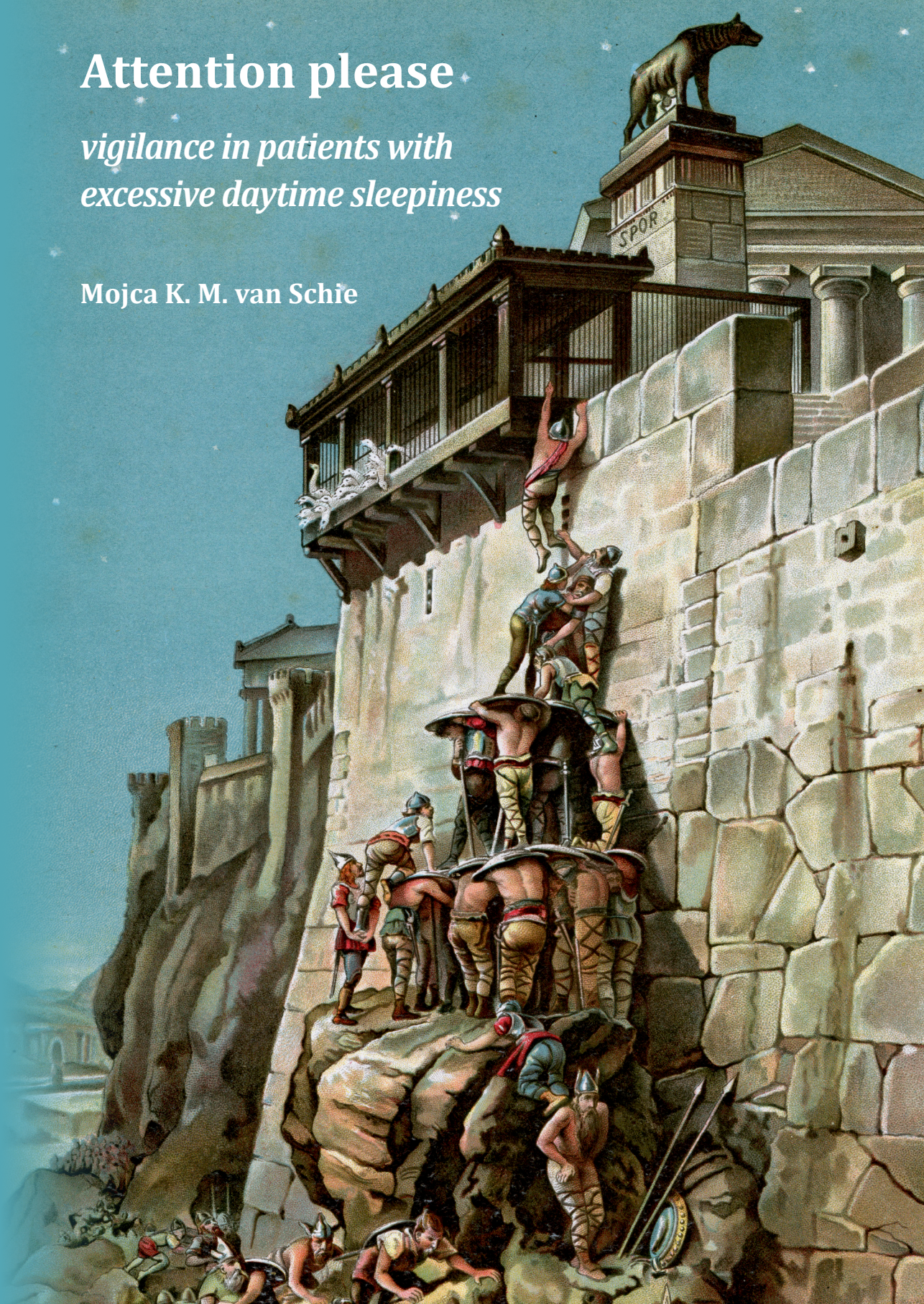
Attention please: vigilance in patients with excessive daytime sleepiness

Schie, M.K.M. van

Attention please

vigilance in patients with excessive daytime sleepiness

Mojca K. M. van Schie



Attention please: vigilance in patients with excessive daytime sleepiness

Mojca K. M. van Schie

Financial support for the publication of this thesis was generously provided by:
Nederlandse Vereniging voor Slaap- en Waak Onderzoek

ISBN	978-94-6332-773-2
Layout	Loes Kema
Cover design	Mojca van Schie
Printed by	GVO Drukkers en Vormgevers

Attention please: vigilance in patients with excessive daytime sleepiness

Proefschrift

ter verkrijging van
de graad van doctor aan de Universiteit Leiden,
op gezag van rector magnificus prof. dr. ir. H. Bijl,
volgens besluit van het college voor promoties
te verdedigen op donderdag 7 oktober 2021
klokke 16.15 uur

door

Mojca Kristina Maria van Schie

geboren te Haarlem
in 1989

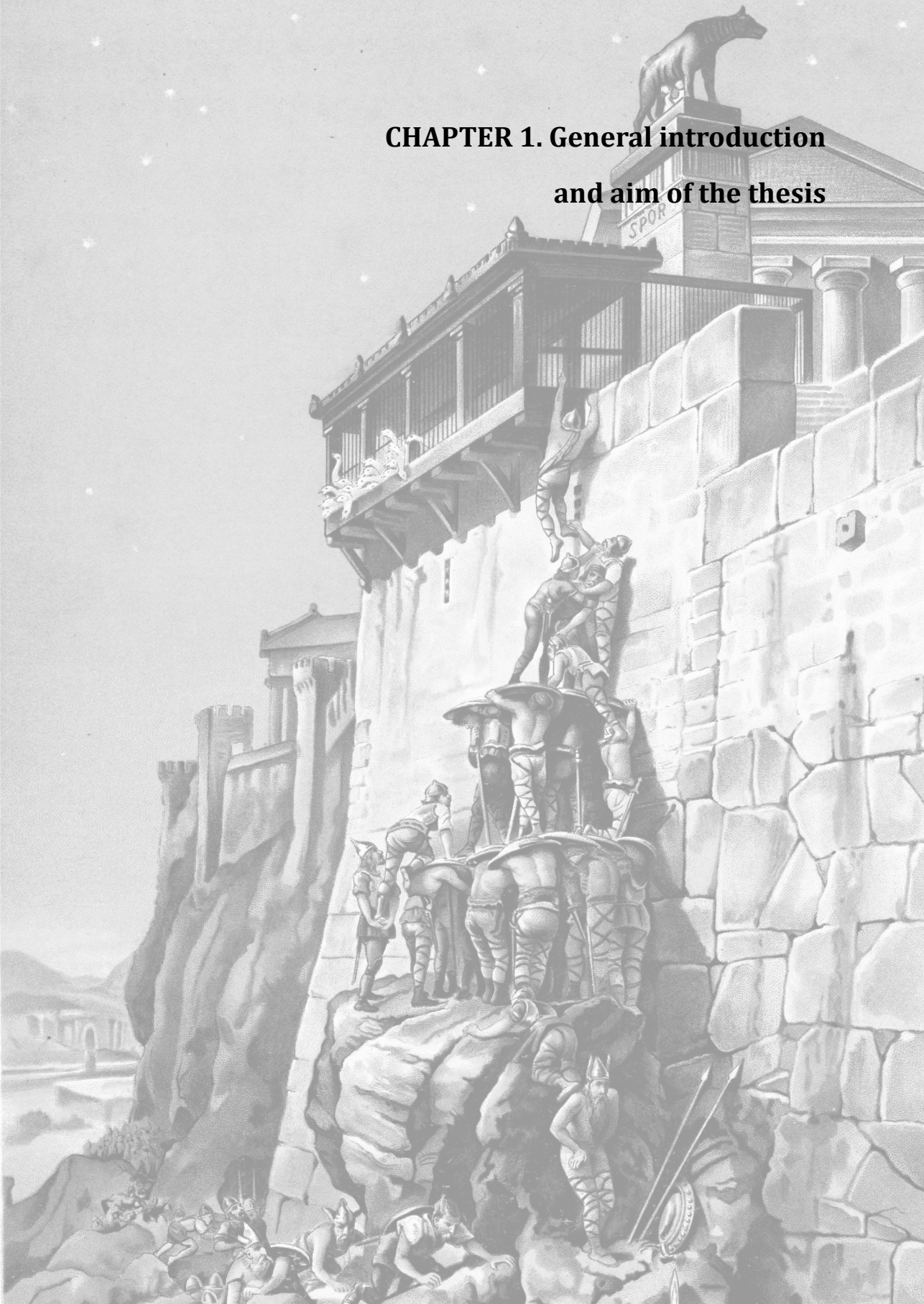
Promotores	prof. dr. J.G. van Dijk prof. dr. G.J. Lammers
Copromotor	dr. R. Fronczek
Promotiecommissie	prof. dr. H.A.M. Middelkoop prof. dr. S. Schwartz (Université de Genève, Genève) prof. dr. Y.D. van der Werf (Amsterdam UMC, Amsterdam)

To my father,
who taught me logic and writing

To my mother,
who inspired my major life choices

To Diederik,
whose support and love were invaluable during completion of this thesis

CONTENTS



CHAPTER 1. General introduction and aim of the thesis

Sleep disorders

Various healthcare professionals are active in the field of sleep medicine: neurologists, pulmonologists, otolaryngologists, dentists, general practitioners and psychologists, all of whom participate in the diagnosis or treatment of sleep disorders. Neurologists are mostly involved in disorders with excessive daytime sleepiness (also referred to as central disorders of hypersomnolence)[1], a disturbed 24-hour sleep-wake cycle (circadian rhythm disorders), and disorders with abnormal behaviour during sleep (so-called parasomnias, e.g. sleep walking, night terror, REM-sleep behaviour disorder). This thesis focuses on disorders characterised by excessive daytime sleepiness.

Disturbed vigilance in excessive daytime sleepiness

Arriving at the diagnosis of a sleep disorders starts with a careful history, which in many cases leads to, if not the diagnosis, a strong suspicion. In this respect much attention has bene given to one symptom, excessive daytime sleepiness (EDS). As this is essentially a subjective phenomenon, it is not surprising that sleep questionnaires constitute a major approach to quantify its severity. One such, the Epworth Sleepiness Scale (ESS)[2], became widely accepted as an important test in the 1990s. The ESS asks respondents to judge the likelihood of falling asleep in certain circumstances, such as 'after lunch', 'in the car when waiting for the traffic light for 5 minutes' or 'when conversing with someone'. Apart from the questionnaire approach, other tests aimed to quantify EDS by measuring physiological variables: the 'Multiple Sleep Latency Test' (MSLT)[3] requires patients to actually fall asleep when given the opportunity to do so, and measures how long it takes patients to do so (MSLT, box 1). The 'Maintenance of Wakefulness Test' (MWT, box 2)[4] measures the ability to resist sleepiness, and also measures the time it takes to fall asleep, but now when patients are asked to not do so. Whether these latter tests assess the same matter as the EDS questionnaires is not a given, and will be discussed below.

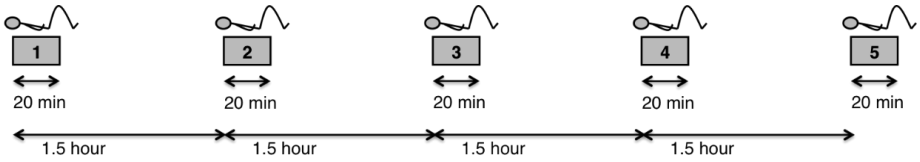
This focus on sleepiness has somewhat overshadowed other important problems that patients with sleep disorders may experience.

Box 1 – Multiple Sleep Latency Test

The MSLT asks subjects to fall asleep in a quiet and dimmed room multiple times on one day. A standardised MSLT consists of four to five twenty-minute sessions with one and a half to two-hour intervals. The subject is asked to stay in bed for the whole session, even when no sleep occurs. Electroencephalograpy for standard sleep recording is performed. Sleep onset is defined as the first 30-second epoch of any sleep stage, including stage I. The primary outcome measure is the average of all four to five sleep-onset latencies obtained

from the separate test sessions over the day. When no sleep occurs, sleep latency is noted as 20 minutes. A secondary outcome measure is the occurrence of sleep-onset rapid eye movement episodes.

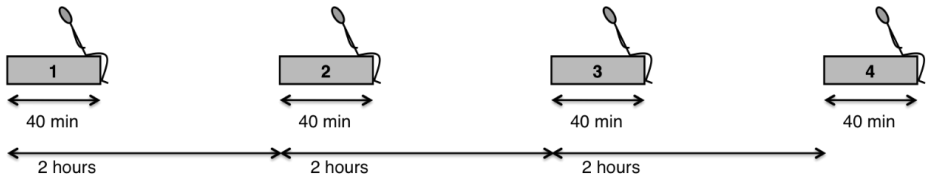
In the studies described in this thesis an MSLT consists of a five-session protocol.



The MSLT was introduced in 1977[5], standardised and accepted in 1986[3]. Test-retest reliability, inter- and intra-rater reliability were all high in a small group of healthy individuals.[6-8] The construct validity of the MSLT to measure sleepiness was based on application of the test to a group of healthy individuals who were sleep deprived[9]. The severity of deprivation during one night proved to be significantly correlated to sleep latency the next day. This suggested that sleep latency, as assessed with an MSLT, indeed represented an objective quantitative marker of sleepiness, at least after sleep deprivation in healthy subjects. The concept was consolidated by the finding that sedative drugs decreased sleep latency[10], and by the finding that the sleep latencies proved abnormally short in patients with narcolepsy and obstructive sleep apnoea[11-13]. No studies were performed to obtain normative data in large population-based cohorts.

Box 2 – Maintenance of Wakefulness Test

The MWT is similar to the MSLT except for one major difference: subjects are instructed to sit still and remain awake while seated in bed in a quiet and dimly-lit room, based on the concept that this task represents the difficulties of daily life of patients with EDS better than their ability to fall asleep. No excessive movement or talking to prevent sleep is allowed. The MWT consists of four 40-minute sessions performed at two-hour intervals. Sleep onset is defined as the first 30-second epoch of any sleep stage, including stage I[14]. Sessions are terminated after 40 minutes if no sleep occurs, or after sleep, defined as three consecutive epochs of stage 1 sleep, or one epoch of any other stage of sleep. A technician has to be present to score sleep continuously. The primary outcome measure is the mean sleep latency (the arithmetic mean of the four sessions).



The MWT was developed in 1982 after it turned out that the MSLT disappointed as a tool to measure treatment efficacy.[4] Although patients felt a subjective improvement after treatment, felt more alert and were better able to stay awake, this was not reflected by the MSLT. Their sleep latency remained short.[15] The MWT proved to detect treatment effects better than the MSLT. However, the correlation with subjective improvement and performance was still only moderate, with the exception of driving performance in obstructive sleep apnoea patients.[16]

One such problem is impaired vigilance. Awake people should become aware of changes in their environment, so they can respond to these changes, if necessary. This capability is called 'vigilance'. Vigilance is a fundamental prerequisite for higher cognitive functions required in daily life, for instance at school or work. Subtle disturbances of vigilance are common in daily life, causing cognitive mishaps, such as forgetting why you went up the stairs, or reading a piece of text more than once without registering the content.

Since vigilance refers to a capability to be aware of internal or external stimuli, wakefulness is a prerequisite for vigilance. When people are asleep or falling asleep, they will not be able to become aware of stimuli either. This state is easily confused with a disturbed vigilance, as the results are the same in that the detection of stimuli is diminished. Caution is warranted to prevent confusing sleepiness with a vigilance disturbance. Vigilance refers to a quality of the awake state only; no inferences can be made about vigilance in people who are not awake.

Type 1 narcolepsy

An excellent example of a condition with severely disturbed vigilance is the primary sleep disorder narcolepsy, type 1.[1] The disorder is caused by loss of hypocretin-producing neurons in the lateral hypothalamus. Type 1 narcolepsy is characterised by severe EDS and cataplexy. The presence of chronic EDS, i.e. daily episodes of an irrepressible need to sleep or daytime lapses into sleep, is mandatory for the diagnosis of narcolepsy. The presence of cataplexy, a sudden partial or complete drop of muscle tone with preserved consciousness triggered by certain emotions, is a prerequisite to diagnose type 1 narcolepsy, but has to be absent in type 2. Additional symptoms of either type include disturbed nocturnal sleep, as well as a range of signs and symptoms that are pathophysiological associated with rapid eye movement (REM) sleep. These are

hypnagogic hallucinations and sleep paralysis.[17] A disturbed vigilance is an additional, largely neglected, symptom of narcolepsy. Note that this does not refer to a consequence of being or falling asleep, but to an aspect of being awake. A disturbed vigilance is directly related to impaired daytime performance and quality of life.[18,19] Examples of the severe vigilance problems experienced by patients with narcolepsy are difficulty in recalling the content of a conversation, finishing a book, or concentrating on studies or work.

Vigilance problems can have life-threatening consequences, e.g. when driving a car or working with potentially dangerous machines under such circumstances. As the consequences of vigilance impairment can be serious, vigilance impairments need to be quantified; something not reflected by the abovementioned questionnaire, nor by the MSLT or MWT.

These currently used sleep tests have other shortcomings. Although the MSLT in particular has contributed to a more precise classification system of sleep disorders, several percent of the normal population may show abnormal test results and therefore fulfil criteria for 'hypersomnia', even though they have no corresponding complaints at all.[20] The MSLT should therefore always be interpreted in the context of a patient's specific symptoms and other test results. As mentioned above, the MSLT is not sensitive to EDS treatment effects, showing it misses important aspects of sleepiness. The MWT and sleep questionnaires have been shown to be more sensitive to detect improvement in pharmacological trials than the MSLT. The MWT is also used in the assessment of safety issues, such as the inability to remain awake during driving. However, there is little evidence of any link between the mean sleep latency as assessed by the MWT and the accident risk in real world circumstances. Moreover, the MWT and sleep questionnaires do not reflect functional improvements during the awake state; while patients reported to be more alert when awake, the sleep latency tests remained unaltered.[14]

Measuring vigilance in sleep disorders

Quantifying how the activities of daily life are impaired by sleep disorders, or how treatment ameliorates such functions, proves to be remarkably difficult. A possible solution is to quantify vigilance itself, since it is a prerequisite for cognitive functions. Descriptions of vigilance measurements in disorders of EDS are limited to a few publications on patients with narcolepsy (see also box 1)[19] and to some contrasting publications on patients with obstructive sleep apnoea syndrome, a sleep-related breathing disorder.[21-25]

Several methods have been proposed to measure vigilance. Examples include subjective visual-analog scales, pupillography [26], quantified electro-encephalography (EEG) [27], brain imaging[28], and a variety of response tasks assessing sustained attention. An ideal test would provide an objective quantification of the level of vigilance while being easy and cheap to administer. These practical requirements are best met by response tasks assessing sustained attention. A conceptual advantage of such tests is that

they also assess the response to a stimulus. In other words, these tests are not limited to the recording of a derivate of brain activity, but are extended to quantify the patient's capability to act upon a trigger. One such test is the Sustained Attention to Response Task (SART) [29], a 4-minute 19-second computer task in which subjects should withhold presses to one out of nine stimuli (box 3).[29,30] This test has been demonstrated capable of quantifying vigilance impairment in narcolepsy patients.[19]

Box 3 - The Sustained Attention to Response Task (SART)

The SART is a go/no-go task in which the no-go target appears unpredictably and rarely, and in which both accuracy and response speed, quantified as reaction time, are important. It lasts 4 minutes and 19 seconds and comprises the numbers 1–9 appearing 225 times in random order and in different sizes in a white font on a black computer screen. Subjects have to respond to the appearance of each number by pressing a button while seated in a dimly-lit room, except when the number is a 3, which occurs 25 times in all. Subjects have to press the button before the next number appears, and are instructed to give equal importance to accuracy and speed in performing the task.[29,30] The primary outcome measure of the SART is the total error score, consisting of, firstly, key presses when no key should be pressed (i.e. after a '3', a so-called 'no-go trial': commission errors), and secondly absent presses when a key should have been pressed (i.e. after anything but a '3', the so-called 'go trials': omission errors).

The SART was initially developed to investigate lapses of sustained attention in individuals with traumatic brain injury[29,30], but appeared to be a promising test in patients with type 1 narcolepsy.[19]

Scope of the present thesis

This thesis covers several steps in the process of validating the SART as a measure of vigilance in patients with excessive daytime sleepiness.

Part I - Measuring vigilance with the SART

Although functional impairments as a result of insufficient vigilance have gradually become recognized in patients with EDS, measurements of vigilance are still scarce. A scrutiny of such publications shows that definitions of 'vigilance' differ between publications. Chapter 2 deals with these different definitions, and proposes a new definition of vigilance. Chapter 3 extends previous vigilance measurements in narcolepsy

through application of the SART to include other patient groups with excessive daytime sleepiness. Chapter 4 analyses various factors possibly influencing SART outcome measures, such as task repetition, napping, time of day, and test instruction.

Part II – The SART as a treatment effect parameter in EDS

The chapters in this part deal with the SART as a tool to measure treatment efficacy in sleep studies. Chapters 5 and 6 describe SART results in narcolepsy patients before and after treatment. Chapter 5 contains a comparison of the SART, ESS and MWT before and after treatment with modafinil, pitolisant or placebo. In Chapter 6, the SART is compared to the Psychomotor Vigilance Test (PVT; see Box 4)[31-33], MWT, and Oxford Sleep Resistance test (OSLER; see Box 4)[34] before and during treatment with sodium oxybate. Chapter 7 deals with the SART to assess continuous positive airway pressure in obstructive sleep apnoea.

Box 4 – Other measurements of vigilance and sleepiness

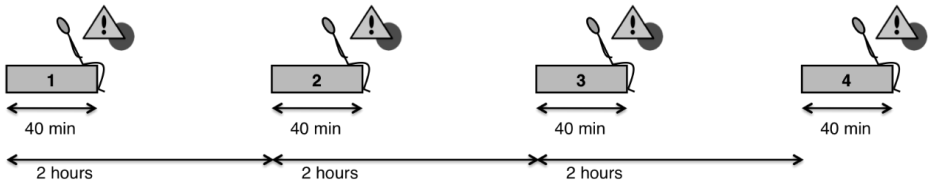
The Psychomotor Vigilance Test (PVT)

The PVT is a reaction time task. Subjects are instructed to press a button as quickly as possible to stop a digital millisecond counter as soon as it starts running, which it does at variable intervals (interstimulus intervals: 2-10 seconds). The task requires continuous attention to detect the randomly occurring stimuli. PVT tests of various durations are available; the best-validated test duration lasts 10 minutes. Outcome measures vary between studies. Frequently used measures are the frequency of lapses, defined as the number of times the subject fails to respond within 500 ms or fails to respond at all; the average reaction time; the average of the 10% longest or shortest reaction times per session. The PVT is widely applied in sleep-deprivation studies.[31-33,35] Only recently some normative data in sleep disorders have become available.[36]

The Oxford Sleep Resistance test (OSLER)

The OSLER is a behavioural version of the MWT. It follows the same schedule and subjects are similarly positioned. Instead of electroencephalographic recording of sleep onset, subjects are required to respond to a non-arousing visual stimulus. The subject's index finger is placed on a sensor. A red light is positioned four to

six feet away at eye level in the frontal visual field. The light flashes regularly for 1 second every 3 seconds. Subjects are instructed to lift their finger from the sensor for 1 second when the red light flashes. Sleep onset is defined as seven consecutive omissions, i.e. non-responding to flashes for ≥ 18 seconds. Similar to the MWT, a session is terminated when sleep onset occurs or after 40 minutes of being awake. The primary outcome measure is the mean of the four sleep onset latencies. The OSLER has the advantage of not requiring constant presence of a technician.



The OSLER has been validated in patients with obstructive sleep apnoea.[34]

REFERENCES

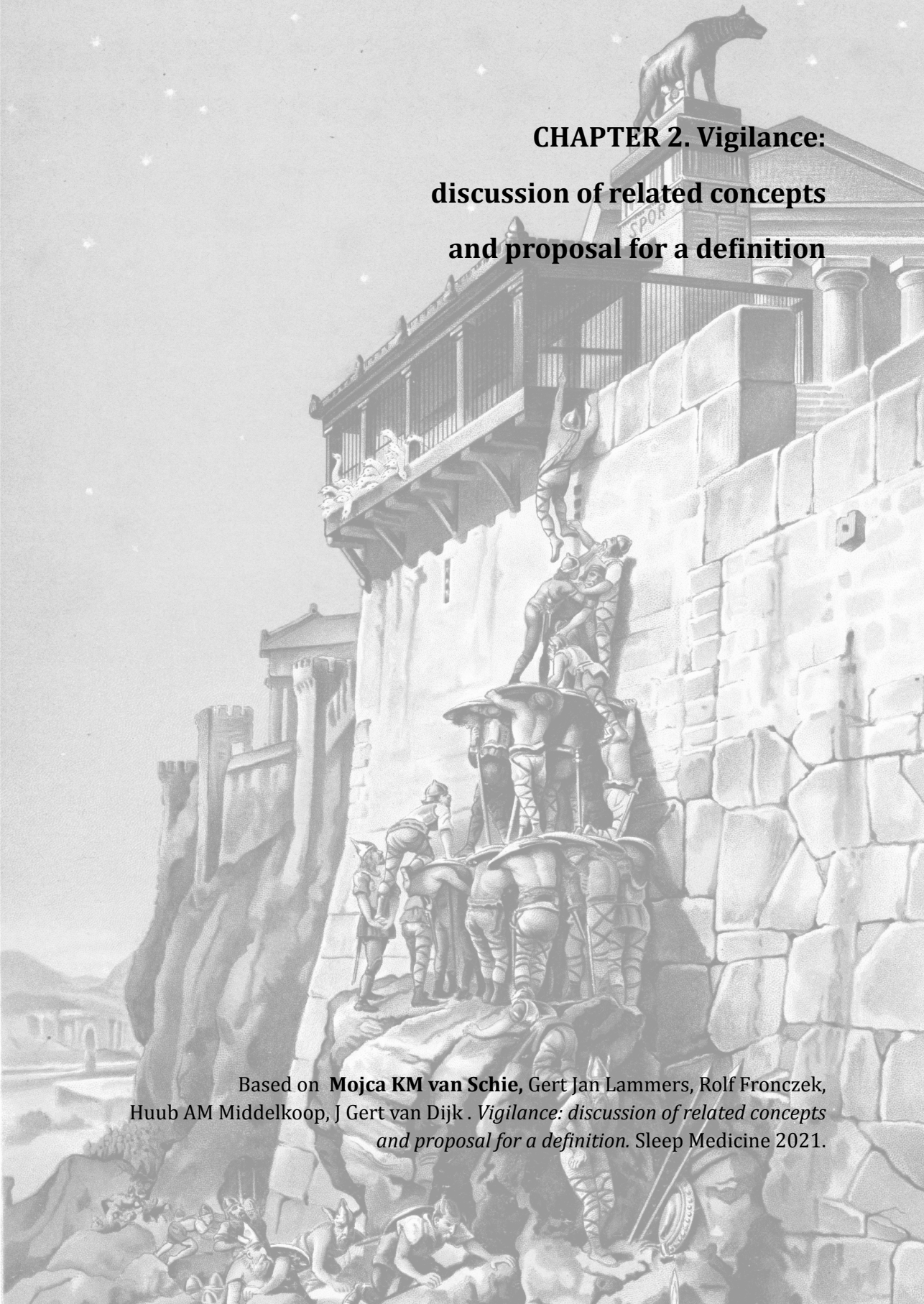
1. American Academy of Sleep Medicine. *International Classification of Sleep Disorders. Diagnostic and Coding Manual.* 3rd ed. Darien, IL: American Academy of Sleep Medicine; 2014.
2. Johns MW. A new method for measuring daytime sleepiness: the Epworth sleepiness scale. *Sleep.* 1991;14(6):540-545.
3. Carskadon MA, Dement WC, Mitler MM, Roth T, Westbrook PR, Keenan S. Guidelines for the multiple sleep latency test (MSLT): a standard measure of sleepiness. *Sleep.* 1986;9(4):519-524.
4. Mitler MM, Gujavarty KS, Browman CP. Maintenance of wakefulness test: a polysomnographic technique for evaluation treatment efficacy in patients with excessive somnolence. *Electroencephalogr Clin Neurophysiol.* 1982;53(6):658-661.
5. Carskadon MA, Dement WC. Sleep tendency: an objective measure of sleep loss. *Sleep Res.* 1977;6:200.
6. Zwyghuizen-Doorenbos A, Roehrs T, Schaefer M, Roth T. Test-retest reliability of the MSLT. *Sleep.* 1988;11(6):562-565.
7. Drake CL, Rice MF, Roehrs TA, Rosenthal L, Guido P, Roth T. Scoring reliability of the multiple sleep latency test in a clinical population. *Sleep.* 2000;23(7):911-913.
8. Benbadis SR, Qu Y, Perry MC, Dinner DS, Warnes H. Interrater reliability of the multiple sleep latency test. *Electroencephalogr Clin Neurophysiol.* 1995;95(4):302-304.
9. Carskadon MA, Dement WC. Effects of total sleep loss on sleep tendency. *Perceptual and motor skills.* 1979;48(2):495-506.
10. Bliwise D, Seidel W, Karacan I, et al. Daytime sleepiness as a criterion in hypnotic medication trials: comparison of triazolam and flurazepam. *Sleep.* 1983;6(2):156-163.
11. Richardson GS, Carskadon MA, Flagg W, Van den Hoed J, Dement WC, Mitler MM. Excessive daytime sleepiness in man: multiple sleep latency measurement in narcoleptic and control subjects. *Electroencephalogr Clin Neurophysiol.* 1978;45(5):621-627.
12. Mitler MM, Van den Hoed J, Carskadon MA, et al. REM sleep episodes during the Multple Sleep Latency Test in narcoleptic patients. *Electroencephalogr Clin Neurophysiol.* 1979;46(4):479-481.
13. Dement W, Carskadon MA, Richardson G. *Excessive daytime sleepiness in the sleep apnea syndrome.* New York: Alan R Liss; 1978.
14. Littner MR, Kushida C, Wise M, et al. Practice parameters for clinical use of the multiple sleep latency test and the maintenance of wakefulness test. *Sleep.* 2005;28(1):113-121.

15. Roth T, Hartse KM, Zorick F, Conway W. Multiple naps and the evaluation of daytime sleepiness in patients with upper airway sleep apnea. *Sleep.* 1980;3(3-4):425-439.
16. Philip P, Sagaspe P, Taillard J, et al. Maintenance of Wakefulness Test, obstructive sleep apnea syndrome, and driving risk. *Ann Neurol.* 2008;64(4):410-416.
17. Scammell TE. Narcolepsy. *The New England journal of medicine.* 2015;373(27):2654-2662.
18. Valley V, Broughton R. Daytime performance deficits and physiological vigilance in untreated patients with narcolepsy-cataplexy compared to controls. *Review of electroencephalograpy and clinical neurophysiology.* 1981;11(1):133-139.
19. Fronczek R, Middelkoop HA, van Dijk JG, Lammers GJ. Focusing on vigilance instead of sleepiness in the assessment of narcolepsy: high sensitivity of the Sustained Attention to Response Task (SART). *Sleep.* 2006;29(2):187-191.
20. Mignot E, Lin L, Finn L, et al. Correlates of sleep-onset REM periods during the Multiple Sleep Latency Test in community adults. *Brain.* 2006;129(Pt 6):1609-1623.
21. K G, AA B, JF G, et al. - Cognitive impairment in obstructive sleep apnea. *D - 0265365.* (- 1768-3114 (Electronic)):- 233-240.
22. L F-S, C B, MR DG, et al. - Cognitive dysfunction in patients with obstructive sleep apnea (OSA): partial. *D - 7605818.* (- 0361-9230 (Print)):- 87-92.
23. SF Q, CS C, WC D, et al. - The association between obstructive sleep apnea and neurocognitive. *D - 7809084.* (- 1550-9109 (Electronic)):- 303-314B.
24. Findley LJ, Barth JT, Powers DC, Wilhoit SC, Boyd DG, Suratt PM. Cognitive impairment in patients with obstructive sleep apnea and associated hypoxemia. *Chest.* 1986;90(5):686-690.
25. Findley L, Unverzagt M, Guchu R, Fabrizio M, Buckner J, Suratt P. Vigilance and automobile accidents in patients with sleep apnea or narcolepsy. *Chest.* 1995;108(3):619-624.
26. Morad Y, Lemberg H, Yofe N, Dagan Y. Pupillography as an objective indicator of fatigue. *Current eye research.* 2000;21(1):535-542.
27. Coenen AM. Neuronal activities underlying the electroencephalogram and evoked potentials of sleeping and waking: implications for information processing. *Neuroscience and biobehavioral reviews.* 1995;19(3):447-463.
28. Olbrich S, Mulert C, Karch S, et al. EEG-vigilance and BOLD effect during simultaneous EEG/fMRI measurement. *Neuroimage.* 2009;45(2):319-332.
29. Robertson IH, Manly T, Andrade J, Baddeley BT, Yiend J. 'Oops!': performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. *Neuropsychologia.* 1997;35(6):747-758.
30. Manly T, Robertson IH, Galloway M, Hawkins K. The absent mind: further investigations of sustained attention to response. *Neuropsychologia.* 1999;37(6):661-670.
31. Dinges DF, Kribbs NB. Performing while sleepy: effects of experimentally-induced

sleepiness In: Monk TH, ed. *Sleep, sleepiness & Performance.* Chister, UK: Wiley; 1991:97-128.

32. Dinges DF, Powell JW. Sleepiness is more than lapsing. *Sleep Res.* 1988;17(84).
33. Dorrian J, Rogers NL, Dinges DF. Psychomotor vigilance performance: a neurocognitive assay sensitive to sleep loss. In: Kushida C, ed. *Sleep deprivation: Clinial Issues, Pharmacology and Sleep Loss Effects.* New York: Marcel Dekker; 2005:39-70.
34. Bennett LS, Stradling JR, Davies RJ. A behavioural test to assess daytime sleepiness in obstructive sleep apnoea. *J Sleep Res.* 1997;6(2):142-145.
35. Basner M, Dinges DF. Maximizing sensitivity of the psychomotor vigilance test (PVT) to sleep loss. *Sleep.* 2011;34(5):581-591.
36. Thomann J, Baumann CR, Landolt HP, Werth E. Psychomotor vigilance task demonstrates impaired vigilance in disorders with excessive daytime sleepiness. *J Clin Sleep Med.* 2014;10(9):1019-1024.

PART I. Measuring vigilance



CHAPTER 2. Vigilance: discussion of related concepts and proposal for a definition

Based on **Mojca KM van Schie,** Gert Jan Lammers, Rolf Fronczek, Huub AM Middelkoop, J Gert van Dijk . *Vigilance: discussion of related concepts and proposal for a definition.* Sleep Medicine 2021.

ABSTRACT

We review current definitions of vigilance to propose a definition, applicable in sleep medicine. As previous definitions contained terms such as attention, alertness, and arousal, we address these concepts too. We defined alertness as a quantitative measure of the mind state governing sensitivity to stimuli. Arousal comprises a stimulus-induced upward change in alertness, irrespective of the subsequent duration of the increased level of alertness. Vigilance is defined as the capability to be sensitive to potential changes in one's environment, i.e. the capability to reach a level of alertness above a threshold for a certain period of time rather than the state of alertness itself. It has quantitative and temporal dimensions. Attention adds direction towards a stimulus to alertness, requiring cognitive control: it involves being prepared to process stimuli coming from an expected direction. Sustained attention corresponds to a state in which some level of attention is purposefully maintained, adding a time factor to the definition of attention. Vigilance differs from sustained attention in that the latter in addition implies a direction to which attention is cognitively directed as well as a specification of duration. Attempts to measure vigilance, however, are often in fact measurements of sustained attention.

INTRODUCTION

A slip of the mind can result in an innocuous failure to perform some daily activity. Examples include putting a cup with a tea bag under the coffee tap of a coffee machine instead of under the hot water spout, or putting pea pods in a pan to cook and throwing the green peas in the waste bin. The behavior resulting from such slips is also referred to as automatic behavior.[1] The reason for such slips is that not enough attention was paid to these simple and ordinary tasks; the cause of this deficit in attention is a lack of vigilance. Not all such slips are innocuous: vigilance problems can have life-threatening consequences, e.g. when train drivers ignore a red signal. An excellent model of severely disturbed vigilance is the primary sleep disorder narcolepsy.[2,3] Patients with narcolepsy suffer considerably in daily life from impaired vigilance. Narcolepsy patients have for instance difficulty in recalling the content of a conversation, finishing reading a book, or keeping focused on a study or work.

In view of its importance for care and research vigilance impairment needs to be quantified, which in turn requires an accurate definition. Without it, different concepts may cause confusion, and impairments may be attributed to incorrect mechanisms or causes. Unfortunately, the definition of vigilance in the scientific literature is far from unambiguous. We reviewed part of the literature and attempted to reach a definition of vigilance that might be useful in the specific context of sleep medicine.

VIGILANCE IN THE LITERATURE

The Merriam-Webster dictionary defines vigilance as 'the quality or state of being vigilant'.[4] In turn, 'vigilant', derived from the Latin word vigilare (to keep watch, to stay awake) is then explained as 'alertly watchful, especially to avoid danger'. This primary definition thus relates vigilant to alert, which is described as 'watchful and prompt to meet danger or emergency' and 'quick to perceive and act'.

Scientists often refer to Mackworth's publications about vigilance decrement for a first description of vigilance.[5,6] He defined vigilance *decrement* as a decline in attention-requiring performance developing and worsening after a prolonged period of time spent on the respective performance. This definition evolved simultaneously with the related 'signal detection theory',[7] which concerned the ability to distinguish relevant from irrelevant (background) stimuli based on certain determinants of how stimuli are detected. Some authors built upon Mackworth's decline in function as the foremost criterion to define vigilance. For instance, Sanders stated that vigilance is defined through the result of a cognitively simple task that is performed less well over time.[8] Although Mackworth's work is dominated by descriptions of decrement of vigilance, he did not

use this decrease in performance as the defining feature of vigilance itself. Instead, he defined vigilance as a state of readiness to detect and respond to small changes in the environment, occurring at irregular times.[9] He had adapted this definition from Head, who described vigilance as a high state of physiological efficiency.[10] Head's description is now regarded the first definition of vigilance in the scientific literature.

At present the term vigilance is widely applied in the psychological field. Examples of definitions include short descriptions as 'the capacity to attend to external stimuli'[11]; other formulations include a time aspect, such as 'the ability to attend over long and generally continuous periods of time for the purpose of detecting and responding to relevant stimuli'.[12]

VIGILANCE IN RELATION TO ATTENTION, AROUSAL, AND ALERTNESS

Attempts to define vigilance often contain the terms alertness, attention, sustained attention, or arousal. The use of any of these three words illustrates differences in interpretation and quantification of vigilance between fields of science. Oken and colleagues,[13] who stated that a first concept of vigilance is used by animal behavior scientists and psychiatrists, interpreting vigilance as being alert for threats or dangers. Psychologists and cognitive neuroscientists define vigilance in a second manner, i.e. as the ability to sustain attention to a task for a period of time. Finally, clinical neurophysiologists and sleep scientists tend to restrict vigilance to the arousal level on the sleep-wake spectrum. These different interpretations of vigilance suggest that reflection on what attention, alertness, and arousal are, is essential before a common definition of vigilance can be attempted.

Alertness

Posner described 'alerting' as achieving and maintaining a state of high sensitivity to incoming stimuli.[14] A recent paper defined alertness not so much as a state, but as the capacity of the mind at a particular moment to respond appropriately to external and internal stimuli.[15] We will return to this difference later.

Alertness is sometimes subdivided in tonic and phasic alertness. Schmidt et al.[12] describe tonic alertness as a slow, diurnal fluctuation in wakefulness and performance, and phasic alertness as a sudden increase in attentiveness, which immediately follows a stimulus requiring a rapid response. Such a stimulus can originate from surrounding factors (external, e.g. a fast approaching car) or from within the body itself (internal, e.g. pain). The term 'tonic alertness', sometimes called intrinsic alertness, has been proposed as a synonym for sustained attention or vigilance by Oken et al.[13] Some researchers intertwine the terms alertness and attention even further: in a paper about intrinsic

and phasic alertness, Sturm and Willmes mention that intensity aspects of attention include both alertness, both intrinsic and phasic, and sustained attention. According to these authors, these intensity aspects form the basis that underlies selective aspects of attention, i.e. orienting and executive attention.[16]

We extract the following concept of alertness from the literature above: alertness is defined as a quantitative description of the state of the mind, characterized by being *sensitive* to incoming stimuli. Following this definition, any method to assess whether incoming stimuli are processed, can be used to quantify alertness. Examples include basic neurophysiological measures such as event-related potentials, but also attention tasks, as alertness is expressed through, and a requisite for attention. Alertness is broken down in two parts: one concerns an intrinsic aspect that slowly fluctuates over time ('tonic alertness'), and the other a stimulus-modulated aspect that may change more quickly ('phasic alertness') (Figure 1A).

Attention and sustained attention

Posner and Petersen[17] proposed a model of attention consisting of three major functions: firstly, orienting to sensory events and directing attention spatially to important stimuli; secondly, detecting signals that need to be thoroughly and consciously processed; and thirdly, the ability to prepare and sustain alertness to process high-priority signals. The latter function, also referred to as sustained attention, is used interchangeably with vigilance by some.[18] Stuss et al.[19] defined sustained attention as a form of executive control that involves monitoring the activation of task-relevant brain areas, re-energizing activated areas when they are low, and inhibiting irrelevant brain areas if they become inappropriately selected. Robertson and Garavan summarized this as the ability to maintain a goal-directed focus in a context of a repetitive, nonarousing nature that provides little external stimulation.[20] In a paper, Robertson and colleagues referred to this capability as endogenous modulation of alertness (self-sustained attention), which they distinguished from exogenously controlled alertness, i.e. alertness driven by factors such as novelty, salience and stimulus change.[21] This distinction is also referred to as top-down (endogenous) versus bottom-up (exogenous) attentional control.[22] Though distinctive aspects, exogenous attentional control can influence endogenous attentional control, and the direction of this influence is depending of stimulus predictability.[23]

The term 'alertness' was mentioned in descriptions of attentional processes in the literature cited above, indicating that attention and alertness are closely related. They are distinguished by the direction towards a specific stimulus that characterizes attention but not alertness (Figure 1B). Hence, attention requires some executive control: being prepared to process incoming stimuli there, where stimuli can be expected (e.g. a task). In other words, alertness corresponds to an implicit sensitivity to unpredictable incoming stimuli, while attention refers to an explicit focus on a certain stimulus.

