video camera methods (PERCLOS)

2.1.2 subjective methods

e.g. scores on the KSS, SSS

VAS of "alertness-sleepiness"

We can draw some tentative conclusions from the above discussion and classification:

- the word *sleepiness* in common English usage means the state of *drowsiness*. However, in sleep medicine and related research it has also come to mean *sleep propensity*
- "sleepiness" is not a general, unidimensional characteristic of each subject that can be measured directly and accurately by any one method.
- measurements of ISP, SSP and ASP in the same subject are only moderately inter-correlated (r = 0.3-0.5 approx).They are not closely related variables, no matter how they are measured
- scores on the KSS, SSS, a VAS of "alertness-drowsiness" and the POMS questionnaire all reflect the subject's level of "sleepiness" at the time to some extent, but those subjective assessments probably do not clearly distinguish feelings and symptoms of *drowsiness* from those of *fatigue*.
- there are valid and very reliable methods (in a test-retest sense) for measuring a subject's *sleep propensity* on all three time-scales considered here – ISP, SSP and ASP (test-retest r = 0.7-0.8 approx for each).
- there are inherent difficulties in making comparisons between measurements of *sleep propensity* and *drowsiness* under different circumstances, on different time-scales, and by different methods.
- which test or combination of tests to use for measuring "sleepiness" will depend on the purpose and circumstances of that assessment. There is no one gold standard test of "sleepiness"

Measuring Excessive Daytime Sleepiness

Any difficulty in defining and measuring "sleepiness" will also be present when defining and measuring EDS. Thus, we probably need new categories of EDS measured on different time-scales, as follows:

Excessive ISP: when a person's sleep propensity at a particular time is so high that it is likely to affect adversely the safety and efficiency of their performance of the task at hand.
Excessive SSP: when a person has a higher-than-normal *situational sleep propensity*, so that it is likely to impair their performance whenever they engage in the particular activity of concern, e.g. whenever they are driving at night after being awake for 18+ hours.
Excessive ASP: when the person's overall sleep propensity in most situations is so high that they that they are likely to doze off inadvertently in many different circumstances in daily life.

It is very important that the reference range of normal values be established for whatever method is used to measure "sleepiness" and EDS. Unfortunately that was not the case for the mean SL in the MSLT until recent years [44]. Too much reliance was placed instead on a "rule of thumb" which suggested that "normal" SLs in the MSLT were in the range of 10 to 20 minutes, SLs between 5 and 10 minutes were in a "diagnostic gray area", and those less than 5 represented "pathological sleepiness" [31]. In fact the normal range has now been established as 1.8 to 19.0 minutes [29]. Some people fall asleep in less than 5 minutes without any complaint of EDS in their daily lives. In the present author's opinion, the "rule of thumb" is quite misleading and should be abandoned. The American Academy of Sleep Medicine now recommends that neither the MSLT nor the MWT should be used alone to diagnose EDS in particular subjects [104]. However, both tests can provide evidence about EDS, albeit not very conclusive evidence in many cases, to be assessed in a broader context of the subject's history and other investigations (e.g. ESS scores).

Many MSLT-SLs recorded from patients with EDS are in the normal range. There is a comparable problem with the MWT, although there is a little less overlap than with the MSLT. These problems have been quantified for the MSLT, MWT and the ESS by calculating the sensitivity and specificity, and the receiver-operator-characteristic (ROC) curve for each test in distinguishing patients with narcolepsy, who by definition have EDS, from normal subjects who do not [44]. The ROC curves indicated that all three methods were reasonably effective in distinguishing the EDS of narcoleptics from normal "sleepiness". However, the ROC curve was higher for the EES than the MWT, which was in turn higher than for the MSLT. There is always a trade-off between high sensitivity and high specificity when choosing a cut-off score between normal and abnormal for such tests. The MSLT had a sensitivity of 80.9% and a specificity of 89.8% when the cut-off score was 5 minutes. With a cut-off score of 8, the sensitivity increased to 94.5% but the specificity fell to 73.3%. When the cut-off score was 3, the sensitivity was only 52.0, but the specificity was 98.3%.

By contrast, the MWT had a sensitivity of 84.3% and a specificity of 98.4% when the cut-off score was 12 minutes. The ESS had a sensitivity of 93.5% and a specificity of 100% when the cut-off score was 11. Other researchers have confirmed the high sensitivity (97%) and specificity (100%) of the ESS with a cut-off score of 14 when distinguishing the "sleepiness" of narcoleptics from that of normal subjects [43]. By this criterion, the MWT is a better objective test of EDS than the MSLT is. However, the ESS is at least as good as the MWT, despite being a subjective test. Criticisms of this conclusion seem to have been based on the mistaken idea that the ESS measures "subjective sleepiness", of which most narcoleptics would complain if asked [40].

