Pupil Staging and EEG Measurement of Sleepiness

Sharon L. Merritt, Ed.D., R.N., Harold C. Schnyders, Ph.D., Minu Patel, M.S., Robert C. Basner, M.D., and William O'Neill, Ph.D.

1Center for Narcolepsy Research, College of Nursing, University of Illinois at Chicago (UIC)
2Department of Medical-Surgical Nursing, UIC
3Center for Sleep and Ventilatory Disorders, UIC*
4College of Engineering, UIC

Corresponding author:

Sharon L. Merritt, R.N., M.S.N., Ed.D.
Associate Professor and Director
Center for Narcolepsy Research, Rm. 215
College of Nursing, M/C 802
845 S. Damen Ave.
Chicago, IL  60612-7350
(312) 996-5176, FAX: (312) 996-7008
email: slm624@uic.edu

*Dr. Basner was Director of the UIC Sleep Center when this study was conducted.
ABSTRACT

The goal of this multi-method, known-groups measurement study was to examine the validity (accuracy) of the pupillometric Alertness Level Test (ALT) as a physiologic measure of sleepiness. The study used a pooled-time series-correlation design with 16 untreated narcolepsy (8F, 8 M), 16 obstructive sleep apnea (OSA) (7 F, 9 M), and 16 healthy control (8 F, 8 M) subjects extensively screened to determine their eligibility. All participants underwent EEG/polysomnography testing using standard Multiple Sleep Latency Test (MSLT) electrode placement concurrent with the 15 min. pupillometric ALT. Data were collected in a quiet dark environment with subjects seated in a comfortable chair with eyes open and staring at a red spot. EEG data were examined to determine if theta power (4-7 Hz) increased during 2-sec. periods of proportional pupil size decreases (pupil Stage 1, 95% or more of maximal pupil size to Stage 4, 65-74% of maximal size). Printed EEG records were visually scored. Self-report sleepiness measures included the Pittsburgh Sleep Quality Index (PSQI), the Profile of Mood States (POMS), and the Epworth Sleepiness Scale (ESS). Within subject groups, theta power ratios significantly increased across pupil stages for the sleep disorder groups but not for controls (theta activity increased 42% for narcoleptic and 36% for OSA subjects). Between subject groups, the amount of theta activity was significantly greater for narcoleptic and OSA subjects than for controls. Visual EEG scoring and self-report measures were usually consistent with objective findings. The ALT is convenient, easily repeatable, and less technically demanding than EEG sleepiness measures, and it deserves more comprehensive testing as a valid measure of sleepiness.

KEY WORDS: Sleepiness, pupillometry, EEG, theta power, pupil stage
INTRODUCTION

Sleepiness can be defined as a variable, relatively extended, and actively regulated reoccurring process that is influenced by levels of arousal and sleep need (Pivik, 1991). Among adequately rested, awake individuals, sleepiness is most prevalent upon awakening, about 12 hours after the mid-sleep period and just prior to consolidated, nighttime sleep. Sleepiness, often due to lifestyle issues (e.g., irregular sleeping habits) or intrinsic factors (e.g., the presence of a sleep disorder such as narcolepsy or obstructive sleep apnea [OSA]), becomes problematic when unintended episodes of sleep arise during waking hours and have a disruptive impact on activities of daily living (National Center for Sleep Disorders Research, 1997).

Physiologic (objective) sleepiness has been defined as how rapidly an individual falls asleep under sleep-promoting situations (Carskadon et al., 1982). The Multiple Sleep Latency Test (MSLT) is the polygraphic EEG "gold standard" for objectively measuring physiologic sleepiness. The MSLT assesses the mean number of minutes people take to fall asleep after being instructed to do so while lying down in a quiet, dark environment with eyes closed during four to five 20-minute daytime nap opportunities scheduled two hours apart. A mean time of <5 minutes is considered pathologic, and >10 minutes is considered normal (Carskadon et al., 1986). The Maintenance of Wakefulness Test (MWT), another EEG technique, is similar to the MSLT except that the individual is seated and instructed to try to stay awake during four to five 20- to 40-minute nap opportunities (Mitler et al., 1982). Recent norms suggest that an ability to stay awake for <11 minutes is abnormal (Doghramji et al., 1997). Both tests do not necessarily determine the cause of sleepiness, are expensive to administer, are relatively invasive, require about 22 hours of testing in a sleep laboratory, are not easily repeated, and require technical expertise. Additionally, microsleeps (the transient periods of decreased wakefulness that can occur prior to sleep onset and are often found in sleepy individuals) are ignored in the standard scoring procedures for both tests (Carskadon et al., 1986, Doghramji et al., 1997).
The ideal measure of sleepiness would be physiologically based, convenient, noninvasive, rapidly administered, and readily repeated throughout the day (Mitler and Miller, 1996). Pupillometry (referred to as "pupillography" by some investigators) is a technique for measuring changes in pupil size that has promise for meeting all of these requirements. Pupillometry could be useful as a screening tool as well as a measure of alertness whenever this outcome is of interest (e.g., in the transportation industry) or for monitoring treatment for sleep disorders or in clinical trials. These are situations for which polygraphic measures are not well suited. However, inconclusive and conflicting results from past research have limited pupillometry's application as a physiologic sleepiness measure. Newman and Broughton (1991) suggested that, foremost, the validity of changes in pupillary behavior as a measure of sleepiness needed to be further assessed.

The purpose of this measurement study was to examine the validity of pupil miosis as a measure of ensuing physiologic sleepiness during the pupillometric alertness level test (ALT). Specifically, we sought to determine if objectively defined sleepiness reflected in EEG theta power measures increased according to proportional pupil diameter decreases among persons with untreated narcolepsy, persons with untreated obstructive sleep apnea (OSA), and healthy controls. People with these two sleep disorders were chosen as subjects for this study because problematic sleepiness during waking hours is a cardinal manifestation of both narcolepsy and OSA and is the most frequent reason people are seen in sleep centers. In U.S. and European populations, 75% of the people seen have OSA, and 20% have narcolepsy (Partinen and Hublin, 2000).