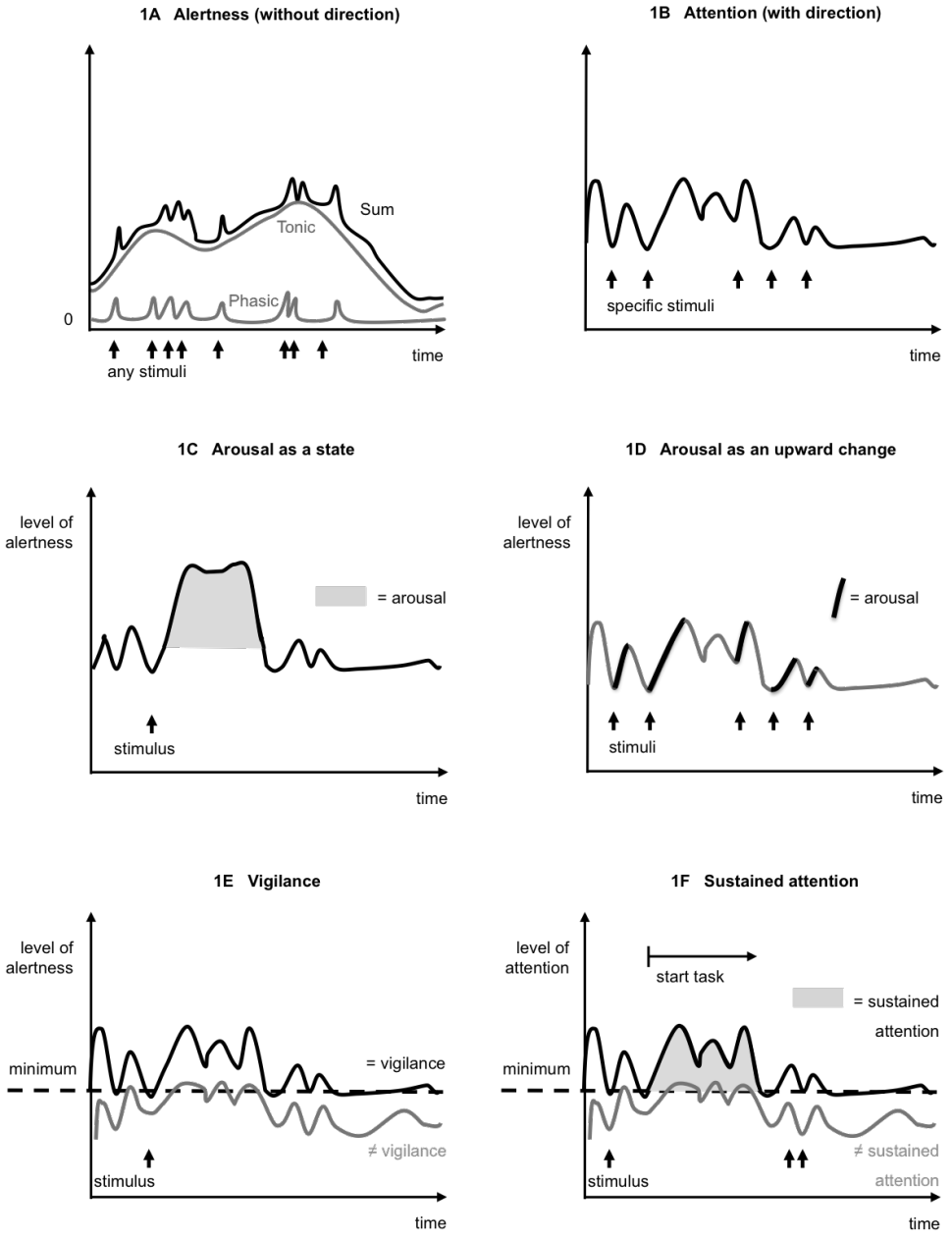


Figure 1 – Illustration of intensity and time aspects of the reviewed terms. A. Alertness: The grey lines represent the tonic and phasic components, i.e. the slowly fluctuating level and stimulus-provoked, temporary increase. The black line combines these aspects; it illustrates the quantitative dimension the state of sensitivity of the mind to incoming stimuli. **B.** Attention: this figure resembles the figure of alertness, as attention also comprises a quantitative dimension, fluctuating over time. Attention differs from alertness in that it has a direction (not drawn): there,

where stimuli can be expected. **C**. Arousal as a state is represented by the shaded part of the figure; it corresponds to the mechanism underlying a prolonged period with a higher level of alertness compared to an implicit baseline. **D.** Arousal as an upward change is represented by the blackened part of the line; it corresponds to the mechanism underlying an upward change in the level of alertness. **E.** Vigilance: the black line corresponds to someone with the capability to maintain a certain level of alertness over a period of time, i.e someone with a normal vigilance. In contrast, the grey line corresponds to someone without such capability, i.e. someone with a disturbed vigilance. **F.** Sustained attention: this figure closely resembles the figure of vigilance: the black line corresponds to someone with the capability to maintain a certain level of attention over a period of time for the purpose of successfully completing a task, represented by the shaded part of the figure. The grey line represents someone without such capability. Vigilance differs from sustained attention in the concepts underlying the y-axis, i.e. the quantitative aspect, which in turn reflects the absence of a direction (alertness) or the presence (attention).

We conclude that sustained attention corresponds to a state in which a certain level of attention is purposefully maintained (Figure 1F), adding a time factor to the definition of attention. This implies that attention can also drop below this particular level. This decrease can be prevented by a stimulating environment, reflected by the arrows in figure 1F, as well as by endogenous modulation of attention.

Arousal

Arousal is probably defined as poorly as vigilance itself, although it is less frequently mixed with definitions of alertness or attention. Arousal is often described as the neurobiological mechanism behind vigilance,[24] with low levels corresponding to sleep, high levels to a vigilant state, and too high levels to a 'hyperaroused' state as proposed in models for insomnia.[25] Probably the most common definition of arousal, used by neurophysiologists, is a sudden activation occurring during sleep.[26] Although both definitions link arousal to sleep, there is an essential difference between the two: the first definition describes arousal as a concept with a certain level, ranging from low to high and relative to an implicit baseline depending on the circumstances (Figure 1C), whereas the second definition is limited to an upward change of state, i.e. from any sleep stage to a lighter sleep stage or to the waking state.

In contrast to authors who link arousal to sleep, Moller et al take a different view: they advocated that a low level of alertness should not be regarded as equalling sleepiness. They regard arousal as a mechanism underlying an upward switch between levels of alertness (Figure 1D) rather than between states of consciousness such as sleep or wake.[15] As such, there is a direct causal relation between arousal and alertness. This relation ties in with the use of the term arousal in descriptions of the neurohormonal

response in situations of acute stress: here too the stimulus results in an increased sensitivity to stimuli and increased preparedness to respond, i.e. an increased alertness.[25] Schmidt et al even decided to unite the terms alertness and arousal.[12]

An unambiguous definition of arousal cannot be distilled from these descriptions. Compared to the scientific literature, the linguistic definition is clearer. The Merriam-Webster dictionary refers to the verb 'to arouse', which is described as 'to awaken from sleep' or 'to rouse to action'.[3] The Shorter Oxford English Dictionary[27] describes arousal as 'the action of arousing or being aroused', and 'to arouse' in turn as 'to raise or stir up from sleep or inactivity', 'to stir into activity', or 'to wake up'. Linguistically, arousal thus refers to an upward change in either the sleep/wake state or the activity spectrum, i.e. an awakening or a change towards action.

If the linguistic lead is followed, i.e. arousal signifies an upward change only, then it would be useful to have a word for a downward change. In literature on animal research, this is sometimes called a dearousal.[28] According to the Merriam-Webster dictionary, the antonym of 'to arouse' is 'to lull'.[4] This is defined as 'to cause (someone) to fall asleep, to become sleepy, or to rest', and, interestingly, 'to cause (someone) to feel safe and relaxed instead of careful and alert' or 'to cause to relax vigilance'. As the antonym of 'to arouse' refers to the cause of a drop of alertness or vigilance, the term 'arousal' and its opposite linguistically purely reflect a change of states, rather than a state itself.

One possible solution to deal with the multiple definitions of arousal would simply be to accept that the term is ambiguous and has multiple definitions. Arousal could then both refer to an upward change between states of wakefulness or alertness, or to such a state itself. This will cause confusion, though.

An alternative approach is to reappraise the definition of arousal in the context of alertness, now that we deduced a definition of alertness from the literature. It is likely that the different definitions of arousal have evolved from its linguistic definition as a result of imprecise definitions of vigilance and alertness. Defined as above, alertness is determined by an intrinsic and extrinsic, stimulus-provoked aspect. As such, arousal as a sudden upward change refers to the mechanism behind a stimulus-caused change in alertness, irrespective of the subsequent duration of the increased level of alertness. In other words, whether the increased alertness lasts for half a second or a half a day does not matter: an upward change is called an arousal. A prolonged period with a high level of alertness as a consequence of an arousal could be referred to as 'an aroused state'. If such an aroused state exceeds the level appropriate to the circumstances, this may be called a hyperaroused state.

We here advocate that arousal be defined as an upward change for three reasons: 1) it is biophysically represented by electro-encephalographic changes, 2) it follows the linguistic definition most closely, and 3) it adds to the definition of alertness, explaining the mechanism behind a stimulus-caused change in alertness.

Vigilance, alertness and arousal

There is a striking resemblance between the descriptions of vigilance and alertness. Consider the following examples: 'watchful and prompt to meet danger' (used to explain the meaning of both terms), 'quick to perceive and act' (alertness), 'readiness to detect and respond to changes in the environment' (vigilance), 'a state of high sensitivity to incoming stimuli' (alertness), and 'the capacity of the mind at that moment to respond appropriately to external and internal stimuli' (both terms).

Is alertness a synonym of vigilance? Firstly, some authors indeed seem to regard it as such based on definitions such as 'maintaining a vigilant or alert state'.[17] Others assume a slightly different definition. Secondly, those who differentiate between tonic and phasic alertness, equate vigilance to the 'tonic', intrinsic fluctuation of alertness.[13] This interpretation of vigilance is narrow and not generally accepted; it excludes phasic, stimulus-provoked alertness. Then again, this stimulus-provoked change in the level of alertness can be referred to by the term arousal. Thus, if vigilance were restricted to the intrinsic fluctuation of alertness, arousal and vigilance would be complementary terms in the description of the spectrum of alertness. Thirdly, some do not equate vigilance to either alertness or tonic alertness. Instead, they consider alertness to be a quantitative measure, whereas vigilance refers to the capacity to maintain a sufficiently high level of alertness, i.e. to maintain a level of alertness above a threshold required to detect unpredictable changes in the environment.[23] The definition of vigilance proposed below corresponds to this last view.

WHAT IS VIGILANCE?

The following three aspects of vigilance can be derived from the literature cited above and should be included in a comprehensive definition. Firstly, being vigilant refers to a quality or state of mind, described by keywords such as 'alertly watchful, especially to avoid danger', 'a readiness to detect and respond to changes in the environment', and 'being alert for threats or danger'. Secondly, vigilance is the extent to which someone is vigilant, so it must contain a quantitative aspect. Thirdly, this level is known to decrease over time in a non-stimulating environment, thus including a temporal dimension. The core elements of the term vigilance as defined above, as well as the position of this term in relation to the terms alertness, sustained attention, and arousal, resulted in the following proposal, also summarized in figure 1E.

Definition

Vigilance is defined as the capability to be aware of relevant, unpredictable changes in one's environment, irrespective of whether or not such changes occur. This capability

has two dimensions. The first is quantitative and refers to the level of alertness that is required for being vigilant. The second dimension follows from the fact that vigilance can change over time: it has a temporal dimension.

IMPLICATIONS & CONSEQUENCES FOR MEASUREMENTS OF VIGILANCE

Vigilance is a prerequisite for being able to pay attention. The ability to maintain a high level of attention over a length of time is summarized by the term sustained attention. Sustained attention differs from vigilance: it is directed towards something, whereas vigilance implies alertness to any possible, relevant new happening, i.e. it does not require a specific focus. The ability to restrict attention to such a task, rather than dividing it, is an aspect of attention itself, not of vigilance. The disorders attention deficit hyperactivity disorder (ADHD) and narcolepsy illustrate this difference. ADHD comprises increased distractibility by lack of continued focus to a task, i.e. lack of sustained attention, whereas narcolepsy comprises a disorder in which vigilance is impaired irrespective of a directional aspect. Though outside the scope of this review, this distinction is reflected by anatomical and biochemical differences between these diseases. The attentional disorder is predominantly a dysfunction of the prefrontal cortex, caudate nucleus and cerebellum, and is mainly dependent on dopaminergic and noradrenergic neurotransmitter systems. In contrast, the vigilance disorder narcolepsy is caused by hypothalamic hypocretin cell loss with broader impact on the monoaminergic systems.

Nevertheless, as vigilance is expressed through attention, measurements of vigilance and attention partially overlap: in fact, measuring sustained attention can be regarded as a method to obtain a quantitative assessment of the underlying construct vigilance, at least in one way. The capability to sustain attention presumes the capability to sustained alertness. The opposite is not necessarily true: the inability to sustain attention can also result from the inability to direct attention.

To illustrate this, note that sustained attention is mainly measured by response tasks in which subjects are asked to appropriately detect changes in the environment by means of appearing or changing stimuli. The performance on such tasks is scored using measurements of accuracy, speed or both. This corresponds to the first part of the definition of vigilance provided above, i.e. it refers to the quantitative dimension of vigilance. Finally, tests of sustained attention mostly account for the temporal dimension by extending the measurements over a certain period of time, so that fluctuations in the level of intrinsic alertness affecting attention can become manifest. This corresponds to the view of Oken et al.[13]

The definition of vigilance proposed in this paper does not include the ability to respond to the stimuli or the nature of such responses. This is in accordance with the literature, where vigilance has been defined as the *readiness* to respond without including the response itself, suggesting that the ability to respond is not part of vigilance. However, vigilance can hardly be assessed without evaluating responses, and the ability to respond may with reason be regarded as the very reason to be vigilant. Such considerations suggest that responsiveness might be included in the definition of vigilance, which would turn vigilance into a much broader concept. The consequence of doing so would be that a failure to respond might be interpreted as impaired vigilance, without any indication whether the failure was due to an impairment of vigilance in the restricted sense as defined above or of responding to a stimulus. As this may once again increase confusion we prefer to exclude responsiveness from the definition of vigilance. Measuring vigilance through sustained attention nevertheless requires responding. One should therefore realize that tests of sustained attention will never be specific for impairments in vigilance or attention, and that the results of such tests should always be interpreted in a broader context, evaluating interference of response characteristics.

Conclusions

The definition of vigilance is linked to definitions of alertness, sustained attention and arousal. Before defining vigilance itself, alertness has to be defined.

- Alertness is the quantification of the state of mind sensitive to incoming stimuli.
- Vigilance is defined as the capability to be aware of potential relevant, unpredictable changes in one's environment, including a quantitative dimension, a sufficient level of alertness, and a temporal dimension.
- Attention adds a direction to this capability.
- Sustained attention adds the prolongation of this capability over time.
- Arousal is a qualitatively distinct concept, which describes a sudden, possibly long-standing, upward change in wakefulness or alertness.

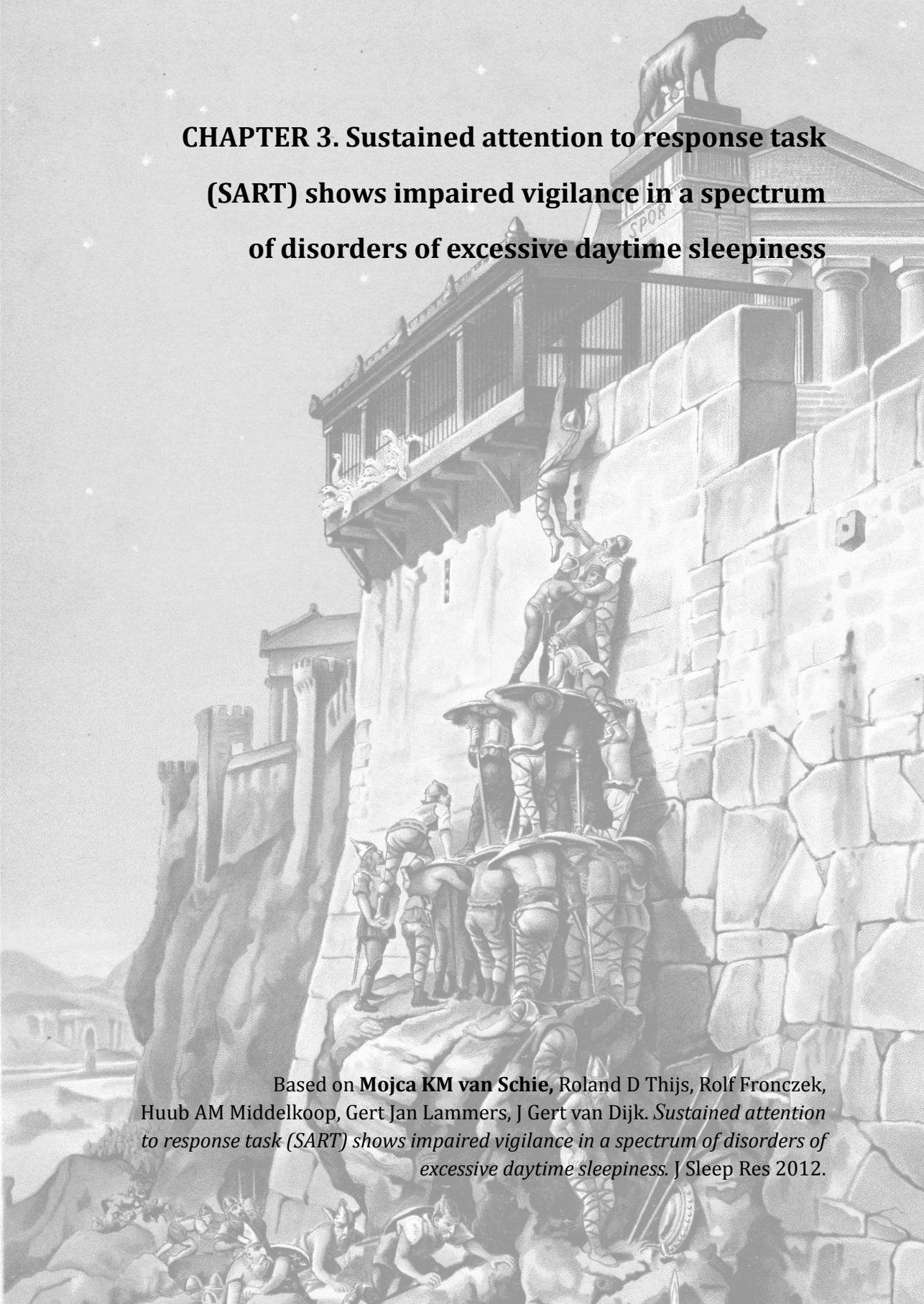
ACKNOWLEDGEMENTS

We thank R.D. Thijs, A. van der Heide, C.E.H.M. Donjacour, and E.P. 't Hart for their advice in the interpretation and definition of the proposed concepts.

REFERENCES

1. Pizza F, Filardi M, Moresco M et al. Excessive daytime sleepiness in narcolepsy and central nervous system hypersomnias. *Sleep Breath 2019*; 24 (2): 605-614.
2. Fronczek R, Middelkoop HA, van Dijk JG, Lammers GJ. Focusing on vigilance instead of sleepiness in the assessment of narcolepsy: high sensitivity of the Sustained Attention to Response Task (SART). *Sleep* 2006; 29 (2): 187-91.
3. Valley V, Broughton R. Daytime performance deficits and physiological vigilance in untreated patients with narcolepsy-cataplexy compared to controls. *Review of electroencephalograpy and clinical neurophysiology* 1981; 11 (1): 133-9.
4. Merriam-Webster.com Dictionary.
5. Mackworth JF, Taylor MM. The d' measure of signal detectability during vigilance-like situations. *Canadian Journal of Psychology* 1963; 17: 302-325.
6. Mackworth NH. Researches on the measurement of human performance. New York: Dover Publications 1961.
7. Green DM, Swets JA. Signal detection theory and psychophysics. New York: NY: Wiley 1966.
8. Sanders AF. Elements of Human Performance: Reaction Processes and Attention in Human Skill. 1st ed. London: Taylor & Francis Ltd 2014.
9. Mackworth JF. Vigilance. *The advancement of Science* 1957; 389-393.
10. Head H. 'Vigilance'; a physiological state of the nervous system. *Brit. J. Psychol.* 1923; 14 (126).
11. Lim J, Dinges DF. Sleep deprivation and vigilant attention. *Ann N Y Acad Sci* 2008; 1129: 305-22.
12. Schmidt C, Collette F, Cajochen C, Peigneux P. A time to think: circadian rhythms in human cognition. *Cognitive Neuropsychology* 2007; 24 (7): 755-89.
13. Oken BS, Salinsky MC, Elsas SM. Vigilance, alertness, or sustained attention: physiological basis and measurement. *Clin Neurophysiol* 2006; 117 (9): 1885-901.
14. Posner MI. Measuring alertness. *Ann N Y Acad Sci* 2008; 1129: 193-9.
15. Moller HJ, Devins GM, Shen J, Shapiro CM. Sleepiness is not the inverse of alertness: evidence from four sleep disorder patient groups. *Exp Brain Res* 2006; 173 (2): 258-66.
16. Sturm W, Willmes K. On the functional neuroanatomy of intrinsic and phasic alertness. *Neuroimage* 2001; 14 (1): S76-84.
17. Posner MI, Petersen SE. The attention system of the human brain. *Annu Rev Neurosci* 1990; 13: 25-42.
18. Parasuraman R, Warm JS, See JE. Brain systems of vigilance: *The attentive brain,* Cambridge, MA: The MIT Press 1998: pp. 221-56.

19. Stuss DT, Shallice T, Alexander MP, Picton TW. A multidisciplinary approach to anterior attentional functions. *Ann N Y Acad Sci* 1995; 769: 191-211.
20. Robertson IH, Garavan H. Vigilant attention: In Gazzaniga M (ed.) *The Cognitive Neurosciences,* 3rd. Cambridge, Massachusetts: Bradford Books, The MIT Press 2004: pp. 631-640.
21. Robertson IH, Manly T, Andrade J, Baddeley BT, Yiend J. 'Oops!': performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. *Neuropsychologia* 1997; 35 (6): 747-58.
22. Sarter M, Givens B, Bruno JP. The cognitive neuroscience of sustained attention: where top-down meets bottom-up. *Brain Res Rev* 2001; 35 (2): 146-60.
23. MacLean KA, Aichele SR, Bridwell DA, Mangun GR, Wojciulik E, Saron CD. Interactions between endogenous and exogenous attention during vigilance. *Atten Percept Psychophys* 2009; 71 (5): 1042-1058.
24. Stahl SM. Disorders of Sleep and Wakefulness and Their Treatment: In Stahl SM (ed.) *Stahl's essential psychopharmacology,* 3rd. New York: Cambridge University Press 2008.
25. Riemann D, Spiegelhalder K, Feige B, Voderholzer U, Berger M, Perlis M, et al. The hyperarousal model of insomnia: a review of the concept and its evidence. *Sleep Med Rev* 2010; 14 (1): 19-31.
26. Kryger MH, Roth T, Dement WC. Principles and practice of sleep medicine. 4th ed. Philadelphia: Elservier Inc. 2005.
27. Little W, Fowler HW, Coulson J, Onions CT. The Shorter Oxford English Dictionary: Oxford university press, London, 1963.
28. Estanislau C, Veloso AWN, Filgueiras GB et al. Rat self-grooming and its relationships with anxiety, dearaousal and perseveration: Evidence for a self-grooming trait. Physiol. Behav. 2019; 209: 112585. doi: 10.1016



CHAPTER 3. Sustained attention to response task (SART) shows impaired vigilance in a spectrum of disorders of excessive daytime sleepiness

Based on **Mojca KM van Schie,** Roland D Thijs, Rolf Fronczek, Huub AM Middelkoop, Gert Jan Lammers, J Gert van Dijk. *Sustained attention to response task (SART) shows impaired vigilance in a spectrum of disorders of excessive daytime sleepiness.* J Sleep Res 2012.

SUMMARY

The Sustained Attention to Response Task (SART) comprises withholding key presses to 1 in 9 of 225 target stimuli; it proved to be a sensitive measure of vigilance in a small group of narcoleptics. We studied SART results in 96 patients from a tertiary narcolepsy referral centre. Diagnoses according to ICSD-2 criteria were narcolepsy with (n = 42) and without cataplexy (n = 5), idiopathic hypersomnia without long sleep time (n = 37), and obstructive sleep apnoea syndrome (n = 12). The SART was administered prior to each of 5 MSLT sessions. Analysis concerned error rates, mean reaction time (RT), RT variability and post-error slowing, as well as the correlation of SART results with mean latency of the Multiple Sleep Latency Test (MSLT) and possible time of day influences. Median SART error scores ranged from 8.4 to 11.1, and mean RTs from 332 to 366 ms. SART error score and mean RT did not differ significantly between patient groups. SART error score did not correlate with MSLT sleep latency. RT was more variable as the error score was higher. SART error score was highest for the first session. We conclude that a high SART error rate reflects vigilance impairment in excessive daytime sleepiness irrespective of its cause. The SART and the MSLT reflect different aspects of sleep/wakefulness and are complementary.

INTRODUCTION

Excessive daytime sleepiness (EDS), an increased tendency or need to fall asleep (International Classification of Sleep Disorders, 2005), is the key symptom of many sleep disorders. In narcolepsy, EDS is accompanied by vigilance impairment, leading to impaired performance in the waking state (Broughton *et al.*, 1982). A previous small study by our sleep lab showed that impaired vigilance in narcolepsy could be quantified with the Sustained Attention to Response Task (SART) ((Fronczek *et al.*, 2006). This test is explained in the methods section; in short, it involves pressing a key when a number (1 to 9) appears on a screen except when that number is a 3. The main outcome is the total error score, consisting of both key presses when no key should be pressed and the reverse. In our previous study the total error score had shown excellent sensitivity (87%) and specificity (100%) in a comparison of 15 patients with narcolepsy and cataplexy and 15 healthy controls. The diagnostic yield was in fact as good as that of the Multiple Sleep Latency Test (MSLT). The mean SART error score was not related to the mean MSLT sleep latency, suggesting that the two approaches measure different aspects: the SART, requiring prolonged attention, likely reflects impaired vigilance, while the MSLT measures the propensity to fall asleep quickly.

Based on these promising results, we investigated vigilance impairment with the SART in a prospective sample of patients with various causes of EDS. The study focused on the mean error score as the mean parameter of interest. Additional questions were the correlation between SART and MSLT results, analysis of reaction time (RT) data and possible time-of-day influences on test results.

MATERIALS AND METHODS

Patients

Our department is a tertiary referral centre for suspected hypersomnias of central origin. The routine work-up comprised a diagnostic interview, nocturnal polysomnography the night before MSLT, SART and MSLT for all patients.

Patients with a complaint of EDS referred for suspected hypersomnias of central origin were included if diagnosed with a sleep disorder based on ICSD-2 criteria (International Classification of Sleep Disorders, 2005). Any other ancillary investigations necessary for ICSD-2 diagnosis were not investigated in the present study. Patients evaluated for driving ability were excluded from this study, as were patients who used stimulants on the day of testing.

In accordance with Dutch law, patients gave oral consent to this routine work-up. The research has not been presented to an Ethics Committee for a review, as Dutch

law states that this is not necessary for research with data that were originally gathered for patient care and afterwards anonymously used by their treating physicians in retrospective case-studies.

The MSLT consisted of five sessions of 20 minutes with a 1.5-hour break in between. The first session started at 09:00 hrs (Carskadon *et al.*, 1986).

Sustained Attention to Response Task

Patients underwent a 4-minute-20-second taking SART session prior to each of five MSLT sessions during one day as described previously (Fronczek *et al.*, 2006). The SART was administered while subjects were seated in front of a computer screen in a quiet room. Before the first session all subjects performed a short version of the SART to become familiar with the test. Between sleep latency tests, participants were allowed to go for short walks in the hospital and eat or drink, but they were not allowed to sleep or use stimulating agents.

In short, the SART comprises the numbers 1 to 9 appearing 225 times in random order and in different sizes in a white font on a black computer screen. Subjects had to respond to the appearance of each number by pressing a button except when the number was a 3, which occurred 25 times in all. Subjects had to press the button before the next number appeared and were instructed to give equal importance to accuracy and speed in performing the task (Manly *et al.*, 1999; Robertson *et al.*, 1997). The primary outcome measure of the SART is the total error score, consisting of, firstly, key presses when no key should be pressed (i.e., after a '3', a so-called 'no-go trial': commission errors) and secondly absent presses when a key should have been pressed (i.e., after anything but a '3', the so-called 'go trials': omission errors).

The appearance of the numbers on a cathode ray tube screen was timed using a dedicated video graphics array switch to avoid delays of uncertain magnitude due to build-up of screen data. The resulting maximal uncertainty was 10 milliseconds, allowing reaction times (RT) to be measured with sufficient accuracy. The following measures of response accuracy were assessed: the number of commission errors, with a maximum of 25; the number of omission errors, with a theoretical maximum of 200 errors. The 'SART error score' represents the sum of the numbers of commission and omission errors.

The following measures of reaction time (RT) were assessed (Picton *et al.*, 2007; Stuss *et al.*, 2003): These were the mean RT in ms, calculated over correct response trials, i.e., key presses after anything but a '3'. RT variability was quantified as the coefficient of variation of RT for correct response trials: this is the standard deviation divided by the mean RT of that test. RT often increases temporarily following a commission error in various choice-response tasks (Dudschig and Jentzsch, 2009; Notebaert *et al.*, 2009; Rabbitt, 1966). This 'post-error slowing' was calculated as follows: The last RT before a commission error was noted as was the first one after it. The difference between the later and the earlier RT was divided by the mean RT of that session.

Statistical analysis

Differences between groups were studied for all six SART outcome measures. For each subject, the means of 5 MSLT sleep latencies and SART scores were computed and used in the analysis. The non-parametric Kruskal-Wallis test was employed to compare SART outcome measures between groups.

The correlation between mean MSLT sleep latency and the SART error score was assessed using Spearman's ρ, since variables were not normally distributed. Correlations of MSLT sleep latency with the numbers of commission and omission errors were assessed.

A higher number of commission errors has previously been demonstrated to correlate with shorter RT (Helton *et al.*, 2010; Manly *et al.*, 1999; Robertson *et al.*, 1997; Shalgi *et al.*, 2007), indicating that errors on the SART should be interpreted along with RT measures. Therefore, the correlation between SART RT measures and SART accuracy measures was assessed using Spearman's ρ. A paired t-test was used to determine whether RT before and after a '3' differed significantly differed, both for a correct response (no key press) and for a commission error.

Effects of testing time on SART error score were evaluated using the Friedman-Test with post-hoc analysis after Conover (Conover, 1980).

RESULTS

Patients

One-hundred twelve patients who were evaluated for EDS between February 2006 and March 2010 met the inclusion criteria. Sixteen of these were excluded due to simultaneous evaluation of driving ability. Ninety-six patients were included, of whom 16 used antidepressants.

Patients were classified in four groups based on diagnosis: narcolepsy with and without cataplexy, idiopathic hypersomnia without long sleep time, and obstructive sleep apnoea syndrome (OSAS) (Table 1). Patient groups differed significantly for the ratio of men to women ($p < 0.02$) and age ($p < 0.01$). The narcolepsy-without-cataplexy group consisted entirely of men. OSAS patients were significantly older than those in other groups. Patient groups did not differ in the proportion of patients who continued the use of anti-depressant medication on the day of MSLT and SART ($p > 0.52$).

SART results

SART and MSLT data for each patient group are presented in Table 2. Neither the median SART error score ($H(3) = 1.66$ $p > 0.64$) nor the numbers of commission ($H(3) = 2.15$ $p > 0.54$) and omission errors ($H(3) = 2.69$ $p > 0.44$) differed significantly between groups.

The same held for mean RT ($p > 0.55$), RT variability ($p > 0.65$) and post-error slowing ($p > 0.41$). After correction for age and gender, SART accuracy measures and RT measures still did not differ between groups. Figure 1 illustrates SART error score findings per group next to findings in healthy controls from our previous study (Fronczek *et al.*, 2006) illustrating the magnitude of the differences in sleep latencies on MSLT and lower error scores on SART compared to all patient groups, while not serving as a direct comparison with the groups in the current study.

Fig. 1a MSLT sleep latencies of patient groups and controls

Fig. 1b SART error scores of patient groups and controls

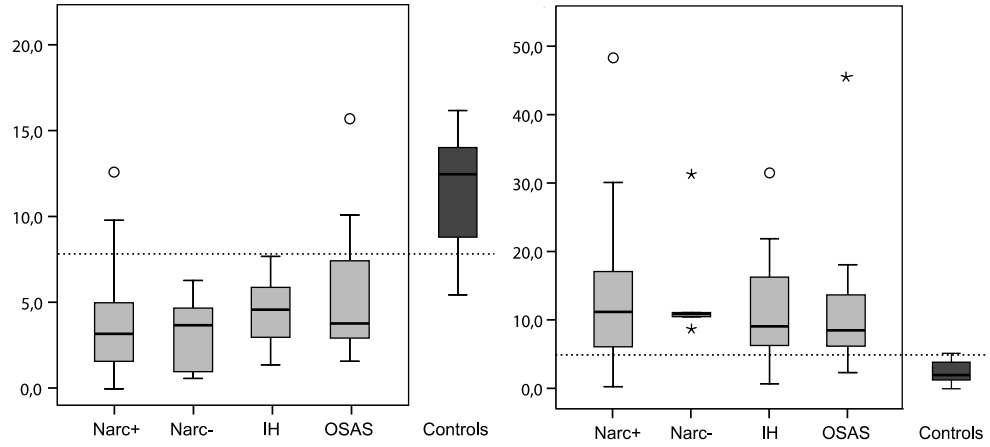


Figure 1– Mean of five Multiple Sleep Latency Test sleep latencies in minutes (a) and Sustained Attention to Response Task error scores (b) on the assessment day for all patients, with separate box plots representing each patient group, and for controls from our pilot study.
Narc+: narcolepsy with cataplexy; Narc-: narcolepsy without cataplexy; IH: Idiopathic Hypersomnia; OSAS: Obstructive Sleep Apnea Syndrome. The horizontal dotted lines represent cut-off values for MSLT (8 minutes) and SART (5 errors) (Fronczek *et al.*, 2006).

Table 1 – Group characteristics

Patient group	N	Age in years (SD)	% of males	Antidepressant use (N)
Narcolepsy with cataplexy	42	37.5 (19.1)	50.0	7
Narcolepsy without cataplexy	5	43.8 (20.3)	100.0	2
Idiopathic Hypersomnia	37	44.2 (15.0)	43.2	5
OSAS	12	57.4 (14.5)	83.3	2
Total	96	42.9 (18.0)	54.2	

Legend: OSAS: Obstructive Sleep Apnea Syndrome.

Table 2 - MSLT and SART outcome measures by patient group

	Narc. + cata (n=42)		**Narc. - cata (n=5)**		**IH (n=37)**		**OSAS (n=12)**	
MSLT	*Median*		*Median*		*Median*		*Median*	
	25th-75th perc.		*25th-75th perc.*		*25th-75th perc.*		*25th-75th perc.*	
latency (min)	3.2	1.5-5.0	3.0	1.2-5.6	4.6	3.0-5.9	3.8	2.9-7.9
REM (n)	2	1-4	4	4-5	0	0-0	0	0-0
SART accuracy meas.	*Median*		*Median*		*Median*		*Median*	
	25th-75th perc.		*25th-75th perc.*		*25th-75th perc.*		*25th-75th perc.*	
comm./25	8.1	4.6-12.6	9.8	8.5-15.5	7.4	4.9-11.3	8.1	6.1-10.3
omiss./200	1.4	0.4-4.4	1.0	0.4-6.2	1.0	0.2-2.7	0.6	0.0-2.3
err.sc./225	11.1	6.0-17.4	10.8	9.5-21.1	9.0	5.9-16.3	8.4	6.1-14.3
SART RT meas.	*Mean*	*(SD)*	*Mean*	*(SD)*	*Mean*	*(SD)*	*Mean*	*(SD)*
mean RT (ms)	337	(83)	332	(74)	359	(82)	366	(87)
RT var. (ms)	0.30	(0.10)	0.32	(0.14)	0.29	(0.07)	0.28	(0.07)
PES (ms)	0.16	(0.22)	0.03	(0.21)	0.11	(0.15)	0.10	(0.17)

Legend: Narc.: Narcolepsy; +/- cata: with/without cataplexy; IH: Idiopathic Hypersomnia; OSAS: Obstructive Sleep Apnea Syndrome; latency: MSLT sleep latency time; min: minutes; REM: number of MSLT sessions in which REM-sleep was recorded; meas: measures; comm.: total number of commission errors; omiss.: total number of omission errors; err.sc.: SART error score; RT var.: RT variability; PES: post-error slowing, perc.: percentiles.

SART-MSLT correlates

No significant correlation was found between SART error score and mean MSLT sleep latency for all 96 patients together, or between the number of SART commission errors and MSLT sleep latency. The number of omission errors, however, was significantly and inversely correlated with MSLT sleep latency ($r_s = -0.24$, $p < 0.02$).

Analysis of SART outcome measures

Greater RT variability was strongly associated with higher SART error score ($r_s = 0.76$, $p < 0.01$). No significant correlation between mean RT and SART error score was found ($r_s = -0.07$, $p = 0.5$), but mean RT was significantly correlated with the number of commission errors ($r_s = -0.38$, $p < 0.01$).

Making errors influenced RT (Figure 2). After a correctly withheld no-go trial (a '3'), a significant decrease in RT occurred ($p < 0.01$). However, when a '3' was incorrectly responded to by a key press, the RT afterwards was increased ($p < 0.01$).

Fig. 2a Reaction times before and after a correct response (no key press)

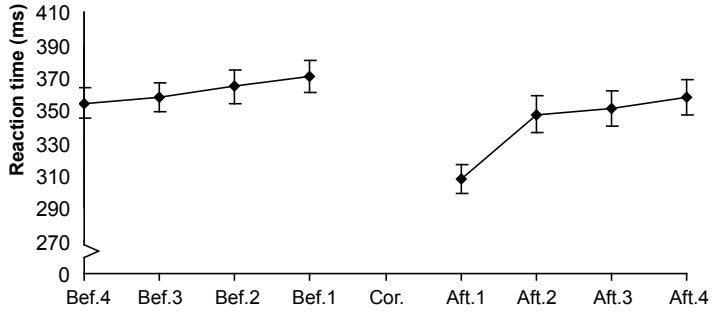


Fig. 2b Reaction times before and after a commission error

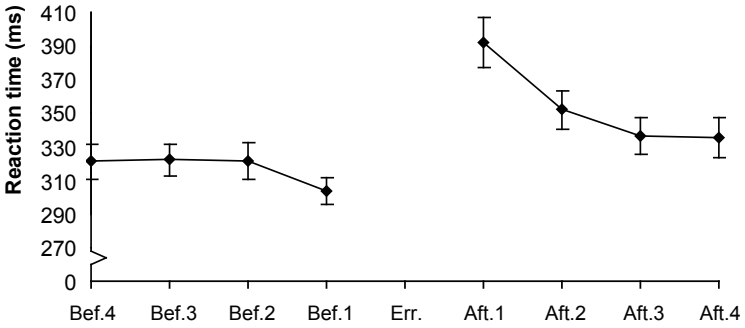


Figure 2– Mean Sustained Attention to Response Task reaction times on 4 trials before and 4 trials after a correct withhold (a) and a commission error (b) on a no-go trial.
Bef.4 = 4th trial before a no-go trial. Aft.1= 1st trial after a no-go trial. Cor. = correct withhold. Err. = commission error. Error bars indicate SEM.

Time-of-day influences on SART error score

SART error scores differed significantly between sessions ($p < 0.02$), with session 1 showing the highest error scores. Post-hoc testing showed that SART error score for the first session differed significantly with that of all but the third sessions. The diurnal variation of SART error score per patient group is shown in figure 3.

SART error score fluctuations between sessions

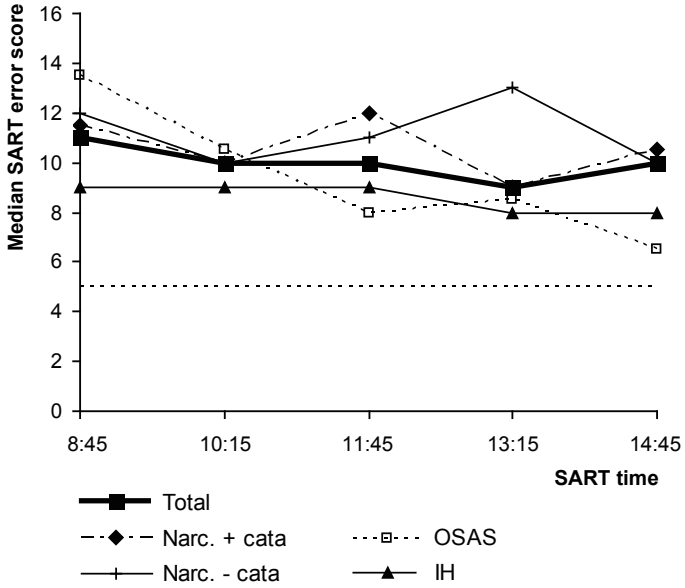


Figure 3 – Median Sustained Attention to Response Task error score at 5 testing times on one day. SART error score was significantly higher at the 8:45 session than at other sessions except 11:45. Narc.: Narcolepsy; +/- cata: with/without cataplexy; IH: Idiopathic Hypersomnia; OSAS: Obstructive Sleep Apnea Syndrome. The horizontal dotted line at SART error score of 5 errors marks the cut-off value for abnormal SARTs previously found in a study of narcoleptics and healthy controls (Fronczek *et al.*, 2006).

DISCUSSION

In this cross-sectional study, we investigated the SART as a tool to measure vigilance in patients with different causes of EDS. The main finding was a high SART error rate in all patient groups. Patient groups did not differ in median SART error score or in other SART outcome measures. This confirms our previous suggestion that the SART error score does not reflect any specific disease entity. Instead, SART error score probably reflects a key symptom of all sleep-related disorders that were studied, i.e., vigilance impairment.

SART-MSLT correlates

Regarding SART-MSLT correlates, we did not find a significant correlation between SART error score and MSLT sleep latency. The lack of a relationship confirms our previous findings (Fronczek *et al.*, 2006) and underlines that the SART and the MSLT measure different phenomena, i.e., the SART is a distinct parameter of disease burden in sleep

disorders. We expanded the study to include commission and omission errors. Somewhat surprisingly, the number of omission errors correlated significantly and inversely with MSLT sleep latency; in other words: those who often failed to press a key after frequently occurring stimuli fell asleep quickly. This correlation resembles the one found between MSLT sleep latency and cumulative increase in performance lapses at the psychomotor vigilance task (PVT) (Carskadon and Dement, 1981; Dinges *et al.*, 1997). SART omission errors may thus have an analogue in PVT attentional lapses. However, the SART, unlike the PVT, requires subjects to decide whether a response is needed or not for each stimulus. As such, the SART, in including response inhibition, assesses an additional component of vigilance (Helton, 2009; Manly *et al.*, 1999; O'Connell *et al.*, 2009).

Impaired SART parameters and sleep disorders

What could be the mechanism for impaired SART parameters in sleep disorders? Molenberghs *et al.* (2009) found indications for decreases in SART scores in patients with frontal brain deficits. These indications of associations between SART parameters and frontal brain regions may reflect the decision-making component of the SART. Falling asleep is associated with inhibiting basal forebrain and brainstem arousal systems, the former being responsible for excitatory projections into, among others, the frontal cortex (Sherin *et al.*, 1998; Steininger *et al.*, 1999; Uschakov *et al.*, 2007). Through these projections, sleep disorders with hypersomnia may affect frontal cortical arousal. In fact, frontal cortical gray matter loss and frontal brain activation changes have been found in various sleep disorders (Ayalon *et al.*, 2006; Ayalon *et al.*, 2009; Brenneis *et al.*, 2005; Kaufmann *et al.*, 2002; Sherin *et al.*, 1998; Steininger *et al.*, 1999; Uschakov *et al.*, 2007). Hence, from a speculative point of view, high error score found in sleep disorders might be explained by changes in frontal functions. However, neuroimaging studies of patients with sleep disorders are needed to help address this hypothesis.

RT measurements in the SART

Correlating SART error score with RT measures showed that errors were more frequent when RT variability was larger. This latter finding may simply reflect impaired vigilance, which can lead to lapses of attention, causing fluctuations in RT as well as an increase in both omission errors (severe response 'lapses') and commission errors (Braver *et al.*, 2003; Duncan *et al.*, 1996; Stuss *et al.*, 1995; Wilkins *et al.*, 1987).

In agreement with the literature (Helton *et al.*, 2010; Manly *et al.*, 1999; Robertson *et al.*, 1997; Shalgi *et al.*, 2007), the number of commission errors was inversely correlated with mean RT. This has previously been explained as the so-called speed-accuracy trade-off, which presumes an influence of task strategy on the SART error score. Therefore, it would be valuable to know whether different instructions affect SART performance in EDS and whether equal importance is given to speed and accuracy.

Time-of-day influences on SART error score

We found diurnal effects on SART performance with the highest SART error score at the first session in the morning. This higher error score may still reflect a brief learning period or an underlying time-of-day effect and thus requires further study.

Strengths and limitations

A strength of this study is the large size of the patient group with several causes of EDS. In addition this study focused on reaction time as well as the more frequently studied accuracy measurements.

The most important limitations are the lack of individually matched controls and limited size of some of the groups. However, we did confirm the high SART error score data in a larger cohort of narcolepsy patients and demonstrated that SART outcome measures do not differ between patients with various causes of EDS.

CONCLUSION

Vigilance, as quantified by the SART, is as impaired in narcolepsy, as in other EDS causes. Yielding different results, SART is complementary to MSLT and does not only provide information about reaction time, as PVT does, but also about the capability of decision-making. The SART is easy to administer, cheap, and takes little time to perform. Combining SART and MSLT has the potential to become an important tool in clinical practice, as the combined approach yields information not only about the propensity to fall asleep, but also about performance in the waking state. Further studies are needed to assess the correlation between subjective impression of vigilance and objective SART outcome, and to probe the ability of the SART to measure treatment effect.