The ESS is now the most commonly used method for measuring EDS, presumably because of its simplicity, ease of administration, low cost, validity and reliability. ESS scores >10 (i.e. between 11 and 24) represent some degree of EDS, particularly for ESS scores >14. It is very unlikely that a patient with narcolepsy would have an ESS <11 [44]. Of course, this does not mean that subjects with higher ESS scores necessarily have narcolepsy.

Who is Too Drowsy to Drive?

As a practical example of the importance of EDS and its measurement in sleep medicine and in the work of regulatory authorities on road safety, etc., we can consider the question "Who is too drowsy to drive?" and propose some tentative answers.

In a large series of drivers, total ESS scores were significantly correlated with their self-reported history of near-miss incidents and actual car crashes [105]. However, it appears that predictions of excessive ISP while driving at a particular time cannot be made accurately for individual drivers from ESS score unless they have a very high ASP (ESS>14) and would be likely to doze off under many circumstances [25].

An experiment by George et al. [106] gives an example of the kind of evidence we have about the MSLT as a measure of fitness to drive. Their subjects included a group of normal sleepers without EDS and a group of patients who had EDS associated with a diagnosis of either narcolepsy or severe obstructive sleep apnea. Simulated driving performance was compared with mean SLs in the MSLT, measured on the same day. As a group, the patients with EDS performed significantly worse than the normal subjects. However, about half of the individuals with EDS drove as well at he normal subjects. Overall, the impaired driving of individuals could not be accurately predicted from their mean SL in the MSLT. This is in accord with a general recommendation that the MSLT-SL should not be used as the sole measure of fitness to drive from the point of view of drowsiness.

Comparable evidence about the MWT as a test of fitness to drive is more equivocal. The standard deviation of lane position when driving in a simulator, a measure of the risk of driving out of the lane, is significantly correlated with the mean SL in the MWT [107]. However, in this and other investigations, the MWT-SL did not clearly distinguish individual drivers with normal performance from those with impaired performance [108]. This is further evidence that it is difficult to make an accurate prediction of drowsiness (i.e. of excessive ISP) in a particular person while driving at some future time on the basis of a measurement of the driver's ISP or SSP measured at another time and in a different situation.

We may conclude tentatively that anyone with a SL of <5 minutes in the MSLT or a SL < 12 minutes in the MWT, as well as a total ESS score >14 is very likely to have excessive ASP, regardless of its cause. They probably should not be granted a driver's license without having had an expert evaluation by a sleep specialist and until any cause(s) for their EDS have been diagnosed and treated successfully. Some such people may not be fit to drive heavy vehicles or buses at all. The fitness of other people to drive with less extreme indications of excessive ASP may be more equivocal, relying on their history of drowsy near-misses and crashes. Assessments of ISP derived from scores on the KSS, SSS or VAS are not appropriate for assessing a person's longer term fitness to drive.

A new method has been developed for measuring the drowsiness of drivers continuously, based on changes in the characteristics of eyelid movements during blinks and longer eyelid closures measured by infrared reflectance oculography. A device (Optalert™, manufactured by Optalert Pty Ltd, Melbourne, Australia) provides a measure of the driver's ISP every minute on the new scale of drowsiness, the JDS [85, 84]. It can detect the first objective signs of drowsiness, often before drivers are aware of it subjectively. As JDS scores increase with *drowsiness*, drivers have greatly increased risks of "performance failure" such as driving out of the lane and perhaps crashing [92]. Consequently, they should not continue driving at the time. The device warns drivers by sounding a loud beep and then issuing a verbal message. This is not intended to keep the driver awake for long, although it can have that effect for a few minutes. It is intended to prompt the driver that he/she should take preventive measures, perhaps by stopping driving as soon as it is safe to do so, having a brief nap and perhaps some caffeine, and thereby avoid falling asleep at the wheel and crashing. There is little to suggest that monitoring the vehicle (the variability of its position on the road or its velocity, etc) adds significantly to the accuracy of risk assessments based on the state of the driver alone.