Narcolepsy is a neurologic disorder in which people experience instability in the sleep-wake cycle that is characterized by problem sleepiness during waking hours and disturbed nighttime sleep. Other symptoms that may be present include cataplexy (momentary loss of voluntary muscle control in intense emotional situations), hypnagogoic/hypnopompic hallucinations (vivid dreaming while falling asleep or waking up) and/or sleep paralysis (brief episodes of an inability to move while falling asleep or waking up) (Overeem et al., 2001). Narcolepsy is diagnosed when (1) an all-night sleep study (PSG) is negative for the presence of another sleep disorder, and (2) MSLT nap study
findings include a mean sleep latency of <5 minutes and ≥2 nap periods during which rapid eye movement sleep (REM) occurred within 15 min. of falling asleep (American Sleep Disorders Association, 1997).

Obstructive sleep apnea is a breathing disorder during sleep that is manifested by loud, irregular snoring, restless sleep, and daytime sleepiness. This disorder is diagnosed when these symptoms are present and an all-night sleep study (PSG) reveals the presence of ≥5 periods per hour of sleep of (1) apnea (a ≥50% decrease in airflow lasting >10 sec.) and/or (2) hypopnea (a ≥30% reduction in airflow lasting >10 sec. with a ≥4% reduction in blood oxygen level). Therefore, an apnea/hypopnea index (AHI) of ≥5 must be present (American Academy of Sleep Medicine Task Force, 1999). The sleepiness associated with OSA is thought to occur because of the arousals from deeper to lighter stages of sleep that happen when apneas and/or hypopneas are experienced.

The pupil is formed by the iris and acts as a diaphragm to control the amount of light reaching the retina. The size of the pupil is determined by the relative activity of two smooth iris muscles. The constrictor pupillae is responsible for pupil miosis and the dilator pupillae for mydriasis. These two functionally opposing muscles are under the control of distinct sympathetic and parasympathetic autonomic neuronal pathways (Szabadi and Bradshaw, 1996). Even under constant accommodation and lighting conditions, the healthy iris undergoes small changes in size (.1-.3 mm oscillations lasting .5-1 second) during wakefulness due to variability in neuronal stimulation of the two iris muscles (Lowenstein et al., 1963). These changes are particularly prominent in moderately bright light. In the dark, the pupils become large and show little oscillation or miosis in alert individuals (Kardon, 1997).

Parasympathetic pupilloconstrictor neurons in the midbrain area of the Edinger-Westphal nucleus (EDW) fire continually and travel via parasympathetic reflex pathways to activate the sphincter muscle and constrict the pupil (Szabadi and Bradshaw, 1996). Central influences (i.e., alertness and emotion) and the peripheral sympathetic branch activate the dilator muscle. In alert
individuals, excitatory impulses arise from the cerebral cortex and travel through the reticular activating system and hypothalamus to cause pupil dilation by (1) inhibiting the EDW and parasympathetic constrictor activity and (2) activating the peripheral sympathetic pathway to the dilator muscle (Szabadi and Bradshaw, 1996). Progressive loss of central sympathetic influences is the basis for pupillary oscillations and miosis. With ensuing sleepiness, central and hypothalamic centers cease to function in an orderly manner, central inhibition of the EDW decreases, sympathetic tone is steadily lost, and a preponderance of parasympathetic activity occurs that is reflected in decreasing pupil size and large, slow pupillary oscillations (Lowenstein et al., 1963; Kardon, 1997).

Lowenstein and Lowenfeld (1963), using electronic video pupillometry, observed that alert individuals seated in a quiet, dark environment with eyes open can maintain a stable, dilated pupil size for 10 minutes or longer. However, in manifestly sleepy people, the pupils demonstrated slowly oscillating, large-amplitude changes that were accompanied by a progressive decrease in size. While pupillometry has had excellent potential for being an easily repeatable and convenient, noninvasive measure of alertness, clear conclusions about this technique have been obscured by the variations in aspects of the pupil that have been studied as well as variations in data collection procedures and the handling and analysis of data (Newman and Broughton, 1991; Pressman and Fry, 1989; Schmidt and Fortin, 1982; Keegan and Merritt, 1996; Kollarits et al., 1982; Pressman et al., 1984; Schmidt, 1982; Mc Laren et al., 1992; Lin et al., 1991; McLaren et al., 2002).

Yoss et al. (1969) standardized the conditions for measuring pupil oscillation and miosis by developing the pupillometric Alertness Level Test (ALT), during which the person, who is instructed to relax but try to stay wake without doing anything special to do so, is seated quietly in a dark room with eyes open and fixated on a small red spot of light. Concurrently, the pupil diameters are measured continuously for up to 15 minutes by infrared video cameras with analog to PC digital image processing. Yoss et al. (1970) also developed a system for scaling levels of wakefulness during 15 minutes of alertness testing by measuring pupil diameter as a proportion of the individual's
maximum diameter (Stage 1 "awake" = 95-100%, Stage 2= 85-94%, Stage 3= 75-84%, and Stage 4 "sleepy" = 65-74% of maximum diameter).

Researchers have found that fluctuations in wakefulness can be examined with EEG measures on active subjects with eyes open and engaged in their usual waking activity (Akerstedt and Gillberg, 1990; Hasan, 1996; Horne and Reyner, 1995; Torsvall and Akerstedt, 1988; Broughton et al., 1998). In these situations, intrusions of alpha and theta activity into the beta activity of active wakefulness have been interpreted as ensuing sleepiness. As a further characterization of a sleepy EEG pattern, Guilleminault (1994) defined lapses into microsleeps with ensuing sleepiness as the appearance of 3- to 14-sec. bursts of theta activity into the awake EEG. In an earlier pupillometry study, Keegan and Merritt (1996) found a significant correlation between variations in pupil size and EEG power measures across the delta through beta bands when data were segmented into 2-sec. windows.

