NEW RESEARCH

JCSM
Journal of Clinical Sleep Medicine

http://dx.doi.org/10.5664/jcsm.3278

The Accuracy of Eyelid Movement Parameters for Drowsiness Detection

Vanessa E. Wilkinson, Ph.D.[1]; Melinda L. Jackson, Ph.D.[1,2]; Justine Westlake, B.A./BAppSci (Hons)[1]; Bronwyn Stevens, BBNSc, PGradDip (Psych)[1]; Maree Barnes, MB.BS[1]; Philip Swann, Ph.D.[3]; Shantha M. W. Rajaratnam, Ph.D.[4,5,6]; Mark E. Howard, MB.BS., Ph.D.[1]

[1]Institute for Breathing & Sleep, Department of Respiratory & Sleep Medicine, Austin Health, Victoria, Australia; [2]Melbourne School of Psychological Sciences, The University of Melbourne, Victoria, Australia; [3]Department Road Safety, Victoria, Australia; [4]School of Psychology and Psychiatry, Monash University, Clayton, Victoria, Australia; [5]Division of Sleep Medicine, Department of Medicine, Brigham and Women's Hospital, Boston, MA; [6]Division of Sleep Medicine, Department of Medicine, Harvard Medical School, Boston, MA

Study Objectives: Drowsiness is a major risk factor for motor vehicle and occupational accidents. Real-time objective indicators of drowsiness could potentially identify drowsy individuals with the goal of intervening before an accident occurs. Several ocular measures are promising objective indicators of drowsiness; however, there is a lack of studies evaluating their accuracy for detecting behavioral impairment due to drowsiness in real time.

Methods: In this study, eye movement parameters were measured during vigilance tasks following restricted sleep and in a rested state (n = 33 participants) at three testing points (n = 71 data points) to compare ocular measures to a gold standard measure of drowsiness (OSLER). The utility of these parameters for detecting drowsiness-related errors was evaluated using receiver operating characteristic curves (ROC) (adjusted by clustering for participant) and identification of optimal cutoff levels for identifying frequent drowsiness-related errors (4 missed signals in a minute using OSLER). Their accuracy was tested for detecting increasing frequencies of behavioral lapses on a different task (psychomotor vigilance task [PVT]).

Results: Ocular variables which measured the average duration of eyelid closure (inter-event duration [IED]) and the ratio of the amplitude to velocity of eyelid closure were reliable indicators of frequent errors (area under the curve for ROC of 0.73 to 0.83, $p < 0.05$). IED produced a sensitivity and specificity of 71% and 88% for detecting ≥ 3 lapses (PVT) in a minute and 100% and 86% for ≥ 5 lapses. A composite measure of several eye movement characteristics (Johns Drowsiness Scale) provided sensitivities of 77% and 100% for detecting 3 and ≥ 5 lapses in a minute, with specificities of 85% and 83%, respectively.

Conclusions: Ocular measures, particularly those measuring the average duration of episodes of eye closure are promising real-time indicators of drowsiness.

Keywords: Behavioral lapses, drowsiness, fatigue, ocular measures, eye blinks

Citation: Wilkinson VE; Jackson ML; Westlake J; Stevens B; Barnes M; Swann P; Rajaratnam SMW; Howard ME. The accuracy of eyelid movement parameters for drowsiness detection. *J Clin Sleep Med* 2013;9(12):1315-1324.

SCIENTIFIC INVESTIGATIONS

Drowsiness as a result of sleep deprivation, circadian effects, or sleep disorders is a major risk factor for motor vehicle and occupational accidents.[1,2] Objective indicators of drowsiness may allow sleepy individuals to be identified in real time in the laboratory, occupational settings, and on the road, with the potential to intervene and prevent accidents. Ideally, indicators of drowsiness for this purpose should be able to detect brief periods of inattention, which may result in an individual failing to respond to hazards in the environment. The lack of validated real-time objective indicators of drowsiness for field research and operational settings, such as driving, restricts both the ability to accurately assess drowsiness in these settings and the development of field-based interventions for drowsiness. Initial laboratory studies have suggested that some ocular measures may be good indicators of drowsiness[3-5]; however, there is a paucity of detailed evaluation of the utility of different ocular measures for this purpose.

Electroencephalography (EEG) is the gold standard method for quantifying sleep state (awake versus sleep). Although EEG changes occur with drowsiness in the wake state (increased alpha and theta activity),[6] they are difficult to measure in field settings due to signal artifact, are not readily amenable to real-time signal processing and are not highly predictive of impaired behavior due to drowsiness.[7] These features have hindered the use of EEG for assessing drowsiness in field settings and the potential to use it to provide real-time drowsiness monitoring.

Changes in the frequency, amplitude, and duration of blinks and episodes of slow eye closure occur in response to increased drowsiness caused by sleep deprivation and circadian rhythm effects.[8-11] While blink duration in rested conditions lasts for less

BRIEF SUMMARY

Current Knowledge/Study Rationale: Objective indicators of drowsiness have the potential to identify drowsy drivers. This study evaluated the accuracy of ocular measures for detecting drowsiness-related behavioral impairment in real-time.

Study Impact: Several ocular measures, particularly those measuring the average duration of episodes of eye closure, were found to be accurate in the real-time detection of behavioral impairment in the laboratory. Ocular measures are promising indicators of real-time drowsiness.

than 200 ms, sleep deprivation results in increased blink duration, episodes of slow eye closure lasting more than 500 ms, and increased proportion of time the eyes are closed.[10,12] The proportion of time the eyes are at least 80% closed (PERCLOS) increases in drowsy participants during task performance and is reported to correlate well with vigilance and simulated driving tasks in the laboratory.[5-7] Technological development has enabled more detailed measurement of eyelid movements in real time. Initial reports suggest that the velocity and amplitude of eyelid movements provide useful indicators of drowsiness and that the use of multiple eyelid closure metrics may improve the prediction of drowsiness.[10,13] There is, however, a paucity of studies evaluating the ability of these measures to detect impaired vigilance as a result of drowsiness. In this study we evaluated the ability of a range of eye movement parameters to detect impaired vigilance (frequent behavioral lapses) following restricted sleep.

METHODS

Participants undertook objective assessment of vigilance and drowsiness with concurrent measurement of eyelid movement parameters following a normal night of sleep and a night restricted to 4 h time in bed in a randomized crossover design.

Participants

Healthy participants aged between 18-70 years were recruited. Participants underwent a medical review and were excluded if they had a chronic medical condition that might affect neurocognitive or motor function or be a contraindication to sleep restriction including sleep disorders and chronic neurological or psychiatric conditions. Participants were also excluded if they were heavy smokers, consumed on average ≥ 5 caffeinated beverages a day, had significant daytime sleepiness (Epworth Sleepiness Scale [ESS] > 11),[14] had a high risk of sleep apnea on a validated screening survey,[15] or had visual impairment which was not corrected with glasses.

The study protocol was approved by the Austin Health Human Research Ethics Committee and was registered on the Australian New Zealand Clinical Trials Registry.

Protocol

Participants attended the Sleep Laboratory, Austin Hospital, Heidelberg, for an initial medical screening, to obtain written informed consent and for familiarization with tasks and fitting of Optalert Drowsiness Measurement System (ODMS) equipment. ODMS glasses were fitted by an experienced researcher trained in this technique to confirm correct measurement of ocular data.

Two separate days of testing were conducted in random order, ≥ 1 week apart. Participants were instructed to maintain an 8-h in bed sleep schedule (22:00-06:00), confirmed by sleep diary, for the week preceding the testing session. A baseline session was performed while participants were in a rested state. On the night prior to the sleep restriction session, participants restricted their time in bed to 4 h (02:00-06:00). Sleep restriction was confirmed via actigraphy (SenseWear Body Monitoring System armband, Pittsburgh, USA) and a sleep diary kept for the preceding week.