In this author's opinion, the best way to assess the risks of EDS while driving for individual drivers is to adopt a multi-level approach, as follows:

- make an assessment of the driver's ASP and history of any drowsy driving incidents and crashes. An MWT may be justified sometimes.
- diagnose and treat any sleep disorders, including insufficient sleep, that may be increasing the driver's ASP
- educate the driver and others about the importance of EDS and it causes, and assess work schedules and other time constraints to reduce the likelihood of EDS while driving
- perhaps make an assessment of the driver's state of alertness/drowsiness on a particular day, to decide whether or not they are fit to begin driving. Passing such a test initially does not guarantee that the driver will remain fit to drive for the next few hours. Such "fitness for duty tests" have not yet been standardized.
- in cases where the risk of having excessive ISP while driving is high (e.g. driving overnight), or where the possible consequences of a drowsy episode are extreme, the drivers' drowsiness can be monitored continuously while driving

Conclusion

It is not a simple matter to measure "sleepiness" or EDS. There is widespread confusion about what "sleepiness" is, and what the many different tests that purport to measure it are actually doing. Perhaps the discussion here will alleviate some of that confusion. There is a great need for more research into the drowsy state which, in this author's opinion, has been sadly neglected and misunderstood. Nevertheless, it is possible to make a reasonably accurate assessment of a subject's "sleepiness" using the methods described here. Some tests are clearly better than others. Which ones to use will depend on the intent and context within which particular assessments are made.

Some new methods for measuring drowsiness objectively show great promise, e.g. the frequency of "lapses" in a psychomotor vigilance test (the PVT), JDS scores made during a standardized vigilance task that also measures psychomotor performance (the JTV), and JDS scores for monitoring drowsiness continually while driving (Optalert ™). These new methods are currently being used to address the problem of drowsy driving in particular. They have not yet been widely used in the clinical practice of sleep medicine, but they could be in the future.

Acknowledgements

Dr Kate Crowley and Dr Andrew Tucker helped with the preparation of this Chapter.

References

[1] Arand, D., Bonnet, M., Hurwitz, T., Mitler, M., Rosa, R., and Sangal, R. B. (2005). The clinical use of the MSLT and MWT. *Sleep, 28*(1), 123-44.

[2] Ohayon, M. M. (2008). From wakefulness to excessive sleepiness: what we know and still need to know. *Sleep Med Rev, 12*(2), 129-41.

[3] Johns, M. (1998). Rethinking the assessment of sleepiness. *Sleep Med Rev, 2*(1), 3-15.

[4] Onions, C. T. (1959). *The Shorter Oxford English Dictionary*. Oxford, UK.

[5] Roth, T., Roehrs, T., Carskadon, M. A., and Dement, W. C. (1989). Daytime sleepiness and alertness. In M. H. Kryger, T. Roth, and W. C. Dement (Eds.), *Principles ad Practice of Sleep Medicine* (pp. 14-23). Philadelphia: W.B. Saunders

[6] Aldrich, M. S. (1989). Automobile accidents in patients with sleep disorders. *Sleep, 12*(6), 487-94.

[7] Harrison, Y. and Horne, J. A. (1996). "High sleepability without sleepiness". The ability to fall asleep rapidly without other signs of sleepiness. *Neurophysiol Clin, 26*(1), 15-20.

[8] Johns, M. W. (2008). The subjective measurement of excessive daytime sleepiness. In S. R. Pandi-Perumal, J. C. Verster, J. M. Monti, M. Lader, and S. Z. Langer (Eds.), *Sleep Disorders: Diagnosis and Therapeutics* (pp. 643-657). London: Informa Healthcare

[9] Borbély, A. A. (1982). A two process model of sleep regulation. *Hum Neurobiol, 1*(3), 195-204.

[10] Folkard, S. and Akerstedt, T. (1989). Towards a model for the prediction of alertness and/or fatigue on different sleep/wake schedules. In J. Oginski, J. Pokorski, and J. Rutenfranz (Eds.), *Contemporary Advances in Shiftwork Research* (pp. 231-240). Krakow: Medical Acedemy

[11] Johns, M. W. (1994). Sleepiness in different situations measured by the Epworth Sleepiness Scale. *Sleep, 17*(8), 703-10.

[12] Johns, M. W. (2002). Sleep propensity varies with behaviour and the situation in which it is measured: the concept of somnificity. *J Sleep Res, 11*(1), 61-7.

[13] Johns, M. W. and Knutson, K. L. (2006). The somnificity of different activities described in the Epworth Sleepiness Scale. *Sleep, 29*(Abstract Supplement), A30.