METHODS

Subjects

Data from 16 untreated narcoleptics (8 F, 8 M), 16 untreated OSA subjects (7 F, 9 M), and 16 controls (8 F, 8 M) were collected at 1400 hours, or about 12 hours after their mid-sleep period to maximize the probability of sleepiness occurring in all three subject groups. To be eligible, subjects needed to (1) provide informed consent for their participation; (2) be between the ages of 18-49; (3) score <16 on the Beck Depression Inventory (BDI), a well validated screening tool with a sensitivity of .92 for detecting depressive symptoms (Beck et al., 1979); (4) provide a copy of their sleep and MSLT studies that documented their diagnosis according to ASDA criteria (American Sleep Disorders Association, 1997); (5) if a normal control, have no symptoms of a sleep disorder by medical history that was further documented by their responses to the Sleep Disorder Questionnaire (Douglas et al., 1994); (6) have no other chronic disease besides narcolepsy or OSA except for essential hypertension under control for OSA subjects (OSA subjects on antihypertensive medications that might influence autonomic function were excluded from participation); (7) abstain from any medications that could influence wakefulness or pupil behavior, and agree to undergo urine
drug screen testing; (8) report a caffeine intake of <500 mg/day (US Modafinil in Narcolepsy Multicenter Group, 1998 and 2000); (9) maintain their usual sleep pattern for 7 days prior to the ALT, and not be employed in rotating or night shift work; and (10) be free of any eye problems or surgery that could affect pupil motility. The procedures for this study were reviewed and approved by the Institutional Review Board of the University of Illinois at Chicago (UIC). Subjects were paid a stipend for their participation as well as travel costs consistent with university policy. The narcoleptic subjects were recruited from the volunteer research participant roster (N = ~300) maintained by the UIC Center for Narcolepsy Research. OSA subjects who appeared to be eligible were recruited consecutively after overnight polysomnography (sleep) testing at the Center for Sleep and Ventilatory Disorders at the UIC Medical Center. Normal control subjects were recruited through advertisements placed in the medical center.

Attempts were made to match control and OSA subjects on the basis of age and gender to narcoleptics, with subjects for the two former groups recruited following the addition of a narcoleptic subject. One-way ANOVA with Least Significant Difference post hoc testing revealed that the OSA subjects were statistically significantly older than the other two groups: F$_2$ = 3.9, p=.027, mean (SD, range) age for OSA subjects was 36.9 (±7.5, 19–47) compared to 30.3 (±8.1, 20–49) for narcoleptics and 30.4 (±7.4, 21–45) for controls. While the OSA subject group was statistically significantly older, none of these subjects could be classified as elderly. Furthermore, absolute and relative pupil diameter (Yoss pupil stage) and the distribution of theta power findings presented later in this manuscript suggest that this age difference did not have a practical effect on the outcomes of this research.

Data Acquisition

After giving informed consent, subjects provided their sleep log record for the seven days prior to the day of data collection, provided a sample for urine drug screening, and completed the BDI and a questionnaire that provided information about sleep/wake pattern, caffeine intake, and
medication use during the past 24 hours. These data were used to further verify eligibility at the time of pupillometry testing.

**Subjective sleepiness measures.** Following Alertness Level Testing (ALT) described in succeeding paragraphs, the remaining validated subjective measures were completed and included the Pittsburgh Sleep Quality Index (PSQI), the Profile of Mood States (POMS), and the Epworth Sleepiness Scale (ESS), and (for controls) the Sleep Disorders Questionnaire (SDQ) to further confirm eligibility. The PSQI is a widely used self-rating questionnaire that assesses sleep quality, latency, duration, efficiency, disturbances, sleep medication, daytime dysfunction due to sleepiness, and the usual hours of sleep per night over the past month (Buysse et al., 1989; Carpenter and Andrykowski, 1998). The POMS is a 65-item mood inventory with items clustered into 6 scales: tension-anxiety, depression-dejection, anger-hostility, vigor, fatigue-inertia, and confusion-bewilderment (McNair et al., 1981). The ESS is a self-report instrument on which respondents rate their chance of falling asleep in recent times in 8 common daily situations (Johns, 1991; Johns, 1992). Significant differences in ESS scores have been found between non-patients and patients with sleep-disordered breathing (Johns, 1993) and narcolepsy (US Modafinil in Narcolepsy Multicenter Group, 1998). The standardized SDQ, which is based on the Stanford Sleep Inventory, provides a detailed summary of sleep habits and patterns of subjects (Douglas, 1993). SDQ scales have been shown to be highly correlated with sleep studies in diagnosing sleep apnea (median scale score = 41 vs. control group score of 18), narcolepsy (median = 38 vs. 21), psychiatric sleep disorders (median = 25 vs. 15), and periodic leg movements (median = 22 vs. 13) (Douglas et al., 1994).

**Pupillometry ALT testing.** The ALT, conducted with a pupillometry system built at Mayo Clinic (Mc Laren et al., 1995), consisted of 1 minute of recording of pupil diameter in the light followed by 14 minutes in a quiet, dark room. Smith and Smith (1999) indicated that measuring the pupil in the dark is a highly reproducible measure. During the ALT, the subjects were seated in a comfortable chair and instructed to try to stay awake but not do anything special to maintain alertness, to try to minimize blinking, and to stare straight ahead at a small red dot projected on the
wall about 2 m from their eyes. Contrary to the concerns raised by some researchers, Hunter et al. (2000) found that pupil size was not affected by lens accommodation when data were collected under constant lighting conditions and far fixation. The video pupil images from two small infrared cameras were processed by an image processor which brightens the outline of the pupil circumference it detects and outputs a voltage proportional to the pupil diameter. The analog pupil diameter data were digitized using an A/D board and downloaded at a rate of 256 Hz to a PC (McLaren et al., 1995). As data were collected, the processed images were observed on two 3-inch TV monitors outfitted with brightness reduction filters, and concurrently videotaped. Offline, pupil diameter data were cleaned of eye blinks, closure, and movement artifacts using a previously validated algorithm (Merritt et al., 1994), averaged, and stored at a rate of 8 Hz on a PC. Concurrent video tape recordings of the pupil guided the removal of artifact not handled by the automated algorithm when necessary.