Participants undertook one 1.5-h performance test battery in the rested state (baseline session) and 2 performance test batteries following one night of sleep restriction (sleep restriction sessions in the morning (SR-AM) and afternoon (SR-PM)). Eighteen participants commenced this sequence with the baseline session prior to the sleep restriction sessions, and 21 participants commenced the sequence with the sleep restriction sessions. Three testing sessions were conducted to sample a range of impairments due to drowsiness and circadian factors. Performance measures were conducted at 09:00 (baseline and SR-AM) and 13:00 (SR-PM). The test battery consisted of objective vigilance tests in addition to exploratory drowsiness and performance questionnaires, which are not presented in this analysis. Testing was conducted in randomized order in a soundproofed room with dim lighting as per maintenance of wakefulness test (MWT) protocol.[16] There was a short break in between tests to check ocular signals and adjust as required. Vigilance testing included the 40-min Oxford Sleep Resistance Test (OSLER), and 10-min psychomotor vigilance task (PVT). Continuous ODMS recording occurred during these testing sessions, in addition to video monitoring to ensure synchronization of ODMS measures with vigilance testing measures.

Outcome Measures and Data Analysis

A two-step process of analysis was undertaken. In Step 1, cutoff values for ocular measures of the ODMS for predicting drowsiness-related impairment were derived using the OSLER as the gold standard. In Step 2, the sensitivity, specificity, and area under the ROC curve were calculated using these cutpoints, using the PVT as a gold standard.

Step 1: Derivation of Cutoff Values

Ocular measures were evaluated for their accuracy in detecting drowsiness-related impairment and cutoff values for detecting impairment were developed. Eyelid movement parameters were measured using the ODMS (Optalert, Sleep Diagnostics Pty Ltd, Melbourne, Australia). This device records 8 ocular variables using infrared light from a light-emitting diode positioned below and in front of the eye on a pair of glasses hardwired to a laptop.[17] ODMS is reported to be fully functional regardless of the position or movement of the person's head, and in sunlight or darkness.[18]

The following ocular variables were calculated as an average on a minutely basis:

- *Inter-Event Duration (IED):* blink duration measured from the point of maximum closing velocity to maximum opening velocity of the eyelid measured in seconds.
- *Percent Time with Eyes Closed (%TEC):* proportion of time eyes are closed, determined from the velocity of eyelid movement during eyelid closure.
- *Blink Total Duration (BTD):* duration of blinks measured in seconds from the start of closing to complete re-opening.
- *Negative Amplitude-Velocity Ratio (–AVR):* the ratio of the maximum amplitude to maximum velocity of eyelid movement for the reopening phase of blinks.
- *Positive Amplitude-Velocity Ratio (+AVR):* the ratio of the maximum amplitude to maximum velocity of eyelid movement for the closing phase of blinks.

- *Percent Long Closures (%LC):* proportion of time eyes are fully closed > 10 ms.
- *Duration of Ocular Quiescence (DOQ):* duration of no movements between eyelid and ocular movement events, including blinks, saccades and smooth pursuit.
- *Johns Drowsiness Scale (JDS)*: a composite measure of drowsiness calculated using weighted values of the other recorded ocular variables. JDS is calculated on a scale from zero to ten, with higher scores indicative of increased drowsiness.[17]

The OSLER was used as the gold standard for indicating drowsiness and determining ocular measure cutoff values. OSLER is a portable, computerized, non-assisted method of monitoring wakefulness through responses to presented stimuli occurring every 3 sec seconds over a 40-min time period, which has been found to be reliable in measuring sleep onset.[19] The OSLER was designed as a simplified version of the maintenance of wakefulness test (MWT), a laboratory-based EEG method of determining an individual's ability to remain awake and the current recommended method of assessing whether people with sleep disorders have the ability to safely drive a vehicle.[20,21] Sleep onset is defined as no response to 7 consecutive OSLER stimuli (absent response for 18-21 sec); however, ≥ 4 consecutive missed stimuli (absent response for 9-12 sec) is strongly predictive of microsleeps (brief occurrences of theta lasting between 3 and 15 sec).[22]

OSLER and concurrent ocular measures were analyzed in 1-min bins. Overall OSLER misses per minute and consecutive misses within each minute were identified. Consecutive misses which crossed a minute bin were attributed to the minute in which the missed signals began. Data from the OSLER was compared to the corresponding time matched ocular variables in one minute bins. Data were excluded from sessions if the ODMS signal quality was poor (low amplitude), and for some instances of failure of time matching data to ocular variables.

Statistical Analysis for Derivation of Cutoff Values

Statistical analyses were conducted using STATA 11 (StataCorp, College Station, TX: StataCorp LP). Receiver operating characteristic (ROC) curve analysis was conducted for each ocular variable, using data from all 3 testing sessions, to assess the ability of each variable to identify missed OSLER signals occurring during any 1-min bin. This analysis was undertaken for (1) ≥ 4 consecutive missed signals on the OSLER in a minute, and (2) ≥ 4 total missed signals on the OSLER in a minute. Data were clustered by participant during statistical analysis to account for the multiple minutes of OSLER replication for each participant.[23,24] Cutoff values to determine a level of drowsiness resulting in drowsiness-related deterioration in vigilance for each ocular variable were determined using the peak of the ROC curve to determine the optimum sensitivity and specificity combination. In addition, high sensitivity cutoff values were determined using the highest sensitivity with a specificity of ≥ 50% and the high specificity cutoff values were determined using the highest specificity with a sensitivity of ≥ 50%. Lastly, logistic regression models were fitted using the ocular variables with the highest discrimination for detection of missed signals on the OSLER (IED, BTD, %TEC, +AVR).

Step 2: Validation of Cutoff Values

The ability of the ocular variables with the greatest discrimination in detecting frequent missed signals on the OSLER (IED, BTD, +AVR, and JDS) were then evaluated for the ability (sensitivity and specificity) of their OSLER determined cutoff levels to detect different frequencies of lapses per minute on the PVT. The PVT is a hand-held reaction time task which assesses sustained attention through measuring reaction time to a visual stimulus, presented at varying intervals approximately 10 times per minute.[25] Impaired attention is a reliable consequence of drowsiness, with the PVT showing results characteristic of decreased attention such as slowed reaction times and increases in errors and lapses.[26-28] A review of 141 articles utilizing the PVT found the 10-min PVT to be the optimal length and the outcome of number of lapses to be the most frequently used and the most reliable measure when evaluating the effect of sleep loss.[29] PVT files were inspected for errors (reaction time < 100 ms) which were removed from the analysis. PVT and concurrent ocular measures were again analyzed in 1-min bins. Lapses (reaction time > 500 ms) were identified and the number of lapses per minute was determined for each minute.

Statistical Analysis for Validation of Cutoff Values

Data from the PVT were compared to the corresponding time matched ocular variables in 1-min bins. The sensitivity and specificity of each ocular variable was calculated for detecting different frequencies of lapses (1 to ≥ 5) in any minute.

RESULTS

Thirty-nine participants were recruited with data utilized from the 33 participants who completed the protocol (29 male, mean age 41.4 [± 12.9], mean BMI 27.5 [± 5.3], and a median ESS score of 5 [IQR 4-8]). The mean hours of sleep prior to the baseline testing was 6.5 (± 1.0) and prior to the sleep restricted testing was 4.0 (± 0.1) (confirmed by actigraphy).