[14] Dinges, D. F. (1989). The nature of sleepiness: causes, contexts and consequences. In A. J. Stunkard and A. Baum (Eds.), *Perspectives in Behavioural Medicine: Eating, Sleepiness and Sex* (pp. 147-179). Hillsdale, New Jersey: Lawrence Erlbaum Associates

[15] *ICSD- International classification of of sleep disorders: diagnostic and coding manual. Diagnostic Classification Steering Committee, Thorpy MJ, Chairman*. 1990, American Sleep Disorders Association: Rochester. Minnesota.

[16] Cole, R. J. (1989). Postural baroreflex stimuli may affect EEG arousal and sleep in humans. *J Appl Physiol, 67*(6), 2369-75.

[17] Bonnet, M. H. and Arand, D. L. (1999). Level of arousal and the ability to maintain wakefulness. *J Sleep Res, 8*(4), 247-54.

[18] De Valck, E., Cluydts, R., and Pirrera, S. (2004). Effect of cognitive arousal on sleep latency, somatic and cortical arousal following partial sleep deprivation. *J Sleep Res, 13*(4), 295-304.

[19] Edgar, D. M., Dement, W. C., and Fuller, C. A. (1993). Effect of SCN lesions on sleep in squirrel monkeys: evidence for opponent processes in sleep-wake regulation. *J Neurosci, 13*(3), 1065-79.

[20] Saper, C. B., Chou, T. C., and Scammell, T. E. (2001). The sleep switch: hypothalamic control of sleep and wakefulness. *Trends Neurosci, 24*(12), 726-31.

[21] Mignot, E., Taheri, S., and Nishino, S. (2002). Sleeping with the hypothalamus: emerging therapeutic targets for sleep disorders. *Nat Neurosci, 5 Suppl*, 1071-5.

[22] Torterolo, P., Yamuy, J., Sampogna, S., Morales, F. R., and Chase, M. H. (2003). Hypocretinergic neurons are primarily involved in activation of the somatomotor system. *Sleep, 26*(1), 25-8.

[23] Sinton, C. M. and McCarley, R. W. (2004). Neurophysiological mechanisms of sleep and wakefulness: a question of balance. *Semin Neurol, 24*(3), 211-23.

[24] Johns, M. W. (1993). Understanding Insomnia. *Australian Family Physician, 22*, 318-328.

[25] Johns, M. W. (2000). A sleep physiologist's view of the drowsy driver. *Transportation Research, Part F 3*, 241-249.

[26] Bonnet, M. H. and Arand, D. L. (1998). Sleepiness as measured by modified multiple sleep latency testing varies as a function of preceding activity. *Sleep, 21*(5), 477-83.

[27] Johns, M. W. (2003). The clinical assessment of daytime sleepiness in patients with obstructive sleep apnea. In M. Fabiani (Eds.), *Surgery for Snoring and Obstructive Sleep Apnea Syndrome* (pp. 283-295). Kugler Publications: The Hague, The Netherlands

[28] Carskadon, M. A. and Dement, W. C. (1982). The multiple sleep latency test: what does it measure? *Sleep, 5 Suppl 2*, S67-72.

[29] Arand, D. L. and Bonnet, M. H. (2008). The multiple sleep latency test and maintenance of wakefulness test: use and procedures. In S. R. Pandi-Perumal, J. C. Verster, J. M. Monti, M. Lader, and S. Z. Langer (Eds.), *Sleep Disorders: Diagnosis and Therapeutics* (pp. 658-668). London: Informa Healthcare

[30] Sangal, R. B., Thomas, L., and Mitler, M. M. (1992). Maintenance of wakefulness test and multiple sleep latency test. Measurement of different abilities in patients with sleep disorders. *Chest, 101*(4), 898-902.

[31] Thorpy, M. J. (1992). The clinical use of the Multiple Sleep Latency Test. The Standards of Practice Committee of the American Sleep Disorders Association. *Sleep, 15*(3), 268-76.

[32] Mitler, M. M., Gujavarty, K. S., and Browman, C. P. (1982). Maintenance of wakefulness test: a polysomnographic technique for evaluation treatment efficacy in patients with excessive somnolence. *Electroencephalogr Clin Neurophysiol, 53*(6), 658-61.

[33] Johns, M. W. (1991). A new method for measuring daytime sleepiness: the Epworth Sleepiness Scale. *Sleep, 14*(6), 540-5.

[34] Foulkes, D. and Vogel, G. (1965). Mental Activity at Sleep Onset. *J Abnorm Psychol, 70*, 231-43.