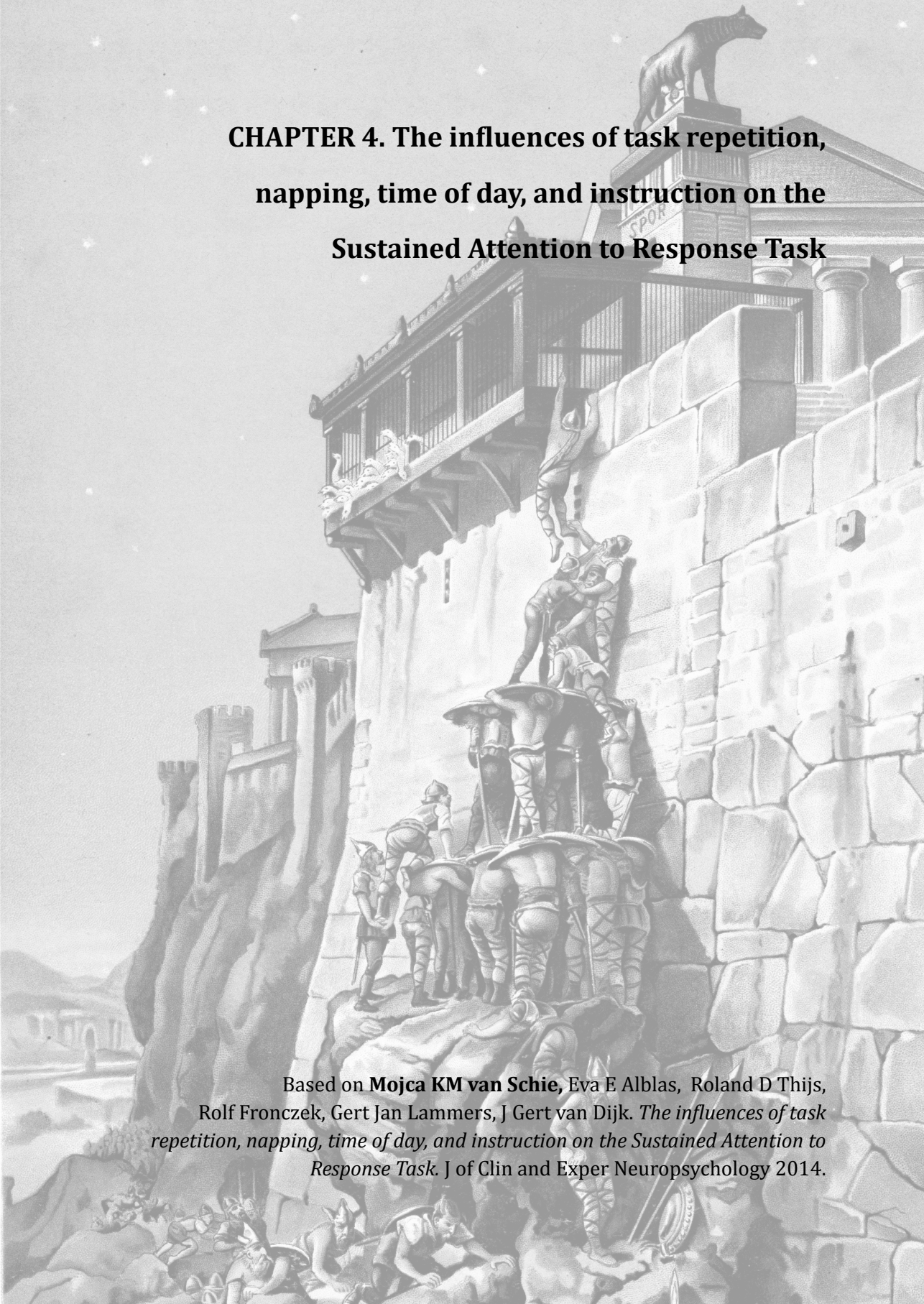
ACKNOWLEDGEMENTS

We thank P.J. van Someren, J.G. van Vliet – de Regt, M.J. Stijl – Pek, F.I. Kerkhof, G.H. van Beukering - Louwes, W.W.M. Huyser, J.J.M. Witteman, and R.H.A.M. Reijntjes for their help in acquiring the data; and N.A. Aziz for statistical advise.

REFERENCES

1. Ayalon, L., Ancoli-Israel, S., Aka, A. A., McKenna, B. S. and Drummond, S. P. Relationship between obstructive sleep apnea severity and brain activation during a sustained attention task. *Sleep*, 2009, 32: 373-381.
2. Ayalon, L., Ancoli-Israel, S., Klemfuss, Z., Shalauta, M. D. and Drummond, S. P. Increased brain activation during verbal learning in obstructive sleep apnea. *Neuroimage.*, 2006, 31: 1817-1825.
3. Braver, T. S., Reynolds, J. R. and Donaldson, D. I. Neural mechanisms of transient and sustained cognitive control during task switching. *Neuron*, 2003, 39: 713-726.
4. Brenneis, C., Brandauer, E., Frauscher, B. et al. Voxel-based morphometry in narcolepsy. *Sleep Med.*, 2005, 6: 531-536.
5. Broughton, R., Low, R., Valley, V., Da, C. B. and Liddiard, S. Auditory evoked potentials compared to performance measures and EEG in assessing excessive daytime sleepiness in narcolepsy-cataplexy. *Electroencephalogr. Clin. Neurophysiol.*, 1982, 54: 579-582.
6. Carskadon, M. A. and Dement, W. C. Cumulative effects of sleep restriction on daytime sleepiness. *Psychophysiology*, 1981, 18: 107-113.
7. Carskadon, M. A., Dement, W. C., Mitler, M. M. et al. Guidelines for the multiple sleep latency test (MSLT): a standard measure of sleepiness. *Sleep*, 1986, 9: 519-524.
8. Conover, W. J. *Practical Nonparametric Statistics*. John Wiley, New York, 1980 (second edition).
9. Dinges, D. F., Pack, F., Williams, K. et al. Cumulative sleepiness, mood disturbance, and psychomotor vigilance performance decrements during a week of sleep restricted to 4-5 hours per night. *Sleep*, 1997, 20: 267-277.
10. Dudschig, C. and Jentzsch, I. Speeding before and slowing after errors: is it all just strategy? *Brain Res.*, 2009, 1296: 56-62.
11. Duncan, J., Emslie, H., Williams, P., Johnson, R. and Freer, C. Intelligence and the frontal lobe: the organization of goal-directed behavior. *Cogn Psychol.*, 1996, 30: 257-303.
12. Fronczek, R., Middelkoop, H. A., van Dijk, J. G. and Lammers, G. J. Focusing on vigilance instead of sleepiness in the assessment of narcolepsy: high sensitivity of the Sustained Attention to Response Task (SART). *Sleep*, 2006, 29: 187-191.
13. Helton, W. S. Impulsive responding and the sustained attention to response task. *J. Clin. Exp. Neuropsychol.*, 2009, 31: 39-47.
14. Helton, W. S., Weil, L., Middlemiss, A. and Sawers, A. Global interference and spatial uncertainty in the Sustained Attention to Response Task (SART). *Conscious. Cogn*, 2010, 19: 77-85.
15. International Classification of Sleep Disorders *Diagnostic and Coding Manual*. American Academy of Sleep Medicine, Westchester, IL, 2005 (second edition).

16. Kaufmann, C., Schuld, A., Pollmacher, T. and Auer, D. P. Reduced cortical gray matter in narcolepsy: preliminary findings with voxel-based morphometry. *Neurology*, 2002, 58: 1852-1855.
17. Manly, T., Robertson, I. H., Galloway, M. and Hawkins, K. The absent mind: further investigations of sustained attention to response. *Neuropsychologia*, 1999, 37: 661-670.
18. Molenberghs, P., Gillebert, C. R., Schoofs, H. et al. Lesion neuroanatomy of the Sustained Attention to Response task. *Neuropsychologia*, 2009, 47: 2866-2875.
19. Notebaert, W., Houtman, F., Opstal, F. V. et al. Post-error slowing: an orienting account. *Cognition*, 2009, 111: 275-279.
20. O'Connell, R. G., Dockree, P. M., Bellgrove, M. A. et al. Two types of action error: electrophysiological evidence for separable inhibitory and sustained attention neural mechanisms producing error on go/no-go tasks. *J. Cogn Neurosci.*, 2009, 21: 93-104.
21. Picton, T. W., Stuss, D. T., Alexander, M. P. et al. Effects of focal frontal lesions on response inhibition. *Cereb. Cortex*, 2007, 17: 826-838.
22. Rabbitt, P. M. Errors and error correction in choice-response tasks. *J. Exp. Psychol.*, 1966, 71: 264-272.
23. Robertson, I. H., Manly, T., Andrade, J., Baddeley, B. T. and Yiend, J. 'Oops!': performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. *Neuropsychologia*, 1997, 35: 747-758.
24. Shalgi, S., O'Connell, R. G., Deouell, L. Y. and Robertson, I. H. Absent minded but accurate: delaying responses increases accuracy but decreases error awareness. *Exp. Brain Res.*, 2007, 182: 119-124.
25. Sherin, J. E., Elmquist, J. K., Torrealba, F. and Saper, C. B. Innervation of histaminergic tuberomammillary neurons by GABAergic and galaninergic neurons in the ventrolateral preoptic nucleus of the rat. *J. Neurosci.*, 1998, 18: 4705-4721.
26. Steininger, T. L., Alam, M. N., Gong, H., Szymusiak, R. and McGinty, D. Sleep-waking discharge of neurons in the posterior lateral hypothalamus of the albino rat. *Brain Res.*, 1999, 840: 138-147.
27. Stuss, D. T., Murphy, K. J., Binns, M. A. and Alexander, M. P. Staying on the job: the frontal lobes control individual performance variability. *Brain*, 2003, 126: 2363-2380.
28. Stuss, D. T., Shallice, T., Alexander, M. P. and Picton, T. W. A multidisciplinary approach to anterior attentional functions. *Ann. N. Y. Acad. Sci.*, 1995, 769: 191-211.
29. Uschakov, A., Gong, H., McGinty, D. and Szymusiak, R. Efferent projections from the median preoptic nucleus to sleep- and arousal-regulatory nuclei in the rat brain. *Neuroscience*, 2007, 150: 104-120.
30. Wilkins, A. J., Shallice, T. and McCarthy, R. Frontal lesions and sustained attention. *Neuropsychologia*, 1987, 25: 359-365.



CHAPTER 4. The influences of task repetition, napping, time of day, and instruction on the Sustained Attention to Response Task

Based on **Mojca KM van Schie,** Eva E Alblas, Roland D Thijs, Rolf Fronczek, Gert Jan Lammers, J Gert van Dijk. *The influences of task repetition, napping, time of day, and instruction on the Sustained Attention to Response Task.* J of Clin and Exper Neuropsychology 2014.

ABSTRACT

Introduction

The Sustained Attention to Response Task (SART) helps to quantify vigilance impairments. Previous studies in which five SART sessions on one day were administered, demonstrated worse performance during the first session compared to the others. The present study comprises two experiments to identify a cause of this phenomenon.

Method

Experiment 1, counting eighty healthy participants, assessed effects of repetition, napping and time of day on SART performance through a between-groups design The SART was performed twice in the morning or twice in the afternoon; half of the participants took a 20-minute nap before the second SART. A strong correlation between error count and reaction time (RT) suggested effects of test instruction. Participants gave equal weight to speed and accuracy in experiment 1; therefore, results of 20 participants were compared to those of 20 additional participants who were told to prefer accuracy (experiment 2).

Results

The average SART error count in experiment 1 was 10.1, the median RT 280 ms. Repetition nor napping influenced error count or RT. Time of day did not influence error count, but RT was significantly longer for morning than afternoon SARTs. The additional participants in experiment 2 had a 49% lower error count and a 14% higher RT than the participants in experiment 1. Error counts reduced by 50% from the first to the second session of experiment 2, irrespective of napping or time of day.

Conclusions

Preferring accuracy over speed was associated with a significantly lower error count. The data suggest that a worse performance in the first SART session only occurs when instructing participants to prefer accuracy, which is caused by repetition, not by napping or time of day.

Note

We advise to instruct participants to prefer accuracy over speed when performing the SART and to include a full practice session.

INTRODUCTION

Vigilance, which can be defined as the capability to be aware of potential changes in one's environment, is a prerequisite for adequate daytime functioning. A low vigilance may lead to cognitive mishaps, e.g. forgetting why you went up the stairs, or reading a piece of text more than once without registering the content. Low vigilance can even be dangerous, e.g. when driving a car.

The sleep disorder narcolepsy is an excellent model of disturbed vigilance (Fronczek, Middelkoop, van Dijk, & Lammers, 2006; Valley & Broughton, 1981; Van Schie et al., 2012). Patients with narcolepsy experience severe vigilance problems in daily life, for instance not being able to recall the content of a conversation, not being able to finish a book, or being unable to concentrate on studies or work. While several wake-promoting drugs aim to improve vigilance and reduce sleepiness, their efficacy depends largely on patient's subjective reports and is difficult to determine objectively.

There are several methods to estimate the level of vigilance, ranging from subjective visual-analog scales through pupillography (Morad, Lemberg, Yofe, & Dagan, 2000) and quantified electro-encephalography (EEG) (Coenen, 1995), to a variety of response tasks assessing sustained attention. Such tests can be regarded as an operationalization of vigilance. One such test is the Sustained Attention to Response Task (SART) (Robertson, Manly, Andrade, Baddeley, & Yiend, 1997), a 4-minute 19-second computer task in which participants should withhold presses to one out of nine stimuli. The SART has shown an ability to quantify vigilance impairment in patients with excessive daytime sleepiness, for instance narcolepsy (Fronczek et al., 2006; Van Schie et al., 2012). Its primary outcome measure is the error count, consisting of key presses when no key should be pressed (commission errors), and absent presses when a key should have been pressed (omission errors).

In previous SART studies in healthy controls and patients with narcolepsy or other sleep disorders such as idiopathic hypersomnia or obstructive sleep apnea, we performed the SART five times per day prior to each session of a Multiple Sleep Latency Test (MSLT). The MSLT measures the tendency to fall asleep and is a routine part of the diagnostic work-up of excessive daytime sleepiness. This test consists of five 20-minute sessions at 1.5-hour intervals, in which participants are requested to try to fall asleep. The average sleep latency of five sessions is the main outcome parameter. Average sleep latency < 8 minutes indicates excessive daytime sleepiness. In this combined SART-MSLT design the SART error score of the first session was higher than of subsequent sessions in healthy controls (Fronczek et al., 2006) and in patients with excessive daytime sleepiness (Van Schie et al., 2012).

This difference may reflect a true change in vigilance level over the day (Manly, Lewis, Robertson, Watson, & Datta, 2002; Schmidt, Collette, Cajochen, & Peigneux, 2007),

or the result of napping between MSLT sessions, or a learning effect. Understanding such sources of variability may improve the reliability of the SART as a tool to measure vigilance, and therefore lower the number of SART sessions needed.

Possible reasons for an increased error score on the first SART session in previous studies were the early hour and the lack of a nap before that session. The lower error scores in later sessions could represent a learning effect, even though this does not seem likely for the following reasons: participants had always received a separate practice session before the first formal SART test; a learning effect is not expected in vigilance tasks; and finally none has been found in the SART before (Manly, Robertson, Galloway, & Hawkins, 1999; McAvinue, O'Keeffe, McMackin, & Robertson, 2005; Robertson et al., 1997).

The design of our first experiment was constructed to unravel the effects of three factors in healthy controls: the influence of repetition, a nap occasion before the SART and time of day.

EXPERIMENT 1

METHODS

Participants

Eighty healthy participants were recruited by advertisement. Reasons for exclusion were a diagnosis of a sleep disorder, any disorder significantly affecting attention, use of psychotropic medication, complaints of sleepiness, an Epworth Sleepiness Scale (ESS) > 10 (Johns, 1991), an irregular sleep pattern including night shift work and traveling through time zones in the two weeks before inquiry, or poor sleep hygiene. The latter was defined as an average night sleep < 6 hours or highly variable bed times (variability in bed time of > 2 hours for ≥ 3 days per week). Twenty men and 60 women with a mean age of 26.2 ± 9.7 years (range 18 to 55 years) were included. The sex ratio was due to recruitment in a faculty with more female students and employees. Participants were assigned to one of four groups matched for age, sex, and level of education. Table 1 shows the baseline characteristics of each group. The groups did not differ significantly in ESS score, sleep duration, or the number of days per week with a deviation from habitual bed time with more than one hour.

The study was approved by the medical ethical committee of Leiden University Medical Centre, and written informed consent was obtained from all participants prior to the study. Participants were given a small incentive to compensate for their time.

Design

The conditions "time of day" and "napping opportunity" led to a two-by-two design with four groups to which participants were assigned quasi-randomly, i.e. they were matched for age, sex and level of education. The two morning groups were tested between 9:15 and 12:15 hours, the two afternoon groups between 13:00 and 16:00 hours. All groups performed two SART sessions with a 1.5-hour break in between the two sessions. A 20-minute nap period similar to an MSLT session was offered to two groups (N+), but not to the two other groups (N-). The N+ groups were after the fact divided into those who actually slept (S+) and who did not (S-).

Upon arrival to the sleep laboratory, participants provided information about their sleep pattern of the last seven days. This was followed by the placement of electrodes for electro-encephalography (EEG, see below) before starting the SART.

Table 1 - Baseline characteristics of the study groups

	AM N+ ***		AM N-	PM N+		PM N-
	S+ (n = 13)	S- (n = 6)	(n = 20)	S+ (n = 16)	S- (n = 4)	(n = 20)
N of males	2	3	5	2	3	5
Age in years	24.6 ± 8.8	29.0 ± 12.8	25.9 ± 10.1	25.2 ± 8.7	30.5 ± 12.9	26.4 ± 10.0
ESS	5.1 ± 3.0	5.0 ± 3.3	5.0 ± 2.0	5.7 ± 2.8	4.8 ± 3.3	4.1 ± 2.9
Average nighttime sleep (hrs.)*	8.7 ± 0.8	8.2 ± 0.5	8.6 ± 0.9	8.5 ± 0.7	8.6 ± 0.7	8.5 ± 0.8
for week days*	8.4 ± 0.9	8.0 ± 0.8	8.3 ± 0.9	8.2 ± 0.7	8.7 ± 0.6	8.2 ± 0.9
for weekend days*	9.3 ± 0.7	8.7 ± 1.2	9.3 ± 1.2	9.2 ± 0.8	8.4 ± 1.7	9.2 ± 1.1
Nr of days per week with deviation from habitual bed time**	1.2 ± 0.9	1.3 ± 0.8	0.9 ± 0.9	0.8 ± 0.9	1.5 ± 0.6	0.9 ± 0.9
last 7 days	1.0 ± 1.3	1.2 ± 1.0	1.3 ± 1.3	1.3 ± 1.0	1.0 ± 0.8	1.5 ± 1.2
Sleep latency nap	10.3 ± 5.5	NA	NA	12.1 ± 4.1	NA	NA

Legend: AM: morning group; PM: afternoon group; N+/-: group with/without napping opportunity; S+/-: slept/remained wake during nap opportunity; ESS: Epworth Sleepiness Scale.
* Calculated from a subjective report of bed times and wake-up times.
** Deviation is defined as > 1.0 hour earlier or later than habitual bed time.
*** EEG recording of one subject in this group is missing due to a technical problem, so that appropriate classification of this subject in either S+ or S- was not possible.

Sustained Attention to Response Task

This test, lasting 4 minutes and 19 seconds, displays the numbers 1 to 9 25 times (225 in total) in random order on a black computer screen. Participants had to respond to the appearance of each number by pressing a button, except for the number 3, which occurred 25 times in all. Participants had to press the button before the next number appeared and were instructed to give equal importance to accuracy and speed in performing the task

(Manly *et al.*, 1999; Robertson *et al.*, 1997). We used the total error count, the average reaction time (RT) and the coefficient of variation (CV) of the RT as SART parameters.

Two SART sessions with a 1.5-hour break were administered to all participants using rooms, body positions, lights and equipment as described previously (Fronczek et al., 2006; Van Schie et al., 2012). Participants performed a short SART training before the first session. The nap occasion was provided to the N+ groups directly after the first SART session. Participants were allowed to go for short walks in the hospital and eat or drink between this nap and the second SART session, or between the two SART sessions for the N- groups. They were instructed to abstain from coffee, coca cola and energy drinks.

Questionnaires

The Stanford Sleepiness Scale (SSS) was administered prior to each SART session to assess the momentary level of sleepiness. Two 100 mm visual-analogue scales (VAS) were presented to the participants following each SART session. Participants were asked to evaluate their own performance concerning (1) accuracy ($VAS_{accuracy}$, from very poor to very good) and (2) response speed (VAS_{RT}, from very slow to very fast). This judgment was compared with their objective performance to obtain an estimation of their approach towards the task, i.e. whether the participants indeed felt they had complied with the instruction to pay equal attention to accuracy and speed.

Electro-encephalography

EEG electrodes were applied to all participants; for the N+ groups to record whether sleep occurred, and for the N- groups to confirm that the participants had remained awake. A total of 9 recording sites (Fpz, Z, Cz, Pz, Oz, Fp1, Fp2, Pg1 and Pg2) were measured with gold-plated 10 mm electrodes placed according to the 10-20 system (Jasper, 1958). Data were acquired using a portable polysomnography recorder (Titanium; Embla Systems, Broomfield, CO) and scored in Somnologica Studio 5.1.1.1684 (Embla Systems). Sleep was scored in 30-second epochs according to the AASM criteria (Littner et al., 2005) by one sleep technician.

Statistical analysis

Data were analyzed using IBM® SPSS® Statistics version 20. The analyses consisted of two parts. Firstly, the accuracy and RT measures of the SART sessions were compared between groups and conditions, i.e. time of day and opportunity to nap, using the paired *t*-test or the Mann-Whitney *U*-test when data were not normally distributed. Multiple testing was accounted for by Holm-Bonferroni correction. Delta scores for all SART accuracy and RT measures, as well as for SSS, were calculated by subtracting the score of the first session from the score of the second one. Correlations between delta scores were assessed using Pearson's r or Spearman's ρ, depending of the distribution of the data.

SART performance of the morning groups was compared to that of the afternoon groups, and SART performance of the N+ groups to that of N- groups. The latter comparison comprised the second session and delta score corrected for the first session. As the nap was provided after the first session, comparing this session was not considered useful to assess the influence of a nap occasion. The comparison between N+ and N- groups was secondarily broken down into a comparison of S+ and S- participants.

Additionally, linear mixed effect models (LMMs) were used to assess the combined effects of the conditions on SART performance, i.e. main effects of repetition, time of day, and napping, as well as their interactions, and to correct the analysis of accuracy measures for RT, CV of RT and SSS. Significance of model parameters was determined after Holm-Bonferroni correction for multiple comparisons.

RESULTS

Participants

EEG recordings confirmed that all participants from the N- groups stayed awake during and between SART sessions. Thirteen participants in the morning N+ group fell asleep during the napping opportunity, compared to 16 in the afternoon N+ group.

No differences in baseline characteristics were observed between those who did and did not fall asleep.

SART performance

SART error count and CV of RT are presented as mean ± SD, and SART RT as median with 25^{th}-75^{th} percentiles. The mean error count of all participants was 10.1 ± 4.5, the median RT was 280 ms (261 – 303 ms) and the mean CV of RT was 0.24 ± 0.06. A significant correlation was found between error count and RT (r_s = -0.53, p < 0.01 for session 1, and r_s = -0.58, p < 0.01 for session 2), as well as between error count and CV of RT (r_s = 0.31, p < 0.01 for session 1, and r_s = 0.34, p < 0.01 for session 2). $VAS_{accuracy}$ was significantly and negatively correlated with SART error count (r_s = -0.43, p < 0.01), and positively with average RT (r_s = 0.28, p < 0.01) and VAS_{RT} (r_s = 0.20, p = 0.01), but not with CV of RT; this indicated that a perceived higher accuracy was associated with a lower error count, higher RT, and, paradoxically, with a higher perceived response speed. Figure 1 presents SART data per group for both error count and RT.

Repetition

SART error count, RT and CV of RT of the second SART session did not differ significantly from those of the first over all participants (mean difference SART error count = 0.86 ± 5.53, *95% C.I.* -0.27 – 1.99, r = 0.17, median difference average RT = 5 ms, -15 – 22 ms, *ns*,

r = -0.14, mean difference CV = 0.01 ± 0.06, *95% C.I* -0.01 – 0.02, *r* = 0.11). Delta scores for accuracy and RT did not differ significantly between the four groups.

Fig. 1a SART error count

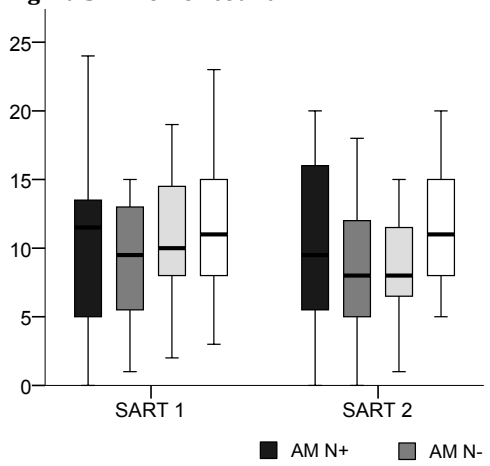


Fig. 1b SART RT

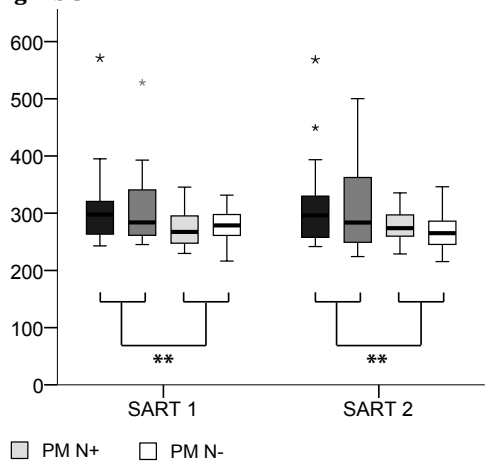


Figure 1 – A. SART error count per group, mean ± S.E. B. SART RT in ms per group.

Legend: AM: morning group; PM: afternoon group; N+/-: group with/without napping opportunity.

** indicate significant differences.

Time of day

The morning and afternoon groups did not differ significantly in SART error count between both sessions (mean difference first session = -1.43 ± 6.99, *95% C.I.* -3.66 – 0.81, *r* = -0.20, mean difference second session = -0.35 ± 6.65, *95% C.I.* -2.48 – 1.78, *r* = -0.05), nor in CV of RT (mean difference first session = 0.01 ± 0.07, *95% C.I.* -0.02 – 0.03, *r* = 0.08, mean difference second session = 0.00 ± 0.10, *95% C.I.* -0.03 – 0.04, *r* = 0.04). However, the morning groups had a longer RT compared to the afternoon groups on both session 1 (median difference average RT = 20 ms, -9 – 67 ms, *p* = 0.01, *r* = -0.40) and 2 (median difference average RT = 28 ms, -13 – 65 ms, *p* < 0.01, *r* = -0.44).

Napping

Neither SART error count (mean difference = 0.65 ± 7.15, *95% C.I.* -1.64 – 2.94, *r* = 0.09) nor average RT (median difference = -15 ms, -46 – 31 ms, *r* = 0.16) or CV of RT (mean difference = 0.02 ± 0.11, *95% C.I.* -0.02 – 0.05, *r* = 0.17) differed significantly between the second sessions of the N+ and N- groups. After correction for the first session by taking the difference with the second session (delta score), SART outcome measures still did not differ between the groups.

Subsequent analysis revealed that these measures did not differ either between the second SART sessions of the participants from the N+ groups who had actually slept (S+), compared to their matched counterparts in the N- group, nor to the participants from the nap groups who had remained awake (S-).

Table 2 – Linear Mixed Models

Model parameters	**Non-adjusted model**		**Adjusted for RT**		**Adjusted for RT*T**	
Basis	***Beta / S.E. / p***					
Intercept	8.88/1.00/0.00	*	18.46/2.17/0.00	*	15.55/2.42/0.00	*
Target factors	***Beta / S.E. / p***					
Time of day (T)	2.78/1.41/0.05		0.34/1.04/0.74		18.08/4.32/0.00	*
Nap occasion (N)	1.58/1.41/0.27		2.02/1.01/0.05		1.98/0.99/0.05	
Test session (S)	N.A.		N.A.		0.04/0.55/0.94	
Interactions	***Beta / S.E. / p***					
T*N	-3.78/1.99/0.06		-2.83/1.44/0.05		-2.66/1.41/0.06	
T*S	N.A		N.A		-1.54/0.78/0.05	
Covariates	***Beta / S.E. / p***					
SSS	N.A.		N.A.		N.A.	
RT	N.T.		-0.06/0.01/0.00	*	-0.05/0.01/0.00	*
CV	N.T.		33.02/4.79/0.00	*	33.74/4.58/0.00	*
RT*T	N.T.		N.T.		-0.06/0.01/0.00	*

Legend: RT: reaction time; SSS: Stanford Sleepiness Scale; CV: coefficient of variation of RT; Beta: regression coefficient derived from the linear mixed model; S.E: standard error of the regression coefficient; N.A: not available, i.e. no significant contribution to the final model; N.T: not tested in the model.

Model building strategy: *Compound symmetry* was chosen as model for the covariance matrix. The model of the mean was created from a saturated model including all target factors and possible interactions between them, followed by removing non-significant parameters as long as the model fit was not significantly impaired. The interactions N*S and T*N*S did not contribute significantly to any of the tested models and were therefore omitted from this table.
Models including the interaction of RT with all three factors were tested, but only the interaction of RT*T resulted in a different model and was therefore displayed. Three-way or higher-order interactions with RT were not modeled. Asterisks flag significant LMM coefficients.

Linear Mixed Models of SART error count corrected for RT

Table 2 presents the model parameters of three LMMs of the combined effects of repetition, time of day and napping on SART error count, firstly unadjusted, secondly adjusted for RT measures, and thirdly including the interaction of RT and time of day. As the latter model fitted our data best, conclusions about the investigated conditions were based upon this model. In line with the correlations mentioned above, the adjusted LMMs indicated a significant effect of CV of RT on SART error count and a significant

inverse effect of RT on SART error count. In other words, the higher the CV or the shorter the RT, the more errors were made ($p < 0.01$). The latter was even more pronounced for the afternoon groups, as demonstrated by the interaction effect of RT and time of day ($p = 0.01$): for every additional 25 ms of RT the error count decreased with 1.25 errors for the morning groups, compared to 2.75 errors for the afternoon groups. As the model including this interaction also showed a significant positive effect of time of day on SART error count ($\beta = 18.04$, $p = 0.01$), combining these findings indicates that the afternoon groups made more errors than the morning groups for RTs below 301 ms, but fewer errors when RT exceeded this value: for an RT of 280 ms, the error count was 2.3 points higher for the afternoon groups compared to the morning groups.

DISCUSSION OF EXPERIMENT 1

SART performance in the first session was similar to performance in the second session for all four groups; no differences were found between N+ and N- groups or between morning and afternoon groups concerning accuracy. A significant speed-accuracy trade-off was observed.

Time of day

The groups performing the SART in the morning had longer RT but preserved accuracy compared to the afternoon groups. Since the expected speed-accuracy trade-off (Wickelgren, 1977) was indeed present, we corrected LMM analyses of SART error count for RT. The final LMM containing the interaction of RT with time of day indicated that the speed-accuracy trade-off was stronger for the afternoon groups: for every additional 25 ms 1.5 error less was made compared to morning groups. The intercept of the model was also higher for the afternoon groups after correction for RT. The size of this difference was rather small though: at a RT of 280 ms, afternoon groups made 2.3 more errors than morning groups. At a RT of 301 ms this difference disappeared.

Napping in between test sessions

The N+ and N- groups did not differ in either error rate or RT. The same was found for S+ participants compared to their matched counterparts in the N- group. As such, a napping opportunity as provided in an MSLT does not interfere with performance of healthy participants on a SART session 70 minutes later. Since our study did not comprise sleep-deprived participants or sleep-disordered patients, no inferences of the influence of a nap for these populations can be made based on this study.

Subjective sleepiness and SART performance

No correlations were found between SART performance and scores of the SSS administered immediately preceding the SART session. This appears to contrast with the study by Manly and colleagues, who found an inverse correlation between errors of commission and SSS rating (Manly et al., 2002). However, their correlation is derived from large differences in both SSS and SART measurements at 13:00 and 19:00 on the one hand versus 01:00 and 07:00 on the other hand. As we did not observe significant differences in SART performance or SSS score between different times of day within our shorter time interval, possible correlations between those measurements would have to be rather strong to reach significance. Moreover, strong correlations were not expected in the present study, because all conditions that produce marked sleepiness were excluded. Despite all this, the lack of a correlation could also indicate that the SSS and SART measure different phenomena: the SSS reflects subjectively experienced sleepiness while the SART reflects sustained attention.

Repetition

No effect of repetition was observed in any group, nor in any of the sleep or time-of-day conditions in which two groups were combined. Although this is consistent with previous research by McAvinue in healthy controls (McAvinue et al., 2005), it contrasts with earlier findings from our study group in data of controls (Fronczek et al., 2006) and patients with excessive daytime sleepiness (Van Schie et al., 2012).

As such, the current experiment failed to disentangle the mechanism responsible for this prior observation, even though the possible contributions of time of day and napping were separated, and two consecutive sessions were administered to SART-naïve participants to be able to catch possible improvements due to a learning effect.

SART error count

In addition to the absence of a difference in SART error count between the first and the second session, the average error count was, at 10.1 errors (median error count at 10.0) rather higher than the median error count of 2.0 from the healthy control participants in our previous study (Fronczek et al., 2006). The currently observed error count even approached the range previously observed for patients with narcolepsy (median error count of 10.6-11.2 (Fronczek et al., 2006; Van Schie et al., 2012)). A possible explanation might reside in task instruction: in the current experiment, participants were instructed to give equal importance to accuracy and speed in performing the task, which is how the instruction was originally proposed by Robertson and colleagues in 1997 (Robertson et al., 1997). In contrast, participants were instructed to perform the SART as accurately as possible in the study by Fronczek (Fronczek et al., 2006).

Until now, we had not considered test instruction critically important for the

SART error count, as healthy participants in earlier studies (Manly et al., 1999; Robertson et al., 1997; Zordan, Sarlo, & Stablum, 2008) who received the original instruction, had similarly low error counts as found in Fronczek's study. When there is a significant speed-accuracy trade-off, error counts may, however, only be compared between studies when RT is taken into account; the RTs of these studies were either similarly high as in Fronczek's study (373 ms, (Robertson et al., 1997) compared to 393 ms, (Fronczek et al., 2006)), or not presented (Manly et al., 1999; Zordan et al., 2008). Based on the current results, a high error count with a low RT of 280 ms, we hypothesized that the instruction difference might have affected the results. Since participants with a long RT made fewer errors, it appeared plausible that changing the instruction towards not paying attention to RTs would lead to longer RTs but fewer errors, as a result of the speed-accuracy trade-off.

A second experiment was therefore conducted to investigate whether the instruction that was given to healthy participants on how to perform the SART influenced SART performance, in particular the size of the error count.

EXPERIMENT 2

METHODS

Participants

Five healthy participants were randomly chosen from each of the four groups of experiment 1, and new participants were recruited to match these 20 controls on age, sex and level of education. The new participants provided written informed consent prior to the study and received the same incentive as the controls had received. Characteristics of the I_M and I_0 groups are presented in Table 3. The groups did not differ significantly in ESS score, sleep duration, or the number of days per week with a deviation from habitual bed time with more than one hour. All participants had either completed or were following higher education at the time of testing (not indicated in the table).

Design

The study design for the new participants was exactly the same as for the participants from experiment 1 (i.e. 5 participants per combination of time of day/napping) with only one difference: participants were instructed to pay attention to accuracy only (modified instruction, I_M), instead of giving equal importance to accuracy and speed in performing the task (original instruction, I_0).

Statistical analysis

Data were analyzed using IBM® SPSS® Statistics version 20. Firstly, the analyses consisted

of paired *t*-tests or a non-parametric counterpart to investigate SART performance between the two instruction groups. Secondly, the analysis concerned LMMs to assess the combined effects of instruction and the other conditions on SART performance, i.e. main effects of instruction, repetition, time of day, and napping, as well as their interactions. SART error count was again corrected for RT measures in the analysis and the Holm-Bonferroni adjustment was used to account for multiple testing. LMM analysis was also used to compare participants' judgments about their performance with their objective performance across instruction groups.

Table 3 - Baseline characteristics of the study groups

	I_0 (N = 20)	I_M (N = 20)
N of males	6	6
Age in years	26.7 ± 11.4	28.0 ± 11.5
ESS	4.9 ± 2.4	5.1 ± 2.7
Average nighttime sleep (hrs.)*	8.5 ± 0.8	8.4 ± 0.6
for weekdays*	8.2 ± 0.9	8.2 ± 0.7
for weekend days*	9.3 ± 1.1	9.0 ± 0.8
Nr of days per week with deviation from habitual bed time**	1.0 ± 0.9	1.1 ± 0.9
last 7 days	1.3 ± 1.2	1.1 ± 0.8

Legend: I_0: original instruction; I_M: modified instruction; ESS: Epworth Sleepiness Scale; Nr: number.
* Calculated from a subjective report of bed times and wake-up times.
** Deviation is defined as > 1.0 hour earlier or later than habitual bed time.

RESULTS

SART performance

The mean error count of the I_0 group was 10.5 ± 4.3, the median RT was 280 ms (261 - 303) and the mean CV of RT was 0.23 ± 0.06 ms. These values were 5.2 ± 2.7, 319 ms (282 - 409 ms) and 0.23 ± 0.07 ms for the I_M group. Figure 2 presents SART data per group for both error count and RT. The error count was significantly lower in the I_M group than in the I_0 group for both session 1 (mean difference = 3.55 ± 7.12, *95% C.I.* 0.22 – 6.88, p = 0.04, r = 0.46) and session 2 (mean difference = 6.90 ± 6.14, *95% C.I.* 4.03 – 9.77, p < 0.01, r = 0.76). SART RT was significantly higher in the I_M group than in the I_0 group for both session 1 (median difference average RT = 35 ms, -2 – 149 ms, p = 0.02, r = 0.54) and session 2 (median difference average RT = 56 ms, 1 – 209 ms, p = 0.02, r = 0.54). CV of RT did not significantly differ between I_M and I_0 groups (mean difference for the first session = -0.03 ± 0.11, *95% C.I.* -0.08 – 0.03, r = 0.16, for the second session = 0.02 ± 0.11, *95% C.I.* -0.03 – 0.08, r = 0.14). $VAS_{accuracy}$ was significantly and negatively correlated with SART error count (-2.93, *C.I.* -3.87 – -1.99, p < 0.01), and positively with VAS_{RT} (0.30, *C.I.* 0.10 – 0.50, p < 0.01), irrespective of instruction; this indicated that a perceived higher

accuracy was associated with a lower error count and a higher perceived response speed in both instruction groups.

Repetition

In contrast to the I_0 group, SART error count decreased from the first to the second session (mean difference = 3.55 ± 3.25, *95% C.I.* 2.03 – 5.07, $p < 0.01$, $r = 0.75$), as did CV of RT (mean difference = 0.04 ± 0.05, *95% C.I.* 0.02 – 0.06, $p < 0.01$, $r = 0.63$). While error count and CV of RT differed between the first and second session of the I_M group, RT did not (median difference average RT = 10 ms, -27 – 43 ms, $r = 0.16$).

Table 4 – Linear Mixed Models

Model parameters	**Non-adjusted model**		**Adjusted for RT**		**Adjusted for RT*I**	
Basis	***Beta / S.E. / p***					
Intercept	10.55/0.95/0.00	*	10.88/1.87/0.00	*	15.78/2.35/0.00	*
Target factors	***Beta / S.E. / p***					
Instruction (I)	-3.55/1.34/0.01	*	-2.98/1.12/0.01		-9.26/2.62/0.00	*
Time of day (T)	N.A.		1.53/1.10/0.17		0.56/0.93/0.55	
Nap occasion (N)	N.A.		3.20/1.33/0.02		3.15/0.99/0.00	*
Test session (S)	-0.20/1.01/0.84		1.14/1.13/0.32		-1.08/0.56/0.06	
Interactions	***Beta / S.E. / p***					
I*N	N.A.		1.22/1.43/0.40		N.A.	
I*S	-3.35/1.42/0.02	*	-1.32/1.16/0.26		N.A.	
T*N	N.A.		-3.96/1.35/0.01	*	-4.09/1.33/0.01	*
T*S	N.A.		-1.73/1.13/0.14		N.A.	
N*S	N.A.		-1.32/1.14/0.26		N.A.	
Covariates	***Beta / S.E. / p***					
SSS	N.A.		N.A.		N.A.	
RT	N.T.		-0.03/0.00/0.00	*	-0.05/0.01/0.00	*
CV	N.T.		32.10/4.87/0.00	*	31.96/4.50/0.00	*
RT*I	N.T.		N.T.		0.02/0.01/0.02	*

Legend: RT: reaction time; SSS: Stanford Sleepiness Scale; CV: coefficient of variation of RT; Beta: regression coefficient derived from the linear mixed model; S.E: standard error of the regression coefficient; N.A: not available, i.e. no significant contribution to the final model; N.T: not tested in the model.

Model building strategy: *Compound Symmetry* was chosen as model for the covariance matrix. The model of the mean was created from a saturated model including all target factors and possible interactions between them, followed by removing non-significant parameters as long as the model fit was not significantly impaired. The interactions I*T, I*T*S, I*N*S, T*N*S, I*T*N and I*T*N*S did not contribute significantly to any of the tested models and were therefore omitted from this table.
Models including the interaction of RT with all three factors were tested, but only the interaction of RT*instruction resulted in a different model and was therefore displayed. Three-way or higher-order interactions with RT were not modeled. Asterisks flag significant LMM coefficients.

Linear Mixed Models of SART error score corrected for RT

Table 4 presents the model parameters of three LMMs of the combined effects of instruction, repetition, time of day and napping on SART error score, firstly unadjusted, secondly adjusted for RT measures, and thirdly including the interaction of RT and instruction. The latter model had the best fit and was used. A significant effect of CV of RT on SART error count and a significant inverse effect of RT on SART error count were observed for both instruction groups, but the latter was less pronounced for the I_M group: for every additional 25 ms, 1.25 errors less were made in the I_0 group compared to 0.75 errors in the I_M group. After correction for RT, LMM indicated that SART error count was still lower in the I_M group than in the I_0 group (*95% C.I.* -14.56 – -3.96): at an RT of 300 ms, for instance, the size of the difference was modeled to be 3.26 errors. The non-significant contributions of the interaction effects of instruction with either or both time of day and napping indicated that the differences between I_0/I_M groups were similar for participants tested in the morning versus afternoon, and with or without napping opportunity. The combined effect of napping with the interaction effect of napping and time of day, irrespective of instruction, indicated that the error count was higher in the morning nap group.

Fig. 2a SART error count **Fig. 2b SART RT**

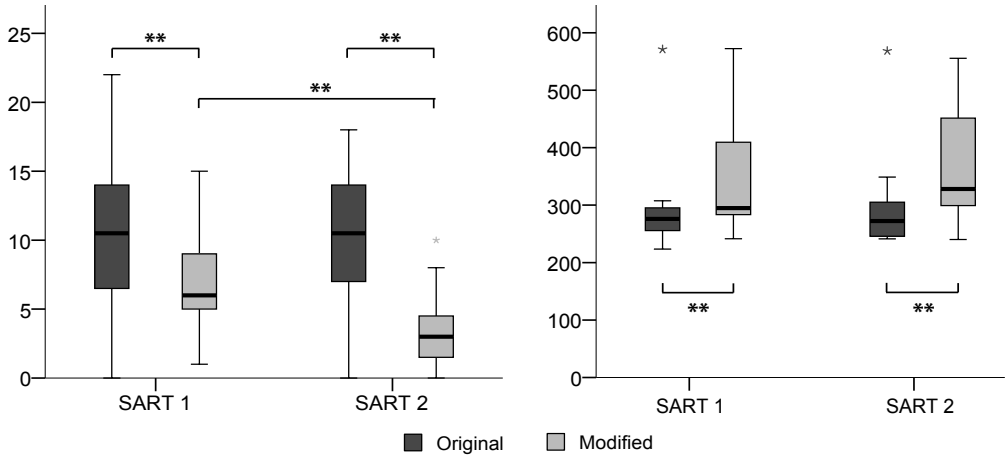


Figure 2– SART performance separated for the original and modified instruction groups. A. Mean SART error count ± S.E. B. SART RT in ms per group. ** indicate significant differences.

DISCUSSION OF EXPERIMENT 2

Participants who had been instructed to only pay attention to accuracy made significantly fewer errors but performed more slowly and with less variation in reaction times than those instructed to pay equal attention to accuracy and speed. The lower error count remained significant after adjusting for RT.

Instruction

The modified instruction was strongly associated with a more than two-fold decrease in SART error score on the second session (effect size = 0.76). As effect sizes > 0.70 correspond to clinically relevant differences, the instruction thus significantly and relevantly influenced SART performance (Cohen, 1988).

The speed-accuracy trade-off was less pronounced but still present for the I_M group, indicating that response strategy may still influence the error count. However, the remaining net effect was small: to reduce the error count by 1 error the I_M group had to prolong responses by 33 ms. As the decrease in error count was accompanied by a small non-significant increase in RT, it is possible that the improved SART performance of the second session in the first experiment is due to participants developing a 'slower but more accurate' strategy. Interestingly, participants were not aware of such a strategy: the higher participants estimated their accuracy, the faster they estimated their response speed. This yielded for both instruction groups. Apart from reaction times per se, participants in the I_M groups showed less variation in their reaction times. This could possibly be interpreted as a learning effect.