OSLER Misses and Ocular Variables: ROC Curve Analysis

A summary of outcome measures for the OSLER and PVT is presented in **Table 1** and ocular variables during these tasks in **Table 2** for baseline, SR-AM, SR-PM, and an overall compilation of all data.

ROC area under the curve (AUC) and 95% confidence intervals (CIs) clustered by participant are presented in **Table 3** for analysis of ocular variables compared to ≥ 4 consecutive missed signals on OSLER (equates to 12 sec) and to ≥ 4 overall missed signals on the OSLER by minute.

The variables measuring blink duration (IED and BTD) had good discriminatory ability for detecting frequent drowsiness-related missed signals, as did the JDS (**Table 3**, **Figures 1** and **2**). The ratio of the amplitude to velocity of eyelid movement during eyelid closure (+AVR) was also an accurate discriminator and was better than the ratio during re-opening at the end of the blink (–AVR). The measures of proportion of time with eyes closed had moderate to poor discriminatory ability (%TEC and %LC).

Logistic regression models fitted for ocular variables with the highest discrimination for detection of drowsiness-related

errors (IED, BTD, %TEC, +AVR) (**Figure 3**) displayed good discriminatory ability in detecting drowsiness via 4 consecutive (ROC AUC = 0.76) and 4 overall (ROC AUC = 0.824) missed signals in one minute. However, the ROC AUC for these models was found to be lower than for the use of the IED variable alone (**Table 3**), and further analysis using these models was not undertaken.

Ocular Variable Cutoff Values for Optimal Sensitivity and Specificity

Cutoff values were chosen from the ROC curves for ocular variables that were moderate to good discriminators for detecting 4 consecutive missed signals and 4 overall missed signals on the OSLER per minute (**Tables 4** and **5**). Three cutoff values were selected for the variables of IED, BTD, %TEC, +AVR,

Table 1—Summary of OSLER measures (sleep latency, misses) and PVT measures (reaction time, lapses)

	OSLER			
	Sleep latency		Misses	
	M ± SD (min)	Mdn (IQR) (min)	M ± SD	Mdn (IQR)
Baseline (n = 17)	37.5 ± 5.1	40 (40-40)	0.05 ± 0.22	0 (0-0)
SR-AM (n = 27)	32.5 ± 10.5	40 (23-40)	0.10 ± 0.30	0 (0-0)
SR-PM (n = 27)	30.5 ± 11.2	39 (23-40)	0.14 ± 0.35	0 (0-0)
Overall* (n = 71)	32.3 ± 10.7	40 (25-40)	0.12 ± 0.32	0 (0-0)

	PVT			
	Reaction Time		Lapses	
	M ± SD (ms)	Mdn (IQR) (ms)	M ± SD	Mdn (IQR)
Baseline (n = 21)	244.4 ± 41.7	236.8 (215.7-265.1)	0.28 ± 0.60	0 (0-0)
SR-AM (n = 26)	234.4 ± 45.9	228.8 (211.3-249.1)	0.22 ± 0.54	0 (0-0)
SR-PM (n = 27)	286.1 ± 203.3	231.9 (208.0-280.9)	0.63 ± 1.29	0 (0-1)
Overall* (n = 75)	263.7 ± 156.1	231.3 (210.2-265.1)	0.45 ± 1.04	0 (0-0)

*Overall analysis used clustering to account for multiple measures per participant. M, mean; Mdn, median; SR-AM, sleep restriction: morning; SR-PM, sleep restriction: afternoon.

Table 2—Summary of ocular variables during OSLER and PVT tasks

	Ocular measures – OSLER				Ocular measures – PVT			
	Baseline (n = 17)	SR-AM (n = 27)	SR-PM (n = 27)	Overall* (n = 71)	Baseline (n = 21)	SR-AM (n = 26)	SR-PM (n = 27)	Overall* (n = 75)
IED								
M ± SD	0.22 ± 0.36	0.34 ± 0.59	0.46 ± 1.12	0.39 ± 0.88	0.14 ± 0.06	0.16 ± 0.37	0.18 ± 0.16	0.17 ± 0.26
Mdn (IQR)	0.13 (0.10-0.20)	0.17 (0.12-0.28)	0.17 (0.10-0.37)	0.17 (0.10-0.30)	0.13 (0.09-0.16)	0.12 (0.09-0.15)	0.14 (0.10-0.18)	0.13 (0.10-0.16)
BTD								
M ± SD	0.48 ± 0.53	0.63 ± 1.44	1.10 ± 7.22	0.84 ± 5.13	0.39 ± 0.15	0.38 ± 0.43	0.43 ± 0.24	0.41 ± 0.33
Mdn (IQR)	0.36 (0.28-0.49)	0.45 (0.32-0.61)	0.40 (0.32-0.68)	0.42 (0.32-0.61)	0.34 (0.29-0.46)	0.34 (0.27-0.40)	0.36 (0.29-0.46)	0.35 (0.29-0.44)
% TEC								
M ± SD	3.12 ± 7.74	3.94 ± 10.10	6.67 ± 19.29	5.18 ± 15.20	1.35 ± 2.83	1.44 ± 6.69	2.26 ± 4.98	1.89 ± 5.64
Mdn (IQR)	0.39 (0.05-1.87)	0.62 (0.07-2.40)	0.91 (0.15-4.92)	0.70 (0.09-3.20)	0.08 (0.02-1.12)	0.19 (0.05-0.97)	0.56 (0.09-1.67)	0.35 (0.06-1.44)
+AVR								
M ± SD	1.5 ± 0.4	1.64 ± 0.60	1.68 ± 0.74	1.64 ± 0.66	1.33 ± 0.35	1.23 ± 0.23	1.38 ± 0.42	1.33 ± 0.36
Mdn (IQR)	1.4 (1.2-1.6)	1.45 (1.26-1.93)	1.44 (1.23-1.89)	1.43 (1.24-1.86)	1.20 (1.10-1.42)	1.19 (1.08 -1.34)	1.24 (1.08 -1.49)	1.21 (1.08-1.43)
JDS								
M ± SD	4.1 ± 1.7	4.9 ± 1.9	4.7 ± 2.2	4.7 ± 2.0	3.3 ± 1.7	3.5 ± 1.4	4.2 ± 1.5	3.9 ± 1.5
Mdn (IQR)	3.8 (2.8-5.1)	5.0 (3.6-6.4)	5.0 (3.2-6.5)	4.8 (3.3-6.4)	3.2 (1.8-4.8)	3.3 (2.4-4.4)	3.9 (3.0-5.5)	3.7 (2.7-5.1)

*Overall analysis used clustering to account for multiple measures per participant. IED, inter-event duration; BTD, blink total duration; %TEC, percent time with eyes closed; +AVR, positive amplitude-velocity ratio; JDS, Johns Drowisness Scale; M, mean; Mdn, median; SR-AM, sleep restriction: morning; SR-PM, sleep restriction: afternoon.

and JDS to demonstrate high sensitivity, high specificity, and intermediate cutoff options for detecting frequent drowsiness-related errors.