**Polysomnography (PSG).** EEG/polysomnography data were recorded concurrently with pupil diameter data using a Grass Instruments Model 8-16D polygraph and the standard MSLT montage: EEG (C3/A2, O1/A2 & P3/O1), EOG (ROC/A1 and LOC/A2), and EMG (bilateral mentalis placement) (Carskadon et al., 1986). To dissipate sleepiness that can ensue during electrode placement, subjects were given a drink of water after electrode placement and took a walk for 10 minutes within the center prior to data collection. Filters for EEG and EOG were set at .3 Hz for high pass and 30 Hz for low pass. EMG filters were set at 10 Hz and 100 Hz, respectively. Amplifiers were calibrated by standard procedures, and inter-electrode impedance was maintained below 5 Kohms (Grass Electrode Impedance Meter EZM4) to insure reliable recordings. PSG data were digitized at 256 Hz and stored with pupillometry data on a PC; data were recorded later on paper at a speed of 10 mm/sec. for visual scoring.

**Design and Data Analysis**

This study used a pooled-time series-correlation design. ANOVA from a previous study comparing the recurrence of pupillary patterns between narcoleptics and controls indicated an effect
size of 1.1 (Keegan et al., 1993). For an alpha level of .05 and beta of .20, a sample size of 16 per group was determined to be sufficient (Cohen, 1988). The independent variables were group (narcoleptic, OSA or control) and Yoss pupil stage (Stage 1–Stage 4). The dependent variable was theta power ratio. The other focal outcomes of interest included selected ratings from the Pittsburgh Sleep Quality Instrument (PSQI), Profile of Mood States (POMS), and the Epworth Sleepiness Scale (ESS), as well as visual scoring of EEG records.

Since the amount of theta wave activity has been shown to increase in value during episodes where people demonstrate increasing sleepiness behaviorally, power spectral density functions (FFT) were computed with 64 Hz data from the C3/A2 lead on 2-sec. epoch windows for the 5-11 minutes of recording for each subject. Data for the first 4 minutes of recording were eliminated from the analysis because the pupil dilates and oscillates when the lights are extinguished and can take 2-3 minutes of darkness to reach a larger stable diameter (i.e., dark adaptation) (Kollarits et al., 1982). The absolute theta power (4-7 Hz) for each 2-sec. EEG epoch was standardized proportionately for each subject by dividing each 2-sec. value by the mean theta power calculated from the thirty 2-sec. epochs of data collected during the first minute of recording in the dark. Printed records of EEG data were reviewed, and epochs with significant muscle, eye closure, or movement artifact were removed from analyses. The mean pupil diameter for each 2-sec. EEG window was calculated and divided by the largest pupil diameter found for that subject during the fourth minute of dark recording, resulting in a standardized, proportional pupil size for each EEG window. The 2-sec. theta power ratio values for each Yoss stage (Stage 1 ≥95% to Stage 4 = 65-74% of maximum pupil diameter) were aggregated for each subject and then aggregated by group (narcoleptic, OSA or control) according to the stage found for the concurrent pupil diameter ratio value (e.g., theta values for all Yoss Stage 2 pupil diameter ratios were combined by group for further analyses regardless of when they occurred during the 5-11 minute time period). ANOVA procedures were used to determine if there were significant differences by group (3 subject groups) and Yoss pupil stage (4 stages), forming 12 group means for analyses. Post hoc Least Significant Difference testing with Bonferroni correction (.05
divided by the number of comparisons undertaken within or between groups [Keppel, 1991]) was conducted to determine which means differed significantly from each other.

Printed EEG records for each subject's recording were visually scored by a registered polysomnography technician blinded to the subject's condition. Criteria for (1) sleepiness and sleep described by Valley and Broughton (1983) and (2) EEG-defined microsleeps by Guilleminault (1994) were used to score the records. The following observable states lasting for 15 sec. or longer of each epoch were used to visually score the thirty 30-sec. epochs (15 min. of recording time) for each subject: (1) awake – beta or alpha rhythm or blocked alpha with eye movement or blink artifacts; (2) stage 1A – slowed (at least 1 Hz slower than baseline, which was defined as the first minute of recording in the light) and/or fragmenting alpha intermixed with a medium voltage mixed-frequency pattern; (3) stage 1B – 4-7 Hz theta activity or other medium-voltage mixed-frequency pattern and less than 20% alpha rhythm that could also be accompanied by sharp vertex waves on the central channel; (4) stage 2 – one or more sleep spindles (11- to 15.5-Hz sigma rhythm lasting more than .5 sec. with 25 $\Phi$V or greater amplitude) and/or $K$ complexes (sharp vertex waves lasting more than .5 sec. associated with a slow wave or spindle) with background activity consisting of medium voltage, mixed frequencies, and less than 20% delta waves. These criteria are similar to the Rechtschaffen and Kales (1968) sleep staging scoring standards, except that stage 1 is divided into two stages offering a somewhat finer stage 1 differentiation (Rogers et al., 1994). Microsleeps were defined as EEG changes to a stage 1 NREM pattern lasting for 3–14 sec. of the epoch (Guilleminault, 1994).

RESULTS

Eligibility

All subjects met the age criterion, although the OSA group was somewhat older. However, there was no significant difference between groups in the mean pupil diameter during the fifth minute of recording after dark adaptation (ANOVA $F_2 = 1.5$, $p=.237$). The expected range for pupil diameter per age decade is as follows: 20 to 29 years of age, 5.6 to 8.6 mm.; 30 to 39, 5.2 to 8.2 mm.; 40 to 49, 4.8 to 7.8 mm (Smith, 1999). When the diameter of each individual was classified according this
range, 63% of OSA subjects’ mean 5 min. pupil diameter fell within the expected range, 31% were smaller than the expected range and 6% larger than the expected range. The results for narcoleptics were 69% in the expected range, 19% smaller and 13% larger while 81% of the controls’ mean 5 min. pupil diameter were in the expected range, 6% smaller and 13% larger indicating that this age difference was not likely to affect pupil-staging results between groups ($\chi^2 = 3.76, p=.44$).