[35] Ogilvie, R. D. (2001). The process of falling asleep. *Sleep Med Rev, 5*(3), 247-270.

[36] Martin, C. M. and Ogilivie, R. D. (2002). A microanalysis of EMG and EEG activity during the sleep onset period (SOP). *Sleep, 25*(Abstract Supplement), A31.

[37] Van Orden, K. F., Jung, T. P., and Makeig, S. (2000). Combined eye activity measures accurately estimate changes in sustained visual task performance. *Biol Psychol, 52*(3), 221-40.

[38] Caffier, P. P., Erdmann, U., and Ullsperger, P. (2003). Experimental evaluation of eye-blink parameters as a drowsiness measure. *Eur J Appl Physiol, 89*(3-4), 319-25.

[39] Johns, M. W., Chapman, R., Crowley, K., and Tucker, A. *A new method for determining when a driver is too drowsy to drive*. in *SENSATION 2nd International Conference: Monitoring sleep and sleepiness with new sensors within medical and industrial applications*. 2007. Chania, Greece.

[40] Chervin, R. D. and Aldrich, M. S. (1999). The Epworth Sleepiness Scale may not reflect objective measures of sleepiness or sleep apnea. *Neurology, 52*(1), 125-31.

[41] Johns, M. W. (1992). Reliability and factor analysis of the Epworth Sleepiness Scale. *Sleep, 15*(4), 376-81.

[42] Carmelli, D., Bliwise, D. L., Swan, G. E., and Reed, T. (2001). Genetic factors in self-reported snoring and excessive daytime sleepiness: a twin study. *Am J Respir Crit Care Med, 164*(6), 949-52.

[43] Parkes, J. D., Chen, S. Y., Clift, S. J., Dahlitz, M. J., and Dunn, G. (1998). The clinical diagnosis of the narcoleptic syndrome. *J Sleep Res, 7*(1), 41-52.

[44] Johns, M. W. (2000). Sensitivity and specificity of the multiple sleep latency test (MSLT), the maintenance of wakefulness test and the epworth sleepiness scale: failure of the MSLT as a gold standard. *J Sleep Res, 9*(1), 5-11.

[45] Engleman, H. M., Kingshott, R. N., Wraith, P. K., Mackay, T. W., Deary, I. J., and Douglas, N. J. (1999). Randomized placebo-controlled crossover trial of continuous positive airway pressure for mild sleep Apnea/Hypopnea syndrome. *Am J Respir Crit Care Med, 159*(2), 461-7.

[46] Mitler, M. M., Harsh, J., Hirshkowitz, M., and Guilleminault, C. (2000). Long-term efficacy and safety of modafinil (PROVIGIL((R))) for the treatment of excessive daytime sleepiness associated with narcolepsy. *Sleep Med, 1*(3), 231-243.

[47] Shen, J., Botly, L. C., Chung, S. A., Gibbs, A. L., Sabanadzovic, S., and Shapiro, C. M. (2006). Fatigue and shift work. *J Sleep Res, 15*(1), 1-5.

[48] Broman, J. E., Bengtson, H., and Hetta, J. (2000). Psychometric properties of a Swedish version of the Epworth Sleepiness Scale. *J Sleep Res, 9*(Supplement 1), 27.

[49] Hagell, P. and Broman, J. E. (2007). Measurement properties and hierarchical item structure of the Epworth Sleepiness Scale in Parkinson's disease. *J Sleep Res, 16*(1), 102-9.

[50] Takegami, M., Suzukamo, Y., Wakita, T., Noguchi, H., Chin, K., Kadotani, H., Inoue, Y., Oka, Y., Nakamura, T., Green, J., Johns, M. W., and Fukuhara, S. (2008). Development of a Japanese version of the Epworth Sleepiness Scale (JESS) based on item response theory. *Sleep Medicine, in press*.

[51] Johns, M. W. and Hocking, B. (1997). Daytime sleepiness and sleep habits of Australian workers. *Sleep, 20*, 844-849.

[52] Manni, R., Politini, L., Ratti, M. T., and Tartara, A. (1999). Sleepiness in obstructive sleep apnea syndrome and simple snoring evaluated by the Epworth Sleepiness Scale. *J Sleep Res, 8*(4), 319-20.

[53] Martikainen, K., Hasan, J., Urponen, H., Vuori, I., and Partinen, M. (1992). Daytime sleepiness: a risk factor in community life. *Acta Neurol Scand, 86*(4), 337-41.