PVT Lapses and Ocular Variables

The sensitivity and specificity of detecting lapses at the selected cutoff values was determined (**Figure 4**). At cutoff

Table 3—ROC area under the curve analysis of missed signals on the OSLER and ocular variables

	Four or more consecutive missed signals by minute		Four or more missed signals overall in one minute	
	AUC	95% CI	AUC	95% CI
Blink duration variables				
IED	0.816	0.729-0.892	0.835	0.758-0.897
BTD	0.733	0.625-0.839	0.767	0.687-0.849
Eyelid closure variables				
% TEC	0.684	0.574-0.802	0.722	0.642-0.806
%LC	0.577	0.530-0.635*	0.648	0.573-0.737
AVR variables				
+AVR	0.743	0.647-0.832	0.760	0.686-0.826
–AVR	0.669	0.561-0.767	0.641	0.529-0.732
Other				
DOQ	0.652	0.545-0.735	0.582	0.477-0.671
JDS	0.744	0.615-0.850	0.770	0.686-0.851

*Unadjusted 95% CIs are presented for %LC (consecutive missed signals). The variable %LC includes a high incidence of 0 values resulting in tied values due to bootstrap sampling procedures. This has been corrected for ties in the variable %LC (overall missed signals) but is unable to be corrected for %LC (consecutive missed signals). 95% CIs for all other variables have been adjusted to account for repeated measures on the participant. IED, inter-event duration; BTD, blink total duration; %TEC, percent time with eyes closed; %LC, percent long closures; +AVR, positive amplitude-velocity ratio; –AVR, negative amplitude-velocity ratio; DOQ, duration of ocular quiescence; JDS, Johns Drowsiness Scale; AUC, area under the curve.

Figure 1—ROC curves of ocular variables for discrimination of consecutive missed signals using inter-event duration (IED), blink total duration (BTD), Johns Drowsiness Score (JDS), and percent long closures (%LC)

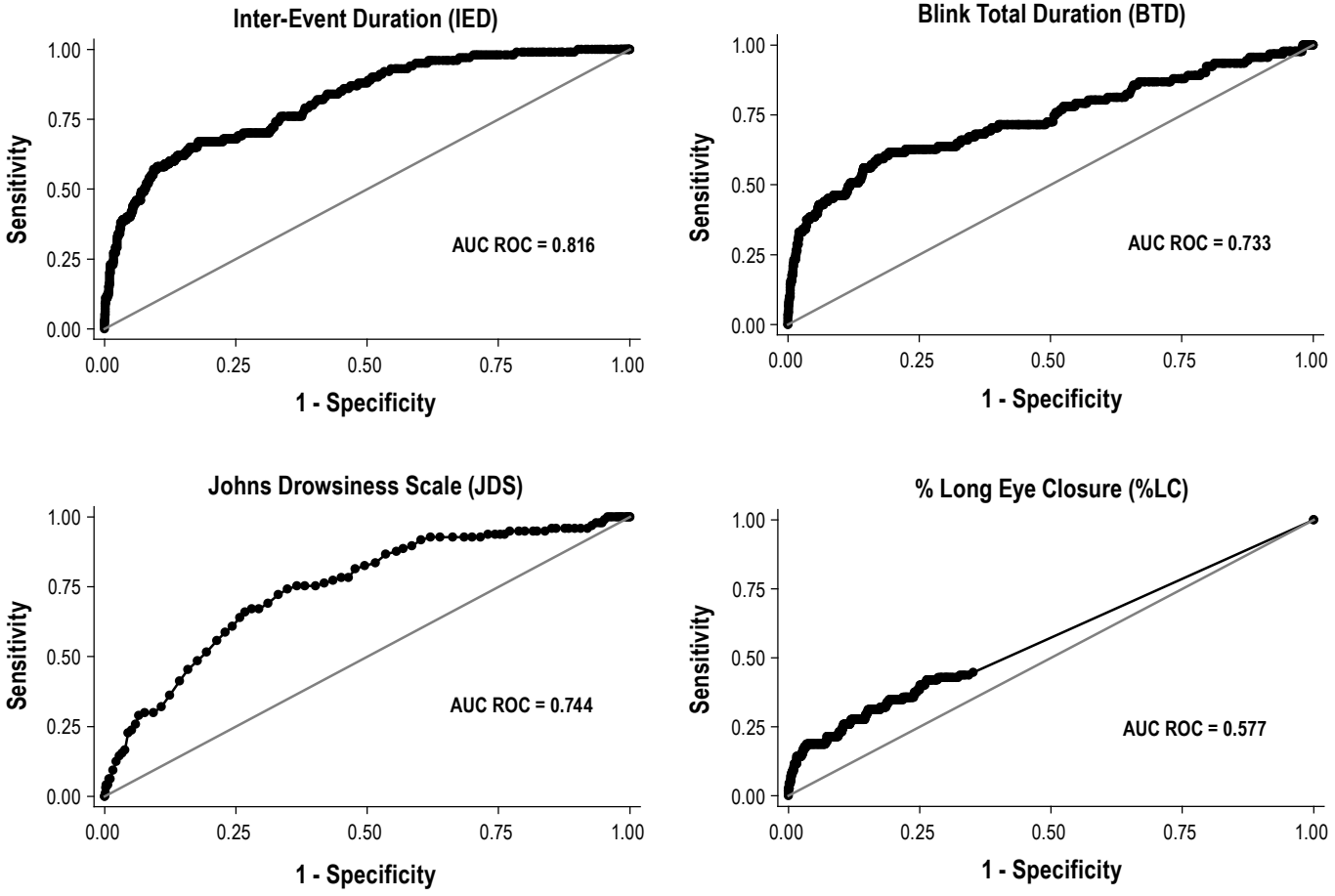


Figure 2—ROC curves of ocular variables for discrimination of overall missed signals using inter-event duration (IED), blink total duration (BTD), Johns Drowsiness Score, (JDS) and percent long closures (%LC)