Effect size of instruction

The error counts of the I_M group resembled those of controls in the study of Fronczek et al: 7.0 in the current experiment compared to the previous 6.0 for the first session, and 3.5 compared to 2.0 for the second session. The modified instruction thus likely accounted for the difference in height of error score between the participants from the first experiment and our previous study, as well as for the effect of repetition. Comparing the error count of our I_0 group to that in studies with the same instruction, the mean error count of the I_0 group (10.5) resembled one of them, a study by 't Hart et al (9.7 errors, (Hart et al., 2012)), but is somewhat higher compared to a third study (5.9, (Zordan et al., 2008)).

GENERAL DISCUSSION

This study investigated the influences of time of day, napping, repetition, and instruction on the performance on two consecutive sessions of the SART in healthy participants.

The aim was to unravel the mechanism responsible for a decrease in SART error count from the previously found marked drop from the first to the second session. Our results demonstrated that such an improvement is only found when participants are instructed to pay attention to accuracy and to ignore response speed. The improvement is likely attributed to an effect of repetition, i.e. a learning effect, although participants were not aware of such an effect, given their own performance judgments. The link to instruction also explained why one of our previous studies did find a learning effect (Fronczek et al., 2006), in contrast to other studies (McAvinue et al., 2005).

SART in sleep medicine

The associations between instruction and error count and between instruction and learning effect need not necessarily hold to the same degree for patients with sleep disorders. The error rate of patients with narcolepsy who received the instruction to pay attention to accuracy only was similar to that of patients with narcolepsy who received the instruction to pay equal attention to both accuracy and speed (Fronczek et al., 2006; Van Schie et al., 2012). It seems likely that patients with narcolepsy already function at maximum task capacity when instructed to pay equal attention to both accuracy and speed: their long RT (mean of 337 ms) suggests that speed was already sacrificed at the expense of accuracy (Van Schie et al., 2012), so that dropping the speed condition would not result in a better accuracy. Their low level of accuracy compared to controls (Fronczek et al., 2006), i.e. their inability to sustain attention to a 4-minute lasting task, may very well reflect the problems patients with narcolepsy face in daily life when trying to follow a conversation or read a book.

Our previous study that used the original instruction (Van Schie et al., 2012), also investigated SART performance in patients with idiopathic hypersomnia and obstructive sleep apnea. Their performance did not significantly differ from that of patients with narcolepsy. It would be interesting to investigate the modified instruction in these conditions as well. Again, patients with these disorders might function at their maximum capacity when instructed to pay equal attention to both accuracy and speed.

Implications for the use of the SART

The results indicate that the SART discriminates better between healthy controls and patients with narcolepsy when the instruction is given to prefer accuracy to speed, than when accuracy and speed are considered equally important. We therefore recommend the "accuracy first" instruction. To minimize the consequences of the learning effect that has been observed when using this instruction, we strongly recommend the use a full practice session, i.e. a 225-trial session instead of the 30-trial session that was used as practice session in both our experiments and the manuscript by Fronczek et al (Fronczek et al., 2006). The rationale for this recommendation is the observation that the higher

error count in the first session of Fronczek's study was followed by stable, lower error counts in the second to the fifth session. In other words, regarding this first session as practice session minimizes the consequences of the learning effect.

The present study also showed that a nap opportunity of 20 minutes more than 1 hour prior to a SART session did not strongly influence the error count of that session in healthy participants. As such, SART sessions obtained from an MSLT design are likely to be suitable for comparison with separate SART sessions from a follow-up occasion. Before doing so, the question whether patients with sleep disorders profit or not from a nap should be answered.

The time of day had no clear effect on the SART error count, and if such an effect exists at all, it is rather small and occurred only following the original instruction. The modified instruction allows a comparison between SART sessions administered at different times of the day during normal working hours.

To conclude, instructing healthy participants to perform the SART as accurately as possible leads to a lower error count with lower between-subject variability, and is thus the preferred instruction to assess the best performance in terms of error count that a subject can achieve.

ACKNOWLEDGEMENTS

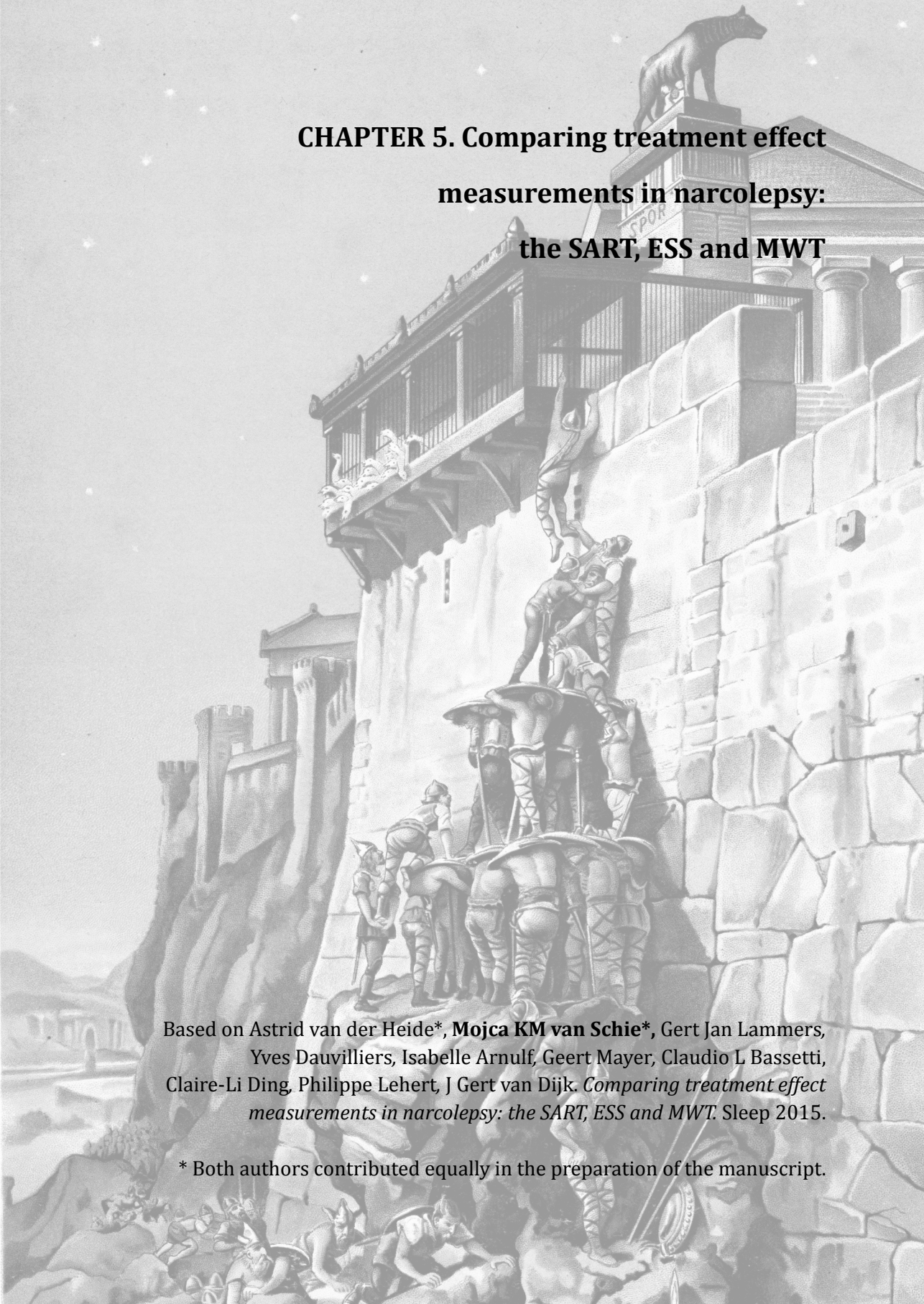
We thank P.P. Buijserd-Amesz, P.J. van Someren, W.W.M. Huyser, A. Janson and S. van Gent for their help in acquiring the data, J.J.M. Witteman for the scoring of sleep and R.H.A.M. Reijntjes for his help in the processing of the data.

REFERENCES

1. Coenen, A. M. (1995). Neuronal activities underlying the electroencephalogram and evoked potentials of sleeping and waking: implications for information processing. *Neurosci Biobehav Rev, 19*(3), 447-463.
2. Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
3. Fronczek, R., Middelkoop, H. A., van Dijk, J. G., & Lammers, G. J. (2006). Focusing on vigilance instead of sleepiness in the assessment of narcolepsy: high sensitivity of the Sustained Attention to Response Task (SART). *Sleep, 29*(2), 187-191.
4. Hart, E. P., Dumas, E. M., Reijntjes, R. H., van der Hiele, K., van den Bogaard, S. J., Middelkoop, H. A., . . . van Dijk, J. G. (2012). Deficient sustained attention to response task and P300 characteristics in early Huntington's disease. *J Neurol, 259*(6), 1191-1198. doi: 10.1007/s00415-011-6334-0
5. Johns, M. W. (1991). A new method for measuring daytime sleepiness: the Epworth sleepiness scale. *Sleep, 14*(6), 540-545.
6. Littner, M. R., Kushida, C., Wise, M., Davila, D. G., Morgenthaler, T., Lee-Chiong, T., . . . Kramer, M. (2005). Practice parameters for clinical use of the multiple sleep latency test and the maintenance of wakefulness test. *Sleep, 28*(1), 113-121.
7. Manly, T., Lewis, G. H., Robertson, I. H., Watson, P. C., & Datta, A. (2002). Coffee in the cornflakes: time-of-day as a modulator of executive response control. *Neuropsychologia, 40*(1), 1-6.
8. Manly, T., Robertson, I. H., Galloway, M., & Hawkins, K. (1999). The absent mind: further investigations of sustained attention to response. *Neuropsychologia, 37*(6), 661-670.
9. McAvinue, L., O'Keeffe, F., McMackin, D., & Robertson, I. H. (2005). Impaired sustained attention and error awareness in traumatic brain injury: implications for insight. *Neuropsychol Rehabil, 15*(5), 569-587. doi: 10.1080/09602010443000119
10. Morad, Y., Lemberg, H., Yofe, N., & Dagan, Y. (2000). Pupillography as an objective indicator of fatigue. *Curr Eye Res, 21*(1), 535-542.
11. Robertson, I. H., Manly, T., Andrade, J., Baddeley, B. T., & Yiend, J. (1997). 'Oops!': performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. *Neuropsychologia, 35*(6), 747-758.
12. Schmidt, C., Collette, F., Cajochen, C., & Peigneux, P. (2007). A time to think: circadian rhythms in human cognition. *Cogn Neuropsychol, 24*(7), 755-789. doi: 10.1080/02643290701754158
13. Valley, V., & Broughton, R. (1981). Daytime performance deficits and physiological vigilance in untreated patients with narcolepsy-cataplexy compared to controls. *Rev Electroencephalogr Neurophysiol Clin, 11*(1), 133-139.

14. Van Schie, M. K., Thijs, R. D., Fronczek, R., Middelkoop, H. A., Lammers, G. J., & Van Dijk, J. G. (2012). Sustained attention to response task (SART) shows impaired vigilance in a spectrum of disorders of excessive daytime sleepiness. *J Sleep Res, 21*(4), 390-395. doi: 10.1111/j.1365-2869.2011.00979.x
15. Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics. *Acta Psychol (Amst), 41*, 67-85.
16. Zordan, L., Sarlo, M., & Stablum, F. (2008). ERP components activated by the "GO!" and "WITHHOLD!" conflict in the random Sustained Attention to Response Task. *Brain Cogn, 66*(1), 57-64. doi: 10.1016/j.bandc.2007.05.005

PART II. Sustained Attention to response Task as a treatment-effect parameter in excessive daytime sleepiness



CHAPTER 5. Comparing treatment effect measurements in narcolepsy: the SART, ESS and MWT

Based on Astrid van der Heide*, **Mojca KM van Schie*,** Gert Jan Lammers, Yves Dauvilliers, Isabelle Arnulf, Geert Mayer, Claudio L Bassetti, Claire-Li Ding, Philippe Lehert, J Gert van Dijk. *Comparing treatment effect measurements in narcolepsy: the SART, ESS and MWT.* Sleep 2015.

* Both authors contributed equally in the preparation of the manuscript.

ABSTRACT

Study Objectives
To validate the Sustained Attention to Response Task (SART) as a treatment effect measure in narcolepsy, and to compare the SART with the Maintenance of Wakefulness Test (MWT) and the Epworth Sleepiness Scale (ESS).

Design
Validation of treatment effect measurements within a randomized controlled trial (RCT).

Patients
95 patients with narcolepsy with or without cataplexy.

Interventions
The RCT comprised a double-blind, parallel-group, multi-center trial comparing the effects of 8-week treatments with pitolisant (BF2.649), modafinil or placebo (NCT01067222). MWT, ESS and SART were administered at baseline and after an 8-week treatment period. The severity of excessive daytime sleepiness and cataplexy was also assessed using the Clinical Global Impression scale (CGI-C).

Measurements and Results
The SART, MWT and ESS all had good reliability, obtained for the SART and MWT using two to three sessions in one day. The ability to distinguish responders from non-responders, classified using the CGI-C score, was high for all measures, with a high performance for the SART (r=0.61) and the ESS (r=0.54).

Conclusions
The SART is a valid and easy to administer measure to assess treatment effects in narcolepsy, enhanced by combining it with the ESS.

INTRODUCTION

While narcolepsy has an undisputed profound impact on daily life,[1] quantifying how it impairs daily life is difficult. The severity of narcolepsy is currently assessed using measures of the ability to stay awake in boring conditions, such as the Maintenance of Wakefulness Test (MWT), or measures of subjective sleepiness, for which the Epworth Sleepiness Scale (ESS) is often used.[2,3] However, sleepiness and sleep propensity are not the only aspects of the burden of narcolepsy. An aspect that is gradually more recognized is the quality of the awake state, for which the ability to sustain attention is an important requisite. The Sustained Attention to Response Task (SART), designed to assess this function, has previously been used in narcolepsy,[4,5] and has shown clear potential to quantify the impairment in function during wake in narcolepsy.

The SART is a go/no-go task in which the no-go target appears unpredictably and rarely, and in which both accuracy and response speed, quantified as reaction time (RT), are important. The SART was developed to investigate lapses of sustained attention in individuals with neurological impairment, and proved to be a useful tool to investigate sustained attention in a number of other clinical conditions, including sleep disorders.[4-6] To date, the validation of the SART as a tool to measure sustained attention in sleep-disordered patients is based on a comparison of SART results between patients with narcolepsy and healthy controls.[4,5] The SART discriminated well between these groups, i.e. it demonstrated good construct validity. Between-subjects variability in SART performance was higher in the narcolepsy group than in the control group. No correlations were found between SART performance and subjective sleepiness (ESS) or between SART performance and the average sleep onset latency during multiple sleep latency tests (MSLT), i.e. the SART showed discriminant validity with these measures of sleepiness/sleep propensity.

As the SART quantifies the impairment of the waking condition in narcolepsy, it should also be a useful tool to measure treatment effects in narcolepsy. Hence, the objective of this study was to validate the SART as a measurement of treatment effect in narcolepsy, and to compare it with the MWT and ESS, two tests frequently used in treatment-effect studies in hypersomnias[7-10] that, however, have never explicitly been validated for their capability to measure treatment effects in narcolepsy. As the initial studies of the SART in sleep disorders have neither assessed the reliability of the test, nor the statistical properties of its outcome measures (i.e. descriptive statistics, statistical distribution of the data), these characteristics were also investigated in this study and compared to those of the ESS and MWT.

METHODS

Subjects

The analysis was conducted on data originating from a double-blind, parallel-group, multi-center trial comparing the effects of eight-week treatment with the experimental drug BF2.649 (pitolisant) to effects of the proven effective drug modafinil and to placebo in narcolepsy (NCT01067222).[11] Inclusion criteria were the presence of narcolepsy with or without cataplexy diagnosed according to the International Classification of Sleep Disorders (ICSD)-2 criteria and a score of ≥ 14 on the Epworth Sleepiness Scale (ESS) during the baseline period.

The trial was conducted in accordance with the International Conference on Harmonization Guidelines for Good Clinical Practice and the Declaration of Helsinki. The protocol was approved by central and local ethics committees and written informed consent was obtained from all subjects prior to the study. The results of this study were published separately.

Design

Eligible patients started with a baseline period of seven days in which they were not allowed to take psychostimulants, medication with sedating properties, tricyclic antidepressants, psychoactive agents, or medication interacting with modafinil. Patients were allowed to take anticataplectic drugs (sodium oxybate and nontricyclic antidepressants). The baseline period was completed by an inclusion visit. Patients continuing to meet the inclusion criteria were randomly assigned to one of three equally sized treatment groups for a total duration of eight weeks, with possible titration after two weeks and, if necessary, also after three weeks. A control visit took place after seven weeks, and an endpoint visit took place after the eight-week treatment period.

The SART and the MWT were performed at the inclusion visit and the endpoint visit (or the last on-study visit). A SART session was administered prior to each of four MWT sessions, starting at 10:00 hrs and at two-hour intervals thereafter. Patients were requested to take their morning treatment and to have a light breakfast before 08:00 hrs, arriving at the trial center around 09:00 hrs. Patients took trial medication and had lunch immediately after the second MWT session. Patients were to refrain from stimulating beverages such as coffee or tea during these visits.

Sustained Attention to Response Task

The SART involved withholding key presses to 1 in 9 target stimuli during a 4-minute 19-second period. A number from 1 to 9 was shown 225 times in white on a black computer screen in a quasi-random way, while patients were seated on a chair in front of a computer screen. The font size was randomly chosen from 26, 28, 36, or 72 points.

Each number was presented for 250 milliseconds, followed by a blank screen for 900 milliseconds. Subjects had to respond to the appearance of each number by pressing a button, except when the number was a 3. Subjects had to press a button before the next number appeared and were instructed to give equal importance to accuracy and speed in performing the task.[6,12]

The primary outcome measure of the SART is the total number of errors, consisting of, firstly, key presses when no key should be pressed (i.e. after a '3', a so-called 'no-go trial') and, secondly, absent presses when a key should have been pressed (i.e. after anything but a '3', the so-called 'go trials'). Errors on no-go trials, with a maximum count of 25, are called commission errors. Errors on go trials, omission errors, have a theoretical maximum count of 200. The sum of commission and omission errors, the total error count, was also analyzed.

Maintenance of Wakefulness Test

The MWT consisted of four 40-minute sessions in a quiet and dimly lit room. Subjects were instructed to stay awake while comfortably seated in a semi-supine position. Movements or vocalizations were not allowed. The session was terminated either when sleep-onset occurred, defined as either three consecutive 30-second epochs of stage 1 sleep or a single 30-second epoch of any other sleep stage, or after 40 minutes of being awake.[13] The mean of the four sleep-onset latencies was considered the primary outcome measure of the MWT.

Epworth Sleepiness Scale

The ESS was administered twice at baseline (at the start of the baseline period and at the inclusion visit) and twice after treatment (at the control visit and the endpoint visit). The two early and the two late measurements were treated as separate sessions in order to assess reliability, i.e. they were not averaged. The four sessions were also separately used in the analysis of treatment efficacy.

Clinical Global Impression

The severity of EDS and of cataplexy was assessed by the local investigator using the Clinical Global Impression of Severity (CGI-S),[14] a 6-point scale, at both baseline visits. Their average value was used for analysis. Any changes in severity of EDS and of cataplexy were measured by the investigator using the Clinical Global Impression of Change (CGI-C) at each follow-up visit.[14] Ratings of this 7-point scale were averaged for the control and endpoint visit to create the final CGI-C score. CGI-S and CGI-C were rated based on a clinical interview before the administration of other scales or tests. The CGI-S and CGI-C scores were linearly transformed into a range from 0 to 4 to enhance comparability, with low values indicating higher severity in the CGI-S and more worsening in the CGI-C.

Statistical analysis

The statistical analyses were carried out with R statistical package (R, version 2.12.2). Unless specified otherwise, we conducted two-sided tests with a significance level of 0.05.

Descriptive statistics

Normality of SART, MWT and ESS outcome measures was assessed by descriptive statistics, parameters of asymmetry and kurtosis, box plots, and the Kolmogorov-Smirnov (KS) test for normality. In case of non-normality, this was repeated for the log transformation of the respective outcome measures. Floor and ceiling effects and homoscedasticity (homogeneity of variance) were tested in subgroups based on age range and gender.

Reliability

A test is considered reliable when within-patient variability is low; no significant change in the test value should occur during a period in which no change is expected, and the value should respond when such a change in condition occurs. We calculated the reliability of each outcome measure with a linear mixed model (see appendix).

Reliability is high when within-patient variability is low compared to the variability of the studied outcome measure. To express this comparison as a number the intra-class correlation coefficient (ICC) of reliability was used.[15] The ICC was estimated from our model as follows: the within-patient variability (squared) was divided by total variability, which is the within-patient variability (squared) plus the variability of the studied outcome measure (squared). The optimal value of the ICC is 1, meaning there is no within-patient variability, and all variability is explained by variability of the studied outcome parameter. An ICC > 0.8 is accepted as indicating good reliability.[16,17]

When one measurement or test session proves to have an insufficiently high reliability, this reliability can be increased by repeating the test.[18] As the SART and MWT were each performed four times on a test day, the ICCs resulting from the first 2 to all 4 sessions were calculated using the Spearman-Brown expression for stepped-up reliability.[19]

Sensitivity

As we aimed to investigate the validity of the SART in the context of narcolepsy, the CGI-C was considered the most appropriate standard to compare SART results with, as it reflects clinically pertinent changes in a patients' condition, assessed in a manner reflecting normal medical practice in a patient-physician interview. We calculated the sensitivity of each outcome measure for treatment efficacy by dividing subjects into responders and non-responders. Such a classification provides two groups that are supposed to differ in the true level of the constructs, but that are quite homogeneous within each group.

The best dichotomy between the categories was found through assessing the linearity of the scale (Logit model between CGI-C and first factor from a confirmatory factor analysis), corroborated with a Rasch Analysis. On this basis, a responder was defined as being 'much' or 'very much' improved on the CGI-C, and all other results were classified as non-responders. This strategy is commonly used in various studies.[20-23] Analysis of covariance (ANCOVA) was used to compare outcome measures between responders and non-responders, corrected for baseline values, age, and sex.

The difference in the mean outcome measure between responders and non-responders was divided by its standard deviation (called 'residual standard deviation') to calculate the so-called Cohen's Effect Size (ES) or standardized mean difference. An ES > 0.5 is considered clinically relevant. If baseline and final values of the same outcome measures are correlated, the residual standard deviation is reduced, leading to a higher ES. A corrected effect size taking into account this correlation was calculated by multiplying the ES by the square root of the coefficient of correlation between the baseline and final values. The effect size was also measured using a linear mixed model, in which the interaction between treatment effect and time provided a more accurate measure of the effect size.

Finally, associations between CGI-C, MWT, ESS, and SART were investigated using factor analysis to demonstrate the contribution of each outcome measure to the CGI-C score.[24,25]

Missing values

Reliability and sensitivity were estimated on the available data set. The trial from which our data originated was considered a pivotal Phase III trial. As such, missing data were rare, with no missing data at baseline and less than 7% at final time. These data were not imputed, but directly handled by the mixed model.[26]

For sensitivity purposes, we repeated our analyses in imputing missing data by using Last observed Carried out Forward techniques (LOCF), Baseline carried forward (BCF) and multiple imputation. We calculated the relative error between the values of the three techniques Qi with our suggested method Q in calculating E= 100* | (Qi-Q)/Q |. These values were 0.8%, 1.3%, and 1.7% respectively. We therefore concluded that imputation of missing data did not change the results.

RESULTS

Subjects and data characteristics

Patient characteristics are summarized in *Table 1*. None of the SART accuracy measures was normally distributed (KS, p<0.001), but after logarithmic transformation commission errors and the total number of errors were normally distributed (KS, p=0.14). No suitable transformation was found that resulted in a normal distribution for omission errors (KS, p<0.001). The ESS showed a slightly platycurtic normal distribution (KS, p>0.55). A ceiling effect was observed for the MWT, caused by the maximum score of 40 minutes; this made the nature of the distribution difficult to define with precision. A log-normal distribution was suspected (KS, p=0.23) and was therefore used in further analysis. As observed more often after log transformations, between-category heteroscedasticity was found for log-MWT.

Table 1 - Patient characteristics

	Responders (N = 51)	**Non-responders (N = 44)**	**P value**
Parameter	**Mean ± SD**	**Mean ± SD**	
Age (years)	38.24 ± 14.08	39.25 ± 15.36	0.737
Sex (Males (%))	28 (55%)	24 (55%)	0.971
Baseline ESS	18.70 ± 2.79	18.13 ± 2.39	0.291
Baseline CGI-S	1.63 ± 1.04	0.93 ± 0.51	<0.001
Baseline SART total errors	15.65 ± 13.69	11.62 ± 7.21	0.079
Baseline MWT sleep latency (min.)	10.6 ± 8.9	13.2 ± 10.6	0.196
Endpoint ESS	9.76 ± 6.56	15.02 ± 4.12	<0.001
Endpoint CGI-C	3.26 ± 0.83	0.91 ± 0.29	<0.001
Endpoint SART total errors	8.77 ± 7.03	11.48 ± 8.91	0.145
Endpoint MWT sleep latency (min.)	23.6 ± 14.6	12.4 ± 11.1	<0.001

ESS: Epworth Sleepiness Scale; CGI-S: Clinical Global Impression of Severity; CGI-C: Clinical Global Impression of Change; SART: Sustained Attention to Response Task; MWT: Maintenance of Wakefulness Test; log: log-transformed; min: minutes; SD: standard deviation.

Reliability

Table 2 presents within-patient variability and variability of the studied measure (i.e. the various SART error counts, ESS, and MWT) as modeled. With the aid of these estimates the ICC was calculated. The ICC was highest for the ESS at 0.83; ICC for log-SART total error count was 0.65, and for log-MWT it was 0.76. The influence of replication is presented in *Table 3*; repeating the test improved the reliability for the MWT to 0.87 for the first two tests and to 0.82 for the first two log-transformed SART commission error counts.

Table 2 – Variability and intra-class coefficient of correlation of SART, MWT and ESS

	Within-patient variability	Variability measure	ICC
SART commission errors (log)	0.14	0.23	0.71
SART omission errors (log)	0.30	0.34	0.56
SART total errors (log)	0.20	0.28	0.65
MWT sleep latency in min. (log)	0.26	0.47	0.76
ESS	1.09	2.45	0.85
SART commission errors	2.85	4.39	0.70
SART omission errors	8.28	7.97	0.48
SART total errors	8.67	10.50	0.59
MWT sleep latency in min.	7.44	13.48	0.77

ICC: intra-class coefficient of correlation, calculated as follows: within-patient variability (squared) divided by the total variability, which is the within-patient variability (squared) plus the variability of the studied outcome measure (squared). The last four rows illustrate that non-log-transformed SART and MWT have a lower ICC compared to their log-transformed match.

Table 3 – Influence of the number of sessions on the intra-class coefficient of correlation

	ICC	Replication		
	1	2	3	4
SART commission errors (log)	0.70	**0.82**	0.88	0.90
SART omission errors (log)	0.56	0.72	0.79	**0.84**
SART total errors (log)	0.65	0.79	**0.85**	0.88
MWT sleep latency in min. (log)	0.76	**0.87**	0.91	0.93
ESS	**0.83**	0.91	0.94	0.95
SART commission errors	0.70	**0.83**	0.88	0.90
SART omission errors	0.48	0.65	0.74	0.79
SART total errors	0.59	0.75	**0.81**	0.85
MWT sleep latency in min.	0.77	**0.87**	0.91	0.93

ICC: intra-class coefficient of correlation. An ICC > 0.80 is regarded as good reliability. The ICC resulting from the first 2 to all 4 sessions was calculated using the Spearman-Brown expression for stepped-up reliability.[19]

Sensitivity

SART, ESS and MWT results differed significantly between the responder group and the non-responder group with lower SART and ESS scores and higher MWT sleep latencies for responders (*Table 4*). The corrected ES was ? 0.5 for all outcome measures except for the SART omission error count (*Table 5*). The highest effect size was seen for the ESS.

Using these results we calculated which sample size would be needed to perform a treatment-effect study based on common assumptions (i.e. α=0.05, 1=β=0.9,

two-sided test). To do so, we performed an analysis of covariance, in which we corrected for the baseline values, age and sex. As the results show, using the log-transformed SART commission error count or the total error count allows studies to be designed with lower numbers of subjects than holds if non-transformed outcome parameters are used.

Table 4 - ANCOVA non-responders versus responders corrected for age and sex

	Delta	SD	P value
SART commission errors (log)	0.13	0.16	<0.001
SART omission errors (log)	0.17	0.31	0.007
SART total errors (log)	0.21	0.20	<0.001
MWT sleep latency in min. (log)	- 0.33	0.32	<0.001
ESS score	6.90	5.20	<0.001
SART commission errors	1.83	3.08	0.007
SART omission errors	5.30	7.77	0.002
SART total errors	7.29	8.77	<0.001
MWT sleep latency in min.	- 13.00	11.50	<0.001

Delta is calculated by subtraction of the values of the parameters of the responders from the non-responders.

Table 5 - Cohen's effect size

	Coefficient of correlation	Effect size	Corr. effect size
SART commission errors (log)	0.81	0.81	0.68
SART omission errors (log)	0.64	0.55	0.41
SART total errors (log)	0.76	1.05	0.85
MWT sleep latency in min. (log)	0.63	1.01	0.88
ESS score	0.34	1.33	1.21
SART commission errors	0.50	0.59	0.50
SART omission errors	0.50	0.68	0.47
SART total errors	0.64	0.83	0.64
MWT sleep latency in min.	0.57	1.13	0.99

Cohen's effect size (ES) was calculated by dividing the difference in the mean outcome measure between responders and non-responders by its standard deviation. The corrected effect size was calculated by multiplying the ES by the square root of the coefficient of correlation between the baseline and final values. An ES > 0.50 is regarded as good.

Comparison of the MWT, ESS and SART

Figure 1 shows the results of the factor analysis in which the CGI-C noted at the final patient visit was compared to the mean change from baseline (or relative variation) of the SART, ESS and MWT. There was a significant correlation between CGI-C and all outcome

measures with the highest correlation for delta log-SART total error count (r=0.606) and the ESS (r=0.535).

Factor analysis of CGI-C, SART, ESS, and MWT

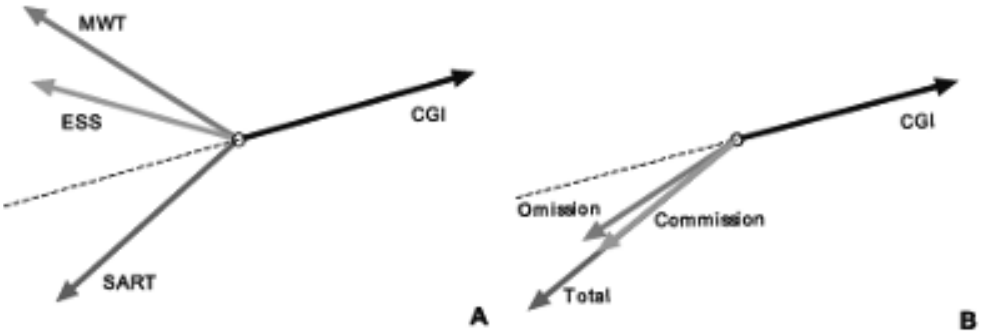


Figure 1 A. Factor analysis of the delta scores of the MWT (log-transformed), ESS, SART (log-transformed total error count) and CGI-C. **B.** Factor analysis of the delta scores of SART outcome measures (all log-transformed), and CGI-C.

The direction of the arrows represents the degree of correlation between the various measures. When arrows point in the exact same direction, they are perfectly positively correlated. Arrows pointing in opposite directions, with an angle between them of 180º indicate a perfect inverse, i.e. negative correlation. Arrows at right angles to one another reflect that the two measures are completely independent. The dashed line represents the 180° opposite of the CGI score.

Figure A shows that the ESS and SART are more parallel to the CGI than the MWT, and the angle between them suggests that they capture different aspects.

Table 6 - Necessitated sample sizes

	N1	N2	N3	N4
SART commission errors (log)	16	14	13	13
SART omission errors (log)	74	58	53	50
SART total errors (log)	13	11	10	10
MWT sleep latency in minutes (log)	16	14	14	13
ESS	13	12	12	12
SART commission errors	64	54	51	50
SART omission errors	71	53	46	43
SART total errors	31	24	23	22
MWT sleep latency in minutes	15	13	13	12

log: log-transformed. N1 until N4 is the sample size necessitated in case of 1/2/3/4 tests or sessions in standard conditions (α=0.05, 1=β=0.9, two-sided test), calculated by an analysis of covariance of the values of the outcome measure, corrected for the baseline values, age and sex.

DISCUSSION

This study demonstrated that the SART is a useful tool to measure treatment efficacy in narcolepsy. Of the various SART outcome measures, the log-transformed total error count proved most sensitive to treatment effects, as established by the CGI-C. The log-transformed commission error count proved the most reliable across sessions performed on the same day; a good reliability of > 0.8 was already reached after performing the SART twice. Performing the SART three times allowed the log-transformed total error count to exceed this threshold as well.

Reliability of the SART, ESS and MWT

Tests are considered reliable when no significant changes in their outcome measures are observed in periods when no change is expected, and when such changes do occur when there is a change in condition. We used the intra-class correlation coefficient to compare reliability of the SART, the ESS and the MWT. As the ESS needed to be administered only once to reach a high level of reliability, the ESS proved the most reliable test. Note that repeated administration of the ESS differed from repeated administration of SART and MWT: the ESS was repeated with an interval of one week, while SART and MWT sessions were repeated on the same day. However, we did not consider this a limitation of high importance, as we aimed at comparing the reliability of SART, MWT and ESS in their usual schedule of administration. The ESS measures experienced sleepiness over the past week(s) or month(s), and is therefore not administered several times per day.

The SART can achieve the same level of reliability as the ESS, but to do so it needed to be administered twice when the log-transformed commission error count was used, and three times when the log-transformed total error count was used.

The MWT reached a similar level of reliability after two sessions, regardless of log transformation. The distribution of the MWT exhibits a ceiling effect meaning that using it for statistical analysis is complex and should be treated with caution.

These results suggest that SART and MWT can measure treatment effects reliably using only the first two or three sessions on one day instead of the four sessions that are conventionally used, given the fact that they are performed at the same time of day as in this study. More than three sessions probably do not relevantly further explain variability and using four tests will be accompanied by higher costs and longer duration. Those who wish to investigate a time of day effect on treatment results might wish to use four or even more tests, but should realize that time of day (morning vs afternoon) did not affect SART performance in a recent study.[27]

Sensitivity of the SART, ESS and MWT

Note that a high reliability of a test does not necessarily mean that it also reflects clinical improvement well. We investigated the latter aspect, sensitivity, for which we used the CGI-C as a gold standard. The ESS, SART and MWT all showed high sensitivity, with highest sensitivity for the ESS. The highest effect size of SART was found for the log-transformed total error count. We also found that the change in clinical condition from baseline to endpoint (CGI-C) was significantly correlated with the changes (delta scores) of all three tests. In fact, of the three studied measurements the change in a SART parameter (log-transformed total errors count) reflected the change in clinical condition most closely, followed by the ESS.

Which aspect of improvement do the various tests reflect?

The SART, ESS and MWT need not reflect the same aspects of the burden of narcolepsy. In fact, in previous studies the SART error count was not related to the ESS, which reflects perceived sleepiness, and the MSLT, which reflects the propensity to fall asleep quickly.[4,5] In these same studies ESS and MSLT results were correlated. We attempted to unravel the correlation between our outcome measures through factor analysis (*Figure 1*). The arrows of the ESS and MWT roughly point in the same direction, which means that changes in MWT and ESS during the study largely reflect the same aspect of the narcolepsy burden. Of these two measures, the treatment response as expressed in the delta ESS score is the better representative of the investigator's impression of treatment response, as the angle between CGI-C and delta ESS is smaller than between CGI-C and MWT. The delta scores of the ESS and SART explain the CGI-C score quite well (i.e. lie close to the 180º opposite of the CGI-C arrow) in a similar magnitude. Interestingly, the delta scores of the ESS and SART form a large angle (close to 90º) among themselves, indicating that they indeed explain different aspects of the CGI-C score. The factor analysis thus shows that the investigator's impression is both based on sustained attention and the ability to stay awake.

The optimal test battery to measure treatment response in narcolepsy

Measures of sleep propensity (MWT) or perceived sleepiness (ESS) on the one hand, and sustained attention on the other hand (SART), are complementary. This study indicates that a combination of the SART with either the MWT or ESS comprises the most suitable combination of the three investigated tests to measure treatment response in narcolepsy. As the MWT and the ESS in part seem to explain the same variability, the question rises which of these tests is best suited to measure treatment effects. The MWT and the SART measure more distinct phenomena than the ESS and SART, as the angle between delta-MWT and delta-SART is closer to 90º in the factor analysis. Another argument in favour of the MWT is that it is easier for a patient to manipulate an ESS result for whatever reason than an MWT result. Then again, there are a number of arguments against the MWT. It

can be manipulated by reducing previous amount of sleep; it is not uniformly carried out: some use 20-minute sessions, others 40-minute ones; some use four, others five sessions; the definition of sleep onset also varies. However, these disadvantages could be overcome by using the protocol recommended in the AASM manual and extensive analysis of sleep prior to the test [13]. Furthermore, the MWT is performed in an artificial setting that need not represent daily life, and, finally, it is time-consuming. Compared to the MWT the ESS is inexpensive, has a high degree of internal consistency and can easily be re-rated for follow-up studies. While these arguments were already known, the present study adds new ones in favor of the ESS over the MWT: it had the highest reliability of all three tests and was more sensitive to subjective treatment efficacy than the MWT. However, when patients have a special interest to perform well, as is the case in for instance the assessment of fitness to drive, including a second measure with a more objective nature than the ESS could be wise.

An interesting characteristic of the SART has to do with the balance between a subjective and objective assessment. An 'objective' test reflects a quantitative test measurement rather than a patient's opinion of disease severity. The SART (and MWT) as objective tests offer the advantage of immunity to manipulation in one direction: it is possible to perform the test worse than one's conditions allows, but not better. However, 'subjective' assessment by patients often forms the primary reason to alter treatment in patient care. The SART has the advantage of objectivity as well as a close relation to subjective changes in severity, reflected in the CGI-C.

We conclude that a single ESS accompanied by two to three SART sessions, depending on the chosen SART outcome parameter, provides a good method to evaluate treatment effects in narcolepsy. This battery comprises two key aspects of narcolepsy, perceived sleepiness and sustained attention, and is easy and cheap to administer.

Detailing SART analysis

Which SART parameter should be used? The factor analysis revealed only minor differences among the various outcome measures of the SART (*Figure 1b*), indicating that they represent the same part of the CGI-C. The highest effect size was found for the total error count. This needs three SART sessions, compared to two for the commission error count. The latter parameter also had a better distribution. The omission error count did not perform as well in terms of distribution and reliability. Still, the total error count did perform well, and, as it contains the omission error count as well, counting omission errors may have a role. The relative importance of omission, commission and total error counts can differ between disorders.[28] We accordingly advise to use the total error count as the primary SART outcome measure.

Reaction time can also be used as a SART parameter, but measuring RT accurately requires special equipment, whereas measuring error counts can be done with standard

personal computers. In the present multicenter study, RTs were not measured.

A different test to measure sustained attention is the Psychomotor Vigilance Task (PVT), which has been used and validated in sleep deprivation studies.[29-31] PVT results in narcoleptics differed from those of healthy controls.[32] The PVT is sensitive to treatment efficacy in obstructive sleep apnea syndrome,[33] but its role in assessing treatment efficacy in narcolepsy would require an assessment similar to the present study, which is currently not available.

Study limitations

Our study is limited to the three measurements of treatment effect that we evaluated, and cannot be used to other potentially useful methods.

Our results are based on data from a study designed to evaluate effects of pitolisant. This means that patients were not selected to represent a typical spectrum of severity of narcolepsy. However, the selection was not limitative, the statistical analysis was prepared before the analysis of the drug trial, and the analysis was conducted independent of the main and secondary endpoints of the trial.

Conclusion

In conclusion, this study shows that the SART, in particular the commission errors and the total error score, is a valid measure to detect treatment effects on sustained attention. A combination of the SART and ESS includes a comprehensive evaluation of treatment effects in narcolepsy, since the ESS represents a subjective estimate of how sleepy patients feel, while the SART is objective in nature. Together they share the advantages of not requiring much time or money, and they correlate well with the clinical global assessment of patient improvement.

ACKNOWLEDGEMENTS

We thank prof. J.C. Schwartz, J.M. Lecomte, E. Deschamps de Paillette, and C. Zerck (Bioprojet, Paris, France) for providing the data for this study.

APPENDIX

Reliability

We calculated the reliability of each outcome measure as the ratio of its observed variability divided by the variability of the true value of the construct that was measured. As this true value cannot be measured directly, it was estimated from our data by means of a linear mixed model. We defined the following linear mixed model to compare the reliability of SART accuracy measures, MWT sleep latency and ESS score:

$$Y(i) = K + Time * [1+N(0, \sigma_e)] + age + gender + \sigma_T$$

In this model, we assumed that the value of the outcome measure (Y) depended on a constant value (K), the variability of the studied outcome measure (σ_T), some random variability expressed as the interaction of within-patient variability (σ_e) with time, and the effects of age and gender. The model contained a random factor for the short time interval in which the value of the outcome measure was not expected to vary within each subject.