[54] Pallesen, S., Nordhus, I. H., Omvik, S., Sivertsen, B., Tell, G. S., and Bjorvatn, B. (2007). Prevalence and risk factors of subjective sleepiness in the general adult population. *Sleep, 30*(5), 619-24.

[55] Åkerstedt, T. and Gillberg, M. (1990). Subjective and objective sleepiness in the active individual. *Int J Neurosci, 52*(1-2), 29-37.

[56] Reyner, L. A. and Horne, J. A. (1998). Falling asleep whilst driving: are drivers aware of prior sleepiness? *Int J Legal Med, 111*(3), 120-3.

[57] Gillberg, M., Kecklund, G., and Akerstedt, T. (1994). Relations between performance and subjective ratings of sleepiness during a night awake. *Sleep, 17*(3), 236-41.

[58] Casagrande, M., Curio, G., Ferrara, M., Tricarico, M., Urbani, L., and Porcu, S. (1999). Subjective ratings of sleepiness: Sensitivity of different measures in revealing time of day effects. *Sleep Research Online, 2*(Supplement 1), 226.

[59] Ingre, M., Akerstedt, T., Peters, B., Anund, A., Kecklund, G., and Pickles, A. (2006). Subjective sleepiness and accident risk avoiding the ecological fallacy. *J Sleep Res, 15*(2), 142-8.

[60] Hoddes, E., Zarcone, V., Smythe, H., Phillips, R., and Dement, W. C. (1973). Quantification of sleepiness: a new approach. *Psychophysiology, 10*(4), 431-6.

[61] Dinges, D. F., Orne, M. T., Whitehouse, W. G., and Orne, E. C. (1987). Temporal placement of a nap for alertness: contributions of circadian phase and prior wakefulness. *Sleep, 10*(4), 313-29.

[62] Stepanski, E. J. (2002). The effect of sleep fragmentation on daytime function. *Sleep, 25*(3), 268-76.

[63] Monk, T. H., Weitzman, E. D., Fookson, J. E., Moline, M. L., Kronauer, R. E., and Gander, P. H. (1983). Task variables determine which biological clock controls circadian rhythms in human performance. *Nature, 304*(5926), 543-5.

[64] Pressman, M. R. and Fry, J. M. (1989). Relationship of autonomic nervous system activity to daytime sleepiness and prior sleep. *Sleep, 12*(3), 239-45.

[65] MacLean, A. W., Fekken, G. C., Saskin, P., and Knowles, J. B. (1992). Psychometric evaluation of the Stanford Sleepiness Scale. *J Sleep Res, 1*(1), 35-39.

[66] Babkoff, H., Caspy, T., and Mikulincer, M. (1991). Subjective sleepiness ratings: the effects of sleep deprivation, circadian rhythmicity and cognitive performance. *Sleep, 14*(6), 534-9.

[67] Herbert, M., Johns, M. W., and Doré, C. (1976). Factor analysis of analogue scales measuring subjective feelings before and after sleep. *Br J Med Psychol, 49*(4), 373-9.

[68] Sauter, C., Asenbaum, S., Popovic, R., Bauer, H., Lamm, C., Klosch, G., and Zeitlhofer, J. (2000). Excessive daytime sleepiness in patients suffering from different levels of obstructive sleep apnoea syndrome. *J Sleep Res, 9*(3), 293-301.

[69] Pilcher, J. J., Pury, C. L., and Muth, E. R. (2003). Assessing subjective daytime sleepiness: an internal state versus behavior approach. *Behav Med, 29*(2), 60-7.

[70] Rosenthal, L., Roehrs, T. A., and Roth, T. (1993). The Sleep-Wake Activity Inventory: a self-report measure of daytime sleepiness. *Biol Psychiatry, 34*(11), 810-20.

[71] Zallek, S. N., Redenius, R., Fisk, H., Murphy, C., and O'Neill, E. (2008). A single question as a sleepiness screening tool. *J Clin Sleep Med, 4*(2), 143-8.

[72] Bennett, L. S., Stradling, J. R., and Davies, R. J. (1997). A behavioural test to assess daytime sleepiness in obstructive sleep apnoea. *J Sleep Res, 6*(2), 142-5.