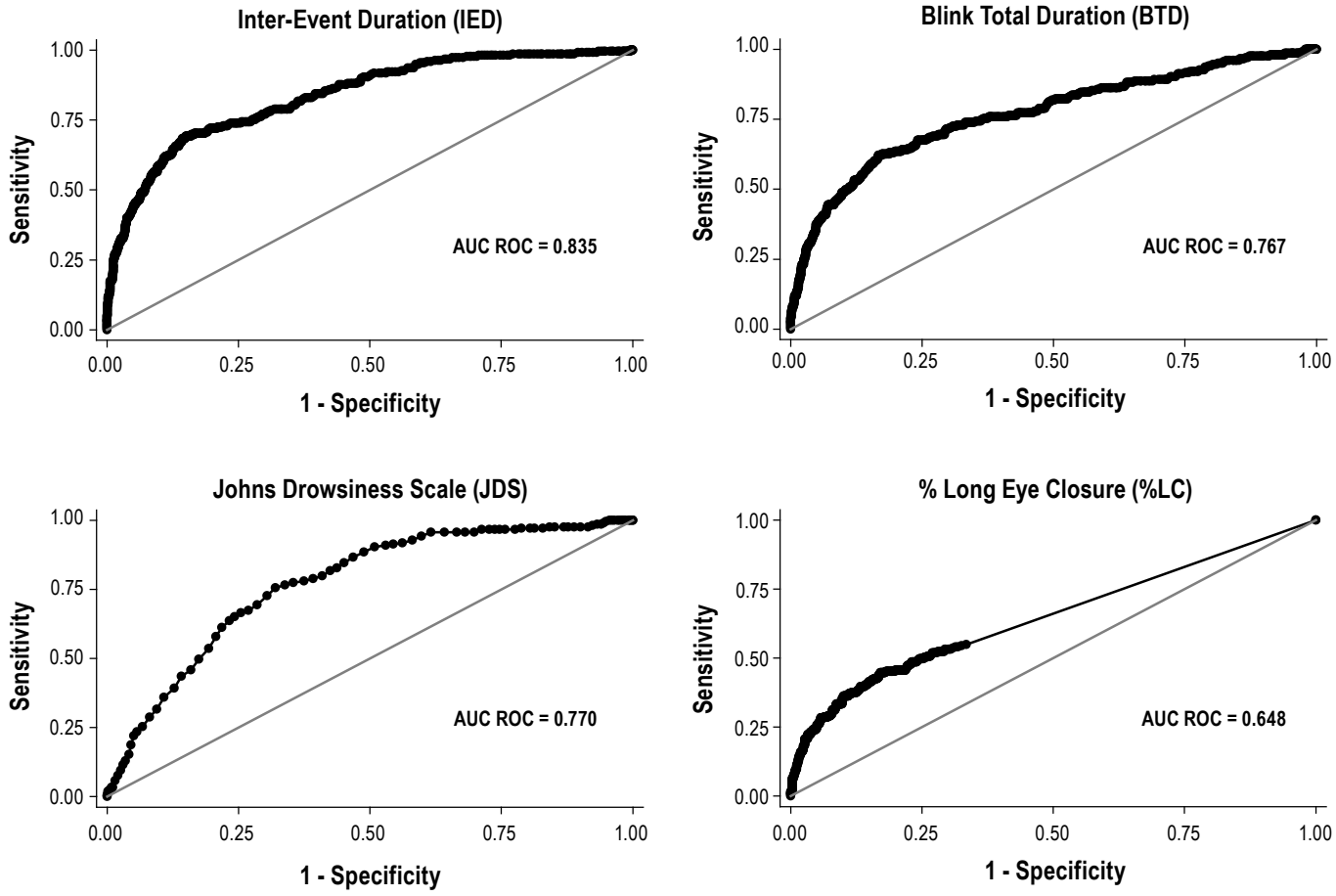
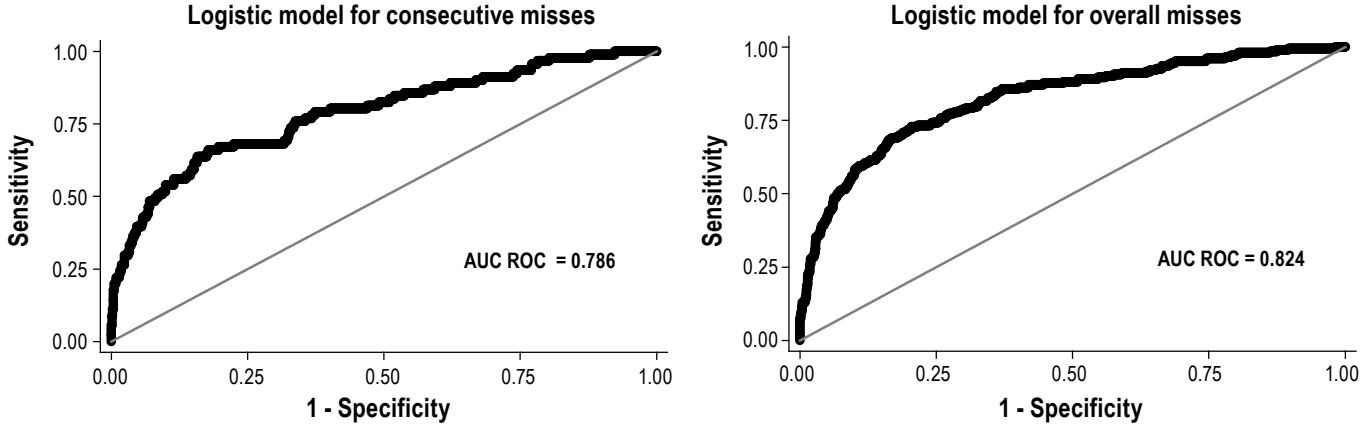


Figure 3—Logistic models using the ocular variables IED (inter-event duration), BTD (blink total duration), %TEC (percent time with eyes closed), and +AVR (positive amplitude-velocity ratio) for four consecutive and overall misses on the OSLER per minute



levels with an optimal balance between sensitivity and specificity (**Table 4**), identification of one lapse had low sensitivity, with increasing sensitivity with increasing number of lapses. IED, BTD, and JDS had good discrimination for ≥ 3 lapses in a minute on the PVT, with a sensitivity and specificity of 71% and 88% for the IED (100% and 86%, respectively, for ≥ 5 lapses). The JDS provided sensitivities of 77% and 100% for detecting 3 and ≥ 5 lapses in a minute, with specificities of 85% and 83%, respectively; +AVR was not sensitive to detecting lapses on the PVT. All 4 variables had high specificity in detecting any number of lapses on the PVT, and specificity did not decrease greatly with increasing number of lapses. Sensitivity of the identification of PVT lapses increased with lower cutoff levels (high sensitivity cutoff); however, this also resulted in a lowering of

Table 4—ROC cutoff values for each ocular variable for four or more consecutive missed signals by minute (including alternative high sensitivity and high specificity cutoff options)

	Optimal sensitivity/specificity			High sensitivity			High specificity		
	Cut-off	Sens (%)	Spec (%)	Cut- off	Sens (%)	Spec (%)	Cut- off	Sens (%)	Spec (%)
IED	0.203	76.00	66.36	0.160	88.00	51.58	0.634	51.00	92.09
BTD	0.462	71.43	59.79	0.412	72.53	50.58	0.804	50.55	87.68
% TEC	0.660	70.33	52.58	0.283	81.32	39.21	3.727	48.35	80.47
+AVR	1.515	76.84	60.21	1.407	80.00	50.13	2.010	50.53	84.08
JDS	5.4	75.26	63.40	4.6	82.47	50.47	6.5	51.55	80.61

IED, inter-event duration; BTD, blink total duration; %TEC, percent time with eyes closed; +AVR, positive amplitude-velocity ratio; JDS, Johns Drowsiness Scale; Sens, sensitivity; Spec, specificity.

Table 5—ROC cutoff values for each ocular variable for four overall missed signals by minute (including alternative high sensitivity and high specificity cutoff options)

	Optimal sensitivity/specificity			High sensitivity			High specificity		
	Cut-off	Sens (%)	Spec (%)	Cut- off	Sens (%)	Spec (%)	Cut- off	Sens (%)	Spec (%)
IED	0.209	77.06	70.45	0.152	90.37	51.13	0.542	51.38	92.57
BTD	0.468	75.37	63.20	0.398	81.77	50.28	0.753	50.74	88.37
% TEC	0.617	74.88	53.24	0.520	77.34	50.34	3.853	50.25	82.94
+AVR	1.567	78.30	66.34	1.402	83.02	50.88	1.990	50.47	85.31
JDS	5.5	75.60	67.97	4.5	88.52	51.16	6.4	53.59	80.63

IED, inter-event duration; BTD, blink total duration; %TEC, percent time with eyes closed; +AVR, positive amplitude-velocity ratio; JDS, Johns Drowsiness Scale; Sens, sensitivity; Spec, specificity.

specificity with the potential to erroneously classify individuals as unacceptably drowsy when they were not actually impaired. Raising cutoff levels to increase specificity (high specificity cutoff) resulted in extremely poor sensitivity in detecting PVT lapses for all variables other than JDS.