Sensitivity

Effect size was also measured using a linear mixed model assuming that the value of the outcome measure depended on a constant value, time, being a responder or not, the interaction of the latter two, age, and gender:

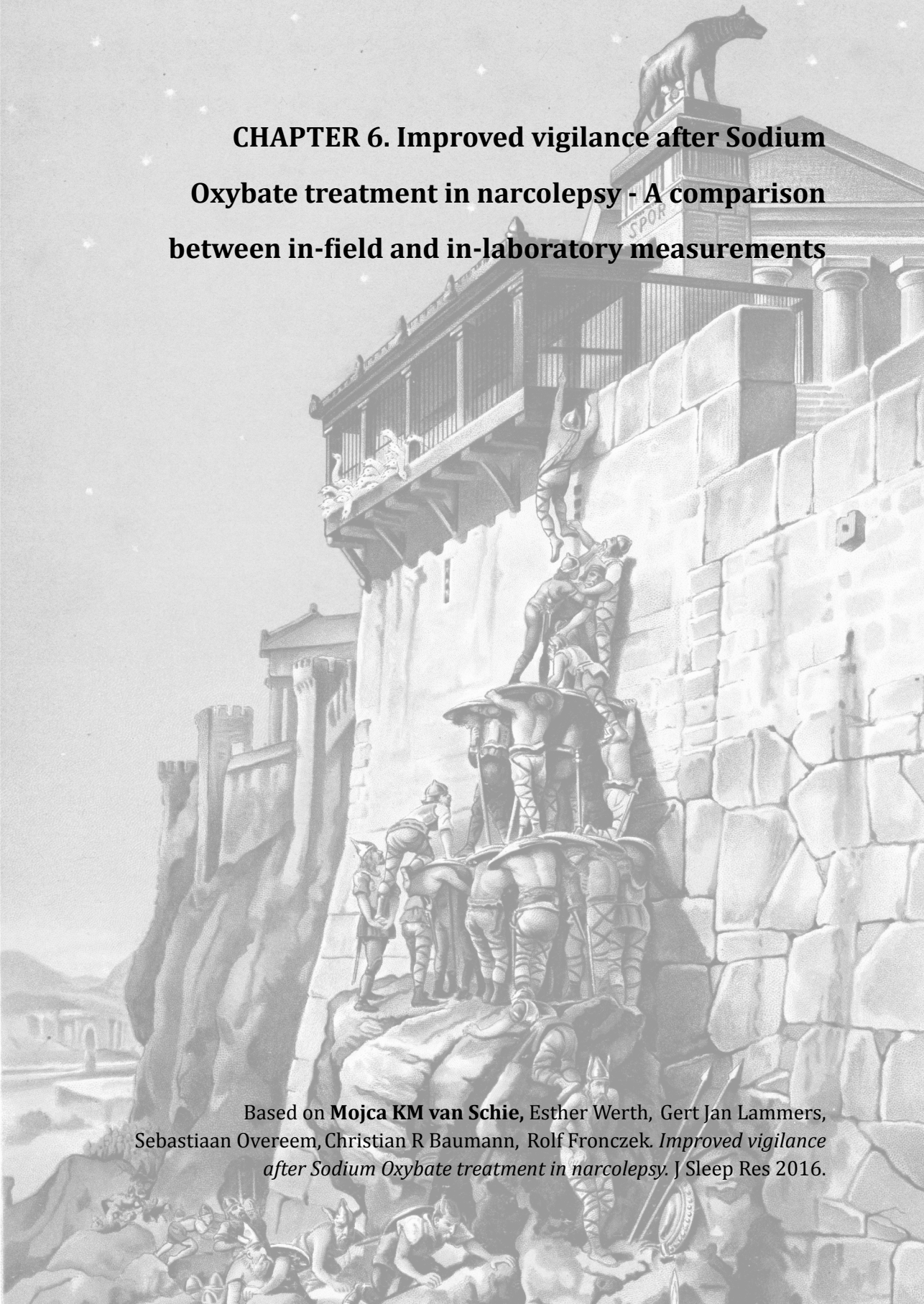
$$Y() = K + Time + Responder + Time * Responder + age + gender$$

REFERENCES

1. Dodel R, Peter H, Spottke A, et al. Health-related quality of life in patients with narcolepsy. Sleep Med 2007;8:733–41.
2. Mitler MM, Gujavarty KS, Browman CP. Maintenance of wakefulness test: a polysomnographic technique for evaluation treatment efficacy in patients with excessive somnolence. Electroencephalogr Clin Neurophysiol 1982;53:658–61.
3. Johns MW. A new method for measuring daytime sleepiness: the Epworth sleepiness scale. Sleep 1991;14:540–5.
4. Fronczek R, Middelkoop HA, van Dijk JG, et al. Focusing on vigilance instead of sleepiness in the assessment of narcolepsy: high sensitivity of the Sustained Attention to Response Task (SART). Sleep 2006;29:187–91.
5. Van Schie MKM, Thijs RD, Fronczek R, et al. Sustained attention to response task (SART) shows impaired vigilance in a spectrum of disorders of excessive daytime sleepiness. J Sleep Res 2012;21:390–5.
6. Robertson IH, Manly T, Andrade J, et al. "Oops!": performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. Neuropsychologia 1997;35:747–58.
7. Arand D, Bonnet M, Hurwitz T, et al. The clinical use of the MSLT and MWT. Sleep 2005;28:123–44.
8. Black J, Houghton WC. Sodium oxybate improves excessive daytime sleepiness in narcolepsy. Sleep 2006;29:939–46.
9. Broughton RJ, Fleming JA, George CF, et al. Randomized, double-blind, placebo-controlled crossover trial of modafinil in the treatment of excessive daytime sleepiness in narcolepsy. Neurology 1997;49:444–51.
10. Johns MW. Reliability and factor analysis of the Epworth Sleepiness Scale. Sleep 1992;15:376–81.
11. Dauvilliers Y, Bassetti C, Lammers G-J, et al. Pitolisant versus placebo or modafinil in patients with narcolepsy: a double-blind, randomised trial. Lancet Neurol 2013;12:1068–75.
12. Manly T, Robertson IH, Galloway M, et al. The absent mind: further investigations of sustained attention to response. Neuropsychologia 1999;37:661–70.
13. Littner MR, Kushida C, Wise M, et al. Practice parameters for clinical use of the multiple sleep latency test and the maintenance of wakefulness test. Sleep 2005;28:113–21.
14. Guy W, National Institute of Mental Health (U.S.).Psychopharmacology Research Branch., Program ECDE. ECDEU Assessment Manual for Psychopharmacology. Rockville, MD.: U. S. Dept. of Health, Education, and Welfare, Public Health Service, Alcohol, Drug Abuse, and Mental Health Administration, National Institute of Mental Health, Psychopharmacology Research Branch, Division of Extramural Research

Programs; 1976.
15. Ferketich S. Focus on psychometrics. Internal consistency estimates of reliability. Res Nurs Health 1990;13:437–40.
16. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika 1951;16:297–334.
17. Lohr KN, Aaronson NK, Alonso J, et al. Evaluating quality-of-life and health status instruments: development of scientific review criteria. Clin Ther 1996;18:979–92.
18. Fleiss JL. The Design and Analysis of Clinical Experiments. New York: John Wiley and Sons; 1986.
19. Lord FM, Novick MR. Statistical theories of mental test scores. Reading, MA: Addison-Welsley Publishing Company; 1968.
20. Bastyr EJI, Price KL, Bril V. Development and validity testing of the neuropathy total symptom score-6: questionnaire for the study of sensory symptoms of diabetic peripheral neuropathy. Clin Ther 2005;27:1278–94.
21. Garralda ME, Yates P, Higginson I. Child and adolescent mental health service use. HoNOSCA as an outcome measure. Br J Psychiatry 2000;177:52–8.
22. Virues-Ortega J, Carod-Artal FJ, Serrano-Duenas M, et al. Cross-cultural validation of the Scales for Outcomes in Parkinson's Disease-Psychosocial questionnaire (SCOPA-PS) in four Latin American countries. ValueHealth 2009;12:385–91.
23. Nelson E, Wasson J, Kirk J, et al. Assessment of function in routine clinical practice: description of the COOP Chart method and preliminary findings. J Chronic Dis 1987;40 Suppl 1:55S–69S.
24. Pinneau S, Newhouse A. Measure of invariance and comparability of factor analysis for fixed variables. Psychometrika 1964;29:271–81.
25. Lee H, Comrey A. Distortions in a commonly used factor analytic procedure. Multiv Behaviour Research 1979;14:301–21.
26. Chakraborty H, Gu H. A Mixed Model Approach for Intent-to-Treat Analysis in Longitudinal Clinical Trials with Missing Values. RTI Press, 2009.
27. Van Schie MKM, Alblas EE, Thijs RD, et al. The influences of task repetition, napping, time of day, and instruction on the Sustained Attention to Response Task. Accepted in J Clin Exp Neuropsychol.
28. Hart EP, Dumas EM, Reijntjes RH, et al. Deficient sustained attention to response task and P300 characteristics in early Huntington's disease. J Neurol 2012;259:1191–8.
29. Dinges DF, Powell JW. Microcomputer analyses of performance on a portable, simple visual reaction task during sustained operations. Behav Res Meth Instr Comp 1985;:652–5.
30. Loh S, Lamond N, Dorrian J, et al. The validity of psychomotor vigilance tasks of less than 10-minute duration. Behav Res Methods Instrum Comput 2004;36:339–46.
31. Basner M, Dinges DF. Maximizing sensitivity of the psychomotor vigilance test (PVT)

to sleep loss. Sleep 2011;34:581–91.
32. Prasad B, Choi YK, Weaver TE, et al. Pupillometric assessment of sleepiness in narcolepsy. Front Psychiatry 2011;2:35–5.
33. Dinges DF, Weaver TE. Effects of modafinil on sustained attention performance and quality of life in OSA patients with residual sleepiness while being treated with nCPAP. Sleep Med 2003;4:393–402.



CHAPTER 6. Improved vigilance after Sodium Oxybate treatment in narcolepsy - A comparison between in-field and in-laboratory measurements

Based on **Mojca KM van Schie,** Esther Werth, Gert Jan Lammers, Sebastiaan Overeem, Christian R Baumann, Rolf Fronczek. *Improved vigilance after Sodium Oxybate treatment in narcolepsy.* J Sleep Res 2016.

ABSTRACT

This two-centre observational study of vigilance measurements assessed the feasibility of vigilance measurements on multiple days using the Sustained Attention to Response Task (SART) and the Psychomotor Vigilance Test (PVT) with portable task equipment, and subsequently assessed the effect of Sodium Oxybate (SXB) treatment on vigilance in narcolepsy patients. Twenty-six narcolepsy patients and 15 healthy controls were included. The study comprised two in-laboratory days for Maintenance of Wakefulness Test (MWT) and Oxford Sleep Resistance test (OSLER), followed by seven-day portable vigilance battery measurements. This procedure was repeated for narcolepsy patients after at least three months of stable treatment with SXB. Narcolepsy patients had a higher SART error count, lower PVT reciprocal reaction time, higher OSLER omission error count adjusted for test duration ($OSLER_{OMIS/MIN}$), and lower OSLER and MWT sleep latency compared to controls (all $P < 0.01$). Treatment with SXB was associated with a longer MWT sleep latency ($P < 0.01$), lower $OSLER_{OMIS/MIN}$ ($P = 0.01$), and a lower SART error count ($P = 0.01$) in narcolepsy patients, but not with absolute changes in OSLER sleep latency or PVT reciprocal reaction time. We concluded that portable measurements of sustained attention as well as in-laboratory OSLER and MWT measurements revealed worse performance for narcoleptic patients compared to controls and that SXB was associated with an improvement of sustained attention and a better resistance to sleep.

INTRODUCTION

Type 1 narcolepsy is a sleep-wake disorder characterized by excessive daytime sleepiness (EDS) and cataplexy. Additional symptoms include disturbed nocturnal sleep, and other rapid eye movement sleep associated symptoms, such as hypnagogic hallucinations, and sleep paralysis and cataplexy (American Academy of Sleep Medicine, 2005). The presence of chronic EDS, that is, the subjective and objective appraisal of the tendency to fall asleep and the ability to stay awake, is mandatory for the diagnosis of narcolepsy. Disturbed vigilance, i.e. a disturbed capability to be aware of internal or external stimuli, is an additional largely neglected symptom of narcolepsy that is directly related to impaired daytime performance and quality of life (Fronczek et al., 2006, Valley and Broughton, 1981). The severe vigilance problems experienced by patients with narcolepsy may be reflected by the inability to recall the content of a conversation, not being able to finish a book, or to concentrate on studies or work.

The Multiple Sleep Latency Tests (MSLT) (Littner et al., 2005) and the Maintenance of Wakefulness Test (MWT) (Littner et al., 2005) are the most commonly used electrophysiological tests for the diagnosis and the quantification of EDS after the initiation of treatment. The MSLT entails the measurement of sleep latency at 4-5 different times on one day while subjects are lying in bed in a quiet, dark room and try to fall asleep. The MSLT thus assesses sleep propensity. The MWT follows a similar schedule, but subjects are requested to try to stay awake instead of trying to fall asleep. The MWT thus assesses the ability to remain awake / resist sleep.

While reports about quality of sleep in narcolepsy are numerous, treatment-effect studies hardly address the quality of wakefulness, even though the value of vigilance measurements has become gradually more recognized (Fronczek et al., 2006, Moller et al., 2006, Van Schie et al., 2012). A study by Weaver (Weaver and Cuellar, 2006) measured changes in quality of life following the administration of sodium oxybate (SXB) in narcolepsy patients. SXB is a strong hypnotic drug, known to be effective in the treatment of disturbed night sleep, EDS, and cataplexy in narcolepsy (The U.S. Xyrem Multicenter Study Group, 2002) In the study by Weaver et al, the nightly administration of SXB produced significant dose-related improvements in the vigilance subscale of the questionnaire (Weaver et al., 1997) that was used. Thus, nocturnal administration of SXB in patients with narcolepsy was associated with clinically relevant improvements in vigilance, an important component of quality of life. Objective tests to measure vigilance, however, have not been applied. The current study was therefore designed to compare vigilance in daily life of narcolepsy patients before and on treatment with SXB by objective measurements.

Several methods are available for this purpose. The most frequently used vigilance measurements are response tasks assessing sustained attention, such as

the Sustained Attention to Response Task (SART) (Robertson et al., 1997) and the Psychomotor Vigilance Test (PVT) (Wilkinson and Houghton, 1982). Such tests only concern external stimuli, not internal stimuli, and measure whether responses are perceived by whether they are acted upon.

SART measurements have been validated in narcolepsy, i.e. have been demonstrated capable of detecting vigilance impairment in narcolepsy compared to healthy controls (Fronczek et al., 2006), while PVT measurements have not yet been validated in this group. The latter however, are widely used in sleep deprivation studies. In contrast, the PVT has been utilized in a portable test version, while SART has not. As direct comparisons of these tests were not available, we decided to combine both tests in a portable task battery. For that purpose, feasibility of portable testing in narcolepsy patients had to be investigated firstly. Since some authors consider the ability to stay awake an aspect of vigilance (Parasuraman et al., 1998), we chose to measure sleep resistance in addition to sustained attention by means of the MWT. Measurements of sustained attention and sleep resistance are combined in the Oxford Sleep Resistance test (OSLER). We therefore considered this test of additional value to a protocol already measuring these aspects, but by different tests in different conditions, that is daily life versus the laboratory. The basic setting for the OSLER is the same as for the MWT, as is its duration. In contrast to the MWT, which is a polysomnographic test and behaviourally undemanding, the OSLER requires continuous monitoring and responding. It is a computerized, non-assisted method for monitoring quality of wakefulness and detecting sleep onset without polysomnography. The OSLER has been validated in patients with obstructive sleep apnea (Bennett et al., 1997). In short, this study comprises the SART, PVT, MWT, and OSLER as objective measurements to compare vigilance in daily life of narcolepsy patients before and on treatment with SXB.

METHODS

Subjects

Subjects were patients with type 1 narcolepsy, diagnosed according to the ICSD-3 criteria (American Academy of Sleep Medicine, 2005). A control group was included to which baseline measurements of sustained attention and sleep resistance were compared. Healthy controls were matched for age and sex. Patients were treatment-naïve and were scheduled to start with SXB or were already using SXB and were prepared to stop medication at least 14 days prior to the study start. The decision for treatment with SXB was part of their therapeutic plan; i.e. no patients were put on SXB treatment for the purpose of participation in this study. Exclusion criteria for both patients and controls were cognitive impairment due to neurological disorders other than sleep-wake

disorders, the use of hypnotics or sleep-wake active drugs other than SXB, and age below 18 or above 70 years.

Twenty-six patients (16 males) were recruited from the narcolepsy outpatient clinics of Leiden University Medical Centre, The Netherlands and University Hospital Zurich, Switzerland between 2007 and 2012. Fifteen controls (8 males) were recruited using notices in local newspapers. The mean age in the patient group was 34.8 years compared to 34.1 years in the control group. Thirteen narcoleptics were available for the second study part. The average of their scheduled treatment dosages was 5.5 g SXB/day.

The protocol was approved by the medical ethical committees of both institutions and written informed consent was obtained from all subjects prior to the study.

Design

The study comprised a nine-day protocol (Figure 1) consisting of two days of in-laboratory sleep resistance tests and a seven-day in-field, i.e. out-of-hospital period of portable vigilance battery measurements. Controls followed the procedure once. Narcolepsy patients followed this procedure before and three months after stable single-drug treatment with the usual therapeutic dose of SXB (4.5 – 9.0 g/day), prescribed by their treating physician.

Figure 1 Overview of study design

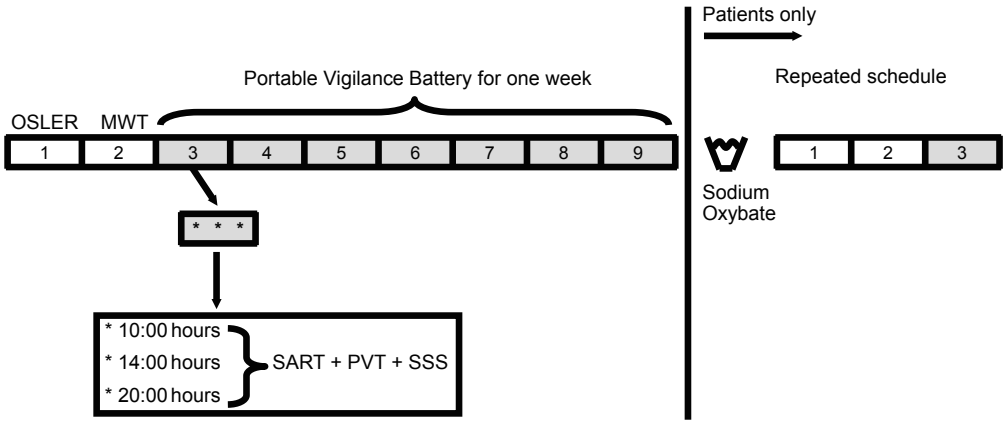


Figure 1: The nine-day protocol consisted of two days of in-laboratory sleep resistance tests and a seven-day in-field, i.e. out-of-hospital period of portable vigilance battery measurements. Controls followed the procedure once. Narcolepsy patients followed this procedure before and three months after stable single-drug treatment with the usual therapeutic dose of SXB, prescribed by their treating physician. SART: Sustained Attention to Response Task; PVT: Psychomotor Vigilance Test; SSS: Stanford Sleepiness Scale; OSLER: Oxford Sleep Resistance; MWT: Maintenance of Wakefulness Test.

Measurements

Vigilance test battery

SART

A number from 1 to 9 was presented 225 times in white on a black computer screen over a 4-minute 19-second period as described previously (Fronczek et al., 2006). Each of the 9 numbers was presented 25 times in a predetermined and quasi-random way so that identical numbers were not clustered. Subjects had to respond to the appearance of each number by pressing a small button, except when the number was a 3. Subjects had to press the button before the next number appeared and were instructed to give equal importance to accuracy and speed in performing the task. The SART error score consisted of the total number of errors, expressed as the sum of the times a key was pressed when no key should have been pressed (i.e. after a '3', the so-called commission errors), and the times when no key was pressed when it should have been (i.e. after anything but a '3', the so-called omission errors).

PVT

Subjects were instructed to press a button as quickly as possible to stop a digital millisecond counter, which started to scroll at variable intervals ranging from 2-10 seconds. Each PVT trial lasted for 10 minutes. We considered the average of the reciprocal RTs (1/RT) the main outcome parameter (Basner and Dinges, 2011), and analyzed the percentage of lapses as secondary outcome parameter.

Subjects received a pocketsize personal digital assistant (PDA) computer to perform SART, PVT and administer the Stanford Sleepiness Scale (SSS) (Hoddes et al., 1973), a seven-point visual-analogue scale to assess momentary EDS, in a 15-minute task battery. Subjects had to take this PDA with them wherever they went during the next seven days. The device could only be turned on during 1-hour-intervals around 10:00 hrs, 14:00 hrs and 20:00 hrs, and gave an acoustic signal at the start of each period. When turned on, instructions appeared on the screen, followed by the SSS. This was followed by a single session of the SART and PVT in random order. Subjects practiced the portable vigilance test battery in the sleep laboratory to assure their familiarity with the device for the start of the ambulatory study part.

MWT

The MWT consisted of four 40-minute sessions in a quiet and dimly lit room according to the AASM recommendations (Littner et al., 2005). The first session started between 1.5 and 3 hours after a participant's usual wake-up time. Consecutive sessions were performed

at two-hour intervals. Subjects were instructed to stay awake while comfortably seated in a semi-supine position. Movements or vocalizations were not allowed. The session was terminated either when sleep onset occurred, defined as three consecutive epochs of stage 1 sleep, or one epoch of any other stage of sleep, or after 40 minutes of being awake. We assessed the mean of the four sleep-onset latencies.

OSLER

The OSLER follows the schedule of the MWT and subjects were similarly positioned. In addition, the participant's dominant hand was placed on a box held in the lap. The index finger was placed on a non-recoil proximity sensor with a sensing distance of 1–2 mm, which transmitted signals of finger contact to a computer. A light-emitting diode was positioned four to six feet away at eye level in the frontal visual field. The light flashed regularly for 1 second every 3 seconds. Subjects were instructed to keep their finger in contact with the button, and to remove the finger for 1 second when the red light flashed. Sleep onset was defined as seven consecutive omissions, i.e. non-responding to flashes for ≥ 18 seconds. The session was terminated when sleep-onset occurred or after 40 minutes of being awake. The mean of the four sleep-onset latencies, a measure of sleep resistance, was considered the primary outcome measure.

By the registration of correct and missed responses before the occurrence of sleep onset, the OSLER may also be interpreted as a measure of sustained attention. We included the following sustained attention outcome measures: the number of omissions per session ($OSLER_{OMIS}$), and the number of omissions per minute test duration ($OSLER_{OMIS/MIN}$).

Actigraphy

Actigraphic data were acquired using wrist actigraphy (on the non-dominant wrist; light sensor data included, Actiwatch, Neurotechnology) (Ancoli-Israel et al., 2003). Determination of the estimated "time in bed" relied on a software algorithm using the activity data recorded by the Actiwatch (Actiwatch Sleep Analysis Version 5, Neurotechnology). The validity of the estimated time in bed detection was improved by using light information and sleep dairy information additionally to set "bed time" and "get-up time". Actigraphic data have been recorded during the week before and week of the in-field study part.

Statistical Analysis

Statistical analysis comprised comparisons of outcome measures between baseline measurements for patients and controls, and before and after treatment with SXB for patients. Data were analyzed using IBM® SPSS® Statistics version 20.

PDA data handling

The in-field nature of the PDA measurements allowed for test sessions being started, but not actually performed due to conflicting demands or technical difficulties at the time of registration. Since the results of all started tests were automatically saved without the possibility of objective verification of their reliability (there was a subjective assessment in the form of a question about having been disturbed while performing the test), unreliable test results had to be filtered out in retrospect before starting the analyses to reduce the type I error in the data obtained from the PDA (Matthias R. Mehl, 2012). The following criteria were defined based on in-laboratory obtained 95th percentile or maximum values for various SART and PVT measurements (Loh et al., 2004, Van Schie et al., 2012) (partially based on unpublished data): SART sessions were regarded unreliable if (1) 50% of RTs were < 200 ms or > 600 ms, (2) if the number of omission errors exceeded 100 for patients or 15 for controls, or (3) if the number of omissions errors was between 40-100 for patients or 10-15 for controls *and* responses to the questions were missing. PVT sessions were filtered out if (1) < 30 responses were recorded (maximum ~ 70-100), (2) if > 50 errors of commission were recorded, or (3) if 50% of responses was > 1000 ms, or 90% of responses > 500 ms. In addition, SART and PVT sessions were regarded unreliable if they concerned abundant sessions (i.e. performed at day 8 or later).

As the number of reliable sessions could differ between subjects from 0 to 21, the outcome measures were separately grouped for time of day within a subject. This resulted in three average values (morning, afternoon, evening) per outcome measure (SART error count, PVT 1/RT) for each subject. To further enhance reliability, only average values with a weight ≥ two original sessions were used in the analyses. 'Average' values consisting of one original session were artificially made missing.

Due to these restrictions, the number of participants with available, reliable PDA data is slightly lower than the total number of participants for certain study parts. Table 1 presents these numbers, as well as the exact number of sessions used in the analyses of the PDA study part. Table 2 illustrates the proportion of PDA sessions available for patients and controls in more detail: firstly irrespective of their reliability, secondly as the proportion of reliable sessions available. The proportions did not differ significantly between patients and controls or before and after treatment with SXB in narcoleptics.

Comparison of narcoleptic patients and controls

Differences between the patient and control groups were primarily analyzed by means of linear mixed effect models (LMMs) of the outcome parameters. The analyses were adjusted for age, time of day and centre, as well as for all two- and three-way interactions between group, time of day and centre. For the PDA measurements, time of day comprised a variable indicating whether the average value was derived from morning, afternoon of evening sessions. "Time of day" for the OSLER and MWT measurements was different

from the PDA measurements and was therefore called differently: since the four sessions were separately analyzed (instead of averaged) to take into account possible fluctuations of sleepiness and vigilance across the day, this variable was called "session".

Table 1 – Group sizes

	Controls	**Narcolepsy patients at baseline**	**Narcolepsy patients on SXB**
SART error count	N = 14 (k = 183)	N = 23 (k = 346)	N = 13 (k = 174)
PVT 1/RT	N = 14 (k = 181)	N = 22 (k = 329)	N = 12 (k = 144)
OSLER	N = 10	N = 22	N = 13
MWT	N = 15	N = 24	N = 13

Legend: SXB: sodium oxybate; N: number of subjects; k: number of sessions; 1/RT: average of reciprocal reaction times. Because of time restraints, 5 controls and 2 narcolepsy patients did not participate in the OSLER study part.

Table 2 – Proportion of PDA sessions performed

	SART at baseline	**SART on SXB**	**PVT at baseline**	**PVT on SXB**
	Proportion of sessions performed: number of sessions performed / 21			
Controls	0.72 (0.60-0.92)	N.A.	0.71 (0.60–0.95)	N.A.
Patients	0.81 (0.67-0.95)	0.62 (0.38-0.90)	0.81 (0.67-0.95)	0.62 (0.38-0.90)
	Proportion of reliable sessions: number of reliable sessions performed / number of sessions performed			
Controls	0.94 (0.84-1.00)	N.A.	0.98 (0.70-1.00)	N.A.
Patients	1.00 (0.89-1.00)	1.00 (0.90-1.00)	1.00 (0.95-1.00)	1.00 (0.85-1.00)

Data are presented as median with 25^{th}–75^{th} percentiles. No significant differences were found between patients and controls or between baseline and post-treatment conditions for patients.
Legend: SXB: Sodium oxybate; NA: not applicable. The number 21 reflects the maximum number of sessions that subjects could have performed, i.e. three sessions per day for seven days.

Thanks to the LMM approach, it was possible to analyze data of subjects with missing values for certain outcome values or time points. Multiple comparisons, inherent to the LMM approach, were accounted for by Bonferroni-Holm adjusted significance levels (adjusted from 0.05) (Holm, 1979). To facilitate the interpretation of *P* values in the context of varying significance levels resulting from this correction, all significant values are marked with an asterisk (*).

Secondarily to the LMMs, the non-parametrical Mann-Whitney U test was used for two-group comparisons in case the assumption of normality had to be dropped, even after data transformation.

Figure 2 Vigilance study parameters from patients vs. controls

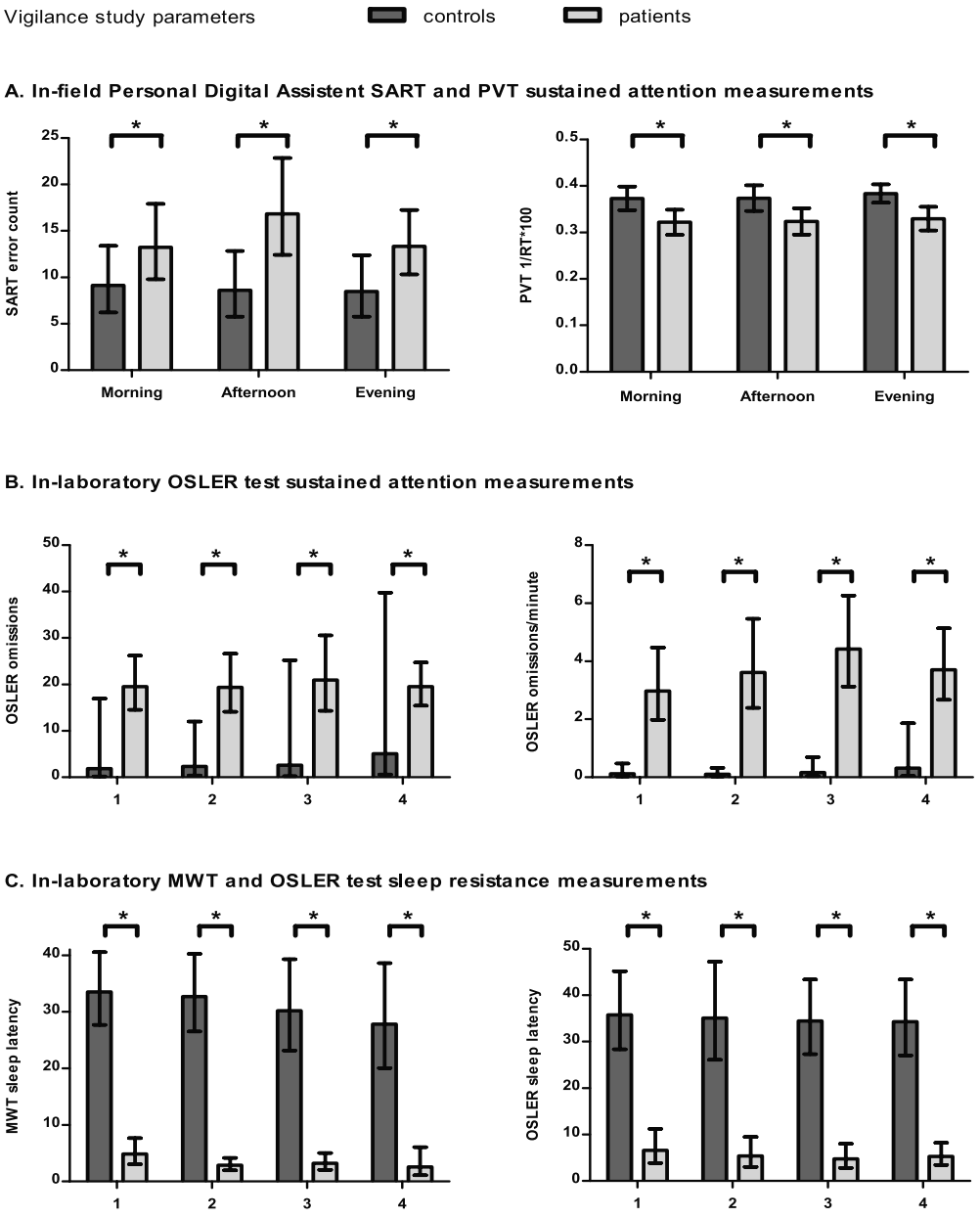


Figure 2: A. Comparisons of seven-day in-field SART and PVT measurements. **B.** Comparisons of one-day in-laboratory MWT and OSLER measurements. Data are presented as geometrical means with 95% confidence intervals. Asterisks (*) flag significant differences between patients and controls. SART: Sustained Attention to Response Task; PVT: Psychomotor Vigilance Test; RT: reaction time; OSLER: Oxford Sleep Resistance; MWT: Maintenance of Wakefulness Test.

Figure 3 Vigilance study parameters before vs. on SXB treatment

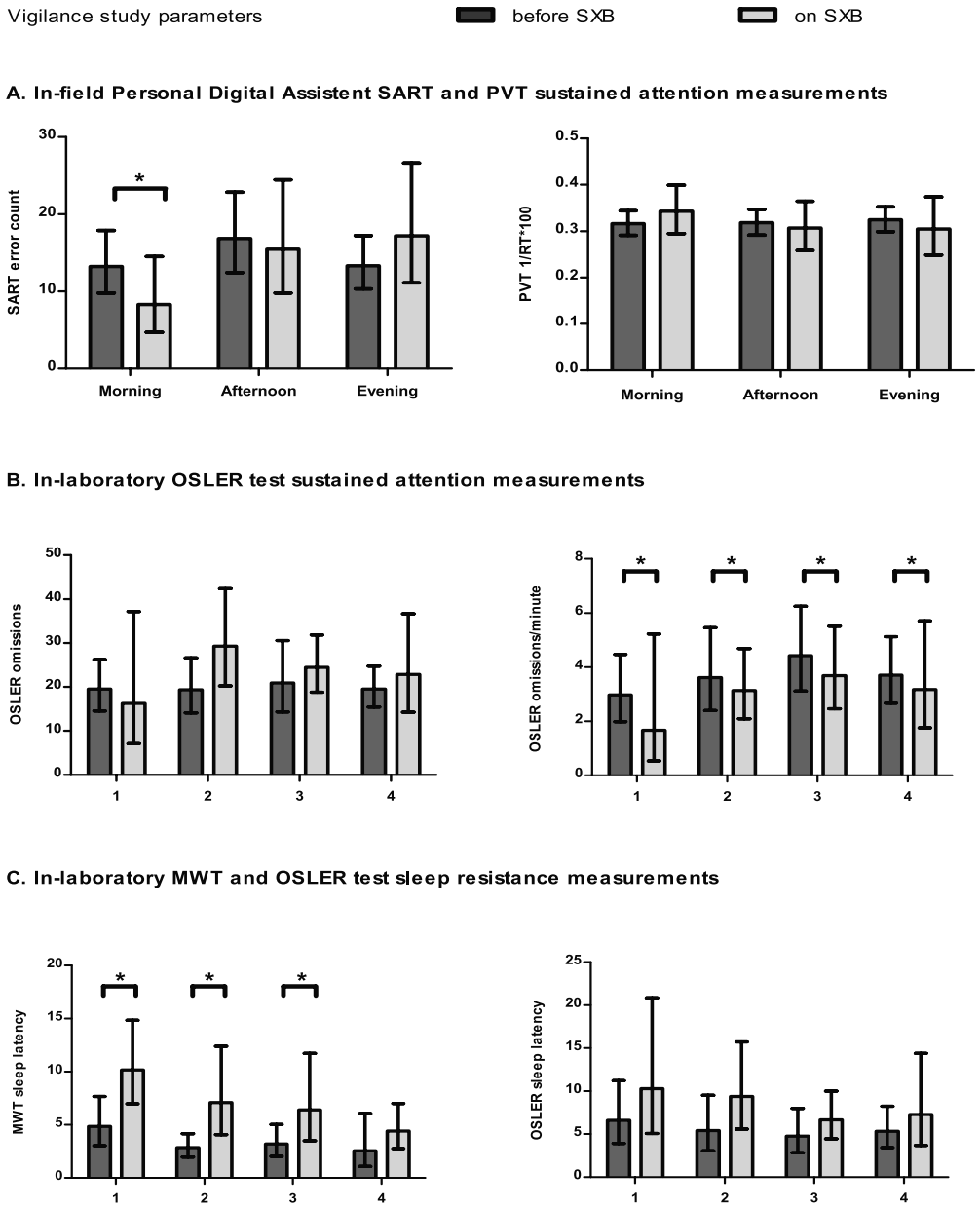


Figure 3: A. Comparisons of seven-day in-field SART and PVT measurements. **B.** Comparisons of one-day in-laboratory MWT and OSLER measurements. Data are presented as geometrical means with 95% confidence intervals. Asterisks (*) flag significant differences between pre-treatment and post-treatment conditions. SART: Sustained Attention to Response Task; PVT: Psychomotor Vigilance Test; RT: reaction time; OSLER: Oxford Sleep Resistance; MWT: Maintenance of Wakefulness Test.

Comparison of narcolepsy patients before and on treatment with SXB

On-treatment data were compared to pre-treatment data of all subjects by LMM analyses similar to those described above. Instead of "group", which was used in the previous comparison, "visit" was analyzed as a factor, which indicated whether measurements were derived from a baseline visit or a post-treatment visit.

Unless specified otherwise, data are presented as median with 25^{th}–75^{th} percentiles in this paper, because most values did not follow a standard normal distribution. Data transformations are indicated in table legends if applicable.

RESULTS

Comparison of narcolepsy patients and controls

Geometrical group means of SART, PVT, OSLER and MWT data are presented in Figure 2. Asterisks flag significant differences resulting from the analyses described below.

Measurements of sustained attention

SART error count was significantly higher in patients than in controls ($P < 0.01$) according to the LMM presented in Table 3. The reciprocal average RT of the PVT was significantly lower in patients compared to controls, i.e. their RT was significantly higher ($P < 0.01$). The proportion of lapses on the PVT was significantly higher in narcolepsy patients (0.10, 0.04-0.19) than in controls (0.02, 0.01-0.05, $P < 0.01$). The average number of omissions on the OSLER was significantly higher in narcoleptics (23.0, 16.0-31.6) than in controls (4.7, 3.4-26.8, $P = 0.01$), which was the same for the number of omissions per minute test duration (4.1, 3.2-6.2 compared to 0.1, 0.1-1.1, $P < 0.01$). Age was inversely correlated with SART error count ($P < 0.01$), but not with PVT or OSLER sustained attention measures.

Measurements of sleep resistance

The MWT sleep latency was 38.5 minutes (23.5-40.0) for controls compared to 4.1 (2.4-5.9) for narcoleptics at baseline ($P < 0.01$). A similar pattern was observed for the OSLER sleep latency, which was 40.0 minutes (33.0-40.0) for controls compared to 8.0 (4.0-15.3) for narcoleptics ($P < 0.01$).

Comparison of narcolepsy patients before and on treatment with SXB

Geometrical group means of SART, PVT, OSLER and MWT data are presented in Figure 3. Asterisks flag significant differences resulting from the analyses described below.

Table 3 – Linear Mixed Models (LMM) of PDA data compared between patients and controls

	SART error count°		**PVT 1/RT*100**	
Cov. matrix	*CS*		*AR1*	
Model parameters				
Basis	***Beta / S.E. / P***			
Intercept	1.04/0.11/0.00	*	0.39/0.02/0.00	*
Target factors	***Beta / S.E. / P***			
Group (G)	0.42/0.11/0.00	*	-0.10/0.02/0.00	*
Time (T)	N.A.		N.A.	
Centre (C)	0.40/0.11/0.00	*	-0.02/0.02/0.36	
Interactions	***Beta / S.E. / P***			
G*C	-0.31/0.14/0.04	*	0.09/0.03/0.00	*
Covariates	***Beta / S.E. / P***			
Age	-0.01/0.00/0.00	*	N.A.	

Asterisks flag significant LMM coefficients. Patients made more errors on the SART and responded slower on the PVT compared to controls.
Legend: Beta: regression coefficient derived from the LMM; S.E: standard error of the regression coefficient; N.A: not available, i.e. no significant contribution to the final model; °: log-transformed parameter; Cov: covariance; CS: compound symmetry; AR1: first-order autoregressive.

Model building strategy: The model of the mean was created from a saturated model including all target factors and possible interactions between them, followed by removing non-significant parameters as long as the model fit was not significantly impaired. The interactions G*T, T*C and G*T*C did not contribute significantly to any of the tested models and were therefore omitted from this table. Target factor coding: Group: 0=controls, 1=patients; Centre: 0=Zurich, 1=Leiden.

Measurements of sustained attention

SXB treatment decreased SART error count according to the LMM presented in Table 4 (P = 0.01). Furthermore, there was a trend towards an interaction of treatment with time of day (P = 0.03), indicating that the significantly lower error count was most pronounced in the morning SART sessions on SXB treatment, but less in the afternoon and evening sessions. PVT 1/RT *per se* was not changed after treatment with SXB, but was significantly lower in the afternoon and evening PVT sessions compared to the morning session on treatment (P = 0.01). The PVT proportion of lapses was not significantly different before (0.10, 0.04-0.19) and on SXB (0.09, 0.03-0.28, P = 0.90). No effect of SXB or any of the other model parameters was found in a LMM analysis of $OSLER_{OMIS}$, as presented in Table 5. In contrast, the number of $OSLER_{OMIS/MIN}$ was significantly decreased on treatment with SXB (P = 0.01). We observed a positive main effect of sessions, indicating that the number of $OSLER_{OMIS/MIN}$ increased during the day. Age was again inversely correlated with SART error count (P < 0.01), as well as $OSLER_{OMIS/MIN}$ and PVT proportion of lapses (R_S -0.323, P < 0.01), but not with PVT 1/RT or $OSLER_{OMIS}$.

Table 4 – Linear Mixed Models (LMM) of baseline versus post-treatment comparisons of PDA data

	SART error count°		**PVT 1/RT*100°**	
Cov matrix	*CS*		*UN*	
Model parameters				
Basis	***Beta / S.E. / P***			
Intercept	1.55/0.13/0.00	*	0.108/0.00/0.00	*
Target factors	***Beta / S.E. / P***			
Visit (V)	-0.18/0.07/0.01	*	0.001/0.00/0.75	
Time (T)	0.01/0.03/0.65		-0.001/0.00/0.32	
Centre (C)	N.A.		0.013/0.01/0.02	
Interactions	***Beta / S.E. / P***			
V*T	0.11/0.05/0.03		-0.005/0.00/0.01	*
V*C	N.A.		N.A.	
T*C	N.A.		0.007/0.00/0.00	*
Covariates	***Beta / S.E. / P***			
Age	-0.01/0.00/0.00	*	N.A.	

Asterisks flag significant LMM coefficients. SXB treatment decreased the number of errors on the SART, but did not alter PVT reciprocal RTs.
Legend: Beta: regression coefficient derived from the LMM; S.E: standard error of the regression coefficient; N.A: not available, i.e. no significant contribution to the final model; °: log-transformed parameter; Cov: covariance; CS: Compound Symmetry; UN: Unstructured.

Model building strategy: The model of the mean was created from a saturated model including all target factors and possible interactions between them, followed by removing non-significant parameters as long as the model fit was not significantly impaired. The interaction V*C did not contribute significantly to any of the tested models and was therefore omitted from this table. Target factor coding: Visit: 0=baseline, 1=post-SXB; Time of day: 0=morning, 1=afternoon, 2=evening; Centre: 0=Zurich, 1=Leiden.

Measurements of sleep resistance

While MWT sleep latency was significantly increased on treatment with SXB ($P < 0.01$), we found no significant effect of treatment on the OSLER sleep latency. However, OSLER sleep latency appeared to decrease across sessions during the day ($P < 0.01$) irrespective of visit (before/after treatment), as no treatment*session interaction was observed. No significant time-of-day effect was found for the MWT sleep latency.

Covariates

Measurements of momentary sleepiness by the SSS did not contribute to any of the tested models and were therefore omitted from all tables.

Age was associated with a minimally higher omission error rate per minute on the OSLER, as well as minimally lower numbers of errors on the SART. Age was not linked to MWT or OSLER sleep latency or to PVT 1/RT.

Table 5 – Linear Mixed Models (LMM) of baseline versus post-treatment comparisons of OSLER and MWT

	OSLER omissions°	**OSLER om/min°**	**OSLER latency°**	**MWT latency°**
Covariance matrix Model parameters	*ARH1*	*UN*	*UN*	*CS*[1]
Basis	***Beta / S.E. / P***			
Intercept	1.36/0.03/ 0.00 *	0.54/0.06/0.00 *	0.88/0.08/0.00 *	0.73/0.09/0.00 *
Target factors	***Beta / S.E. / P***			
Visit (V)	N.A.	-0.11/0.03/0.01 *	N.A.	0.39/0.09/0.00 *
Centre (C)	N.A.	-0.18/0.04/0.00 *	0.07/0.10/0.47	0.05/0.12/0.67
Session (S)	N.A.	0.05/0.01/0.00 *	-0.07/0.01/0.00 *	-0.02/0.02/0.28
Interactions	***Beta / S.E. / P***			
V*C	N.A.	-0.24/0.03/0.00 *	N.A.	N.A.
S*C	N.A.	N.A.	0.06/0.01/0.00 *	-0.04/0.02/0.05
Covariates	***Beta / S.E. / P***			
Age		0.00/0.00/0.05 *	N.A.	N.A.

Asterisks flag significant LMM coefficients. SXB treatment increased MWT but not OSLER sleep latency and decreased the number of OSLER omissions per minute.
Legend: om: omissions; min: minute; Beta: regression coefficient derived from the linear mixed model; S.E: standard error of the regression coefficient; N.A: not available, i.e. no significant contribution to the final model; N.T: not tested in the model; °: log-transformed parameter; [1]: no convergence was reached with any other covariance matrix; ARH1: heterogeneous first-order autoregressive; UN: unstructured; CS: compound symmetry.

Model building strategy: The model of the mean was created from a saturated model including all target factors and possible interactions between them, followed by removing non-significant parameters as long as the model fit was not significantly impaired. The interactions V*S and V*S*C did not contribute significantly to any of the tested models and were therefore omitted from this table. Target factor coding: Visit: 0=baseline, 1=post-SXB; Centre: 0=Zurich, 1=Leiden; Session: 0-3 for the first-fourth session.

SART error counts were higher in Leiden compared to Zurich ($P < 0.01$), especially for controls ($P = 0.04$). Controls in Leiden also had higher PVT reaction times ($P < 0.01$).

There were no site differences for the comparison of SART error count before and after SXB treatment, nor were there main effects of center for PVT 1/RT. On the opposite, a centre*time-of-day interaction effect was observed for PVT 1/RT in narcolepsy patients, i.e. there was a time-of-day effect on PVT 1/RT in Leiden, irrespective of the administration of treatment ($P = 0.01$). Patient versus control comparisons of OSLER and MWT measures did not differ across study sites. The same was found for pre- and on-treatment comparisons of $OSLER_{OMIS}$ and OSLER and MWT sleep latency in narcoleptics. However, there was a significant main effect of centre ($P < 0.01$), as well as a

centre*visit interaction ($P < 0.01$) for $OSLER_{OMIS/MIN}$ in narcolepsy patients, which means that the number of omissions per minute test duration was lower in Leiden, especially on treatment with SXB. The described time-of-day effect on OSLER sleep latency in narcolepsy was less pronounced in Leiden, as indicated by a session*centre interaction ($P < 0.01$).

There was no significant difference in time in bed (night) between controls and narcolepsy at baseline and there was no significant difference in time in bed (night) between narcolepsy patients at baseline and narcolepsy patients during SXB treatment.

DISCUSSION

We investigated sustained attention and sleep resistance in type 1-narcolepsy patients before and during SXB treatment. Sustained attention was measured in normal daily life using the PVT and the SART; and in the sleep laboratory using the $OSLER_{OMIS/MIN}$. Sleep resistance was measured in the sleep laboratory using the MWT and the OSLER sleep latency. Pre-treatment data of narcolepsy patients were compared to data from a matched group of healthy controls. The investigated measurements consistently indicated lower sustained attention and decreased sleep resistance in patients compared to controls. SXB treatment was associated with a better resistance to sleep and a small improvement of sustained attention, i.e. improved wakefulness.

Feasibility of a portable vigilance task battery

Narcolepsy patients and healthy control subjects performed on average 70% of portable SART and PVT sessions in this study, which required subjects to pay attention to three test sessions per day for seven consecutive days and to simultaneously ignore competitive obligations. Narcolepsy patients did not significantly differ from controls in the proportion of sessions performed, indicating the feasibility for these patients to comply with such a demanding protocol. Moreover, over 95% of tests were considered reliable. A shorter protocol is likely to enhance compliance even more.