Gender was fairly well matched in spite of an estimated male-to-female OSA prevalence of 2:1 among adults between 30 to 60 years of age (Phillipson, 1993; Young et al., 1993). Mean (SD, range) scores on the Beck Depression Inventory (BDI) by subject group were as follows: narcoleptics 6.4 (±5.4, 0–15), OSA 7 (±4.9, 1–15), controls 2.5 (±3.0, 0–11). A review of all-night sleep (PSG) and MSLT nap studies provided the following data. (1) Narcoleptic subjects had no evidence of another concurrent sleep disorder on PSG, an average time to sleep across their nap opportunities of 3.4 (±2.6, 1-8) minutes with 2 to 5 SOREMPs per subject and reported experiencing cataplexy (momentary loss of voluntary muscle control in intense emotional situations), hypnagogic/hypnopompic hallucinations (vivid dreaming while falling asleep or waking up), and/or sleep paralysis (brief episodes of an inability to move while falling asleep or waking). (2) OSA subjects had a mean sleep disordered breathing apnea/hypopnea index (AHI) of 33.3 (±30, 9.4–129.2) that was associated with frequent arousals from deeper to lighter sleep stages (arousal index = 35.6±33.6, 3–132.4) and no PSG evidence of another concurrent sleep disorder. Three OSA subjects could be classified as mild (5–15 AHI), 7 as moderate (15–30 AHI), and 6 as severe (>30 AHI) (American Academy of Sleep Medicine, 1999). In terms of treatment status, 1 narcoleptic had never taken stimulant medication, 2 had been off stimulants for 1 year or longer, 1 for 3 weeks, 5 for 2 weeks, and 7 for one week. Fourteen of the OSA subjects were waiting to be equipped for home continuous positive airway pressure treatment (CPAP). Two OSA subjects who had personally discontinued treatment due to intolerance for 3 and 7 days respectively prior to participation were included because sleepiness can return after just one night of sleeping without CPAP (Kribbs et al.,
On self-ratings of their daytime sleepiness (none to severe), 80% of narcoleptics rated themselves as moderately sleepy and 20% severely sleepy; OSA subjects rated themselves as mildly sleepy (18%), moderately sleepy (55%), or severely sleepy (27%). Other than hypertension under control for three OSA subjects, none of the subjects in the sleep disorder or control groups had a history or clinical evidence of another concurrent medical disorder. Sleep Disorder Questionnaire scale scores for control subjects were below the empirically established cut-point scores derived from Receiver Operating Curve analyses for the presence of a particular sleep disorder (Douglas et al., 1994). (1) Sleep apnea scale scores ranged from 12–29 compared to suggested cut-point scores of 36 for M and 32 for F. (2) Narcolepsy scale scores ranged from 15-28 compared to 30 for M and 31 for F. (3) Psychiatric sleep disorder scale scores ranged from 9–18 compared to 19 for M and 21 for F. (4) Periodic leg movement disorder scale scores ranged from 9–16 compared to a cut-point of 21 for both genders. All subjects tested negative for prescription or recreational drugs that could affect alertness on urine drug screen analyses and were classified according to daily caffeine intake as follows: (1) none, (2) light <125 mg/day, (3) moderate 125-250 mg/day, or (4) heavy 250-500 mg/day (Mitler et al., 1998). There were no significant differences between subject groups in caffeine use \[\chi^2_e = 1.95, p= .924\]. Eighty-one percent of the narcoleptic and control group subjects, and 88% of OSA subjects reported no or light caffeine use. Since caffeine withdrawal can result in drowsiness (Nehlig, 1999), subjects were allowed to continue their usual rising morning caffeine intake but discontinued use by 1000 hours on the day of testing. All reported maintaining their usual sleep schedule for the past week, worked a straight day or evening shift (if employed), and denied having any eye problems or surgery that could interfere with pupil motility.

**Subjective rating scales**

The respective standardized instruments were scored for each subject according to the instructions provided for their use. ANOVA results showed no significant differences between subject groups in the self-reported hours of sleep per night for the past month (Table 1). However, paired student t-tests showed that narcoleptic and control subjects reported sleeping about 40 minutes
less than their usual sleep time during the 24 hours preceding ALT testing. Kruskal-Wallis $\chi^2$ results revealed significant differences in the distribution of scores between the 3 groups on the sleep quality, sleep disturbances, and daytime dysfunction due to sleepiness scales of the Pittsburgh Sleep Quality Index (PSQI) (Table 1). Mann-Whitney U post hoc testing with Bonferroni probability adjustment ($p \leq .016$) demonstrated that OSA subjects experienced poorer subjective sleep quality and greater sleep disturbances, and narcoleptics experienced more daytime dysfunction compared to controls. The narcoleptic and OSA distributions were not significantly different.

[Table 1 here]

Repeated measures ANOVA results for selected Profile of Mood States (POMS) scales relevant to this study are presented in Table 2. Least Significant Difference post hoc testing with Bonferroni correction ($p \leq .016$) revealed that, compared to controls, (1) narcoleptic and OSA subjects reported experiencing significantly less vigor (i.e., energy level), and (2) narcoleptics experienced significantly more fatigue (i.e., weariness and inertia).

One-way ANOVA with post hoc Least Significance Difference testing and Bonferroni correction ($p \leq .016$) was used to determine if the Epworth Sleepiness Scale (ESS) ratings differed significantly from each other by subject group (Table 2). Compared to controls, both narcoleptic and OSA subjects rated themselves as more sleepy in sleep-promoting situations. No significant differences were found between the sleep disorder groups on either the POMS scales or the ESS.

[Table 2 here]

Visual EEG scoring

Results from the visual scoring procedure can be found in Table 3. The proportion of time spent awake during the 15 minutes of the ALT was significantly less ($p=.007$) for narcoleptics compared to controls and approached significance for OSA subjects compared to controls ($p=.09$) [ANOVA $F_2=3.9$, $p=.026$, Least Significant Difference testing with Bonferroni correction $p \leq .016$]. The sleep disorders groups were not significantly different from each other. When the research scoring criteria of 3 consecutive epochs of stage 1 sleep or the first epoch of any other sleep stage...
were applied (Mitler et al., 2000), 2 narcoleptic subjects fell asleep during the ALT at 9.5 and 15 minutes respectively, and 3 OSA subjects fell asleep: 1 at 11 minutes, and 2 at 12 minutes.

Table 3 here

### Pupil staging and proportional theta

Graphs portraying the theta power ratio and pupil diameter ratio results for 2 subjects during the ALT are presented in Figure 2. When the results of the narcoleptic versus the control subject are visually compared, the increases in theta power (Panel A, upper graph) and decreases in pupil size (Panel B, upper graph) can be seen readily in the narcoleptic subject. The control subject, whose EEG confirmed that alertness had been maintained throughout the test, showed small changes in theta power and was able to sustain a relatively stable pupil diameter for the 15-min. ALT recording period.