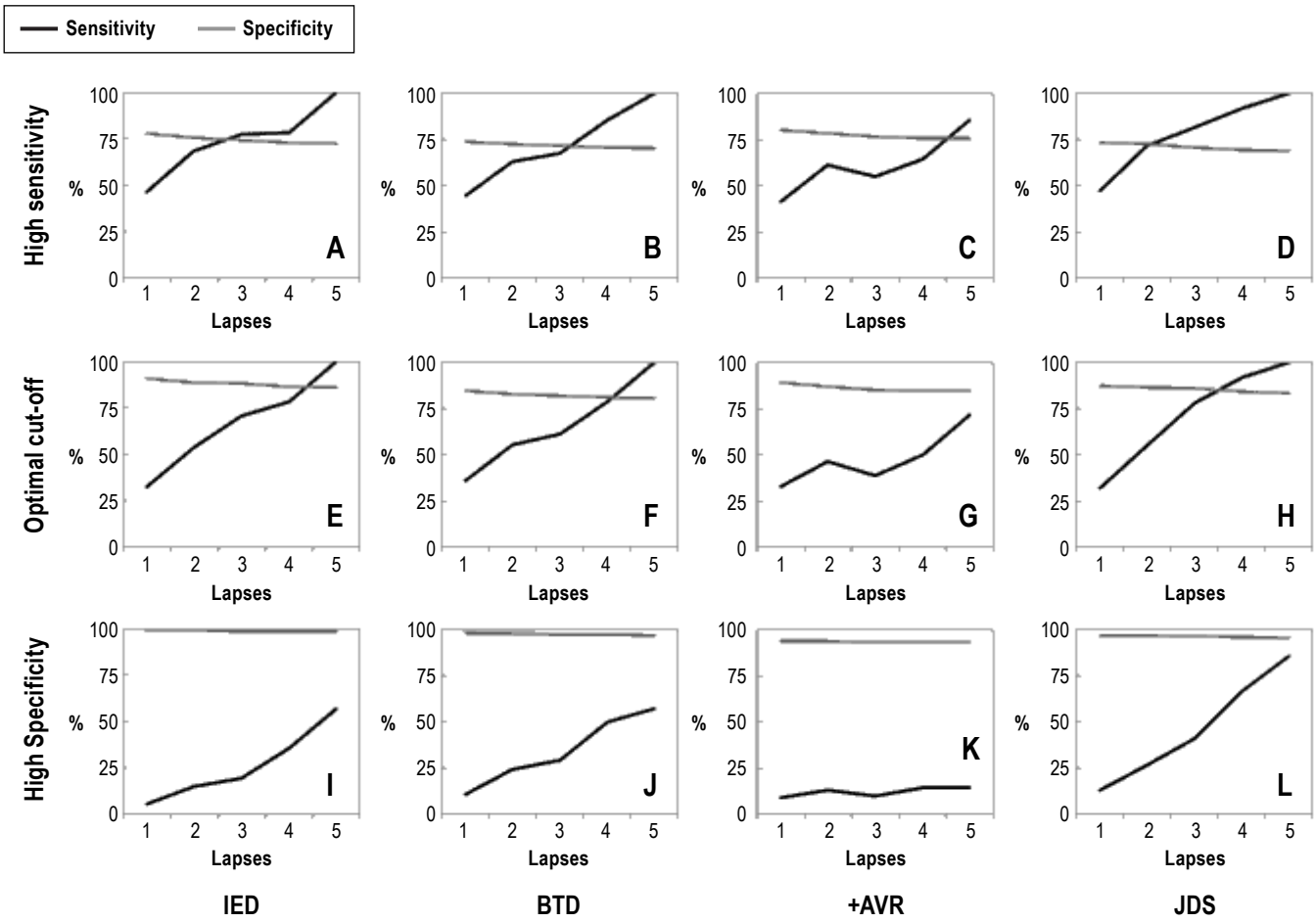
DISCUSSION

In this sleep restriction paradigm, measurement of eyelid movements accurately detected frequent episodes of failure to respond to visual signals during vigilance tasks. The average duration of episodes of eye closure (IED and BTD) provided the best discrimination from the primary measures, with the ratio of the amplitude to the velocity of eyelid movement during eyelid closure also providing good discrimination. These results support the use of ocular measures for identifying people who are impaired as a result of drowsiness.

Although blink duration and the proportion of time with eyes closed increase during circumstances designed to increase drowsiness, there is only limited work attempting to relate these physiological metrics to the behavioral changes that occur with restricted sleep. Reliable automated measures of blink duration suitable for on-road driving have previously not been available and manual determination of blink duration has been measured from EEG/electrooculography signals, video recordings, or stationary infrared recordings.[13,30,31] Mean blink duration and proportion of blinks of prolonged duration increase during monotonous tasks and are related to subjectively reported drowsiness.[13] Blink duration is also increased in untreated obstructive sleep apnea patients and reduces following treatment both in the laboratory and during on road driving.[32,33] The variable IED, a measure of eyelid closure duration between the points of maximum closing and re-opening velocity of the eyelid, has previously only recently been reported. This is a measure similar to blink duration, and proved to be most accurate at detecting drowsiness-related errors. Both blink duration and IED were recently shown to increase in duration after more than 24 hours of wakefulness and during the circadian nadir but the effect of milder sleep restriction, as used in our study, has not been described.[34]

The variable IED accurately identified drowsiness-related errors with an ROC curve area under the curve (AUC) of over 0.8 for detecting frequent missed signals during the OSLER task (AUC = 0.816, 95% CI 0.715-0.886, in the analysis of four consecutive missed signals and 0.834, 95% CI 0.757-0.896, in the analysis of four overall missed signals per minute). Total blink duration (BTD), the ratio of the amplitude to velocity of eyelid movements during eyelid closure (+AVR) and the Johns Drowsiness Score (JDS) were all moderately accurate at detecting frequent missed signals, with AUC of 0.733 to 0.767 for four or more consecutive missed signals and four or more overall missed signals per minute. The percentage of time with eyes closed (%TEC) had a moderate ability to identify frequent missed signals (AUC = 0.683 and 0.721), while other individual measures of eyelid movements (–AVR, %LC, and DOQ) had fair to poor ability to detect frequent drowsiness-related errors in both analyses of OSLER data. Several of these variables have recently been reported to have moderate ability to predict increased lapse frequency and slowing of reaction time, with AUC on ROC curves of between 0.62 to 0.74 for BTD, IED, %LC, AVR, and JDS.[34] These values are slightly lower than those identified in our study, perhaps due to the comparison utilizing ocular data collected prior to the vigilance

Figure 4—Sensitivity and specificity of the ocular variables correctly identifying increasing numbers of drowsiness-related PVT lapses (1–5) using OSLER-derived cutoff values at high sensitivity **(A)** IED, **(B)** BTD, **(C)** +AVR and **(D)** JDS; optimal cutoff values **(E)** IED, **(F)** BTD, **(G)** +AVR, **(H)** JDS; and high specificity cutoff values **(I)** IED, **(J)** BTD, **(K)** +AVR **(L)** JDS



Sensitivity = black, specificity = gray. IED, inter-event duration; BTD, blink total duration; %TEC, percent time with eyes closed; +AVR, positive amplitude-velocity ratio; JDS, Johns Drowsiness Scale.

testing rather than during the testing, and the comparison being over a longer time frame.

The percentage long eye closure and percentage of time with eyes closed variables are similar to PERCLOS (the proportion of time eyes are > 80% closed) that has been found to be good at discriminating between alert and drowsy states in some previous studies. The current analysis found that these variables had a moderate discriminatory power for detecting frequent drowsiness-related errors. %LC was poor at detecting sequential drowsiness-related errors, with a maximum sensitivity of 44.64% for detecting four consecutive missed signals on the OSLER, a strong indicator of brief sleep periods on EEG.[21] It proved to be more accurate at detecting four misses in total and %TEC, which includes all episodes of eye closure irrespective of duration, was more accurate than %LC. The proportion of time with eyes closed (PERCLOS) was moderately accurate at identifying behavioral lapses in drowsy participants in previous studies with improved accuracy when averaged over longer time periods.[5,7] It has been considered as a potential measure for real-time monitoring of drowsiness, although some laboratory studies have also found other drowsiness detection methods more reliable than PERCLOS.[35]

In this study we found measures of eyelid closure duration, such as IED and BTD, and the ratio of amplitude to velocity of eyelid movements, to be better predictors of behavioral lapses than the percentage of time with eyes closed, producing the highest sensitivities and specificities. For example, the optimal cutoff value for IED, derived from the OSLER data, achieve a sensitivity of 71% for detecting three behavioral lapses on the PVT task, increasing to 100% for detecting five lapses while maintaining good specificity (88% and 86% respectively). Increasing sensitivity at the expense of lowering specificity results in a higher false positive rate, but also a higher likelihood of identifying episodes of drowsiness-related impairment. In applied settings such as on-road driving, it may be deemed more important to have a low rate of false negatives (high sensitivity), despite a greater false positive rate, to reduce the risk of missing episodes of drowsiness.