Quantifying sustained attention and sleep resistance in narcolepsy

Narcolepsy patients had a lower level of sustained attention compared to controls on SART and PVT measurements, as well as on the respective aspects of the OSLER. In other words, both in-field and in-laboratory measurements consistently indicated impaired sustained attention in patients compared to controls.

Compared to previous in-laboratory data, portable SART error count data were approximately 3-4 points higher (Fronczek et al., 2006) and portable PVT RTs were approximately 40-50 ms faster (Dimitrova et al., 2011) for both narcolepsy patients and

controls.

The large difference in MWT sleep latencies between patients and controls is consistent with previous research (Arand et al., 2005, Doghramji et al., 1997). This study indicated that the OSLER was capable of measuring similarly large differences in sleep latency between narcolepsy patients and controls as the MWT.

Sodium oxybate for the treatment of impaired sustained attention and sleep resistance

In line with previous research (The U.S. Xyrem Multicenter Study Group, 2002), we found longer MWT sleep latencies during SXB treatment. On the contrary, OSLER sleep latency was not significantly longer after SXB treatment.

Both in-field (SART) and in-laboratory ($OSLER_{OMIS/MIN}$) error counts were lower during SXB treatment compared to the baseline measurements, whereas PVT reciprocal RT or proportion of lapses was not associated with SXB treatment. Interestingly, the SART and OSLER error counts exhibited a time-of-day variation with the highest performance measured in the morning, i.e. following a night with SXB administration, whereafter performance decreased during the day. As such, this pattern differs from the effect of SXB on cataplexy, which comprises a longer period of time before effects occur. As SXB acts at night to improve sleep duration and has an elimination half-life of 0.5-1.0 hour (The U.S. Xyrem Multicenter Study Group, 2002), the observed improvement of sustained attention in the morning might result from the improved nocturnal sleep duration and stability, for an alerting effect of SXB itself is not expected. This would also clarify why the improvements did not last longer than a few hours: the decreasing sustained attention paralleled the decreasing sleep resistance as the day advanced. An alternative explanation for the observed time-of-day variation might reside in the suppression of dopaminergic neurotransmission by SXB (Maitre, 1997). While SXB has largely disappeared from our body upon awakening, its suppression of dopamine has ended (Donjacour et al., 2011). Hence, from a speculative point of view, it could allow for dopamine to be released in initially high quantities, which positively affects wakefulness.

Partly due to the observed time-of-day variation, the size of the overall differences between treatment conditions for SART and $OSLER_{OMIS/MIN}$ measurements was fairly low (both measurements < 1 error difference), as was the size of the difference for MWT sleep latency (3 minutes). The relatively low mean dosage of SXB used by the study participants might as well have contributed to the low size of differences between treatment conditions. Another explanation may reside in the duration of exposure to SXB treatment. The three months of stable treatment required in this study should suffice to assess a clinical meaningful improvement, but a longer period might be necessary to reach the maximum response (Bogan et al., 2015). Nevertheless, the observed differences in this study may reflect an important clinical effect, similar to observations in modafinil

treatment effect studies: in these studies, differences in MWT sleep latency and Epworth sleepiness scale were also small, but associated with significant clinical improvements measured by clinical global impression scales (The U.S. Xyrem Multicenter Study Group, 1998, 2000). Unfortunately, we did not administer patient-rated clinical effect scales.

Covariates

The covariate age contributed significantly to the models of SART and OSLER sustained attention measurements, but the observed size of this contribution was very small.

This study included populations from Zurich and Leiden, and cultural characteristics could have influenced behavioral measurements. The main differences between the study sites were a worse PDA performance in both patients and controls in Leiden opposed to a better OSLER sustained attention performance of Leiden patients; and a clearer time-of-day effect on the PVT opposed to a less clear time-of-day effect on OSLER sleep resistance in narcolepsy patients at the Leiden site. However, the main findings of the study, i.e. the differences between patients and controls on the one hand and before and on SXB treatment on the other hand were present in both populations.

Limitations

Since actigraphic data of time in bed did not differ between the investigated study groups, the duration of time in bed is unlikely to have confounded our results. It might be considered a shortcoming that we did not assess the chronotypes of our participants, since the possibility of a selection bias between patients and controls in case of different chronotypes between these groups cannot be excluded. However, we did account for the possibility of different chronotypes among our participants by adapting the start of the first session of MWT and OSLER test measurements to a person's regular wake time, as well as by the careful chosen times of the portable vigilance tests, starting not too early (10:00 hours) and finishing not too late (20:00 hours) during the day. Moreover, the comparison between narcolepsy patients at baseline with those during SXB treatment is made within subject, excluding bias due to chronotype differences. Differences in daytime naps, use of caffeinated drinks, and participants' jobs (sedentary/active) might have influenced our results, as we did not register these as covariates during the in-field study phase. This was because we aimed at assessing differences between patients and controls and patients before and during therapy, while living their lives as they normally would do, without restrictions on life style habits or influencing them by asking questions about those habits. In other words, we improved external validity of our data at the expense of internal validity.

The OSLER as a sustained attention task

While the OSLER has been designed as a behavioural and cost-effective alternative to MWT measuring sleep resistance (Bennett et al., 1997), we additionally investigated two behavioural outcome measures: $OSLER_{OMIS}$ and $OSLER_{OMIS/MIN}$. Both measures reflect the behavioural level of vigilance before falling asleep, turning the OSLER into a sustained attention task. In fact, both OSLER and PVT can be considered simple RT tasks, i.e. tasks in which every stimulus requires an active response. While the timing of the stimulus varies from 2-10 seconds in the PVT, it comes at fixed times in the OSLER, creating a monotonous situation.

The number of $OSLER_{OMIS}$ was less sensitive to differences in sustained attention following SXB therapy than the number of $OSLER_{OMIS/MIN}$. This resulted from the simultaneous occurrence of a higher error count and consequently, shorter test duration, as the time to occurrence of seven consecutive errors was shortened. Subjects who could sustain attention for a longer period made a similar number of errors compared to less vigilant subjects (post-hoc analyses), since the duration of their OSLER sessions was longer. Therefore, correcting for test duration would provide a more sensitive measure of sustained attention. This was indeed demonstrated in this study. The sustained attention aspect of the OSLER, reflected by the $OSLER_{OMIS/MIN}$, was even more sensitive to the effects of SXB than the aspect of sleep resistance.

Conclusion

Ambulatory administration of SART and PVT on a PDA was feasible in both narcolepsy patients and controls. PDA SART and PVT measurements as well as in-laboratory OSLER and MWT measurements revealed worse performance for narcolepsy patients compared to controls. In line with previous research, SXB treatment was associated with a better resistance to sleep, measured by the MWT. Moreover, SXB treatment was associated with a small improvement in sustained attention, which was quantified by both OSLER and SART but not PVT. The SART and OSLER offer solutions for a less time- and manpower-consuming evaluation of treatment effects in patients with narcolepsy than PVT and MWT.

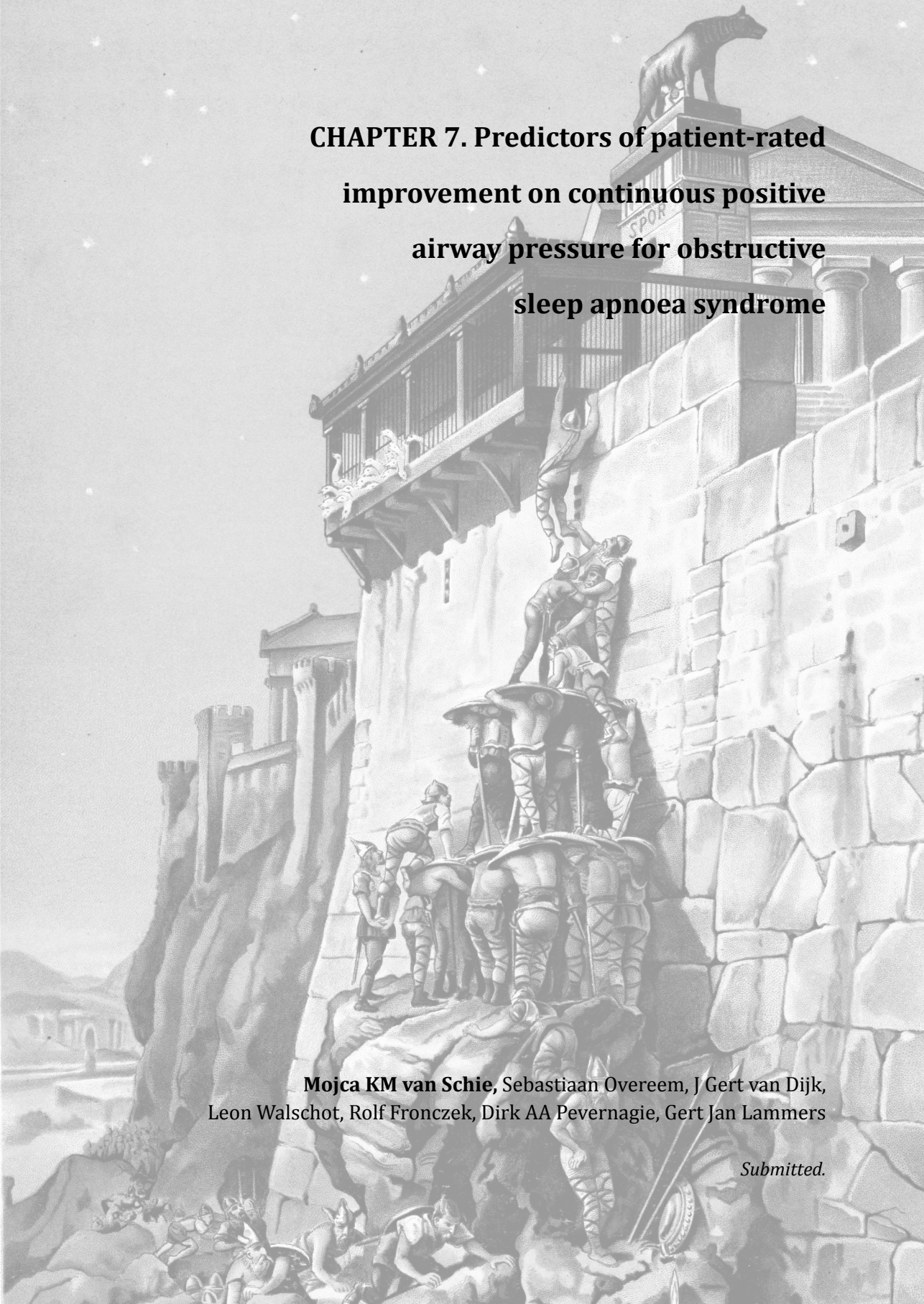
ACKNOWLEDGEMENTS

We thank J.G. van Dijk, C. Bassetti en R. Khatami for their involvement in study design; M. Bach, J. Meier, S. Weber, J.G. van Vliet – de Regt, C.E.H.M. Donjacour, and A. van der Heide for their help in acquiring the data; R.H.A.M. Reijntjes for data processing; and T. Stijnen for statistical advice. We thank UCB Pharma Europe for the research grant that facilitated this study.

REFERENCES

1. American Academy of Sleep Medicine *International Classification of Sleep Disorders. Diagnostic and Coding Manual*. American Academy of Sleep Medicine, Westchester, IL, 2005 (2nd edn edition).
2. Ancoli-Israel, S., Cole, R., Alessi, C., Chambers, M., Moorcroft, W. and Pollak, C. P. The role of actigraphy in the study of sleep and circadian rhythms. *Sleep*, 2003, 26: 342-92.
3. Arand, D., Bonnet, M., Hurwitz, T., Mitler, M., Rosa, R. and Sangal, R. B. The clinical use of the MSLT and MWT. *Sleep*, 2005, 28: 123-44.
4. Basner, M. and Dinges, D. F. Maximizing sensitivity of the psychomotor vigilance test (PVT) to sleep loss. *Sleep*, 2011, 34: 581-91.
5. Bennett, L. S., Stradling, J. R. and Davies, R. J. A behavioural test to assess daytime sleepiness in obstructive sleep apnoea. *J Sleep Res*, 1997, 6: 142-5.
6. Bogan, R. K., Roth, T., Schwartz, J. and Miloslavsky, M. Time to response with sodium oxybate for the treatment of excessive daytime sleepiness and cataplexy in patients with narcolepsy. *J Clin Sleep Med*, 2015, 11: 427-32.
7. Dimitrova, A., Fronczek, R., Van Der Ploeg, J. *et al.* Reward-seeking behavior in human narcolepsy. *J Clin Sleep Med*, 2011, 7: 293-300.
8. Doghramji, K., Mitler, M. M., Sangal, R. B. *et al.* A normative study of the maintenance of wakefulness test (MWT). *Electroencephalogr Clin Neurophysiol*, 1997, 103: 554-62.
9. Donjacour, C. E., Aziz, N. A., Frolich, M. *et al.* Sodium oxybate increases prolactin secretion in narcolepsy patients and healthy controls. *Eur J Endocrinol*, 2011, 164: 363-70.
10. Fronczek, R., Middelkoop, H. A., Van Dijk, J. G. and Lammers, G. J. Focusing on vigilance instead of sleepiness in the assessment of narcolepsy: high sensitivity of the Sustained Attention to Response Task (SART). *Sleep*, 2006, 29: 187-91.
11. Hoddes, E., Zarcone, V., Smythe, H., Phillips, R. and Dement, W. C. Quantification of sleepiness: a new approach. *Psychophysiology*, 1973, 10: 431-6.
12. Holm, S. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* 1979, 6: 65-70.
13. Littner, M. R., Kushida, C., Wise, M. *et al.* Practice parameters for clinical use of the multiple sleep latency test and the maintenance of wakefulness test. *Sleep*, 2005, 28: 113-21.
14. Loh, S., Lamond, N., Dorrian, J., Roach, G. and Dawson, D. The validity of psychomotor vigilance tasks of less than 10-minute duration. *Behav Res Methods Instrum Comput*, 2004, 36: 339-46.

15. Maitre, M. The gamma-hydroxybutyrate signalling system in brain: organization and functional implications. *Progress in neurobiology*, 1997, 51: 337-61.
16. Matthias R. Mehl, T. S. C. *Handbook of Research Methods for Studying Daily Life*. The Guilford Press, New York, NY, 2012 (1st Edition edition).
17. Moller, H. J., Devins, G. M., Shen, J. and Shapiro, C. M. Sleepiness is not the inverse of alertness: evidence from four sleep disorder patient groups. *Exp Brain Res*, 2006, 173: 258-66.
18. Parasuraman, R., Warm, J. S. and See, J. E. Brain systems of vigilance. In), *The attentive brain*. The MIT Press, Cambridge, MA, 1998: 221-56.
19. Robertson, I. H., Manly, T., Andrade, J., Baddeley, B. T. and Yiend, J. 'Oops!': performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. *Neuropsychologia*, 1997, 35: 747-58.
20. The U.S. Modafinil in Narcolepsy Multicenter Study Group Randomized trial of modafinil for the treatment of pathological somnolence in narcolepsy. . *Ann Neurol*, 1998, 43: 88-97.
21. The U.S. Modafinil in Narcolepsy Multicenter Study Group. Randomized trial of modafinil as a treatment for the excessive daytime somnolence of narcolepsy. *Neurology*, 2000, 54: 1166-75.
22. The Us Xyrem Multicenter Study Group. A randomized, double blind, placebo-controlled multicenter trial comparing the effects of three doses of orally administered sodium oxybate with placebo for the treatment of narcolepsy. *Sleep*, 2002, 25: 42-9.
23. Valley, V. and Broughton, R. Daytime performance deficits and physiological vigilance in untreated patients with narcolepsy-cataplexy compared to controls. *Review of electroencephalograpy and clinical neurophysiology*, 1981, 11: 133-9.
24. Van Schie, M. K., Thijs, R. D., Fronczek, R., Middelkoop, H. A., Lammers, G. J. and Van Dijk, J. G. Sustained attention to response task (SART) shows impaired vigilance in a spectrum of disorders of excessive daytime sleepiness. *Journal of Sleep Research*, 2012, 21: 390-5.
25. Weaver, T. E. and Cuellar, N. A randomized trial evaluating the effectiveness of sodium oxybate therapy on quality of life in narcolepsy. *Sleep*, 2006, 29: 1189-94.
26. Weaver, T. E., Laizner, A. M., Evans, L. K. *et al.* An instrument to measure functional status outcomes for disorders of excessive sleepiness. *Sleep*, 1997, 20: 835-43.
27. Wilkinson, R. T. and Houghton, D. Field test of arousal: a portable reaction timer with data storage. *Hum Factors*, 1982, 24: 487-93.



CHAPTER 7. Predictors of patient-rated improvement on continuous positive airway pressure for obstructive sleep apnoea syndrome

Mojca KM van Schie, Sebastiaan Overeem, J Gert van Dijk, Leon Walschot, Rolf Fronczek, Dirk AA Pevernagie, Gert Jan Lammers

Submitted.

ABSTRACT

Study Objectives

To investigate whether vigilance predicted patient-rated improvement after start of continuous positive airway pressure (CPAP) for obstructive sleep apnea syndrome (OSAS) better than parameters of breathing, sleepiness and well-being

Methods

This study comprised a prospective observational treatment-effect study of CPAP in 30 OSAS patients with an apnea-hypopnea index (AHI) >15. Vigilance, assessed through a sustained attention to response task (SART), sleepiness, measured using questionnaires, and well-being, measured with visual-analog scales, were measured during two pre-treatment visits and one after 8 weeks of CPAP. Improvement was scored on the patient-rated Clinical Global Impression of Change (PCGI-C)

Results

A linear mixed model analysis of CPAP effect indicated an improvement of all breathing indices; the AHI decreased from 41.1±24.4 to 4.1±4.3 ($p<0.001$). The Epworth Sleepiness Scale (ESS) decreased from 14.4±4.2 to 7.9±4.8 ($p<0.001$). The 100mm- visual-analog scale (VAS) of physical exhaustion decreased by 6.4 mm ($p=0.009$). No significant difference was observed in the other VAS ratings, nor in the error score on the SART. Eighty percent of patients considered themselves improved on the PCGI-C. This improvement correlated with improvement of breathing indices and the ESS.

Conclusions

The large majority of OSAS patients considered themselves improved after 8-week CPAP treatment. This improvement was best predicted by a decrease of the breathing disturbance indices. Patients' sleepiness also improved significantly. Vigilance did not predict patient-rated improvement. This study did not provide better predictors of subjective improvement after CPAP.

INTRODUCTION

Obstructive sleep apnea syndrome (OSAS) is a sleep-related breathing disorder characterized by apneas and hypopneas during sleep, associated with desaturations and sleep disruption. These may lead to daytime symptoms impairing general well-being, including excessive daytime sleepiness (EDS), decreased vigilance, fatigue, mood disturbances, and cognitive complaints.[1]

The severity of OSAS is traditionally quantified with the apnea-hypopnea index (AHI), i.e. the number of apneas and hypopneas per hour of sleep. Continuous positive airway pressure (CPAP), the most frequently used treatment for moderate to severe OSAS, aims to reduce the AHI and consequently improve symptoms. CPAP improves symptoms of OSAS in the majority of patients, though depending on patients' adherence.[2] AHI reduction is considered an important efficacy parameter of CPAP treatment.[3,4] However, this focus on the AHI has been criticized for two main reasons. Firstly, the pathophysiological consequences of OSAS result from the severity of oxygen desaturation rather than the number of apneas or hypopneas itself, implying that the severity of the breathing disturbance will be reflected better by the oxygen desaturation index (ODI[5-8]). Secondly, improving the AHI with CPAP does not alleviate all symptoms,[9-12] indicating that some symptoms may not be a direct consequence of a reversible sleep-related breathing disturbance.

In addition to diminishing the AHI, CPAP has been described to decrease daytime sleepiness, measured by the Epworth Sleepiness Scale (ESS),[13,14] especially in a subgroup with a baseline AHI >15. CPAP in OSAS is also found to improve cognitive functions,[15] in particular attentional functions.[16] The most significant improvements were observed with tests of divided or sustained attention, more than held for classical vigilance tests involving responses to infrequently occurring stimuli.[17] Several vigilance and sustained attention tests have been used in OSAS, both to describe baseline functions and to assess efficacy of CPAP treatment. Validated tests include the Oxford Sleep Resistance (OSLER) test,[18] Psychomotor Vigilance Test (PVT[19]), Steer-Clear,[20] and Sustained Attention to Response Task (SART[21]). The SART demonstrated impaired vigilance in patients with various sleep disorders, including OSAS.[22] It has not yet been used to evaluate CPAP efficacy. It has, however, proved to correlate well with patient-rated treatment efficacy in narcolepsy patients.[23]

Efficacy of CPAP is usually quantified as an improvement of the AHI and other breathing indices, and through patients' reports. Although some correlations between decrease in AHI and self-reported daytime functioning have been described,[24] a substantial number of studies reported absent correlations between AHI, and measures of well-being or daytime functioning such as sleepiness, vigilance, mood, quality of life, or driving simulator performance, following CPAP treatment.[25]

This lack of a clear relation between improved AHI and subjective improvement is puzzling. We hypothesized that subjective improvement after CPAP treatment would be related to improvement of daytime functioning. Unfortunately, it is not obvious which parameters best reflect daytime functioning. We therefore designed this study concerning CPAP in OSAS to compare patient-rated clinical global improvement to parameters of vigilance, sleepiness, well-being, and breathing disturbances. We hypothesized that vigilance improvement might be the best candidate to reflect patient-rated improvement, since vigilance is a prerequisite for daytime functioning. Earlier studies[15-18,20] yielded contrasting findings, perhaps due to a variety of vigilance tests. We therefore decided to investigate vigilance by means of the SART, which has been shown to correlate well with patient-rated treatment efficacy in narcolepsy patients, as mentioned above.[23]

MATERIAL AND METHODS

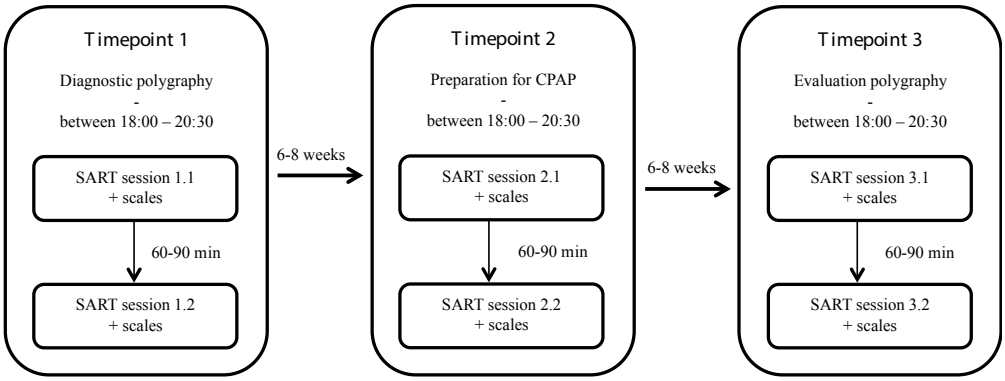
Patients

Study inclusion comprised two steps. Patients referred to the tertiary referral center Kempenhaeghe between June 2011 and June 2013 were screened for eligibility if suspected to have OSAS and aged between 18 and 70 years old. A diagnostic polygraphy/polysomnography was scheduled. OSAS was based on the ICSD-2 criteria[26]. Those with an AHI > 15/hour were candidates for CPAP and were included in the study. Patients with significant comorbidity or coexisting sleep disorders were excluded. The study was approved by the local medical ethical committee, and written informed consent was obtained from all patients prior to the study.

Design

Data were obtained from three overnight visits in the routine work-up and treatment for obstructive sleep apnea (Figure 1). Although the study comprised therapy, this was not part of the study design. There were two pre-treatment visits with 6-8 weeks in between: the diagnostic polygraphy or polysomnography (timepoint 1) and the CPAP titration night to achieve optimal fixed pressure (timepoint 2). One visit assessed the situation after eight weeks of fixed-pressure CPAP (timepoint 3). Vigilance tests and subjective scores were taken at each timepoint. Perceived improvement with CPAP treatment was scored at timepoint 3. Patients were instructed to refrain from caffeine during all visits.

Figure 1. Study design



Diagnostic polygraphy: either polygraphy or polysomnography (see Table 1); scales: visual-analog scales.

Patient- and partner-rated Clinical Global Impression of Change

The CGI-C is a seven-point visual-analogue scale ranging from (1) 'very strong decrease of complaints' to (7) 'very strong increase of complaints'.[27] Though originally developed as a physician-rated scale, previous work by Forkmann et al indicated a moderate to good agreement of a patient-rated version of the CGI-C in comparison to the doctor-rated version.[28] As we aimed to investigate determinants of subjective improvement in well-being, we chose the patient-rated version of the CGI-C as the gold standard. Patients and their partners rated the scale (patient-version called PCGI-C from here on) at visit 3. Furthermore, patients rated the 16-point efficacy index,[27] which combines a score for the impression of change due to the treatment with a score for the inconvenience or adverse effects caused by the treatment. The efficacy index ranges from (1) 'marked improvement without side effects' to (16) 'unchanged or worse symptoms and side effects that outweigh therapeutic effect'.

Determinants

Vigilance

Vigilance was measured through measurement of sustained attention using the SART, a Go/No-Go paradigm characterized by responding to frequent Go trials and withholding responses to infrequent No-Go trials.

The SART was administered while subjects were seated in front of a computer screen in a quiet room. This 4-minute-19-second taking test comprises the numbers 1 to 9 appearing 225 times in random order on a black computer screen. Subjects had to respond to the appearance of each number by pressing a button, except for the number 3, which occurred 25 times in all. Subjects had to press the button before the next number appeared and were instructed to give equal importance to accuracy and

speed in performing the task.[21] Two SART sessions with a 1,5-hour break in between were performed between 18:00 and 22:00 hours on each timepoint.

The primary outcome measure of the SART is the total error score, consisting of key presses when no key should be pressed (i.e. commission errors), and absent presses when a key should have been pressed (i.e. omission errors). The secondary outcome measure is the reaction time, the average time in milliseconds between the appearance of any number and the subject's response. Reaction times could be measured with sufficient accuracy by using a cathode ray tube screen, which was timed using a dedicated video graphics array switch to avoid delays of uncertain magnitude due to build-up of screen data.

Sleepiness

The Epworth Sleepiness Scale served as a general indication of sleepiness during the past month, measured at timepoints 1 and 3. Stanford Sleepiness Scale (SSS) measurements indicated the momentary level of sleepiness and were administered prior to each SART session at timepoints 1, 2 and 3.[29]

Well-being

Patients used seven visual-analog scales (VAS), as previously used in a sleep-restriction study,[30] prior to each SART session at all three visits, assessing the momentary level of general well-being (I feel very bad to very good), daytime alertness (sleepy to alert), stress (stressed to calm), happiness (unhappy to happy), health (sick to healthy), physical exhaustion (physically exhausted to energetic) and mental exhaustion (mentally exhausted to sharp).

Breathing disturbance indices

Apneas were defined as decrements in airflow of at least 90% from baseline for at least 10 seconds.[31] Hypopneas were defined as decrements in airflow of ≥ 50% from baseline for at least 10 seconds, accompanied by a desaturation ≥ 3% from pre-event baseline or an arousal. The sum of apneas and hypopneas per hour formed the AHI. The number of apneas per hour was calculated to obtain the apnea index (AI). The number of desaturations ≥ 3% and ≥ 4% from pre-event baseline per hour were calculated to obtain the oxygen desaturation indices, respectively ODI-3% and ODI-4%. Breathing indices were obtained at timepoint 1 and 3.

Statistical analysis

Data were analyzed using IBM® SPSS® Statistics version 23.

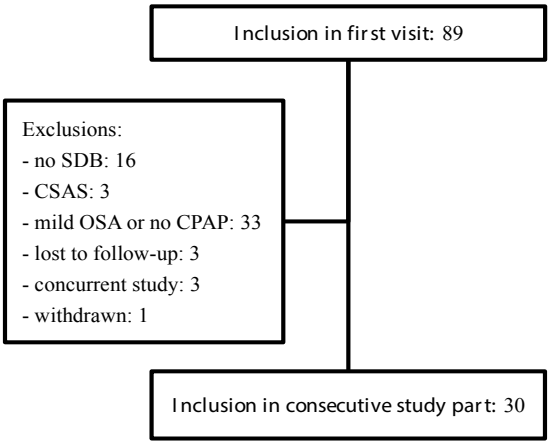
A linear mixed effect model was used to compare trends in parameters of vigilance, sleepiness, well-being, and breathing disturbance before and after treatment, taking into account all repeated measurements separately. Significance was set at the p=0.01 level to

correct for multiple comparisons. Only parameters with a statistically significant change after treatment were used in the subsequent correlation analysis. For this analysis, delta scores for vigilance, sleepiness, well-being, and breathing disturbance were calculated by subtracting the average score before treatment from the score on treatment. Correlations between these delta scores, the PCGI-C scores and the efficacy index were assessed using Pearson's r or Spearman's ρ. Significance was again set at the 0.01 level to correct for multiple comparisons.

RESULTS

Ninety patients were considered eligible. Thirty fulfilled the criteria after polygraphy/ polysomnography and were included (Figure 2, Table 1).

Figure 2. Patient inclusion



SDB: sleep-disordered breathing; CSAS; central sleep apnea syndrome

CPAP compliance data of the month previous to timepoint 3 were available for 28 patients. The median of average CPAP compliance per night was 6:42 hours, and the median percentage of nights with CPAP use > 4 hours was 95%. Eighty percent of patients and 72% of partners reported that patients were much or very much improved on the PCGI-C. No patients or partners considered the patients worsened. Average pre-treatment and post-treatment values of breathing disturbance indices, parameters of sleepiness and vigilance, and VAS scores are displayed in table 1. Average pre-treatment SART error score indicated that pre-treatment vigilance was only moderately disturbed, in contrast to sleepiness and breathing disturbances.[32] Table 2 contains the results of the

repeated-measurements analysis of CPAP on all outcome parameters. CPAP significantly and decreased mean AHI, AI and ODI to normal values. Simultaneously, mean ESS score decreased to a normal value. The VAS rating for physical exhaustion also decreased significantly after CPAP. No significant differences were found for SART error count or reaction time, SSS score, or the other VAS ratings.

Table 1 - Characteristics of the patient group

	Before treatment (N = 30)		**After treatment (N = 30)**	
Patient characteristics				
Mean age (years)	55 ± 8			
Sex (n)	M: 27 (90%)			
	F: 3 (10%)			
Mean BMI	31.3 ± 5.3			
Diagnostic PG/PSG (n)	PG: 11; PSG: 19*			
Test characteristics	*Mean*	*(SD)*	*Mean*	*(SD)*
Breathing disturbances				
AHI	41.1	(24.4)	4.1	(4.3)
AI	22.5	(19.7)	1.3	(2.7)
ODI-3%	29.6	(23.7)	4.3	(4.1)
ODI-4%	36.8	(24.9)	1.9	(2.6)
Sleepiness				
ESS	14.4	(4.2)	7.9	(4.8)
SSS°	4.5	(0.9)	4.8	(1.0)
Vigilance°				
SART error score	11.7	(7.1)	10.1	(6.1)
SART RT (ms)	312	(61)	308	(70)
Well-being (VAS)°				
General well-being	64.3	(19.4)	67.3	(19.0)
Daytime alertness	51.8	(19.9)	58.4	(19.6)
Stress	66.6	(22.2)	70.4	(19.6)
Happiness	71.2	(18.8)	72.8	(20.8)
Health	64.6	(21.6)	66.3	(21.5)
Physical exhaustion	55.8	(19.1)	62.2	(19.6)
Mental exhaustion	54.2	(19.7)	58.9	(19.9)
Mean of VAS	428.4	(125.9)	456.4	(127.2)

Legend: n: number; SD: standard deviation; BMI: body-mass index; PG: polygraphy; PSG: polysomnography; AHI: apnea/hypopnea index; AI: apnea index; ODI: oxygen-desaturation index; ESS: Epworth Sleepiness Scale; SSS: Stanford Sleepiness Scale; SART: Sustained Attention to Response Task; RT: reaction time; ms: milliseconds; VAS: visual-analog scales; * Two patients had to come to the clinic twice for timepoint 1 because of an unreliable polygraphy/polysomnography. The baseline breathing disturbance indices were derived from the second timepoint '1' because of the unreliability of the first. °Average of the 4 pre-treatment measurements (timepoint 1 and 2) and the 2 on-treatment measurements (timepoint 3) respectively.

Table 2– Linear Mixed Models of CPAP on all outcome parameters

Modeled parameter	**Intercept** ***Baseline condition***		**Coefficient** ***CPAP effect***	
Breathing	***Beta / S.E. / p***			
AHI	41.1 / 3.07 / 0.000	*	-37.0/ 1.99 / <0.001	*
AI	22.5 / 2.46 / 0.000	*	-21.2/ 1.60 / <0.001	*
ODI_3%	37.0 / 3.11 / 0.000	*	-32.7/ 1.99 / <0.001	*
ODI_4%	29.7 / 2.94 / 0.000	*	-27.8/ 1.92 / <0.001	*
SART	***Beta / S.E. / p***			
Error score	11.8 / 1.19 / 0.000	*	-1.7 / 0.68 / 0.015	
Reaction time	311 / 11.0 / 0.000	*	-2.8 / 6.17 / 0.656	
Sleepiness	***Beta / S.E. / p***			
ESS	14.6 / 0.69 / 0.000	*	-6.8 / 0.46 / <0.001	*
SSS	4.5 / 0.16 / 0.000	*	0.3 / 0.13 / 0.021	
VAS	***Beta / S.E. / p***			
General well-being	64.4 / 3.21 / 0.000	*	2.6 / 2.15 / 0.223	
Daytime alertness	51.7 / 3.19 / 0.000	*	6.6 / 2.66 / 0.015	
Stress	66.3 / 3.54 / 0.000	*	3.5 / 2.32 / 0.136	
Happiness	71.0 / 3.34 / 0.000	*	1.6 / 2.10 / 0.454	
Health	64.3 / 3.62 / 0.000	*	1.7 / 2.43 / 0.494	
Physical exhaustion	55.8 / 3.17 / 0.000	*	6.4 / 2.39 / 0.009	*
Mental exhaustion	53.6 / 3.34 / 0.000	*	5.0 / 2.43 / 0.040	

Legend: AHI: apnea-hypopnea index; AI: apnea index; ODI: oxygen desaturation index with either 3 or 4% cut-off value; ESS: Epworth Sleepiness Scale; SSS: Stanford Sleepiness Scale; VAS: visual-analog scales; Beta: regression coefficient derived from the linear mixed model; S.E: standard error of the regression coefficient; N.A: not available, i.e. no significant contribution to the final model; N.T: not tested in the model. Compound symmetry was chosen as a model for the covariance matrix. Asterisks flag significant LMM coefficients.

The PCGI-C score and the efficacy index were significantly correlated to all breathing disturbance indices. The better the PCGI-C score was, the more improved were the breathing disturbance indices (Table 3). Partners' CGI-C score was significantly correlated to delta-AHI, only. Delta-ESS itself was significantly correlated to delta-AHI, -AI, and -ODI-4% (not shown in the table), indicating that the lower (i.e. the better) the ESS score, the more improvement of the other outcome measures.

Table 3 – Correlations of outcome parameters with patient-rated improvement

	Patient CGI-C	Partner CGI-C	Efficacy index
AHI	r=0.59 **	r=0.58 **	r=0.51 *
AI	r=0.60 **	r=0.29	r=0.48 *
ODI-3%	r=0.59 *	r=0.45	r=0.41
ODI-4%	r=0.64 **	r-0.36	r=0.46
ESS	r=0.45	r=0.26	r=0.43
VAS			
- Physical exh.	r=-0.22	-0.31	r=-0.22

Legend: CGI-C: clinical global impression of change; AHI: apnea-hypopnea index; AI: apnea index; ODI: oxygen desaturation index with either 3 or 4% cut-off value; ESS: Epworth Sleepiness Scale; exh: exhaustion. Asterisks flag significant correlation coefficients with *: p≤0.01, **: p≤0.001. The outcome parameters used in this correlation analysis are the outcome parameters for which a statistically significant change following CPAP treatment was found with the Linear Mixed Models analysis shown in table 2.

DISCUSSION

We investigated changes in vigilance, sleepiness, well-being, and indices of breathing disturbance after 8-week CPAP treatment in OSAS patients, as well as the correlation between these changes and patient-rated improvement on the PCGI-C. In contrast to our hypothesis, there was no significant change in vigilance as assessed with the SART, possibly because SART performance was only moderately disturbed in this study at baseline. In other words, vigilance as assessed with the SART was not a sensitive indicator of baseline impairment. However, other parameters did show patient-rated improvement. We observed a substantial improvement in breathing disturbance indices, implicating that obstructive sleep apnea as causal factor was well controlled. In addition, we observed a substantial improvement of excessive daytime sleepiness measured by the ESS, and a small improvement in the VAS subscale of physical exhaustion. Eighty percent of patients reported themselves much or very much improved on PCGI-C. This improvement correlated well with the improved breathing disturbance indices but only moderately and not statistically significant to ESS. It did not correlate to the VAS of physical exhaustion either. This study therefore showed that changes of AHI and other parameters assessing breathing disturbance best reflected patient-rated improvement.

The correlation coefficients of the correlations between the PCGI-C and the breathing indices were all large, whereas those between PCGI-C and sleepiness were moderate. These results, as well as the correlations observed between sleepiness on the one hand with breathing disturbance indices on the other hand, contrast with previous studies in which breathing disturbance indices, especially AHI, did not correlate with subjective estimates of daytime functioning.[3,8,17,25,33] A possibly contributing factor might

be that our patient group was preselected on the criterion AHI > 15, appeared to be relatively severely affected in terms of AHI, and that CPAP adherence was high. Moreover, patients with comorbid sleep disorders were excluded. There have been some indications in previous studies that the treatment effect of CPAP differs between OSAS severity groups based on AHI, with more improvement on sleepiness but less improvement on vigilance for groups with AHI > 30 as compared to groups with lower AHI (range varying across studies, mostly 5-10 or 5-15), for which the opposite yields. [3,4,34] This could apply to our study as well. It might also explain the inconsistency of the literature regarding sustained attention in OSAS: Some studies assessing Steer-Clear performance found a significant difference in obstacle hit between OSAS patients and controls,[35] while others did not or did so only in specific OSAS severity groups based on AHI.[4] A treatment-related improvement of Steer-Clear performance was found in some studies. PVT results also differed across studies, some showing improvement following CPAP,[3,36] while others did not. One recent publication by Guaita et al. did not find a change in SART performance, but pre-treatment error count was already relatively low, as in our study.[37]

Study limitations

This was an observational treatment-effect study without a placebo treatment group. Therefore, the relevance of small improvements remains uncertain. The large effect size of the improvements of breathing disturbance indices and ESS score is, however, in the same range as was found in the CPAP treatment group in placebo-controlled CPAP treatment-effect studies.[24] The strong correlation between the improved objective breathing disturbance indices and our patient-rated gold standard is reassuring: it excludes the possibility that patients only found themselves improved as a consequence of medical attention. Diurnal influences could have affected our results, since vigilance and sleepiness measurements have only been taken during evening hours. Although these tests were administered at similar times across visits to minimize the consequences of time-of-day performance fluctuations, recent work in narcolepsy showed the possibility of a treatment-induced time-of-day effect with worst performance in the evening.[38] If a similar mechanism would apply to OSAS patients and CPAP as well, our study could have missed a relevant improvement of vigilance in the morning. Another limitation of this study concerns the use of non-validated VAS instead of validated quality of life measurements, although a momentary rating of well-being prior to each SART session would not have been feasible or meaningful with validated quality of life measurements. Nevertheless, discriminative validity of these VAS remains unknown.

Conclusions

The majority of OSAS patients considered themselves improved after 8-week CPAP treatment. This improvement was best predicted by a large and clinically relevant

decrease of the breathing disturbance indices AHI, AI, and ODI-3% and ODI-4%. Patients' sleepiness also improved significantly. Vigilance did not significantly improve and, as such, did not predict patient-rated improvement. This study therefore did not provide better predictors of subjective improvement after CPAP.

ACKNOWLEDGEMENTS

We thank P. van Mierlo for advice in study design; P.P. Buijserd-Amesz for help in patient inclusion; the sleep laboratory nurses at Kempenhaeghe for their help in acquiring data; and R.H.A.M. Reijntjes for data processing.

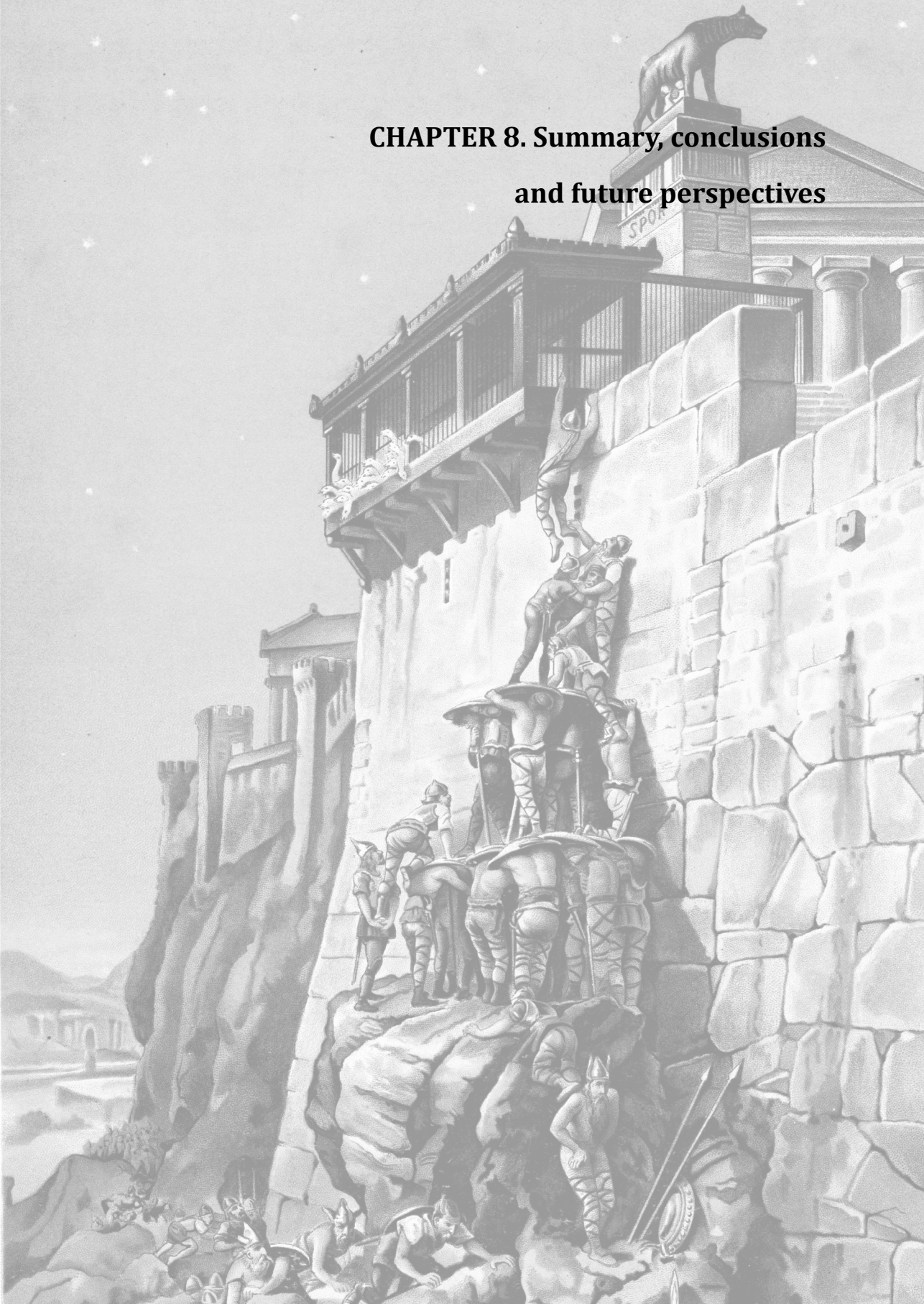
REFERENCES

1. Vaessen TJ, Overeem S, Sitskoorn MM. Cognitive complaints in obstructive sleep apnea. *Sleep Med Rev.* 2015;19:51-58.
2. McMahon JP, Foresman BH, Chisholm RC. The influence of CPAP on the neurobehavioral performance of patients with Obstructive Sleep Apnea Hypopnea Syndrome: A Systematic Review. *WMJ.* 2003;102(1):36-43.
3. Barnes M, Houston D, Worsnop CJ, et al. A randomized controlled trial of continuous positive airway pressure in mild obstructive sleep apnea. *Am J Respir Crit Care Med.* 2002;165(6):773-780.
4. Engleman HM, Kingshott RN, Wraith PK, Mackay TW, Deary IJ, Douglas NJ. Randomized placebo-controlled crossover trial of continuous positive airway pressure for mild sleep Apnea/Hypopnea syndrome. *Am J Respir Crit Care Med.* 1999;159(2):461-467.
5. Bedard MA, Montplaisir J, Richer F, Malo J. Nocturnal hypoxemia as a determinant of vigilance impairment in sleep apnea syndrome. *Chest.* 1991;100(2):367-370.
6. Roehrs T, Merrion M, Pedrosi B, Stepanski E, Zorick F, Roth T. Neuropsychological function in obstructive sleep apnea syndrome (OSAS) compared to chronic obstructive pulmonary disease (COPD). *Sleep.* 1995;18(5):382-388.
7. Shpirer I, Elizur A, Shorer R, Peretz RB, Rabey JM, Khaigrekht M. Hypoxemia correlates with attentional dysfunction in patients with obstructive sleep apnea. *Sleep Breath.* 2012;16(3):821-827.
8. Quan SF, Chan CS, Dement WC, et al. The association between obstructive sleep apnea and neurocognitive performance-the Apnea Positive Pressure Long-term Efficacy Study (APPLES). *Sleep.* 2011;34(3):303-314B.
9. Castiglioni P, Lombardi C, Cortelli P, Parati G. Why excessive sleepiness may persist in OSA patients receiving adequate CPAP treatment. *Eur Respir J.* 2012;39(1):226-227; author reply 227-228.
10. Dinges DF, Weaver TE. Effects of modafinil on sustained attention performance and quality of life in OSA patients with residual sleepiness while being treated with nCPAP. *Sleep Med* 2003;5(5):393-402.
11. Chapman JL, Kempler L, Chang CL, et al. Modafinil improves daytime sleepiness in patients with mild to moderate obstructive sleep apnoea not using standard treatments: a randomised placebo-controlled crossover trial. *Thorax.* 2014;69(3):274-279.
12. Vernet C, Redolfi S, Attali V, et al. Residual sleepiness in obstructive sleep apnoea: phenotype and related symptoms. *Eur Respir J.* 2011;38(1):98-105.
13. Johns MW. A new method for measuring daytime sleepiness: the Epworth sleepiness scale. *Sleep.* 1991;14(6):540-545.