Figure 2 here

When the artifact-free 2-sec. mean pupil diameters were staged according to the Yoss criteria, about 80% of the available samples for narcoleptics were Stages 1 and 2 (85% or greater of the initial pupil size), 69% for OSA subjects, and about 90% for controls (Figure 3). In spite of their being older, OSA subjects demonstrated substantial variability in their relative pupil size (Yoss pupil stage distribution). ANOVA procedures by condition (narcoleptic, OSA, or control subject) and Yoss pupil stage (4 stages) revealed significant differences in mean theta power ratio by group and Yoss stage (F (11)=27.1, p=.000) (Figure 3). Post hoc Least Significant Difference testing with Bonferroni correction (p≤.008 within groups and P≤.016 between groups) was conducted to determine which means differed significantly from each other. Within groups, (1) the theta power ratios for pupil Stages 3 and 4 were significantly higher than Stages 1 and 2, and Stage 2 was significantly higher than Stage 1 for narcoleptics. (2) Theta ratios for Stages 3 and 4 were significantly higher than Stages 1 and 2 for OSA subjects. (3) Theta ratios by pupil stage were not significantly different from each other for controls. For narcoleptics, the amount of theta activity increased about 42% from pupil Stages 1–4 and about 36% between these same stages for OSA subjects. Between the groups,
(1) the amount of theta activity was significantly higher in Stages 1-3 for narcoleptics compared to OSA subjects, who in turn showed significantly more theta activity than controls, and (2) both narcoleptic and OSA subjects showed significantly more theta activity at Stage 4 than did controls.

Subjects also rated their current level of sleepiness on a visual analog scale (VAS) (10 cm line, 0=sound asleep to 10 cm=as awake as can be) immediately before and right after the ALT. On the before VAS, narcoleptics rated themselves as significantly more sleepy (mean=4.5±2.1) than both OSA (6.8±2.6) and controls (7.9±1.3) (F\textsubscript{2}=11.7, p=.000), and as significantly more sleepy (3.5±1.9) than controls (6.2±1.9) on the post-ALT VAS measure (F\textsubscript{2}=5.1, p=.003). OSA subjects (4.6±3.1) showed a trend towards significance (p=.07) on the post-VAS compared to controls.

[Figure 3 here]

DISCUSSION

Whether or not pupillary miosis measured during the pupillometric ALT is a measure of sleepiness has been controversial. Tyron (1975) listed factors that can influence absolute pupil size which must be considered by investigators including: (1) with increasing age, pupil diameters decrease, and become more variable in size especially in elderly adults; (2) intensity and color of light stimuli affect pupil size; (3) alerting or relaxing experimental conditions induces changes in pupil size. Inconsistent findings between earlier studies may be due to the use of size as an outcome variable when there were varying conditions within and between experiments and/or the subjects included elderly adults. When pupil size is measured in the dark, as is done in the ALT, Smith (1999) concluded that variability in measurement between laboratories is about 3%. The Yoss et al. (1970) method of pupil staging provides a method for comparing the relative amount of pupil miosis over time between individuals rather than absolute pupil size. Furthermore, sleep promotion situations, such as are present during the ALT, unmask physiologic sleepiness, they do not cause it (Roehrs et al., 2000). By using the ALT with a known-groups approach in this study, we found significant differences between controls compared to narcoleptic and OSA subjects across a variety
of subjective and objective measures that provide key evidence that a relative decrease in pupil size is a measure of sleepiness.

The subjective and visual scoring results are interesting in that they are similar to the objective pupil/EEG findings. In terms of their sleep pattern, there were no differences between subject groups in the usual amount of sleep they reported obtaining per night for the past month, although both narcoleptic and control subjects reported somewhat shorter sleep durations in the 24 hours before testing. On the Pittsburgh Sleep Quality Index, narcoleptic subjects rated themselves as experiencing more daytime dysfunction due to sleepiness than controls. They also independently rated their daytime sleepiness as moderate to severe. In contrast, OSA subjects varied in their sleepiness ratings from mild to severe and did not report experiencing more sleepiness dysfunction than controls, in spite of the fact that they rated their sleep quality as less and their sleep disturbances as more than controls. Some studies of sleepiness in OSA subjects have found that the participants were unaware of the extent of their sleepiness (Chervin and Guilleminault, 1996; Engleman et al., 1997). One could speculate that the OSA subjects may have been less aware than the narcoleptic subjects of how their difficulties with daytime sleepiness affected their functioning.

On the POMS, both of the sleep disorder groups reported experiencing less vigor and more fatigue than the controls. Likewise, on the ESS, both sleep disorder groups rated themselves as more likely to fall asleep in sleep-promoting situations.

Based on visual scoring of the EEG data for each 30-sec. epoch of the 15 min. of recording time, narcoleptics were awake during the ALT about 74% of the recording time compared to 91% of the time for controls. While OSA subjects were scored as awake 81% of the time, the difference between controls and the OSA group approached but did not achieve significant difference.

Yoss (1970) reported behavioral signs of decreasing alertness concurrent with pupil miosis and increasing pupil stage. As the diameter decreased to less than 85% of maximum size, subjects were observed to progress from feeling tired to feeling drowsy at Stages 3 and 4. Eyelid ptosis was noted to increase from occasional mild droopiness in Stage 2 to frequent mild ptosis at Stage 3 and
frequent mild ptosis with occasional severe droopy eyelids at Stage 4. The pupil staging data are similar to those of the EEG visual scoring findings outlined in Table 3 for two subject groups, in that approximately 80% of the pupil samples for narcoleptics and 90% for controls were scored as awake, while 69% were scored as awake samples for OSA subjects.