Sensitivity of accurately detecting PVT lapses at the selected cutoff values for IED, BTD, +AVR, and JDS increased with

increasing number of lapses, without a large decrease in specificity. Single instances of PVT lapse could be due to non-drowsiness-related events such as distraction. This is supported by the low sensitivity but high specificity found when testing all ocular variables at this number of lapses. The frequency of PVT lapses has been studied under a variety of circumstances, which can help in considering clinically relevant levels of drowsiness. For example, participants averaged five lapses in ten minutes at a blood alcohol level of 0.05% in one study.[36] Lapse frequencies of 8 and 16 in 10 minutes have been described for 24 and 72 hours of wakefulness, respectively.[37] Hence, lapse frequencies of two and certainly three or more in a minute would indicate marked drowsiness. IED, BTD, and JDS were all able to detect this frequency of lapses with high sensitivity and specificity. The selected lower cutoff levels for these variables provided sensitivities of 67% to 81% for detecting three lapses in a minute, increasing to 100% for detecting five or more lapses. The specificities were reasonable at 70% to 74% for detecting three or more lapses at this high sensitivity cutoff. Higher cutoff levels resulted in a higher specificity but low sensitivity which would indicate a high rate of false negatives. Monitoring drowsiness in an applied setting would require an appropriate balance, however the fact some of these metrics can achieve a good sensitivity for detecting a moderate frequency of drowsiness-related errors increasing to a very high sensitivity with very frequent errors, while maintaining a low false positive rate suggests that they have the potential to be used for drowsiness monitoring.

AVR for eyelid movements is a measure of the velocity of eyelid movements relative to the amplitude of the upper eyelid movement. AVR increases with drowsiness,[17] particularly for the eyelid re-opening phase. These ratios have been reported to have low inter-subject variability, and hence potentially reduce the need for individual calibration.[17,38] In the current study the +AVR (eyelid closure phase), had an ROC area under the curve that was similar to the measures of eyelid closure duration (IED and BTD) and the JDS. However, it tended to have a lower sensitivity than the other measures for detecting behavioral lapses at a range of cutoffs while maintaining a good specificity.

The use of logistic models to fit a more accurate measure of drowsiness using several of the recorded ocular variables, although producing good discriminatory ability, did not improve detection of missed OSLER signals beyond the use of individual ocular variables. IED was found to have more accurate discrimination alone than a logistic model using a combination of ocular measures.

The protocol allowed for a mixture of rested and moderate sleep restriction conditions and demonstrated that several ocular variables have good ability to detect drowsiness-related errors on two psychophysiological tasks. A number of factors might alter these outcomes in different settings. While four hours sleep restriction is a relatively realistic level of sleep loss experienced in the real world, a greater level of sleep restriction and associated drowsiness might alter the discriminatory power of different ocular variables in detecting behavioral lapses. For example the speed of eyelid movements or blink duration might increase prior to appreciable increases in the percent of time with eyes closed. While we found that blink duration measures were better predictors of performance within this paradigm of mild sleep restriction; others with more severe sleep deprivation have found that percent of time with eyes closed is a better predictor.[34] There was some individual variability in ocular measures despite the same level of sleep restriction. This may be due to individual variability in responses to sleep loss and may also be due to baseline variability in psychophysiological measures, such as differences in ocular muscle responses. The applicability of results from the current study may be limited to the study tasks and the laboratory setting. In our study, participants were instructed to look directly ahead and sit still while performing the tasks. Although our findings suggest that ocular measurements may be a useful indicator of drowsiness during driving-related performance some caution needs to be exercised in extrapolating these results to other tasks and settings such as on-road driving where factors such as head and vehicle movement may affect eyelid measures. Of the vigilance tasks utilized in this study, the PVT, at 10-min duration, could be used in practical situations to assess driver drowsiness,[39] such as at a roadside testing stop. However, it is not suitable for the continuous monitoring of drowsiness as can be achieved with ocular measures.

To be functional in monitoring driver drowsiness, a device must be portable and able to acquire, process, and produce feedback to the driver before drowsiness reaches a level when deterioration in attention may lead to accidents. In this study, several ocular variables were reliable indicators of drowsiness-related deterioration in vigilance in the laboratory setting. Ocular variables which measured the duration of ocular events; IED (duration of eyelid events) and BTD (duration of blinks) were the most reliable in detecting drowsiness and lapses, with the ratio of velocity to amplitude of eyelid closure also a reliable indicator. These are promising measures for real-time drowsiness monitoring. Further research should evaluate their utility during a variety of tasks, in different environments (including on-road in vehicle validation) and under a variety of sleep restriction conditions.

ABBREVIATIONS

–AVR, negative amplitude-velocity ratio
+AVR, positive amplitude-velocity ratio
%LC, percent long closures
%TEC, percent time with eyes closed
AUC, area under the curve
BTD, blink total duration
CI, confidence interval
DOQ, duration of ocular quiescence
EEG, electroencephalography
ESS, Epworth Sleepiness Scale
IED, inter-event duration
JDS, Johns Drowsiness Scale
MWT, maintenance of wakefulness test
M, mean
Mdn, median
ODMS, Optalert Drowsiness Measurement System
OSLER, Oxford Sleep Resistance Test
PERCLOS, proportion of time eyes are more than 80% closed
PVT, psychomotor vigilance task
ROC, receiver operating characteristic
SR-AM, sleep restriction: morning
SR-PM, sleep restriction: afternoon