[73] Liistro, G. and Rodenstein, D. (2008). The Osler test. In S. R. Pandi-Perumal, J. C. Verster, J. M. Monti, M. Lader, and S. Z. Langer (Eds.), *Sleep Disorders: Diagnosis and Therapeutics* (pp. 669-670). London: Informa Healthcare

[74] Hori, T. (1985). Spatiotemporal changes of EEG activity during waking-sleeping transition period. *Int J Neurosci, 27*(1-2), 101-14.

[75] Williams, H. L., Granda, A. M., Jones, R. C., Lubin, A., and Armington, J. C. (1962). EEC-frequency and finger pulse volume as predictors of reaction time during sleep loss. *Electroencephalogr Clin Neurophysiol, 14*, 64-70.

[76] Thomas, M., Thorne, D., Sing, H., Redmond, D., Balkin, T., Wesensten, N. J., Russo, M., Welsh, A., Rowland, L., Johnson, D., Aladdin, R., Cephus, R., Hall, S., and Belenky, G. (1998). The relationship between driving accidents and microsleeps during cumulative partial sleep deprivation. *Journal of Sleep Research, 7*(Suppl 2), 275.

[77] Mitler, M. M., Miller, J. C., Lipsitz, J. J., Walsh, J. K., and Wylie, C. D. (1997). The sleep of long-haul truck drivers. *N Engl J Med, 337*(11), 755-61.

[78] Porte, H. S. (2004). Slow horizontal eye movement at human sleep onset. *J Sleep Res, 13*(3), 239-49.

[79] Lockley, S. W., Cronin, J. W., Evans, E. E., Cade, B. E., Lee, C. J., Landrigan, C. P., Rothschild, J. M., Katz, J. T., Lilly, C. M., Stone, P. H., Aeschbach, D., and Czeisler, C. A. (2004). Effect of reducing interns' weekly work hours on sleep and attentional failures. *N Engl J Med, 351*(18), 1829-37.

[80] Chapman, R., Johns, M. W., and Crowley, K. (2006). In the drowsy state, errors of omission in a visual reaction-time test occur with eyes open or closed. *Sleep, 29*(Abstract Supplement), A364.

[81] Wilkinson, R. T. and Houghton, D. (1982). Field test of arousal: a portable reaction timer with data storage. *Human Factors, 24*(4), 487-93.

[82] Dinges, D. F. and Powell, J. W. (1985). Microcomputer analysis of performance on a portable, simple visual RT task during sustained operations. *Behavior Research Methods, Instruments, & Computers, 17*, 652-655.

[83] Dorrian, J., Rogers, N. L., and Dinges, D. F. (2005). Psychomotor vigilance performance: neurocognitive assay sensitive to sleep loss. In C. A. Kushida (Eds.),

Sleep Deprivation: Clinical issues, pharmacology, and sleep loss effects (pp. 39-70). New York: Marcel Dekker

[84] Johns, M. W., Tucker, A., Chapman, R., Crowley, K., and Michael, N. J. (2007). Monitoring eye and eyelid movements by infrared reflectance oculography to measure drowsiness in drivers. *Somnologie, 11*, 234-242.

[85] Johns, M. W. and Tucker, A. (2005). The amplitude-velocity ratio of blinks: changes with drowsiness. *Sleep, 28*(Suppl), A122.

[86] Tucker, A. and Johns, M. W. (2005). The duration of eyelid movements during blinks: changes with drowsiness. *Sleep, 28*(Supp), A122.

[87] Johns, M. W., Crowley, K., Chapman, R., Tucker, A., and Hocking, C. (2008). The effect of blinks and saccadic eye movements on visual reaction-times. *Percept Psychophys, in press.*

[88] Johns, M. W., Crowley, K. E., Chapman, R. J., Tucker, A. J., and Hocking, C. A. (2008). The Test-Retest Reliability of an Ocular Measure of Drowsiness. *Sleep, 31*((Suppl)), A118.

[89] Michael, N., Johns, M., Owen, C., and Patterson, J. (2008). Effects of caffeine on alertness as measured by infrared reflectance oculography. *Psychopharmacology (Berl), 200*(2), 255-60.

[90] Crowley, K., johns, M. W., Chapman, R. J., Tucker, A. J., and Patterson, J. (2008). An Ocular Measure of Drowsiness and the EEG: Changes with Sleep Deprivation. *Sleep, 31*((Suppl)), A119.

[91] Stephan, K., Hosking, S., Regan, M., Verdoorn, A., Young, K., and Haworth, N., *The relationship between driving performance and the Johns Drowsiness Scale as measured by the Optalert system*, in *Monash University*. 2006, Monash Accident Research Centre: Melbourne. p. Report No 252.