Based on the subjective and visual scoring data, one could hypothesize that the narcoleptic subjects might be the sleepiest group on the EEG measure. Our findings imply that this was the case, since the narcoleptic subject group showed the greatest increase in the amount of theta activity with decreasing pupil size. Horne and Reyner (1995) stated that there is good evidence that rising theta EEG activity is a sign of increasing sleepiness. Under constant illumination, pupil dilation during wakefulness is the result of centrally derived sympathetic stimulation and central nervous system supranuclear inhibition of parasympathetic activity (Kardon, 1997). During ensuing sleepiness, there is a progressive loss of central influences, and the pupils oscillate and become increasingly more miotic. The central derivation of this pupil behavior has been inferred from observations that (1) pupil fluctuations in humans are synchronous and the same in both pupils (Stark, 1968) and (2) in sympathectomized animals, stimulation of the cortex or diencephalon caused pupil dilation but abolished the light reflex (Lowenstein and Loewenfeld, 1962). The increases in theta activity with the progressive decreases in pupil size in the sleep disorders groups suggest that the sleep-promoting conditions of ALT testing unmasked the sleepiness present in these two groups. The non-significant changes in theta activity in controls imply that these subjects were able to maintain a higher degree of alertness during the 15 minutes of ALT testing, an inference which also is consistent with the visual scoring and subjective rating results. Finally, the significant increases in the amount of theta activity with progressive decreases in pupil size in the sleep disorder groups compared to controls provides empirical evidence that the pupil behavior during the ALT is a measure of increasing sleepiness.

Recent reviews of available techniques for measuring sleepiness concluded that the current literature on pupillometry is inconclusive (Kessler and Rodenstein, 2001) and that currently pupillometry cannot be considered a clearly reliable technique for assessing sleepiness (Curcio et al.,
Our findings of increasing EEG theta power with decreasing pupil size suggest that pupillometry deserves further examination as a valid, objective measure of sleepiness, particularly further comparison with the MSLT which is the current polygraphic "gold standard" for measuring physiologic sleepiness. To date, only one study comparing pupillometry to the MSLT can be found in the published literature. Using 49 hypersomnolent patients 18-78 years of age undergoing sleep center evaluation, McLaren et al. (2002) found (1) no significant Spearman's correlation ($r=.14$) between mean pupil diameter during ALT testing at 0900 hrs and the mean sleep latency (SL) of four MSLT naps during which data were collected every two hours starting at 1000 hrs, and (2) a significant negative Spearman's correlation ($r=-.44$) between the amount of pupil fluctuation and the four-nap MSLT mean SL. Healthy participants underwent pupillometry but no sleep or nap testing. The findings could be less than robust because of the variability in age of the subjects. Additionally, since pupillometry was conducted only once at 0900 hrs. and then compared to a mean SL from nap opportunities that were presented throughout the day starting at 1000 hrs., data were collected under experimental conditions that could have affected the SL outcome but did not affect the pupillometry findings. Sleep latency on the MSLT has been found to vary based on time of day (Carskadon and Dement, 1987), as has pupil size (Merritt et al., 1998) and amount of pupillary oscillation (Merritt et al., 2000). Finally, experimental fatigue, which is known to affect pupil findings (Tyron, 1975), could have influenced the SL results. Studies that have consistent experimental conditions across subject groups and include alert healthy normal subjects as well as sleepy individuals need to be conducted to compare the relative changes in pupil size and the amount of pupillary oscillation to MSLT sleep latency findings. Presently, the interpretation of pupillometric findings as a measure of sleepiness is limited because the only normative data available come from a study of 144 healthy male policemen who reported a normal sleep pattern, denied experiencing any sleep disturbances, and only underwent pupillometry testing once in the morning (Wilhelm et al., 1999). In this study, a significant Spearman's rank correlation ($r=.31$) was found between self-ratings on the Stanford Sleepiness Scale and pupillary measures.
Increasingly, sleepy U.S. drivers are being recognized as a public health problem, with 51% of respondents in a representative U.S. poll admitting they drove while sleepy (National Sleep Foundation, 2002). This suggests that there is a need for an objective sleepiness screening measure in the transportation industry. In this age of evidence-based health care, the outcomes of treatment for sleep disorders have been questioned, particularly the amelioration of sleepiness in OSA patients (Ross et al., 1999; Wright et al., 1997). Desktop pupillometers that are operated with a laptop PC are now available so that pupillometric sleepiness testing could be repeatedly undertaken in any quiet, dark environment. With more comprehensive study comparing pupillary outcomes to MSLT sleep latency findings throughout the day, pupillometric alertness level testing and its associated measures could become recognized as a cost-effective, useful measure of physiologic sleepiness.

ACKNOWLEDGMENTS

The authors would like to thank (1) Ms. Maureen Smith for her assistance in coordinating the recruitment of OSA subjects with other personnel in the UIC Center for Sleep and Ventilatory Disorders, and (2) Mr. Glenn Clark for visually scoring the PSG/EEG records. This research was supported by NIH grant R15 NR04030 and Mr. J. A. Piscopo. The authors wish to thank Kevin Grandfield for editing the manuscript.
REFERENCES


Figure 1. Autonomic neural control of pupil diameter. Solid lines depict excitatory connections between brain areas, broken lines depict inhibitory ones. Reprinted by permission of Sage Publications from Szabadi E, Bradshaw CM. (1996) Autonomic pharmacology of $\alpha_2$-adenoceptors. J Psychopharmacology, 10 (Supplement 3): 6-18.
Figure 2. Theta power ratio and pupil diameter ratio during the ALT for a narcoleptic and control subject.
**Number of artifact-free 2 sec theta samples (total and for each Yoss stage) for 5 – 11 minutes of recording by subject group** (number of possible samples: C = 3360; OSA = 3330; N = 3285):

- **Controls** = 2390, Stage 1 - 1333, Stage 2 - 805, Stage 3 - 195, Stage 4 - 57
- **OSA** = 2261, Stage 1 - 685, Stage 2 - 875, Stage 3 - 438, Stage 4 - 263
- **Narcoleptics** = 2433, Stage 1 - 1167, Stage 2 - 768, Stage 3 - 315, Stage 4 - 183

**Within Subjects** (p ≤ .008)  
- C: n. s.
- O: Stages 3 & 4 > Stages 1 & 2
- N: Stages 3 & 4 > Stages 1 & 2

**Between Subjects** (p ≤ .016)  
- Stage 1: N > O > C
- Stage 2: N > O > C
- Stage 3: N > O > C
- Stage 4: N & O > C

**Stage 2 > Stage 1**

Figure 3. Mean (±SD) theta power ratio by Yoss pupil stage for control (N=16), OSA (N=16) and narcoleptic (N=16) subjects for the 5 to 11 minute recording period.
Table 1.