REFERENCES

1. Barger LK, Lockley SW, Rajaratnam SMW, Lanrigan CP. Neurobehavioral, health, and safety consequences associated with shift work in safety-sensitive professions. *Curr Neurol Neurosci Rep* 2009;9:155-64.
2. Connor J, Norton R, Ameratunga S, Robinson E, Wigmore B, Jackson R. Prevalence of driver sleepiness in a random population-based sample of car driving. *Sleep* 2001;24:688-94.
3. Johns MW, Chapman R, Crowley K, Tucker A. A new method for assessing the risks of drowsiness while driving. *Somnologie (Berl)* 2008;12:66-74.
4. Schleicher R, Galley N, Briest S, Galley L. Blinks and saccades as indicators of fatigue in sleepiness warnings: looking tired? *Ergonomics* 2008;51:982-1010.
5. Wierwille WW, Ellsworth LA. Evaluation of driver drowsiness by trained raters. *Accid Anal Prev* 1994;2:571-81.
6. Strijkstra AM, Beersma DG, Drayer B, Halbesma N, Daan S. Subjective sleepiness correlates negatively with global alpha (8–12 Hz) and positively with central frontal theta (4–8Hz) frequencies in the human resting awake electroencephalogram. *Neurosci Lett* 2003;340:17-20.
7. Dinges DF, Mallis M, Maislin G, Powell JW. *Evaluation of techniques for ocular measurements as an index of fatigue and the basis for alertness management.* U.S. Department of Transportation, National Highway Traffic Safety Administration, Contract No. DTNH22-93-D-07007, 1998.
8. Akerstedt T, Peters B, Anund A, Kecklund G. Impaired alertness and performance driving home from the night shift: a driving simulator study. *J Sleep Res* 2005;14:17-20.
9. Barbato G, De Padova V, Paolillo AR, Russo E, Ficca G. Increased spontaneous eye blink rate following prolonged wakefulness. *Physiol Behav* 2007;90:151-4.
10. Morris TL, Miller JC. Electrooculographic and performance indices of fatigue during simulated flight. *Biol Psychol* 1996;42:343-60.
11. Tucker A, Johns M. The duration of eyelid movements during blinks: Changes with drowsiness. *Sleep* 2005;28(Abstract Supplement):A122.
12. Sirevaag EJ, Stern JA. Ocular measures of fatigue and cognitive factors. In: Backs RW, Boucsein W, eds. *Engineering psychophysiology: issues and applications*. Hillsdale, NJ: Lawrence Erlbaum Associates, 2000:269-86.
13. Caffier PP, Erdmann U, Ullsperger P. Experimental evaluation of eye-blink parameters as a drowsiness measure. *Eur J Appl Physiol* 2003;89:319-25.
14. Johns M. A new method for measuring daytime sleepiness: The Epworth sleepiness scale. *Sleep* 1991;14:540-5.
15. Maislin G, Pack AI, Kribbs NB, et al. A survey screen for prediction of apnea. *Sleep* 1995;18:158-66.
16. 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.
17. Johns M, Tucker A, Chapman R, Crowley K, Michael N. Monitoring eye and eyelid movements by infrared reflectance oculography to measure drowsiness in drivers. *Somnologie (Berl)* 2007;11:234-42.
18. Johns M, Tucker A. The amplitude-velocity ratios of eyelid movements during blinks: Changes with drowsiness. *Sleep* 2005;28(Abstract Supplement):A122.
19. Bennett LS, Stradling JR, Davies RJ. A behavioural test to assess daytime sleepiness in obstructive sleep apnoea. *J Sleep Res* 1997;6:142-5.
20. Banks S, Catcheside P, Lack LC, Grunstein RR, McEvoy RD. The Maintenance of Wakefulness Test and driving simulator performance. *Sleep* 2007;28:1381-5.
21. 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.
22. Priest B, Brichard C, Aubert G, Liistro G, Rodenstein DO. Microsleep during a simplified maintenance of wakefulness test. A validation study of the OSLER test. *Am J Respir Crit Care Med* 2001;163:1619-25.
23. Janes H, Longton G, Pepe M. Accommodating covariates in ROC analysis. *Stata J* 2009;9:17-39.
24. Pepe MS, Longton G, Janes H. Estimation and comparison of receiver operating characteristic curves. *Stata J* 2009;9:1-16.
25. Dinges DF, Powell JW. Microcomputer analyses of performance on a portable, simple visual RT task during sustained operations. *Behav Res Methods Instrum Comput* 1985;17:652-5.
26. Goel N, Rao H, Durmer JS, Dinges DF. Neurocognitive consequences of sleep deprivation. *Semin Neurol* 2009;29:320-39.
27. Banks S, Dinges DF. Behavioral and physiological consequences of sleep restriction. *J Clin Sleep Med* 2007;3:519-28.
28. Dorrian J, Rogers NL, Dinges D. Psychomotor vigilance performance: Neurocognitive assay sensitive to sleep loss. In: Kushida CA, ed. *Sleep deprivation: clinical issues, pharmacology and sleep loss effects.* New York: Marcel Dekker, 2005:39-70.
29. Basner M, Dinges DF. Maximizing sensitivity of the psychomotor vigilance test (PVT) to sleep loss. *Sleep* 2011;34:581-91.
30. Ingre M, Akerstedt T, Peters B, Anund A, Kecklund G. Subjective sleepiness, simulated driving performance and blink duration: Examining individual differences. *J Sleep Res* 2006;15:47-53.
31. Papadelis C, Chen Z, Kourtidou-Papadeli C, et al. Monitoring sleepiness with on-board electrophysiological recordings for preventing sleep-deprived traffic accidents. *Clin Neurophysiol* 2007;118:1906-22.
32. Caffier PP, Erdmann U, Ullsperger P. The spontaneous eye-blink as sleepiness indicator in patients with obstructive sleep apnoea syndrome - A pilot study. *Sleep Med* 2005;6:155-62.
33. Hakkanen H, Summala H, Partinen M, Tihonen M, Silvo J. Blink duration as an indicator of driver sleepiness in professional bus drivers. *Sleep* 1999;22:798-802.
34. Anderson C, Chang AM, Sullivan JP, Ronda JM, Czeisler CA. Assessment of drowsiness based on ocular parameters detected by infra-red reflectance oculography. *J Clin Sleep Med* 2013;9:907-20.
35. Sommer D, Golz M. Evaluation of PERCLOS based current fatigue monitoring technologies. *Conf Proc IEEE Eng Med Biol Soc* 2012;2010:4456-9.
36. Howard ME, Jackson ML, Kennedy GA, Swann P, Barnes M, Pierce RJ. The interactive effects of extended wakefulness and low-dose alcohol on simulated driving and vigilance. *Sleep* 2007;30:1334-40.
37. Van Dongen HP, Maislin G. The cumulative cost of additional wakefulness: dose-response effects on neurobehavioral functions and sleep physiology from chronic sleep restriction and total sleep deprivation. *Sleep* 2003;26:117-26.
38. Johns MW. The amplitude-velocity ratio of blinks: A new method for monitoring drowsiness. *Sleep* 2003;26(Abstract Supplement):A51-2.
39. Jackson ML, Croft RJ, Kennedy GA, Owens K, Howard ME. Cognitive components of simulated driving performance: Sleep loss effects and predictors. *Accid Anal Prev* 2013;50:438-44.

SUBMISSION & CORRESPONDENCE INFORMATION

Submitted for publication January, 2013
Submitted in final revised form September, 2013
Accepted for publication September, 2013
Address correspondence to: Dr. Vanessa Wilkinson, Institute for Breathing & Sleep, Austin Health, PO Box 5555, Heidelberg, Victoria, Australia 3084; Tel: (613) 94965390; Fax: (613) 9496 5124; Email: vanessa.wilkinson@austin.org.au

DISCLOSURE STATEMENT

This was not an industry supported study. This project was supported by VicRoads, the state road and traffic authority in the state of Victoria, Australia. Dr. Howard has received research support from ResMed Foundation, Prevention Express, and Mining CRC. Part of Dr. Barnes salary is paid by involvement in sponsored clinical trial for Apnex Medical. Dr. Rajaratnam has served as a consultant (through service agreements with Monash University) to Vanda Pharmaceuticals, Philips Respironics, Edan-Safe, The Australian Workers' Union, and National Transport Commission, and has received research grants and/or unrestricted educational grants from Vanda Pharmaceuticals, Takeda Pharmaceuticals North America, Philips Lighting, Philips Respironics, Cephalon, and ResMed Foundation, and reimbursements for conference travel expenses from Vanda Pharmaceuticals. His institution has received equipment donations or other support from Optalert, Compumedics, and Tyco Healthcare. He has also served as an expert witness and/or consultant to shift work organizations. Dr Rajaratnam presently serves on the Board of Directors of the Australasian Sleep Association, and has previously served on the Board of Directors of the Sleep Health Foundation. The other authors have indicated no financial conflicts of interest.