[92] Johns, M. W., Chapman, R., Crowley, K., and Tucker, A. (2008). A new method for assessing the risks of drowsiness while driving. *Somnologie, 12*, 66-74.

[93] Wierwille, W. W. and Ellsworth, L. A. (1994). Evaluation of driver drowsiness by trained raters. *Accident: Analysis and Prevention, 26*(5), 571-81.

[94] Dinges, D. F., Mallis, M. M., Maslin, G., and Powell, J. W., *Evaluation of techniques for ocular measurement as an index of fatigue and the basis for alertness management*, in *DOT HS 808 762*. 1998, U.S. Department of Transportation National Highway Traffic Safety Administration.

[95] Wilhelm, B., Giedke, H., Ludtke, H., Bittner, E., Hofmann, A., and Wilhelm, H. (2001). Daytime variations in central nervous system activation measured by a pupillographic sleepiness test. *J Sleep Res, 10*(1), 1-7.

[96] McLaren, J. W., Hauri, P. J., Lin, S. C., and Harris, C. D. (2002). Pupillometry in clinically sleepy patients. *Sleep Med, 3*(4), 347-52.

[97] Russo, M., Thomas, M., Thorne, D., Sing, H., Redmond, D., Rowland, L., Johnson, D., Hall, S., Krichmar, J., and Balkin, T. (2003). Oculomotor impairment during chronic partial sleep deprivation. *Clin Neurophysiol, 114*(4), 723-36.

[98] Balkin, T. J., Bliese, P. D., Belenky, G., Sing, H., Thorne, D. R., Thomas, M., Redmond, D. P., Russo, M., and Wesensten, N. J. (2004). Comparative utility of

instruments for monitoring sleepiness-related performance decrements in the operational environment. *J Sleep Res, 13*(3), 219-27.

[99] Sangal, R. B., Mitler, M. M., Sangal, J. A. M., and US Modafinil in Narcolepsy Multicenter Study, G. r. (1997). MSLT MWT and ESS: indices of sleepiness in 522 drug-free patients with narcolepsy. *Sleep Research, 26*, 492.

[100] Kim, H. and Young, T. (2005). Subjective daytime sleepiness: dimensions and correlates in the general population. *Sleep, 28*(5), 625-34.

[101] Frey, D. J., Badia, P., and Wright, K. P., Jr. (2004). Inter- and intra-individual variability in performance near the circadian nadir during sleep deprivation. *J Sleep Res, 13*(4), 305-15.

[102] Johnson, L. C., Spinweber, C. L., Gomez, S. A., and Matteson, L. T. (1990). Daytime sleepiness, performance, mood, nocturnal sleep: the effect of benzodiazepine and caffeine on their relationship. *Sleep, 13*(2), 121-35.

[103] Mitler, M. M., Walsleben, J., Sangal, R. B., and Hirshkowitz, M. (1998). Sleep latency on the maintenance of wakefulness test (MWT) for 530 patients with narcolepsy while free of psychoactive drugs. *Electroencephalogr Clin Neurophysiol, 107*(1), 33-8.

[104] Littner, M. R., Kushida, C., Wise, M., Davila, D. G., Morgenthaler, T., Lee-Chiong, T., Hirshkowitz, M., Daniel, L. L., Bailey, D., Berry, R. B., Kapen, S., and Kramer, M. (2005). Practice parameters for clinical use of the multiple sleep latency test and the maintenance of wakefulness test. *Sleep, 28*(1), 113-21.

[105] Powell, N. B., Schechtman, K. B., Riley, R. W., Guilleminault, C., Chiang, R. P., and Weaver, E. M. (2007). Sleepy driver near-misses may predict accident risks. *Sleep, 30*(3), 331-42.

[106] George, C. F., Nickerson, P. W., Hanly, P. J., Millar, T. W., and Kryger, M. H. (1987). Sleep apnoea patients have more automobile accidents. *Lancet, 2*(8556), 447.

[107] Sagaspe, P., Taillard, J., Chaumet, G., Guilleminault, C., Coste, O., Moore, N., Bioulac, B., and Philip, P. (2007). Maintenance of wakefulness test as a predictor of driving performance in patients with untreated obstructive sleep apnea. *Sleep, 30*(3), 327-30.

[108] Banks, S., Catcheside, P., Lack, L. C., Grunstein, R. R., and McEvoy, R. D. (2005). The Maintenance of Wakefulness Test and driving simulator performance. *Sleep, 28*(11), 1381-5.