Mean (SD) usual hours of sleep and selected Pittsburgh Sleep Quality Index (PSQI) scale score distributions by subject group

<table>
<thead>
<tr>
<th>PSQI instrument scale</th>
<th>Subject Group</th>
<th>Narcoleptic (Nar)</th>
<th>OSA (O)</th>
<th>Control (C)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean (SD)*</td>
<td>Mean (SD)*</td>
<td>Mean (SD)*</td>
</tr>
<tr>
<td>Usual hours of sleep per night</td>
<td></td>
<td>8.2±1.2</td>
<td>6.9±1.4</td>
<td>7.7±1.1</td>
</tr>
<tr>
<td>Sleep Quality</td>
<td></td>
<td>N (%) of grp</td>
<td>N (%) of grp</td>
<td>N (%) of grp</td>
</tr>
<tr>
<td>Very good</td>
<td></td>
<td>3 (18.8 %)</td>
<td>1 (6.3 %)</td>
<td>4 (25 %)</td>
</tr>
<tr>
<td>Fairly good</td>
<td></td>
<td>9 (56.3 %)</td>
<td>5 (31.3 %)</td>
<td>12 (75 %)</td>
</tr>
<tr>
<td>Fairly bad</td>
<td></td>
<td>3 (18.8 %)</td>
<td>8 (50 %)</td>
<td>---</td>
</tr>
<tr>
<td>Very bad</td>
<td></td>
<td>1 (6.3 %)</td>
<td>2 (12.5 %)</td>
<td>---</td>
</tr>
<tr>
<td>Sleep Disturbances</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not in past month</td>
<td></td>
<td>---</td>
<td>---</td>
<td>1 (6.3 %)</td>
</tr>
<tr>
<td>&lt; once a week</td>
<td></td>
<td>10 (62.5 %)</td>
<td>5 (31.3 %)</td>
<td>14 (87.5 %)</td>
</tr>
<tr>
<td>1-2 a week</td>
<td></td>
<td>6 (37.5 %)</td>
<td>9 (56.3 %)</td>
<td>1 (6.3 %)</td>
</tr>
<tr>
<td>3 or &gt; a week</td>
<td></td>
<td>---</td>
<td>2 (12.5 %)</td>
<td>---</td>
</tr>
<tr>
<td>Daytime Dysfunction</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td></td>
<td>1 (6.3 %)</td>
<td>3 (18.8 %)</td>
<td>4 (25 %)</td>
</tr>
<tr>
<td>Once or twice</td>
<td></td>
<td>5 (31.1 %)</td>
<td>5 (31.3 %)</td>
<td>10 (62.5%)</td>
</tr>
<tr>
<td>Once or twice a week</td>
<td></td>
<td>4 (25 %)</td>
<td>8 (50 %)</td>
<td>2 (12.5 %)</td>
</tr>
<tr>
<td>Three or &gt; times a week</td>
<td></td>
<td>6 (37.5 %)</td>
<td>---</td>
<td>---</td>
</tr>
</tbody>
</table>

*Paired sample student t-tests between usual hours of sleep and mean hours of sleep in the past 24 hours were significantly different for narcoleptics [7.4±1.6, $t_{15} = -2.704$, $p = .02$] and controls [7.0±, $t (15) = -2.38$, $p = .03$] but not for OSA subjects [7.1±2.8, $t_{15} = .212$, $p = .84$].
Table 2.

Selected Profile of Mood States (POMS) scales and Epworth Sleepiness Scale (ESS) means (SD) by subject group

<table>
<thead>
<tr>
<th>Instrument and Scale</th>
<th>Subject Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>(possible score range)</td>
<td>Narcoleptic (N)</td>
</tr>
<tr>
<td>Profile of Mood States*</td>
<td></td>
</tr>
<tr>
<td>Vigor (0-32)</td>
<td>13.5 (5.9)</td>
</tr>
<tr>
<td>Fatigue (0-28)</td>
<td>11.8 (7.0)</td>
</tr>
<tr>
<td>Epworth Sleepiness Scale*</td>
<td>15.4 (2.7)</td>
</tr>
<tr>
<td>(0-24)</td>
<td></td>
</tr>
</tbody>
</table>

* POMS Repeated Measures ANOVA  $F_{10} = 4.2$, p .000, post hoc Least Significant Difference test with Bonferroni correction $p \leq .016$; ESS ANOVA  $F_{2} = 13.1$, p .000, post hoc Least Significant Difference test with Bonferroni correction $p \leq .016$
Table 3.

Visual Scoring Results for Total EEG Recording Period (15 minutes) by Subject Group

<table>
<thead>
<tr>
<th>Subject Group</th>
<th>Awake*</th>
<th>Microsleeps</th>
<th>Stage 1A Sleep</th>
<th>Stage 1B Sleep</th>
<th>Stage 2 Sleep</th>
</tr>
</thead>
<tbody>
<tr>
<td>Narcoleptic</td>
<td>73.6%</td>
<td>17.2%</td>
<td>6.2%</td>
<td>1.3%</td>
<td>1.8%</td>
</tr>
<tr>
<td>OSA</td>
<td>80.7%</td>
<td>12.2%</td>
<td>5.3%</td>
<td>1.5%</td>
<td>.4%</td>
</tr>
<tr>
<td>Control</td>
<td>91.3%</td>
<td>7.7%</td>
<td>.8%</td>
<td>.21%</td>
<td>------</td>
</tr>
</tbody>
</table>

*Number of epochs without microsleep(s) or sleep EEG pattern divided by the number of recording epochs; F<sub>2</sub> = 3.9, p ≤ .026; post hoc Least Significant Difference test with Bonferroni correction (p ≤ .016), narcoleptic < controls, p .007. Based on standard sleep latency criteria (3 consecutive epochs of stage 1 sleep or first epoch of any other sleep stage), 1 narcoleptic subject fell asleep with eyes closed at 9.5 min and 1 at 15 min; 1 OSA subject fell asleep with eyes closed at 11 min and 2 at 12 min.