14. Weaver TE, Maislin G, Dinges DF, et al. Relationship between hours of CPAP use and achieving normal levels of sleepiness and daily functioning. *Sleep.* 2007;30(6):711-719.
15. Kielb SA, Ancoli-Israel S, Rebok GW, Spira AP. Cognition in obstructive sleep apnea-hypopnea syndrome (OSAS): current clinical knowledge and the impact of treatment. *Neuromolecular Med.* 2012;14(3):180-193.
16. Aloia MS, Ilniczky N, Di Dio P, Perlis ML, Greenblatt DW, Giles DE. Neuropsychological changes and treatment compliance in older adults with sleep apnea. *J Psychosom Res.* 2003;54(1):71-76.
17. M. O, Duchna HW, Leidag M, et al. Driving simulator and neuropsychological corrected testing in OSAS before and under CPAP therapy. - *Eur Respir J.* 2005;26(5):898-903.
18. Bennett LS, Stradling JR, Davies RJ. A behavioural test to assess daytime sleepiness in obstructive sleep apnoea. *J Sleep Res.* 1997;6(2):142-145.
19. Wilkinson RT, Houghton D. Field test of arousal: a portable reaction timer with data storage. *Hum Factors.* 1982;24(4):487-493.
20. Findley LJ, Fabrizio MJ, Knight H, Norcross BB, LaForte AJ, Suratt PM. Driving simulator performance in patients with sleep apnea. *The American review of respiratory disease.* 1989;140(2):529-530.
21. Robertson IH, Manly T, Andrade J, Baddeley BT, Yiend J. 'Oops!': performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. *Neuropsychologia.* 1997;35(6):747-758.
22. Van Schie MK, Thijs RD, Fronczek R, Middelkoop HA, Lammers GJ, Van Dijk JG. Sustained attention to response task (SART) shows impaired vigilance in a spectrum of disorders of excessive daytime sleepiness. *Journal of Sleep Research.* 2012;21(4):390-395.
23. van der Heide A, van Schie MK, Lammers GJ, et al. Comparing Treatment Effect Measurements in Narcolepsy: The Sustained Attention to Response Task, Epworth Sleepiness Scale and Maintenance of Wakefulness Test. *Sleep.* 2015;38(7):1051-1058.
24. Siccoli MM, Pepperell JC, Kohler M, Craig SE, Davies RJ, Stradling JR. Effects of continuous positive airway pressure on quality of life in patients with moderate to severe obstructive sleep apnea: data from a randomized controlled trial. *Sleep.* 2008;31(11):1551-1558.
25. Kingshott RN, Vennelle M, Hoy CJ, Engleman HM, Deary IJ, Douglas NJ. Predictors of improvements in daytime function outcomes with CPAP therapy. *Am J Respir Crit Care Med* 2000(161):866-871.
26. American Academy of Sleep Medicine. *International Classification of Sleep Disorders. Diagnostic and Coding Manual.* 2nd ed. Westchester, IL: American Academy of Sleep Medicine; 2005.

27. Guy W. *ECDEU Assessment Manual for Psychopharmacology* Rockville, MD: U.S. Department of Health, Education, and Welfare; 1976.
28. Forkmann T, Scherer A, Boecker M, Pawelzik M, Jostes R, Gauggel S. The Clinical Global Impression Scale and the influence of patient or staff perspective on outcome. *BMC psychiatry.* 2011;11:83.
29. Hoddes E, Zarcone V, Smythe H, Phillips R, Dement WC. Quantification of sleepiness: a new approach. *Psychophysiology.* 1973;10(4):431-436.
30. Dinges DF, Pack F, Williams K, et al. Cumulative sleepiness, mood disturbance, and psychomotor vigilance performance decrements during a week of sleep restricted to 4-5 hours per night. *Sleep.* 1997;20(4):267-277.
31. Iber C A-IS, Chesson AL, Quan S. *The AASM manual for the scoring of sleep and associated events: rules, terminology and technical specification.* 1st ed. Westchester, IL: American Academy of Sleep Medicine; 2007.
32. van Schie MK, Alblas EE, Thijs RD, Fronczek R, Lammers GJ, van Dijk JG. The influences of task repetition, napping, time of day, and instruction on the Sustained Attention to Response Task. *Journal of clinical and experimental neuropsychology.* 2014;36(10):1055-1065.
33. Crawford-Achour E, Dauphinot V, Martin MS, et al. Protective Effect of Long-Term CPAP Therapy on Cognitive Performance in Elderly Patients with Severe OSA: The PROOF Study. *J Clin Sleep Med.* 2015;11(5):519-524.
34. Marshall NS, Barnes M, Travier N, et al. Continuous positive airway pressure reduces daytime sleepiness in mild to moderate obstructive sleep apnoea: a meta-analysis. *Thorax.* 2006;61(5):430-434.
35. Munoz A, Mayoralas LR, Barbe F, Pericas J, Agusti AG. Long-term effects of CPAP on daytime functioning in patients with sleep apnoea syndrome. *Eur Respir J.* 2000;15(4):676-681.
36. Bhat S, Gupta D, Akel O, et al. The relationships between improvements in daytime sleepiness, fatigue and depression and psychomotor vigilance task testing with CPAP use in patients with obstructive sleep apnea. *Sleep Med.* 2018.
37. Guaita M, Salamero M, Vilaseca I, et al. The Barcelona Sleepiness Index: A New Instrument to Assess Excessive Daytime Sleepiness in Sleep Disordered Breathing. *J Clin Sleep Med.* 2015;11(11):1289-1298.
38. van Schie MK, Werth E, Lammers GJ, Overeem S, Baumann CR, Fronczek R. Improved vigilance after sodium oxybate treatment in narcolepsy: a comparison between in-field and in-laboratory measurements. *J Sleep Res.* 2016;25(4):486-496.

CHAPTER 8. Summary, conclusions and future perspectives



GENERAL INTRODUCTION AND AIM OF THE THESIS

Patients with sleep disorders often experience problems in daily life due to impaired vigilance. Type 1 narcolepsy, a disorder caused by hypocretin deficiency, is an excellent example of a disorder of severely disturbed vigilance. Chapter 1 offered an introduction to the measurement of vigilance impairment in sleep disorders. While several methods have been proposed to measure vigilance impairment, only few have been applied to sleep disorders and none have explicitly been validated in these patient groups. Promising results have been observed in a study of Sustained Attention to Response Task (SART) measurements in patients with type 1 narcolepsy.[1] The SART is a 4-minute 19-second lasting go-/no-go task assessing sustained attention.[2,3] This thesis covers several steps in the validation process of the SART as a means to quantify vigilance in sleep disorders.

PART I – MEASURING VIGILANCE

This part addresses basic aspects of vigilance measurements in sleep disorders. Chapter 2 deals with different definitions of vigilance with the aims of discussing the various concepts involved and arriving at a new definition. Chapter 3 extends previous vigilance measurements in narcolepsy by means of the SART to include other sleep disorders than narcolepsy. Various aspects possibly influencing SART outcome are analysed in Chapter 4, such as task repetition, napping, time of day, and test instruction. These chapters will now be discussed separately and briefly.

Challenges in defining vigilance

Chapter 2 addresses the differences between variously proposed definitions of the concept 'vigilance'. All the identified variants of the definition of vigilance proved to be linked to aspects of alertness, sustained attention and arousal; in turn, all these concepts were themselves the subject of variable interpretations. We proposed a new definition of vigilance; it is defined as the capability to be aware of potential changes in one's environment, including a quantitative dimension, expressed as a level of alertness, and a temporal dimension. Attention adds a goal-directed focus to this capability. Sustained attention refers to prolongation of this capability over time. To disentangle the various related concepts, we found it necessary to also specify what we mean by arousal. To do so we upheld its linguistic meaning, so we define arousal as an upward change in alertness taking place in a short time.

Vigilance is defined as the capability to be aware of potential changes in one's environment, including a quantitative and a temporal dimension.

Sustained attention to response task (SART) shows impaired vigilance in a spectrum of disorders of excessive daytime sleepiness.

Chapter 3 described a cross-sectional study of the SART as a tool to measure vigilance in patients with different causes of excessive daytime sleepiness: 42 patients with type 1 narcolepsy, 5 with type 2 narcolepsy, 37 with idiopathic hypersomnia, and 12 with obstructive sleep apnoea syndrome (OSAS). The main finding was that the SART error rate was high in all four patient groups. The median SART error count did not differ between groups, nor did other SART descriptors. This result confirmed the previous suggestion that a high SART error count was not specific for any disease entity. Instead, a high error count probably reflects a key symptom of the disorders in question, i.e. a vigilance impairment. The previously reported absence of a correlation between the SART error count and sleep latency in the multiple sleep latency task (MSLT) already implied that vigilance and sleepiness represent different phenomena.[1] The SART error count proved inversely correlated with reaction time variability, and the number of commission errors was inversely correlated with mean reaction time. These relations have previously been explained as the so-called 'speed–accuracy trade-off'.[2-5] We found diurnal effects on SART performance, with the highest SART error count at the first session in the morning. This higher error score may reflect a brief learning period or an underlying time-of-day effect.

Vigilance, as quantified by the SART, is as impaired in type 1 narcolepsy as it is in other causes of excessive daytime sleepiness. The absence of a correlation between SART and MSLT measurements indicates that the two tests measure separate constructs.

The influences of task repetition, napping, time of day, and instruction on the Sustained Attention to Response Task

Chapter 4 dealt with the possible influences of time of day, napping, repetition, and instruction on the performance on two consecutive sessions of the SART in 100 healthy subjects. The aim was to unravel the mechanism responsible for a decrease in SART error count from the first to the second SART session (See also Chapter 3). The first part of the study comprised a 2x2 design with two time-of-day groups (morning or afternoon) and two nap groups (20-minute nap one hour before the second SART session versus no nap) in 80 healthy subjects. Twenty additional subjects took part in the second study

part, to provide a match for 20 subjects from the first study part; they followed the same procedure, but with a different test instruction.

The results demonstrated that an improvement from the first to the second SART session was only found when subjects were instructed to pay attention to accuracy and to ignore response speed. The improvement is likely attributed to an effect of repetition, i.e. a learning effect, although subjects were not aware of such an effect, given their own performance judgments. This repetition effect was sufficiently minimized after a full 4-minute 19-second practice session. The effect of the specific instruction given probably also explains why some previous studies reported a learning effect[1], in contrast to other studies.[6]

This study also demonstrated that the SART error count was significantly higher in healthy subjects who were instructed to pay equal attention to accuracy and speed than in those instructed to pay attention to accuracy only and to ignore response speed. The 'accuracy first' instruction led to a lower error count with lower between-subject variability. As such, it is the preferred instruction to use when it is importance to assess the best error count. The 'accuracy first' instruction also yielded the largest difference in error score between narcolepsy patients and controls.

In addition, this study also showed that a nap opportunity of 20 minutes more than one hour prior to a SART session did not influence the error count of that session in healthy subjects. The time of day had no clear effect on the SART error count; if such an effect existed at all, it was small and occurred only following the instruction to pay equal attention to accuracy and speed. The 'accuracy first' instruction allowed a comparison between SART sessions administered at different times of the day during normal working hours.

The 'accuracy first' instruction is best suited to assess the lowest error count a subject can achieve. To minimize the consequences of the learning effect that has been observed when using this instruction, we strongly recommend the use a full 4-minute 19-second practice session. Time of day during regular working hours and a nap occasion, as is present in the MSLT, do not influence SART results in healthy controls.

PART II - SUSTAINED ATTENTION TO RESPONSE TASK AS A TREATMENT-EFFECT PARAMETER IN DISORDERS CHARACTERISED BY EXCESSIVE DAYTIME SLEEPINESS

The chapters in this part dealt with the SART as a parameter of treatment efficacy in sleep studies. Chapters 5 and 6 concern measurements in type 1-narcolepsy patients. These chapters differ in two ways: the treatment that was investigated (pitolisant versus sodium oxybate) and the other outcome measurements in addition to the SART (ESS and MWT versus PVT, OSLER and MWT). Chapter 7 comprises the SART as a parameter of treatment efficacy parameter of continuous positive airway pressure in obstructive sleep apnoea.

Comparing treatment effect measurements in narcolepsy: the SART, ESS and MWT

Chapter 5 described the validation of the SART as a parameter of treatment efficacy in narcolepsy, comparing its performance to that of the Maintenance of Wakefulness Test (MWT) and Epworth Sleepiness Scale (ESS). This study was performed within a randomized controlled, double-blind, parallel-group, multicentre trial comparing the effects of 8-week treatments with pitolisant, modafinil, or placebo in ninety-five patients with type 1 or 2 narcolepsy.[7] MWT, ESS, and SART were administered at baseline and after an 8-week treatment period. The severity of excessive daytime sleepiness and cataplexy was also assessed using the Clinical Global Impression scale (CGI-C). Both reliability and sensitivity of SART, MWT and ESS as compared to the CGI-C were addressed. The SART, MWT, and ESS all had good reliability, obtained for the SART and MWT using two to three sessions in one day. The log-transformed SART commission error count proved the most reliable SART parameter across sessions performed on the same day; a good reliability, of > 0.8, was already reached after two SART sessions. Performing the SART three times allowed the log-transformed total error count to exceed this threshold.

The ability to distinguish responders from non-responders, classified using the CGI-C score, was high for all measures, with a high performance for the log-transformed total error count ($r = 0.61$) and the ESS ($r = 0.54$). A subsequent factor analysis indicated that changes in MWT and ESS during the study largely reflected the same aspect of the narcolepsy burden, whereas the SART reflected a completely different aspect. The factor analysis showed that the investigator's impression is both based on sustained attention and the ability to stay awake.

This study showed that the SART, in particular the number of commission errors and the total error score, is a valid measure to detect treatment effects in type 1 narcolepsy. A combination of the SART and ESS provides a comprehensive evaluation of treatment effects: while the ESS represents a subjective estimate of how sleepy patients feel, the SART is objective in nature. Together they share the advantages of requiring little time and money and correlating well with the clinical global assessment of patient improvement.

Improved vigilance after sodium oxybate treatment in narcolepsy

Chapter 6 reports a two-centre observational study of vigilance measurements in 26 patients with type 1 narcolepsy and 15 healthy controls. The aim of this study was to assess the feasibility of vigilance measurements on multiple days using the SART and the Psychomotor Vigilance Test (PVT) with portable equipment, and subsequently to assess the effect of sodium oxybate treatment on vigilance in narcolepsy patients. The study concerned two measurement of the MWT and Oxford Sleep Resistance test (OSLER), followed by seven-day portable vigilance battery measurements. This procedure was repeated for narcolepsy patients after at least three months of stable treatment with sodium oxybate. Ambulatory administration of SART and PVT proved feasible in both narcolepsy patients and controls. These, as well as OSLER and MWT measurements, revealed worse performance by narcolepsy patients compared to controls. Sodium oxybate treatment was associated with a better resistance to sleep measured by the MWT. Moreover, treatment was associated with a small improvement in sustained attention, quantified by both OSLER and SART but not PVT.

Portable measurements of sustained attention as well as in-laboratory OSLER and MWT measurements revealed worse performance for narcolepsy patients compared to controls. Sodium oxybate treatment was associated with an improvement of sustained attention and a better resistance to sleep. The SART and OSLER offer solutions for a less time- and manpower-consuming evaluation of treatment effects in patients with narcolepsy than PVT and MWT.

Predictors of patient-rated improvement on continuous positive airway pressure for obstructive sleep apnea syndrome

Chapter 7 describes a prospective observational treatment-efficacy study of continuous positive airway pressure (CPAP) in OSAS. The study aimed at investigating which parameters best predicted patient-rated improvement. Candidate parameters included

breathing indices (the apnoea-hypopnoea index -AHI), questionnaires of sleepiness (ESS and the Stanford Sleepiness Scale), the SART, and visual-analogue scales of several aspects of well-being. Improvement was scored using the patient-rated Clinical Global Impression of Change (PCGI-C). Thirty OSAS patients with an AHI >15 were investigated during two pre-treatment visits and one visit after 8 weeks of CPAP. A marked improvement after CPAP was observed for all breathing parameters, as well as the ESS, whereas only marginal improvements were observed for SART performance and some visual-analogue scales. Eighty percent of patients considered themselves improved on the PCGI-C. This improvement correlated well with improvement of breathing parameters and the ESS: patients who considered themselves much or very much improved, also had the most improved breathing parameters and ESS score. No correlation was observed between PCGI-C and SART error score.

The majority of OSAS patients considered themselves improved after 8-week CPAP treatment. This improvement was best predicted by improved sleep-related breathing indices, as well as excessive daytime sleepiness scored by ESS. SART measurements of vigilance improved only marginally, not correlated to patient-rated clinical global improvement.

FUTURE PERSPECTIVES

The studies described in this thesis contribute to the validation of the SART as a parameter of treatment efficacy in sleep disorders with excessive daytime sleepiness. The SART proved to be robustly resistant against external influences such as time of day or taking a nap. We identified the optimal test instruction, improving discrimination between patients and healthy individuals. Additionally, the SART was able to detect treatment effects and correlated well with patient-rated improvement in type 1 narcolepsy, and proved to be complementary to measurements of sleepiness.

Nevertheless, additional studies are needed to further validate its use in other sleep disorders. Our OSAS study (Chapter 7) illustrates that its validity differs between sleep disorders. Our studies should preferably be replicated in other sleep laboratories, if possible using larger patient groups. This is especially important for the observational study of baseline SART measurements in various sleep disorders (Chapter 3), since the number of patients in this study varies between each reported group. Such a study should preferably include an individually matched control group.

First priority: replication of Chapter 4 in patients with various sleep disorders

In Chapter 4 we had investigated various aspects that may influence the SART, such as time of day, napping in between SART sessions, repetition effect and test instruction. Their effect should be investigated in patients with sleep disorders other than narcolepsy. The reason for this is that the associations between a specific instruction and the error count, or between the instruction and learning effect, need not be equally strong for patients and controls, nor for other sleep disorders. Type 1 narcolepsy is the only sleep disorder patient group for which a study with each test instruction has been performed. The error rate of patients with type 1 narcolepsy who received the instruction to pay attention to accuracy only was similar to that of patients with narcolepsy who received the instruction to pay equal attention to both accuracy and speed[1,8]. It seems likely that patients with narcolepsy function at their maximum capacity when instructed to pay equal attention to both accuracy and speed: their long RT suggests that they did not in fact pay equal attention to both aspects, but that they had sacrificed speed to maintain some accuracy.[8] In that case dropping the speed condition altogether would not improve accuracy. Their low level of accuracy[1], i.e. their inability to sustain attention for four minutes, quite probably reflects the problems narcoleptics face in daily life when trying to follow a conversation or read a book. Nevertheless, a direct comparison of test instruction in a group of narcolepsy patients should be preferred.

Moreover, such comparisons are required for other sleep disorders. Chapter 3 reported on SART performance in patients with idiopathic hypersomnia and obstructive sleep apnoea in addition to patients with narcolepsy. Subjects were instructed to pay equal attention to speed and accuracy. No control group was included. SART error count was visually compared to that of a historical healthy control group. In retrospect, this group turned out to be instructed to prefer accuracy over speed, so this comparison has its problems. Moreover, the study in Chapter 4 demonstrated that healthy controls might reach error counts similar to narcolepsy patients when instructed to pay equal attention to speed and accuracy as a result of the speed-accuracy trade-off. As a result, no inference may be made regarding the abnormality of the SART error count in patients with idiopathic hypersomnia and obstructive sleep apnoea in Chapter 3. These patients may possibly also function at maximum capacity when instructed to pay equal attention to both accuracy and speed. One indication that this held for narcolepsy patients is their long reaction time, observed even when subjects were instructed to pay equal attention to both accuracy and speed. The reaction times of healthy control subjects are significantly shorter when performing the SART with this test instruction. The reaction times of patients with idiopathic hypersomnia and obstructive sleep apnoea as described in Chapter 3 lie in between those of healthy controls and narcolepsy patients. As such, the two explanations for the abnormality of their error counts remain possible. Investigating

the SART error count while instructing these patients to prefer accuracy to speed is therefore essential to understand whether vigilance impairment is indeed present in various sleep disorders.

The results of Chapter 4 were obtained after the studies in Chapters 5, 6 and 7 had been commenced. In the three latter studies, the SART had been administered with the instruction to pay equal attention to accuracy and speed. These studies comprise within-subject comparisons instead of a cross-sectional measurement. Narcolepsy patients seem to perform the SART poorly with both test instructions, so it is unlikely that the differences between pre-treatment and on-treatment conditions in Chapters 5 and 6 would have been even higher in case the instruction to prefer accuracy over speed had been provided. This explanation is not likely to hold true for Chapter 7. In that chapter, no clear improvement in SART performance had been observed in OSAS patients following CPAP treatment. Since a direct baseline comparison of OSAS patients and healthy controls was not performed, it remains unknown whether their baseline performance was abnormal. Indeed, their relatively high error rate could be the result of a strategy, since the patients were instructed to pay equal attention to accuracy and speed. A direct comparison with healthy controls remains therefore needs to be performed, preferably with the "accuracy first" instruction, as this optimises performance. If such a study would indicate a baseline difference between OSAS patients and healthy controls, then the CPAP treatment effect study would subsequently have to be replicated with the "accuracy first" instruction.

Second priority: SART accuracy parameters in future research

The SART parameters presented in this thesis are divided into accuracy measurements (commission error count, omission error count and total error count) and reaction time measurements (for instance average reaction time and reaction time variability). Only minor differences among the various accuracy measures of the SART were reported in Chapter 5, indicating that they reflect the same phenomenon. The highest effect size was found for the total error count. Statistically, the commission error count, i.e. the count of key presses when no key should have been pressed) was more reliable: this parameter was judged to yield reliable results when only two SART sessions were performed, compared to three for the total error count. The omission error count, i.e. the count of absent presses when a key should have been pressed), did not perform as well in terms of distribution and reliability. Still, the total error count did perform well, and, because it contains the omission error count as well, counting omission errors may have a role. The relative importance of omission, commission, and total error counts can differ between disorders.[9] The total error count was therefore chosen as primary outcome measure throughout this thesis. Nevertheless, it remains prudent to assess these error types separately when trying to replicate the findings of this thesis, especially when different patient groups are studied.

The SART in relation to other tests of sustained attention

Reaction times are recorded when performing the SART, as with other tests of sustained attention. We limited the number of reaction time parameters to two. The main value of the SART is to provide information about someone's capability to appropriately detect a change in stimuli to which an alternative response is needed, a function that is expressed through the error rate, more than through reaction time only. The influence of reaction time is minimised when using the "accuracy first" instruction. A practical disadvantage of a focus on reaction time is that measuring it accurately requires special equipment to exclude inaccuracies due to technical factors such as the monitor refresh rate or processing a mouse click. These factors cause inaccuracies that can vary over time, and therefore have the ability to unpredictably influence reaction time results. In contrast, measuring an error count only requires a standard personal computer. In case the main interest concerns simple reaction time measurements, then the PVT may suffice (preferably with special equipment for the reasons described above). This test has primarily been used and validated in sleep deprivation studies.[10-12] Studies in sleep disorders are scarce. Recently, baseline PVT results in patients with narcolepsy, idiopathic hypersomnia and behaviourally induced insufficient sleep syndrome were shown to differ from those of healthy controls.[13] In addition, the PVT is sensitive to treatment efficacy in obstructive sleep apnoea syndrome,[14] but its role in assessing treatment efficacy in narcolepsy was not proved in the study described in Chapter 6 of this thesis. In that study, the SART outperformed the PVT as a parameter of treatment effect measure in narcolepsy.

In the same study, we investigated the OSLER as a sustained attention task by adding behavioural outcome measures reflecting the level of vigilance before falling asleep. In fact, both the OSLER and PVT can be considered simple reaction time tasks. While the timing between stimuli varies from 2-10 seconds in the PVT, the stimuli of the OSLER are presented at a monotonous fixed rate. The sustained attention aspect of the OSLER, reflected by the parameter '$OSLER_{OMISSIONS/MINUTE}$', was more sensitive to the effects of sodium oxybate treatment than the PVT. Both the SART, measured in daily life, and OSLER, measured in the sleep laboratory, demonstrated capable of measuring treatment effects in narcolepsy. A direct comparison of both measures in the sleep laboratory would be interesting to assess the position of the OSLER as a sustained attention task.

Limitations of the SART

A major limitation of the SART is that a poor performance is not always due to a vigilance problem. Other explanations of poor performance include cognitive and motor problems. Cognitive disturbances that affect attention come into play, and any motor problem that affects reaction time may do so too. Accordingly, the importance of the SART is not that it is highly and purely sensitive to vigilance disturbances, but lies in the quantification of such impairments and its ability to compare situations, such as before and during

treatment. It can be applied to quantify vigilance similarly as the MWT is applied to quantify sleepiness.

Performance on the SART is presumably influenced by motivational and environmental factors. If a test subject decides to perform poorly, the SART will show an abnormal response, so a poor performance may be the result of fraud. It is however nearly impossible to perform the SART abnormally well, so the test is resistant to fraud when it is important to perform well, as is the case for driving ability. The SART is therefore quite robust in this context, in contrast to the MWT, in which patients can use tricks to stay awake. Hence, the SART is suitable in situations where patients have a vested interest in performing well.

Position of the SART in future research and patient care

The studies in this thesis hopefully contributed to underline the importance of vigilance for patients with sleep disorders. We feel this is worthwhile, as advocated in the introduction in this thesis, as a vigilance impairment impacts functioning in daily life and may have serious safety implications, for instance regarding driving or working heavy machinery. In addition to the importance of vigilance impairment for individual patients, there are practical reasons to focus on measurements of vigilance rather than sleepiness. Currently used objective measurements of sleepiness such as MSLT and MWT require a laboratory setting. Clinical experience indicates that some patients feel their daytime functioning improved substantially due to stimulant drugs whereas their MWT improves hardly if at all. The non-arousing circumstances in the laboratory may impair the validity to the test to assess problems in the real world.

This thesis expands the understanding of essential practical prerequisites for a reliable and valid use of the SART in future studies and patient care, summarized by the following recommendations:

1. Five SART sessions prior to each of five MSLT sessions are recommended to quantify the level of vigilance for diagnostic purposes.
2. For other purposes we recommend to administer at least two SART sessions with 1.0-1.5 hour in between, preceded by a full training session. The time of day during regular working hours does not influence outcome.
3. Instructions to the patient should consist of the following: “A number from 1 to 9 will be shown 225 times in random order. You have to respond to the appearance of each number by pressing a button, except when the number is a 3. You have to press the button before the next number appears, but note that accuracy is more important than speed.”
4. The following data should be recorded: the number of times a key was pressed when a 3 was presented (commission errors), the times when no key was pressed when it should have been (omission errors), and preferably also the reaction time of every correct press.

5. The SART error score is the preferred outcome measure and consists of the total number of errors, expressed as the sum of the commission and omission errors.

Recommendations 2 and 3 require further study in patients with sleep disorders, among which type 1 and 2 narcolepsy, idiopathic hypersomnia and obstructive sleep apnoea syndrome. Afterwards, vigilance measurements by means of the SART should form a cornerstone of future treatment efficacy studies of new drugs designed to improve daytime functioning. As an objective, functional outcome measure that correlates well with perceived improvement, the SART should at least be positioned as equally important as measurements focusing on sleepiness.

Vigilance measurement, by means of the SART, deserves to be more widely applied in studies of treatment efficacy in sleep disorders.

In addition to the role of the SART in assessing efficacy of drugs, it may well have a role in the assessment of fitness to drive. Dutch legislation on fitness to drive in patients with sleep disorders is currently based upon the opinion of a patient's functioning by an independent physician, supported by additional investigations depending on the sleep disorder. Patients with obstructive sleep apnoea syndrome should for instance have an apnoea-hypopnoea index under 15. However, functional outcome measures are not required. Patients with narcolepsy and idiopathic hypersomnia should have a mean sleep latency above eight minutes on the MWT and an ESS score under 11. In other words, they have to fulfil the requirements of two sleepiness measurements; no quantification of vigilance is required. Moreover, both measures of sleepiness have important disadvantages. Validity in real-world circumstances is not guaranteed. There is no evidence that MWT performance is a reliable predictor of risk of accident, although it does correlate to driving simulator performance in narcolepsy and OSAS.[15,16] In addition, the MWT is a costly investigation. In contrast, the ESS is a cheap and self-administered questionnaire on daily life situations, which does not share these disadvantages. However, the ESS is susceptible to voluntary efforts to perform better than usual, and is therefore not suitable as solitary additional test. As such, there is a need for novel parameters to support the physician's assessment of fitness to drive in patients with excessive daytime sleepiness. A crucial step in validation of the SART as a clinical tool in sleep medicine should therefore be to address the question whether the SART, both cheap and objective in nature, is a more reliable estimator of the risk of accident in the real-world situation than the MWT. To answer this question, SART and MWT performance should preferably be compared to a driving test on the road, taken by a driving instructor or examiner.

Driving simulator performance may aid to the comparison, but should not be the only test of real driving performance.

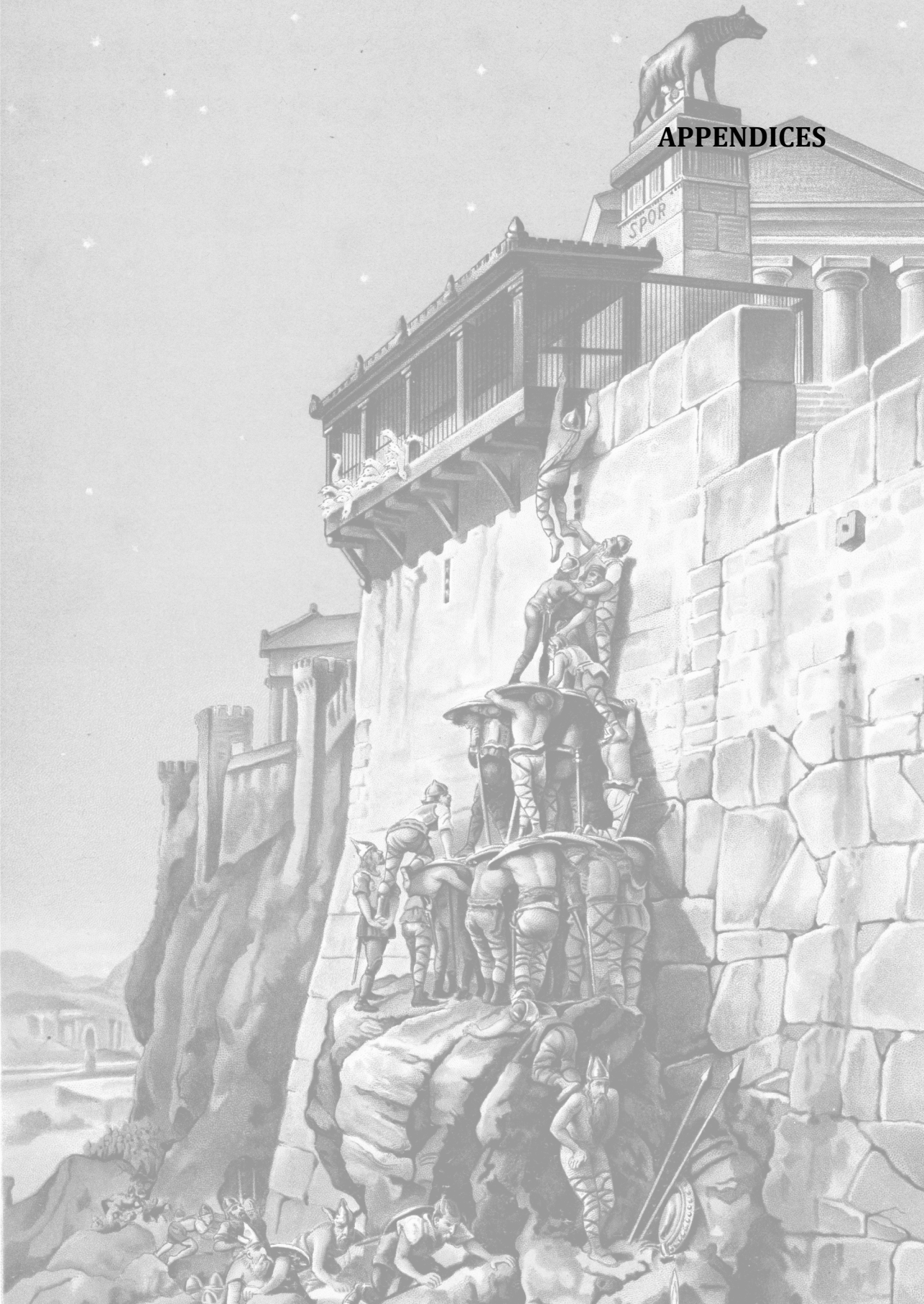
The SART may prove suitable as a tool to assess fitness to drive in patients with sleep disorders. In addition to its validity to quantify treatment effects, it does not share some major disadvantages of the currently applied tests: it is cheap, easy to administer, and quite robust against attempts to perform better than in a real life situation.

REFERENCES

1. Fronczek R, Middelkoop HA, van Dijk JG, Lammers GJ. Focusing on vigilance instead of sleepiness in the assessment of narcolepsy: high sensitivity of the Sustained Attention to Response Task (SART). *Sleep.* 2006;29(2):187-191.
2. Robertson IH, Manly T, Andrade J, Baddeley BT, Yiend J. 'Oops!': performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. *Neuropsychologia.* 1997;35(6):747-758.
3. Manly T, Robertson IH, Galloway M, Hawkins K. The absent mind: further investigations of sustained attention to response. *Neuropsychologia.* 1999;37(6):661-670.
4. Helton WS, Weil L, Middlemiss A, Sawers A. Global interference and spatial uncertainty in the Sustained Attention to Response Task (SART). *Conscious Cogn.* 2010;19(1):77-85.
5. Shalgi S, O'Connell RG, Deouell LY, Robertson IH. Absent minded but accurate: delaying responses increases accuracy but decreases error awareness. *Exp Brain Res.* 2007;182(1):119-124.
6. McAvinue L, O'Keeffe F, McMackin D, Robertson IH. Impaired sustained attention and error awareness in traumatic brain injury: implications for insight. *Neuropsychological Rehabilitation.* 2005;15(5):569-587.
7. Dauvilliers Y, Bassetti C, Lammers GJ, et al. Pitolisant versus placebo or modafinil in patients with narcolepsy: a double-blind, randomised trial. *Lancet neurology.* 2013;12(11):1068-1075.
8. Van Schie MK, Thijs RD, Fronczek R, Middelkoop HA, Lammers GJ, Van Dijk JG. Sustained attention to response task (SART) shows impaired vigilance in a spectrum of disorders of excessive daytime sleepiness. *Journal of Sleep Research.* 2012;21(4):390-395.
9. Hart EP, Dumas EM, Reijntjes RH, et al. Deficient sustained attention to response task and P300 characteristics in early Huntington's disease. *Journal of neurology.* 2012;259(6):1191-1198.
10. Dinges DF, Kribbs NB. Performing while sleepy: effects of experimentally-induced sleepiness In: Monk TH, ed. *Sleep, sleepiness & Performance.* Chister, UK: Wiley; 1991:97-128.
11. Dinges DF, Powell JW. Sleepiness is more than lapsing. *Sleep Res.* 1988;17(84).
12. Dorrian J, Rogers NL, Dinges DF. Psychomotor vigilance performance: a neurocognitive assay sensitive to sleep loss. In: Kushida C, ed. *Sleep deprivation: Clinial Issues, Pharmacology and Sleep Loss Effects.* New York: Marcel Dekker; 2005:39-70.
13. Thomann J, Baumann CR, Landolt HP, Werth E. Psychomotor vigilance task demonstrates impaired vigilance in disorders with excessive daytime sleepiness. *J Clin Sleep Med.* 2014;10(9):1019-1024.

14. Dinges DF, Weaver TE. Effects of modafinil on sustained attention performance and quality of life in OSA patients with residual sleepiness while being treated with nCPAP. *Sleep Med* 2003;5(5):393-402.
15. Philip P, Sagaspe P, Taillard J, et al. Maintenance of Wakefulness Test, obstructive sleep apnea syndrome, and driving risk. *Ann Neurol.* 2008;64(4):410-416.
16. Philip P, Chaufton C, Taillard J, et al. Maintenance of Wakefulness Test scores and driving performance in sleep disorder patients and controls. *Int J Psychophysiol.* 2013;89(2):195-202.

APPENDICES



A. SAMENVATTING, CONCLUSIES EN TOEKOMSTPERSPECTIEVEN

INTRODUCTIE

Patiënten met slaapstoornissen ervaren vaak problemen in het dagelijks leven als gevolg van een gestoorde 'vigilantie' ofwel waakzaamheid. Dit is bij uitstek het geval bij patiënten met narcolepsie type 1, een ziekte die wordt veroorzaakt door een tekort aan de neurotransmitter hypocretine in de hersenen. Een sterk gestoorde vigilantie is een kernsymptoom van deze ziekte. Hoe een dergelijke vigilantiestoornis bij patiënten met slaapstoornissen kan worden gemeten wordt uitgelegd in **hoofdstuk 1.** Hoewel er verschillende mogelijkheden zijn om vigilantie te meten, wordt slechts een handvol van deze tests binnen de slaapgeneeskunde toegepast. Geen van deze tests is gevalideerd binnen patiënten met slaapstoornissen. Een eerste onderzoek naar de Sustained Attention to Response Task (SART, letterlijk vertaald 'proef van volgehouden aandacht tot een reactie') bij patiënten met narcolepsie type 1 toonde veelbelovende resultaten.[1] De SART, een 4 minuten en 19 seconden durende computertaak, meet aanhoudende aandacht.[2,3] De test valt in de categorie 'go /no go': de testpersoon moet telkens beslissen of een stimulus al dan niet een respons vergt. Dit proefschrift beschrijft verscheidene stappen op weg naar validatie van de SART als kwantificatie van vigilantie bij patiënten met slaapstoornissen.

DEEL I – HET METEN VAN VIGILANTIE

Dit deel beschrijft enkele basale aspecten van het meten van vigilantie bij patiënten met slaapstoornissen. **Hoofdstuk 2** bevat een overzicht van de verschillende definities van het begrip vigilantie die circuleren in de wetenschappelijke literatuur en heeft tot doel om tot een algemeen acceptabelle definitie te komen. In **hoofdstuk 3** breidden wij het eerdere pilot-onderzoek van het meten van vigilantie met behulp van de SART uit, van alleen narcolepsie type 1 naar enkele andere slaapstoornissen, die ook gepaard gaan met overmatige slaperigheid overdag. In **hoofdstuk 4** wordt van diverse aspecten (leereffect, tijd van de dag, het doen van een dutje overdag en testinstructie) beschreven of ze SART-resultaten beïnvloeden. Hieronder worden de bevindingen van elk van deze hoofdstukken nog eens kort samengevat.

Uitdagingen bij het definiëren van vigilantie

In **hoofdstuk 2** worden verschillen beschreven tussen definities van vigilantie die circuleren in de wetenschappelijke literatuur. Daarbij bleek dat al deze definities verwezen naar een of meerdere van de nauw verwante begrippen alertheid, volgehouden aandacht

en 'arousal'. Deze begrippen bleken echter op hun beurt ook weer variabel te worden gedefinieerd. Wij stelden in dit hoofdstuk een nieuwe definitie van het concept vigilantie voor, namelijk als het vermogen om zich bewust te worden van potentiële veranderingen in de omgeving. Deze capaciteit heeft zowel een kwantitatief aspect, uitgedrukt in niveau van alertheid, als een temporeel aspect, namelijk een verandering in de tijd. Wanneer alertheid gepaard gaat met een bepaald focus, spreekt men van aandacht in plaats van alertheid. Volgehouden aandacht verwijst dan naar de capaciteit om een bepaald niveau van aandacht enige tijd vol te houden. Bij het definiëren van 'arousal' volgden wij de linguïstische definitie: een in korte tijd optredende, opwaartse verandering in het niveau van alertheid.

> Vigilantie is het vermogen om zich bewust te worden van potentiële veranderingen in de omgeving en omvat zowel een kwantitatieve als temporele dimensie.

De 'sustained attention to response task' (SART) toont aan dat er sprake is van gestoorde vigilantie bij een spectrum van slaapstoornissen die alle gepaard gaan met overmatige slaperigheid overdag.

Hoofdstuk 3 beschrijft een dwarsdoorsnede van de SART als maat voor het kwantificeren van vigilantie bij patiënten met verschillende oorzaken van overmatige slaperigheid overdag: 42 patiënten met narcolepsie type 1, 5 met narcolepsie type 2, 37 met idiopathische hypersomnie (IH) en 12 met het obstructiefslaapapneusyndroom (OSAS) worden beschreven. Elke patiënt onderging onder meer vijf SART-sessies, waarvan de resultaten werden gemiddeld. De belangrijkste bevinding was dat in al deze patiëntgroepen een hoge SART-foutscore werd gemeten. Deze mediane foutscore verschilde niet wezenlijk tussen de groepen en dat gold ook voor de andere onderzochte uitkomstmaten. Dit bevestigt de eerdere suggestie dat een hoge SART-foutscore niet specifiek is voor een bepaalde ziekte. Integendeel, een hoge SART-foutscore reflecteert waarschijnlijk een kernsymptoom van de beschreven slaapstoornissen: een gestoorde vigilantie. Dat vigilantie echt een ander fenomeen is dan slaperigheid, werd al eerder verondersteld, omdat er geen enkele correlatie werd gevonden tussen de SART-foutscore en de slaaplatentie in de multipele slaaplatentietest (MSLT; **zie ook hoofdstuk 1**).[1] Verder bleek de SART-foutscore in ons onderzoek omgekeerd evenredig aan de variabiliteit in reactietijd, en het totaal aantal onterecht-gedrukt-fouten ('commissiefouten') bleek omgekeerd evenredig aan de gemiddelde reactietijd. Deze bevindingen worden in de literatuur uitgelegd als de 'speed-accuracy trade-off', de wisselwerking tussen enerzijds snelheid en anderzijds nauwkeurigheid.[2-5] Tot slot stelden we vast dat de SART-foutscore in de eerste sessie in de ochtend hoger was dan in de andere sessies: dit zou het gevolg kunnen zijn van ofwel een leereffect ofwel een tijd-van-de-dag-effect.

Met de SART werd eenzelfde mate van vigilantiestoornis gevonden bij patiënten met andere oorzaken van overmatige slaperigheid overdag als eerder werd gevonden (en nu opnieuw is bevestigd) bij patiënten met narcolepsie type 1. De afwezigheid van een correlatie tussen SART en MSLT-maten geeft aan dat die twee tests echt verschillende fenomenen meten.

De invloed van leereffect, een dutje doen, tijd van de dag en testinstructie op de Sustained Attention to Response Task

In hoofdstuk 4 beschrijven wij ons onderzoek naar de potentiële invloed van een leereffect, een dutje doen, de tijd van de dag en de testinstructie op de resultaten van twee opeenvolgende SART-sessies bij 100 gezonde proefpersonen. Dit onderzoek werd opgezet om te achterhalen waarom de SART-foutscore bij de tweede sessie afneemt ten opzichte van de eerste sessie (zie **hoofdstuk 3**). Het eerste deel van dit onderzoek bestond uit een 2x2-ontwerp: er waren twee tijd-van-de-dag-groepen (ochtend en middag) en twee dutjesgroepen (wel of geen gelegenheid tot het doen van een maximaal 20 minuten durend dutje een uur voorafgaand aan de tweede SART-sessie) met in totaal 80 gezonde proefpersonen (20 per groep). In deel twee van het onderzoek werden 20 extra proefpersonen geworven, gekoppeld aan 20 personen uit het eerste deel. Zij verrichten de SART eveneens in de hierboven beschreven opzet, maar met een andere testinstructie voor beide SART-sessies. Uit het onderzoek bleek dat de afname van de SART-foutscore bij de tweede sessie ten opzichte van de eerste sessie alleen werd gezien als de proefpersonen de instructie kregen dat alleen accuratesse belangrijk was, maar reactietijd niet. Dat betekent dat deze daling van de foutscore waarschijnlijk het gevolg is van een leereffect, ook al bleken de deelnemers zich daar niet zelf bewust van (zij hadden hun eigen prestaties bij beide sessies moeten beoordelen). Dit leereffect verdwijnt na het doen van een volledige SART-oefensessie van 4 minuten en 19 seconden. Deze sterke invloed van testinstructie verklaart waarom sommige eerdere onderzoeken wel een leereffect vonden[1] en andere niet.[6]

Verder bleek de SART-foutscore significant hoger te zijn als proefpersonen de instructie kregen om evenveel aandacht te besteden aan accuratesse als aan snelheid, dan wanneer zij werden geïnstrueerd dat alleen accuratesse belangrijk was. In het laatste geval werd op groepsniveau een lagere SART-foutscore gevonden met minder variatie tussen verschillende proefpersonen in die groep. Dit leidt tot de aanbeveling om personen te instrueren om alleen aandacht aan accuratesse te besteden, als het voor het resultaat van de test belangrijk is dat zij een zo laag mogelijke SART-foutscore behalen. Dit is ook de enige instructie waarbij een duidelijk verschil zichtbaar is in prestatie tussen patiënten met narcolepsie type 1 en gezonde proefpersonen.

Tot slot toonde dit onderzoek dat de SART-resultaten noch worden beïnvloed

door tijd van de dag, noch door het krijgen van de gelegenheid om gedurende maximaal 20 minuten een dutje te doen een uur voor een volgende SART-sessie. Als er toch een subtiel, met dit onderzoek onderschat effect van tijd van de dag zou zijn, dan geldt dit alleen voor proefpersonen die de instructie kregen dat accuratesse en reactietijd even belangrijk waren. Het gebruik van de instructie dat alleen accuratesse belangrijk is, maakt het derhalve mogelijk om betrouwbaar SART-sessies met elkaar te vergelijken die zijn gemaakt op verschillende tijdstippen van de dag (binnen kantoortijden).

Als van belang is wat iemands' laagst mogelijke foutscore is, is het het best om de instructie te gebruiken dat accuratesse belangrijk is en reactietijd niet. Om het gevolg van een leereffect te minimaliseren, is het bij gebruik van deze instructie wel belangrijk dat proefpersonen een volledige oefensessie (dat wil zeggen, met een duur van 4 minuten en 19 seconden) hebben doorlopen. Zowel de tijd van de dag (binnen kantoortijden) als het krijgen van de gelegenheid tot het doen van een dutje (zoals gebeurt bij een MSLT-onderzoek) hebben geen invloed op SART-resultaten van gezonde vrijwilligers.

DEEL II - DE SUSTAINED ATTENTION TO RESPONSE TASK ALS BEHANDELEFFECTPARAMETER BIJ SLAAPSTOORNISSEN DIE GEPAARD GAAN MET OVERMATIGE SLAPERIGHEID OVERDAG

De hoofdstukken in dit gedeelte gaan over het valideren van de SART als behandeleffectparameter bij slaaponderzoek. Bij **hoofdstuk 5 en 6** gaat het daarbij om patiënten met narcolepsie. Deze hoofdstukken verschillen in twee opzichten van elkaar: de onderzochte behandeling (pitolisant *versus* natriumoxybaat) en de andere gebruikte uitkomstmaten naast de SART: Epworth Sleepiness Scale (ESS; zie ook **hoofdstuk 1**) en Maintenance of Wakefulness Test (MWT; zie ook **hoofdstuk 1**) *versus* Psychomotor Vigilance Test (PVT), Oxford Sleep Resistance Test (OSLER; zie ook **hoofdstuk 1**) en MWT. **Hoofdstuk 7** gaat over de SART als behandeleffectparameter van continuepositievedrukbeademing (CPAP) bij patiënten met OSAS.

Een vergelijking van behandeleffectparameters bij narcolepsie: de SART, ESS en MWT

In **hoofdstuk 5** wordt de validatie van de SART als behandeleffectparameter bij narcolepsie beschreven, waarbij de SART-uitkomsten worden vergeleken met die van de MWT en ESS. Dit onderzoek vond plaats als gerandomiseerd, gecontroleerd,

dubbelblind, parallel-groep 'multi-centre' onderzoek waarin de effecten van een achtweekse behandeling met het medicijn pitolisant, modafinil of een placebo werden vergeleken bij 95 patiënten met narcolepsie type 1 of 2.[7] Patiënten ondergingen de MWT, ESS en SART voorafgaand aan en na afloop van de achtweekse behandelperiode. Aanvullend werd verbetering van de ernst van overmatige slaperigheid overdag en van kataplexie bepaald met de Clinical Global Impression-schaal (CGI-C). De precisie en de betrouwbaarheid van de SART, MWT en ESS werden bepaald met de CGI-C-score als goudstandaard. De precisie van alle drie de tests was goed, waarbij voor de SART en MWT wel twee tot drie sessies op een dag nodig waren. De meest precieze SART-uitkomstmaat was de logaritmisch getransformeerde SART-commissiefoutscore (d.w.z. de optelsom van keren dat ten onrechte is gereageerd); bij deze uitkomstmaat werd al na twee sessies een precisie boven de 0,8 bereikt. De totale foutscore bereikte deze graad van precisie na drie sessies. Alle drie de uitkomstmaten waren goed in staat om patiënten die verbeterden op de behandeling (bepaald middels CGI-C-score) te onderscheiden van hen die niet verbeterden. De logaritmisch getransformeerde totale foutscore (r = 0.61) en de ESS (r = 0.54) correleerden het best met de uitkomst. Bij een vervolgens verrichte factoranalyse bleken veranderingen in scores op de MWT en ESS globaal hetzelfde aspect van de ziektelast van narcolepsie te vertegenwoordigen, terwijl veranderingen op de SART een compleet ander aspect vertegenwoordigden. De factoranalyse toont derhalve aan dat de klinische indruk van de mate van verbetering die de onderzoeker heeft, zowel gebaseerd is op het kunnen volhouden van de aandacht als het wakker kunnen blijven.

De SART is een valide uitkomstmaat om behandeleffecten te detecteren bij patiënten met narcolepsie. Dat geldt met name voor de totale foutscore en de commissiefoutscore. Een combinatie van SART- en ESS-resultaten waarborgt een veelomvattende evaluatie van het behandeleffect: waar de ESS een subjectieve inschatting van de mate van slaperigheid vertegenwoordigt, is de SART een objectieve maat. Beide tests hebben als voordeel dat ze weinig tijd en geld kosten om te verrichten en goed correleren met de algehele, klinische inschatting van de mate waarin in een patiënt al dan niet is verbeterd.

Verbetering van vigilantie na behandeling van patiënten met narcolepsie met het middel natriumoxybaat

In **hoofdstuk 6** beschrijven wij een observationeel onderzoek naar vigilantiemetingen bij 26 patiënten met narcolepsie type 1 en 15 gezonde proefpersonen, dat in twee centra werd uitgevoerd. Het doel van dit onderzoek was ten eerste om de haalbaarheid te onderzoeken van meerdaagse vigilantiemetingen op een draagbaar apparaat, waarbij

gebruik werd gemaakt van zowel de SART als de PVT; en ten tweede om het effect van behandeling met natriumoxybaat op vigilantie te onderzoeken bij patiënten met narcolepsie. Het onderzoek bestond uit twee achtereenvolgende, poliklinische dagen waarin deelnemers de MWT en OSLER ondergingen, gevolgd door een zevendaagse periode van vigilantiemetingen op een draagbaar apparaat in de thuissituatie. Bij narcolepsiepatiënten werd dit schema na drie maanden behandeling met een stabiele dosis natriumoxybaat nog eens herhaald. Het doen van herhaalde vigilantiemetingen in de thuissituatie op een draagbaar apparaat bleek bij zowel gezonde proefpersonen als patiënten met narcolepsie haalbaar. Bij alle tests (MWT, OSLER, SART en PVT) scoorden patiënten met narcolepsie slechter dan gezonde proefpersonen. Na behandeling met natriumoxybaat verbeterde de slaaplatentie op de MWT. Ook werd een kleine verbetering in volgehouden aandacht gezien bij zowel de OSLER als de SART, maar niet bij de PVT.

Zowel bij metingen van volgehouden aandacht op een draagbaar apparaat in de thuissituatie als bij poliklinische OSLER- en MWT-metingen presteerden patiënten met narcolepsie slechter dan gezonde proefpersonen. Na behandeling met natriumoxybaat werd bij patiënten met narcolepsie een verbetering van volgehouden aandacht en van het vermogen om slaap te weerstaan gezien. De SART en OSLER kunnen beide worden gebruikt als een minder tijd en mankracht kostend alternatief voor het meten van behandeleffect bij patiënten met narcolepsie dan de PVT en MWT.

Voorspellers van door patiënten gescoord behandeleffect van continue-positievedrukbeademing bij het obstructiefslaapapneusyndroom

Hoofdstuk 7 bevat de resultaten van een prospectief, observationeel behandeleffectonderzoek naar CPAP bij matig tot ernstig OSAS. Doel van het onderzoek was om te onderzoeken welke parameters het voorspellendst zijn voor het door patiënt ervaren, dus subjectieve effect van behandeling. De parameters die wij hebben onderzocht, zijn ademhalingsmaten (zoals de apneu-hypopneu-index – AHI), vragenlijsten over slaperigheid (ESS en Stanford Sleepiness Scale), de SART en visueel-analoge schalen over diverse aspecten van welbevinden. De goudstandaard, verbetering na behandeling, was de door patiënten gescoorde schaal Clinical Global Impression of Change (PCGI-C). Aan het onderzoek namen dertig patiënten met matig of ernstig OSAS (AHI > 15) deel. Gegevens van twee onderzoeksdagen vóór start van CPAP en één onderzoeksdag acht weken na start van CPAP werden verzameld. Na CPAP-behandeling toonden alle ademhalingsparameters een duidelijke verbetering, net als de ESS-score. Er werden slechts marginale verbeteringen gezien in SART-score en op enkele visueel-analoge

schalen. Tachtig procent van de deelnemers scoorde zichzelf als verbeterd op de PCGI-C. Deze PCGI-C-score correleerde goed met de ademhalingsparameters en de ESS: patiënten die over zichzelf veel tot erg veel verbetering rapporteerden, hadden ook de grootste afname van ademstops, desaturaties en van slaperigheid. Er werd geen correlatie gevonden tussen de PCGI-C-score en de SART-foutscore.

De meerderheid van de onderzochte patiënten met OSAS ervoer een duidelijke verbetering op hun klachten na een achtweekse CPAP-behandeling. De beste voorspellers van deze verbetering waren de ademhalingsparameters en overmatige slaperigheid overdag, gescoord op de ESS. De SART-foutscore verbeterde slechts minimaal en was niet gecorreleerd aan de PCGI-C-score.

TOEKOMSTPERSPECTIEVEN

De in dit proefschrift beschreven onderzoeken dragen bij tot de validatie van de SART als behandeleffectparameter bij slaapstoornissen die gepaard gaan met overmatige slaperigheid overdag. De SART blijkt een robuuste test te zijn, wat betreft mogelijk storende invloeden zoals tijdstip van de dag of het doen van een dutje vooraf. Wij hebben vastgesteld welke testinstructie het best werkt om patiënten te onderscheiden van gezonde proefpersonen. Verder bleek de SART goed in staat om het behandeleffect te meten bij een tweetal medicamenteuze behandelingen bij narcolepsie type 1, en ook bleek deze SART-verbetering sterk gecorreleerd aan de door deze patiënten ervaren mate van klinische verbetering. Daarbij waren de SART-resultaten complementair aan slaperigheidsmaten.

Ondanks bovenstaande bevindingen is aanvullend onderzoek nodig om het gebruik van de SART te valideren bij andere slaapstoornissen. Ons OSAS-onderzoek (**hoofdstuk 7**) toont immers aan dat de validiteit van de SART kan verschillen tussen slaapstoornissen. Het verdient de voorkeur dat ons onderzoek in andere slaapcentra wordt herhaald, zo mogelijk gebruik makend van grotere patiëntgroepen. Dit is met name van belang voor het observationele onderzoek naar de SART-uitkomstmaten bij diverse slaapstoornissen in de situatie vóór start van enige behandeling (**hoofdstuk 3**), aangezien de groepen van verschillende slaapstoornissen uit zeer wisselende aantallen patiënten bestonden. Een dergelijk vervolgonderzoek omvat dan bij voorkeur een gematchte controlegroep van gezonde vrijwilligers.

Prioriteit 1: herhalen van het in hoofdstuk 4 beschreven onderzoek bij patiënten met slaapstoornissen

In **hoofdstuk 4** onderzochten we diverse aspecten die SART-resultaten zouden kunnen beïnvloeden, zoals tijd van de dag of het doen van een dutje. De potentiële invloed van deze aspecten zou ook bij verschillende slaapstoornissen anders dan narcolepsie moeten worden onderzocht. De reden hiervoor is dat zowel de gevonden associatie tussen het effect van testinstructie en de SART-foutscore als de associatie tussen de testinstructie en een leereffect niet noodzakelijkerwijs even sterk hoeven te zijn bij patiëntengroepen als bij gezonde vrijwilligers. Dit zou zelfs kunnen verschillen tussen diverse patiëntgroepen. Van slaapstoornissen die gepaard gaan met overmatige slaperigheid overdag is narcolepsie type 1 de enige aandoening waarvoor onderzoeksresultaten met beide testinstructies zijn beschreven. De SART-foutscore van deze patiëntengroep verschilde niet tussen beide testinstructies.[1,8] Het lijkt aannemelijk dat patiënten met narcolepsie al op hun maximale capaciteit functioneren als ze geïnstrueerd worden om zowel aandacht te geven aan accuratesse als snelheid: de lange gemiddelde reactietijd suggereert dat ze in de praktijk toch niet evenveel aandacht gaven aan beide aspecten; het is mogelijk dat ze hun reactiesnelheid al hebben moeten opofferen om nog enige accuratesse te behouden.[8] Het geven van de instructie dat reactiesnelheid niet belangrijk is, komt de accuratesse dan weinig meer ten goede. De lage accuratesse[1], oftewel het onvermogen om de aandacht ruim vier minuten vol te houden, reflecteert waarschijnlijk de problemen die narcolepsiepatiënten in het dagelijks leven ervaren als ze bijvoorbeeld een conversatie proberen te volgen of een boek te lezen. Het is overigens wel aan te bevelen om de invloed van beide testinstructies nog eens te onderzoeken in één groep patiënten met narcolepsie.

Dergelijke vergelijkingen zijn ook voor andere slaapstoornissen hard nodig. In **hoofdstuk 3** werd immers beschreven dat de SART-prestaties van patiënten met IH en OSAS niet wezenlijk verschilden van patiënten met narcolepsie. Bij dit onderzoek werden de deelnemers echter geïnstrueerd om zowel aandacht te besteden aan accuratesse als aan snelheid en er was geen directe controlegroep. De SART-foutscore werd visueel vergeleken met een groep historische controles. Achteraf bleek dat deze groep de instructie heeft gekregen dat alleen accuratesse belangrijk is, dus deze vergelijking is problematisch. Bovendien werd in **hoofdstuk 4** gezien dat gezonde vrijwilligers als gevolg van de 'speed-accuracy trade-off' een SART-foutscore hadden die bij benadering vergelijkbaar was met die van patiënten met narcolepsie, wanneer de gezonde vrijwilligers de instructie hadden gekregen om evenveel aandacht te besteden aan accuratesse en snelheid. Dit betekent dat er eigenlijk geen conclusie kan worden getrokken over de (ab-) normaliteit van de SART-foutscore bij patiënten met IH en OSAS in **hoofdstuk 3**. Het is natuurlijk mogelijk dat ook deze patiënten al op hun maximale capaciteit functioneerden bij het maken van de SART. Een aanwijzing hiervoor bij de patiënten met narcolepsie was de desalniettemin lange reactietijd die werd gezien als zij de instructie kregen dat zowel

accuratesse als snelheid belangrijk was. De reactietijden van gezonde vrijwilligers zijn immers significant korter bij dezelfde testinstructie. De reactietijden van patiënten met IH en OSAS zoals in hoofdstuk 3 beschreven, liggen echter tussen die van gezonde controles en patiënten met narcolepsie in. Er blijven daardoor twee interpretaties mogelijk over de abnormaliteit van hun foutscores. Het opnieuw onderzoeken van de SART-foutscore bij deze patiënten, maar nu met de instructie dat alleen accuratesse belangrijk is, is derhalve essentieel om vast te kunnen stellen of er ook bij deze patiëntgroepen sprake is van een gestoorde vigilantie.

De bevindingen die in **hoofstuk 4** worden beschreven, werden chronologisch echter pas verkregen nadat de onderzoeken beschreven in **hoofdstuk 5, 6 en 7** reeds van start waren gegaan. Bij deze laatste drie onderzoeken werden de deelnemers geïnstrueerd om evenveel aandacht te besteden aan accuratesse als aan snelheid. De onderzoeken betreffen echter vergelijkingen van prestaties op verschillende tijdpunten bij dezelfde deelnemers, in plaats van vergelijkingen tussen verschillende personen/groepen. Omdat patiënten met narcolepsie ongeacht de testinstructie slecht presteren op de SART, is het niet waarschijnlijk dat de verschillen in prestatie tussen het moment vóór start van behandeling en het moment tijdens behandeling, zoals beschreven in **hoofdstuk 5 en 6**, nog groter hadden kunnen zijn als de patiënten de instructie hadden gekregen dat alleen accuratesse belangrijk was. Deze uitleg gaat waarschijnlijk niet op voor **hoofdstuk 7**. In dat hoofdstuk werd geen duidelijke verbetering gezien in de SART-foutscore van patiënten met OSAS voor en na CPAP-behandeling. Omdat er echter geen directe baseline-vergelijking was tussen patiënten met OSAS en gezonde vrijwilligers, blijft het onduidelijk óf hun prestatie vóór start van de behandeling überhaupt abnormaal was. De relatief hoge foutscore kan immers het gevolg zijn van de gevolgde strategie, omdat ze werden geïnstrueerd dat zowel accuratesse als snelheid belangrijk waren. Een directe vergelijking met gezonde controlepersonen blijft derhalve noodzakelijk, bij voorkeur met de instructie dat alleen accuratesse belangrijk is, omdat deze instructie de prestaties optimaliseert. Als bij een dergelijk onderzoek verschillen worden gevonden tussen patiënten met OSAS en de gezonde controles, zou ook opnieuw moeten worden onderzocht of er wel een CPAP-behandeleffect zichtbaar is bij gebruik van de instructie dat alleen accuratesse belangrijk is.

Prioriteit 2: SART-uitkomstmaten gericht op accuratesse in toekomstig onderzoek

De in dit onderzoek gepresenteerde SART-parameters worden onderverdeeld in enerzijds maten die accuratesse beschrijven (de totale foutscore of de specifieke commissie- en omissiefoutscores), en anderzijds reactietijdmaten (bijvoorbeeld de gemiddelde reactietijd of de variabiliteit in reactietijd). In **hoofdstuk 5** werden slechts minimale verschillen beschreven in precisie en betrouwbaarheid tussen de verschillende maten

van accuratesse, wat erop wijst dat ze ongeveer hetzelfde fenomeen vertegenwoordigen. De hoogste betrouwbaarheid werd gevonden voor de totale foutscore. Statistisch bekeken was de commissiefoutscore, d.w.z. het aantal keer dat er onterecht op een toets is gedrukt, echter preciezer: al na twee SART-sessies werd een voldoende precies resultaat gezien, terwijl voor de SART-foutscore drie sessies nodig zijn om deze precisie te behalen. De omissiefoutscore, d.w.z. het aantal keer dat onterecht níet is gedrukt op een toets, is op het gebied van verdeling en precisie een minder goede uitkomstmaat. Dat de totale foutscore het echter wel goed doet, betekent dat de omissiefouten toch belangrijk zijn, omdat de totale foutscore immers uit de som van het aantal commissie- en omissiefouten bestaat. Het relatieve belang van omissiefouten, commissiefouten en de totale foutscore kan verschillen tussen stoornissen.[9] Om deze reden is in dit proefschrift steeds de totale foutscore als primaire uitkomstmaat gebruikt. Het blijft echter raadzaam om de commissiefouten en omissiefouten apart bij te houden als wordt gepoogd de bevindingen in dit proefschrift te repliceren, met name als verschillende groepen patiënten worden bestudeerd.

De SART in relatie tot andere maten van volgehouden aandacht

Tijdens het maken van de SART worden reactietijden opgeslagen, net als hij andere tests van volgehouden aandacht. Wij hebben ervoor gekozen het aantal beschreven reactietijdparameters te beperken tot twee. De belangrijkste waarde van de SART ligt immers in het vaststellen van iemands capaciteit om correct een verandering op te merken in de aangeboden stimuli, waarop een alternatieve respons nodig is. Deze capaciteit komt veel beter tot uitdrukking in de SART-foutscore dan in reactietijdmaten alleen. Als de instructie wordt gebruikt dat alleen accuratesse belangrijk is, wordt de invloed van reactietijd geminimaliseerd. Een praktisch nadeel van een focus op reactietijdmaten is dat het nauwkeurig meten hiervan speciale apparatuur vereist: variabele onnauwkeurigheden die worden veroorzaakt door technische factoren zoals schermverversingssnelheid en het verwerken van een klik met een computermuis moeten worden geminimaliseerd, omdat deze factoren de reactietijdmaten op een niet van tevoren te voorspellen wijze kunnen beïnvloeden. Daarentegen vereist het meten van foutscores alleen standaard apparatuur zoals een personal computer of een draagbaar digitaal apparaat. Als in een bepaalde situatie eenvoudige reactietijdmetingen het meest van belang zijn, dan voldoet de PVT (wel bij voorkeur met gebruikmaking van speciale benodigdheden om de hierboven beschreven redenen). Deze test is echter voornamelijk gebruikt en gevalideerd bij slaapdeprivatieonderzoek.[10-12] Onderzoeken bij slaapstoornissen zijn schaars. Recent is beschreven dat PVT-resultaten van onbehandelde patiënten met narcolepsie, IH en het insufficiënteslaapsyndroom slechter zijn dan van gezonde controles.[13] Daarnaast is bekend van de PVT dat deze gevoelig is om behandeleffect te meten bij OSAS,[14] maar in het in **hoofdstuk 6** beschreven onderzoek wordt geen rol voor de PVT gevonden in het

meten van behandeleffect bij patiënten met narcolepsie. De SART bleek hier een betere maat voor.

In hetzelfde onderzoek beschreven wij de waarde van OSLER als een maat van volgehouden aandacht, door niet te kijken naar slaaplatentie, maar een gedragsmatige uitkomstparameter te berekenen die het niveau van vigilantie meet voorafgaand aan het in slaap vallen. De PVT verschilt van de OSLER in het stimulusaanbod: bij de PVT varieert deze van 2-10 seconden, terwijl de stimuli bij de OSLER continu worden aangeboden met een vast en korter interval. Het gebruik van de OSLER als maat voor volgehouden aandacht, weergegeven in de parameter aantal omissies per minuut, was sensitiever voor het detecteren van behandeleffect van natriumoxybaat dan de PVT. Zowel de SART, ambulant gemeten terwijl mensen zich 'in het dagelijks leven' bevonden, als de OSLER, gemeten in de gecontroleerde omstandigheden van een slaaponderzoekscentrum, bleken goed in staat dit behandeleffect bij patiënten met narcolepsie te meten. Een directe vergelijking van beide uitkomstmaten in de setting van een slaaponderzoekscentrum zou interessant zijn om de waarde van de OSLER als test van volgehouden aandacht te bepalen.

Beperkingen van de SART

Een belangrijke beperking van de SART is dat een slechte score niet per definitie het gevolg hoeft te zijn van een gestoorde vigilantie. Zowel cognitieve stoornissen als motorische problemen kunnen het SART-resultaat negatief beïnvloeden, respectievelijk door gestoorde aandachtsfuncties of een gestoord vermogen om (snel) te reageren. Bijgevolg ligt de waarde van de SART niet in het zeer selectief detecteren van een gestoorde vigilantie, maar in kwantificatie van de beperkingen en het vervolgen hiervan in de tijd of in verschillende omstandigheden zoals voor en na behandeling. De SART kan worden toegepast om vigilantie te kwantificeren zoals de MWT wordt toegepast om slaperigheid te kwantificeren. Het ligt voor de hand dat SART-resultaten worden beïnvloed door motivationele of omgevingsfactoren. Als degene die de SART moet maken moedwillig niet zijn of haar best doet, zal de uitkomst waarschijnlijk gestoord zijn. Met andere woorden, een slecht resultaat op de SART kan het gevolg zijn van fraude. Het omgekeerde is echter vrijwel onmogelijk: de SART abnormaal goed doen, terwijl je die capaciteit eigenlijk niet hebt. De SART is dus bestand tegen fraude als het gaat om omstandigheden waarin het van belang is om zo goed mogelijk te presteren, zoals het geval is bij rijbewijskeuringen. In deze context is de SART robuuster dan de MWT, waarbij patiënten trucjes kunnen gebruiken om te proberen wakker te blijven. De SART is daarom een geschikt onderzoek voor situaties waarin patiënten er een belang bij hebben om goed te presteren.

De plaats van de SART in toekomstig onderzoek en toekomstige patiëntenzorg

Hopelijk draagt dit proefschrift bij tot meer besef van het belang van vigilantie bij slaapstoornissen die gekenmerkt worden door overmatige slaperigheid overdag: een gestoorde vigilantie heeft immers een negatieve invloed op het dagelijks functioneren en is van belang bij diverse afwegingen op het gebied van veiligheid, zoals het besturen van voertuigen of het bedienen van gevaarlijke apparatuur. Naast het belang van meten van een gestoorde vigilantie voor de individuele patiënt, zijn er praktische argumenten om meer aandacht te geven aan het meten van vigilantie in plaats van slaperigheid. Voor huidige, objectieve slaperigheidstests zoals de MSLT en MWT is een slaaponderzoeksfaciliteit vereist. Veel somnologen hebben ervaren dat sommige patiënten vinden dat hun functioneren substantieel is verbeterd door het gebruik van stimulantia, terwijl hun MWT nauwelijks verbeterde. De weinig stimulerende omgeving van een slaaponderzoeksfaciliteit kan een beperking opleveren voor de externe validiteit van het onderzoek; het is de vraag of de test een voorspellende waarde heeft voor het vaststellen van problemen die zich in 'het echte leven' voordoen.

Dit proefschrift heeft een uitgebreide kennis opgeleverd van essentiële, praktische voorwaarden voor het gebruik van de SART in toekomstig onderzoek en toekomstige patiëntenzorg. De volgende aanbevelingen vatten onze bevindingen samen:

1. Voor het diagnosticeren van een gestoorde vigilantie adviseren wij vijf SART-sessies te laten verrichten voorafgaand aan elke van vijf MSLT-sessies.
2. Voor andere doeleinden adviseren wij om ten minste twee SART-sessies met 1-1,5 uur tussentijd te laten verrichten, waarbij de eerste sessie wordt voorafgegaan door een volledige oefensessie. Binnen kantoortijden beïnvloedt het tijdstip van de dag de uitkomst van de SART niet.
3. De instructie aan degene die de SART ondergaat, moet ongeveer als volgt luiden: "U krijgt achtereenvolgens 225 keer een cijfer van 1 tot 9 te zien in een willekeurige volgorde. U moet na ieder cijfer zo snel mogelijk op de toets drukken, behalve als u het cijfer 3 ziet. U moet hebben gedrukt voordat het volgende cijfer in beeld verschijnt, maar weest u zich ervan bewust dat het belangrijker is om correct te drukken dan om dit heel snel te doen."
4. De volgende gegevens moeten worden opgeslagen: het aantal keer dat toch gedrukt is terwijl het cijfer 3 werd getoond (commissiefouten), het aantal keer dat niet werd gedrukt terwijl een ander cijfer dan 3 werd gepresenteerd (omissiefouten), en bij voorkeur ook de reactietijd van iedere juiste respons.
5. De aanbevolen primaire uitkomstmaat is de SART-foutscore, die bestaat uit de som van het aantal commissie- en omissiefouten.

Aanbeveling 2 en 3 vereisen nog wel aanvullend onderzoek bij patiënten met slaapstoornissen, waaronder narcolepsie, IH en OSAS. Als dit heeft plaatsgevonden, zouden vigilantiemetingen middels de SART de hoeksteen moeten vormen van toekomstige behandeleffectstudies waarbij nieuwe medicijnen worden getest die het dagelijks functioneren moeten verbeteren. Omdat de SART een objectieve, functionele uitkomstmaat is die goed correleert met ervaren verbetering, zou de SART ten minste als even belangrijk moeten worden gezien als uitkomstmaten van slaperigheid.

Het meten van vigilantie met de SART dient ruim toegepast te worden bij behandeleffectonderzoeken bij slaapstoornissen.

Naast een rol als behandeleffectparameter in medicatiestudies zou de SART wellicht een rol kunnen gaan spelen bij rijbewijskeuringen. De huidige Nederlandse wetgeving op het gebied hiervan bij patiënten met slaapstoornissen stelt dat de rijgeschiktheid momenteel wordt gebaseerd op het oordeel van een onafhankelijk, deskundig arts; dit oordeel moet afhankelijk van de slaapstoornis worden ondersteund door aanvullend onderzoek. Patiënten met OSAS moeten bijvoorbeeld een AHI onder de 15 hebben. Functionele uitkomstmaten zijn echter niet vereist. Patiënten met narcolepsie en IH moeten een ESS-score onder de 11 hebben. Met andere woorden, zij moeten voldoen aan de vereisten voor een niet afwijkend slaaponderzoek of niet afwijkende slaapvragenlijst; kwantificatie van vigilantie is niet vereist. De wél vereiste uitkomstmaat heeft een aantal nadelen. Hoewel de ESS een goedkoop vragenlijstonderzoek is, is deze subjectief van aard en daardoor kwetsbaar voor het moedwillig onderschatten van de eigen beperkingen, d.w.z. een gunstiger resultaat voor te stellen dan daadwerkelijk het geval zou zijn. De ESS is daarom niet geschikt als enige aanvullend onderzoek. Uit bovenstaande volgt dat het nuttig zou zijn als er nieuwe, objectieve parameters zouden komen die steun kunnen bieden aan het oordeel van de arts. De MWT is daarvoor vanwege een aantal nadelen niet geschikt: het kostenaspect speelt daarin een rol, maar de validiteit in de echte wereld is ook niet gegarandeerd. Er is geen bewijs dat de score op de MWT een betrouwbare voorspeller is van het risico op ongelukken, al is wel bekend dat het resultaat bij patiënten met narcolepsie en OSAS correleert met de prestaties in een rijsimulator.[15,16] Een cruciale volgende stap in het validatieonderzoek van de SART als klinische test in de slaapgeneeskunde zou daarom het volgende moeten zijn: het onderzoeken of de SART, die zowel goedkoop als objectief is, een betrouwbare voorspeller kan zijn voor het risico op ongelukken in de echte wereld. Om deze vraag te beantwoorden, zouden SART-resultaten moeten worden vergeleken met een rijtest op de weg, afgenomen door een rij-instructeur of -examinator. Rijsimulatorprestaties kunnen ook helpen, maar moeten

niet als enige worden gebruikt om de daadwerkelijke rijvaardigheden te toetsen.

De SART zou een geschikte methode kunnen zijn om de rijvaardigheid van patiënten met slaapstoornissen te onderzoeken. De test is immers bewezen valide in het vaststellen van behandeleffecten en deelt enkele belangrijke nadelen van de huidige gebruikte tests niet: de SART is goedkoop, eenvoudig om af te nemen, en robuust tegen pogingen om beter te presteren dan in het dagelijks leven mogelijk zou zijn.

REFERENTIES

1. Fronczek R, Middelkoop HA, van Dijk JG, Lammers GJ. Focusing on vigilance instead of sleepiness in the assessment of narcolepsy: high sensitivity of the Sustained Attention to Response Task (SART). *Sleep.* 2006;29(2):187-191.
2. Robertson IH, Manly T, Andrade J, Baddeley BT, Yiend J. 'Oops!': performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. *Neuropsychologia.* 1997;35(6):747-758.
3. Manly T, Robertson IH, Galloway M, Hawkins K. The absent mind: further investigations of sustained attention to response. *Neuropsychologia.* 1999;37(6):661-670.
4. Helton WS, Weil L, Middlemiss A, Sawers A. Global interference and spatial uncertainty in the Sustained Attention to Response Task (SART). *Conscious Cogn.* 2010;19(1):77-85.
5. Shalgi S, O'Connell RG, Deouell LY, Robertson IH. Absent minded but accurate: delaying responses increases accuracy but decreases error awareness. *Exp Brain Res.* 2007;182(1):119-124.
6. McAvinue L, O'Keeffe F, McMackin D, Robertson IH. Impaired sustained attention and error awareness in traumatic brain injury: implications for insight. *Neuropsychological Rehabilitation.* 2005;15(5):569-587.
7. Dauvilliers Y, Bassetti C, Lammers GJ, et al. Pitolisant versus placebo or modafinil in patients with narcolepsy: a double-blind, randomised trial. *Lancet neurology.* 2013;12(11):1068-1075.
8. Van Schie MK, Thijs RD, Fronczek R, Middelkoop HA, Lammers GJ, Van Dijk JG. Sustained attention to response task (SART) shows impaired vigilance in a spectrum of disorders of excessive daytime sleepiness. *Journal of Sleep Research.* 2012;21(4):390-395.
9. Hart EP, Dumas EM, Reijntjes RH, et al. Deficient sustained attention to response task and P300 characteristics in early Huntington's disease. *Journal of neurology.* 2012;259(6):1191-1198.
10. Dinges DF, Kribbs NB. Performing while sleepy: effects of experimentally-induced sleepiness In: Monk TH, ed. *Sleep, sleepiness & Performance.* Chister, UK: Wiley; 1991:97-128.
11. Dinges DF, Powell JW. Sleepiness is more than lapsing. *Sleep Res.* 1988;17(84).
12. Dorrian J, Rogers NL, Dinges DF. Psychomotor vigilance performance: a neurocognitive assay sensitive to sleep loss. In: Kushida C, ed. *Sleep deprivation: Clinial Issues, Pharmacology and Sleep Loss Effects.* New York: Marcel Dekker; 2005:39-70.
13. Thomann J, Baumann CR, Landolt HP, Werth E. Psychomotor vigilance task demonstrates impaired vigilance in disorders with excessive daytime sleepiness. *J Clin Sleep Med.* 2014;10(9):1019-1024.

14. Dinges DF, Weaver TE. Effects of modafinil on sustained attention performance and quality of life in OSA patients with residual sleepiness while being treated with nCPAP. *Sleep Med* 2003;5(5):393-402.
15. Philip P, Sagaspe P, Taillard J, et al. Maintenance of Wakefulness Test, obstructive sleep apnea syndrome, and driving risk. *Ann Neurol.* 2008;64(4):410-416.
16. Philip P, Chaufton C, Taillard J, et al. Maintenance of Wakefulness Test scores and driving performance in sleep disorder patients and controls. *Int J Psychophysiol.* 2013;89(2):195-202.

B. DANKWOORD

Velen hebben op enigerlei wijze een belangrijke bijdrage geleverd aan dit proefschrift. Een aantal van hen wil ik hier noemen. In de eerste plaats wil ik alle patiënten en gezonde vrijwilligers bedanken die hebben deelgenomen aan de onderzoeken. Zonder hen had dit proefschrift niet tot stand kunnen komen.

Mijn bijzondere dank gaat uit naar mijn promotores Gert van Dijk en Gert Jan Lammers en mijn co-promotor Rolf Fronczek. Dankzij jullie begeleiding is dit langlopende promotietraject uiteindelijk toch tot een goed einde gekomen. Gert, dank voor je vele wijze woorden; of het nu ging om de opzet van een onderzoek, een discussie over een moeilijke analyse, taalkundige wijzigingen waardoor de artikelen naar een hoger niveau werden getild, of de omgang met de bij het onderzoeksveld horende frustraties: altijd waren ze raak. Gert Jan, zonder het in mij gestelde vertrouwen én je vermogen om jonge mensen te enthousiasmeren voor de wereld van de slaapgeneeskunde, was dit proefschrift niet geschreven. Rolf, als begeleider van het laatste uur heb jij me door moeilijke perioden gecoacht; zonder je luisterend oor had ik vermoedelijk meer moeite gehad om dit traject tot een goed einde te brengen.

Geen proefschrift zonder data: graag bedank ik Robert Reijntjes voor de beschikbaarheid van de SART en SART-data. Verder bedank ik de verschillende medewerkers van het KNF-, poli- en stafsecretariaat die van belang zijn geweest voor de voortgang van het onderzoek. Daarnaast wil ik de KNF-laboranten die hebben bijgedragen aan het verkrijgen van de in dit proefschrift beschreven SART- en slaaponderzoeken hartelijk danken: Paul van Someren, Jannie van Vliet – de Regt, Marion Stijl – Pek, Fabian Kerkhof, Geertrui van Beukering - Louwes, Wesseline Huyser en Judith Witteman. Voorts speelden Eva Alblas en Leon Walschot ieder een belangrijke rol bij de dataverzameling van een specifiek hoofdstuk.

Een bijzondere plek in dit kader is voor Pauline Buijserd-Amesz: Pauline, zonder jouw hulp bij het werven van deelnemers aan diverse onderzoeken was die fase niet zo vlot verlopen. Maar bovenal: zonder de talloze gezellige momenten die samenhingen met jouw aanwezigheid en vele cappuccino's, was dit promotietraject niet geweest zoals het nu was.

Graag bedank ik de overige co-auteurs voor hun scherpe blik en wil ik Ahmad Aziz en Roland Thijs specifiek noemen: jullie hebben mij op respectievelijk statistiek- en schrijfgebied gegidst gedurende mijn eerste voorzichtige stappen in wetenschapsland.

Ik kijk met een warm gevoel terug op alle belevenissen met de onderzoeksgroep, door Claire ook wel ‘team Lammers’ genoemd. Gert Jan, Claire, Astrid, Rolf, Wisse, Gert, Roland, Pauline, Mink en Jari: van jullie allen draag ik voor altijd bijzondere herinneringen met me mee. Terugkijkend op een promotietraject zijn dit soort herinneringen misschien wel de eerste die je weer te binnen schieten.

Vrienden, collega’s, familie en schoonfamilie: bedankt voor jullie aanhoudende en motiverende interesse in de voortgang van dit onderzoekstraject. Speciale dank gaat uiteraard uit naar mijn ouders, voor het warme thuis waarin ik me altijd veilig heb gevoeld en kon groeien en bloeien. Zonder deze basis had ik hier niet gestaan. Tot slot Diederik: dank dat ik al zo lang mijn leven met je mag delen, en misschien ook wel dank voor de vele plagende opmerkingen of dit proefschrift er ooit nog zou komen. Sara, dank voor de vreugde die je dagelijks in mijn leven brengt.

C. CURRICULUM VITAE

Mojca van Schie was born in Haarlem on 17 January 1989 as a daughter of a Dutch father and Slovenian mother. Her secondary education was at Atheneum College Hageveld in Heemstede, where she graduated cum laude in 2006 with Latin as optional subject. She proceeded directly to the University of Leiden to study medicine. She participated in an Erasmus exchange program to Karolinska Instituted in Stockholm in 2007. She finished her "doctoraal" cum laude at the University of Leiden in 2010. During the first years of the study, she worked as a student assistent in microscopy and volunteered in several committees of the local and national departments of the International Federation on Medical Students' Associations (IFMSA), including a position as vice-president on external affairs for IFMSA-Leiden from 2009-2010. She started her PhD training at the department of Neurology of the Leiden University Medical Centre (LUMC) as part of LUMC's excellent student program. Part of the PhD training was performed at the Brain and Behaviour Laboratory of the University of Geneva under supervision of prof. S. Schwartz (2011-2012). She proceeded to the clinical rotations in 2013 and obtained her medical degree cum laude in 2014. She worked for a year as a neurology intern at Medisch Centrum Haaglanden-Bronovo in The Hague before she started her neurology training at the LUMC in 2016. Mojca lives in Haarlem with her husband Diederik Brandwagt and their daughter Sara (2019).

D. LIST OF PUBLICATIONS

M.K.M. van Schie, R.D. Thijs, R. Fronczek, H.A.M. Middelkoop, G.J. Lammers, J.G. van Dijk. Sustained attention to response task (SART) shows impaired vigilance in a spectrum of disorders of excessive daytime sleepiness. J Sleep Res 2012 Aug;21(4):390-5.

M. van Dijk, E. Donga, M. Van Schie, G.J. Lammers, E. van Zwet, E. Corssmit, J. Romijn, J.G. Van Dijk. Impaired sustained attention in adult patients with type 1 diabetes is related to diabetes per se. Diabetes Metab Res Review 2014 Feb;30(2):132-9.

M.K.M. van Schie, E.E. Alblas, R.D. Thijs, R. Fronczek, G.J. Lammers, J.G. van Dijk. The influences of task repetition, napping, time of day, and instruction on the Sustained Attention to Response Task. J of Clin and Exper Neuropsychology 2014;36(10):1055-65.

M. van Schie*, A. v/d Heide*, G.J. Lammers et al. Comparing treatment effect measurements in narcolepsy: the SART, ESS and MWT. Sleep 2015 Jul 1;38(7):1051-8.

M. van Schie, E. Werth, G.J. Lammers, S.Overeem, C. Baumann, R. Fronczek. Improved vigilance after Sodium Oxybate treatment in narcolepsy. J Sleep Res. 2016

M. van Schie, G.J. Lammers. Sustained Attention to Response Task (SART) - a measure of daytime vigilance. In: H. Schulz, P. Geisler, A. Rodenbeck. Kompendium Schlafmedizin. Landsberg: Ecomed Medizin, 2015.

P. van Vliet, A.E. Berden, M.K.M. van Schie, J.A. Bakker, C. Heringhaus, I.F.M. de Coo, M. Langeveld, M.A. Schroijen, M.S. Arbous. Peripheral neuropathy, episodic rhabdomyolysis, and hypoparathyroidism in a patient with mitochondrial trifunctional protein deficiency. JIMD reports 2017.

M.K.M. van Schie, R. Fronczek. Fatale familiaire insomnie. Nervus 2019.

M.K.M. van Schie, G.J. Lammers, R. Fronczek, H.A.M. Middelkoop, J.G. van Dijk. Vigilance: discussion of related concepts and proposal for a definition. Sleep Medicine 2021; https://doi.org/10.1016/j.sleep.2021.04.038.

V. Sterpenich, M.K.M. van Schie, M. Catsiyannis, A. Ramyead, S. Perrig, H.D. Yang, D. Van De Ville, S. Schwartz. Reward biases spontaneous neural reactivation during sleep. Nature Communications 2021;12:4162.

E. ABBREVIATIONS

AASM	American Academy of Sleep Medicine
ADHD	attention deficit hyperactivity disorder
AHI	apnea-hypopnea index
AI	apnea index
ANCOVA	Analysis of Covariance
CGI-C	Clinical Global Impression of Change
CPAP	continuous positive airway pressure
EDS	excessive daytime sleepiness
EEG	electro-encephalography
ES	Cohen's effect size
ESS	Epworth Sleepiness Scale
ICC	Intra-class correlation coefficient
ICSD	International Classification of Sleep Disorders
IH	Idiopathic Hypersomnia
I_M	modified instruction
I_0	original instruction
KS	Kolmogorov-Smirnov
LMM	linear mixed effect model
LMMs	linear mixed effect models
MSLT	Multiple Sleep Latency Test
MWT	Maintenance of Wakefulness Test
N+/-	group with/without napping opportunity
ODI-3%	oxygen-desaturation index based on desaturations ≥ 3% of baseline
ODI-4%	oxygen-desaturation index based on desaturations ≥ 4% of baseline
OMIS	omissions
OMIS/MIN	omissions per minute test duration
OSAS	obstructive sleep apnea syndrome
OSLER	Oxford Sleep Resistance test
PCGI-C	patient-rated clinical global impression of change
PDA	personal digital assistant
PVT	Psychomotor Vigilance Test
RT	reaction time
RTs	reaction times
S+/-	participants who did/did not fall asleep during the nap occasion
SART	Sustained Attention to Response Task
SSS	Stanford Sleepiness Scale
SXB	sodium oxybate
VAS	visual-analogue scales