



# The validity of the pupillographic sleepiness test at shorter task durations

Jessica E. Manousakis<sup>1</sup> · Jordan Maccora<sup>1</sup> · Clare Anderson<sup>1</sup>

Accepted: 4 November 2020 / Published online: 23 November 2020  
© The Psychonomic Society, Inc. 2020

## Abstract

The pupillographic sleepiness test (PST) is an accurate predictor of alertness failure and performance impairment across sleep deprivation. At 11 min in duration, the task is considered too long to be used in occupational or roadside settings. We therefore investigated the predictive capacity of the PST at seven shortened test durations. Eighteen healthy young adults (aged  $21.4 \pm 3.2$  years, 10 men) underwent 40 h of continuous wakefulness, completing an 11-min PST and a 10-min psychomotor vigilance task (PVT) every 2 h. Waking electroencephalography was recorded and scored for microsleeps during PVTs. The PST was divided into eight equal 82-s blocks and the predictive capacity of the pupillary unrest index (PUI) calculated at descending PST durations by systematically removing blocks. PUI increased significantly with time awake for all test durations ( $p < .0001$ ), with a similar amplitude of PUI observed for test durations of 5.5 min and longer. While all test durations accurately predicted PVT impairment (AUC: 0.72–0.86,  $p < .001$ ) and microsleep (AUC: 0.74–0.84,  $p < .0001$ ), 5.5 min was the shortest duration where accuracy remained high across level and type of impairment (AUC: 0.79–0.86). For the 5.5-min duration, the positive predictive value (PPV) and negative predictive value (NPV) were on average 50.1% and 89.4%, respectively, and were comparable to the full 11-min task (PPV: 49.2%; NPV: 91%). The PST can be shortened to 5.5 min without compromising accuracy in detecting performance impairment or physiological drowsiness. The PST is an ideal candidate for fitness-for-duty or fitness-to-drive testing, and future studies should examine its predictive capacity, at shorter durations, against operationally relevant outcomes.

**Keywords** Pupillary instability · Fitness for duty · Alertness · Sustained attention · Vigilance · Drowsiness · Sleepiness · Fatigue

## Introduction

Drowsiness is a leading cause of motor vehicle crashes (MVCs), alongside alcohol intoxication and speeding (Australian Transport Council, 2011; Connor et al., 2002). Worldwide estimates suggest approximately 20% of serious MVCs are caused by drowsiness (Blanco, Biever, Gallagher, & Dingus, 2006; Connor et al., 2002), with insufficient sleep resulting in a two- to fifteen-fold increase in crash risk (Tefft, 2018). Occupations that require individuals to work shifts, particularly night shifts, are of major concern due to the synergistic impact of sleep loss and circadian misalignment on alertness (Anderson et al., 2012). Accordingly, shift workers are at a heightened risk of drowsiness-related MVCs

(Crummy, Cameron, Swann, Kossmann, & Naughton, 2008; Lee et al., 2016), particularly professional drivers (Stevenson et al., 2014) or those working in healthcare settings (Anderson et al., 2018; Barger et al., 2005; Ftouni et al., 2012; Mulhall et al., 2019). A clear strategy for alleviating drowsiness-related MVCs is the development of tools to monitor and identify the drowsy state (Wolkow, Rajaratnam, Anderson, Howard, & Mansfield, 2019). While a number of devices have been developed and validated in recent years (Dawson, Searle, & Paterson, 2014), many of these are continuous monitoring devices providing a warning of the increasing drowsy state (Anderson, Chang, Sullivan, Ronda, & Czeisler, 2013; Anderson et al., 2018; Ftouni et al., 2013; Ftouni et al., 2012; Mulhall et al., 2019).

Single-time-point predictions of drowsiness-related crash risk, such as tests designed for fitness to drive or fitness for duty, would allow for early interventions. One such test is the pupillographic sleepiness test (PST), an 11-min test of pupil diameter fluctuations (Wilhelm, Wilhelm, Ludtk, Streicher, & Adler, 1998b). In total darkness, spontaneous oscillations of pupil diameter reflect changes in autonomic nervous system

✉ Clare Anderson  
clare.anderson@monash.edu

<sup>1</sup> Turner Institute for Brain and Mental Health, School of Psychological Sciences, Monash University, Clayton, Victoria 3800, Australia

activity (Oken, Salinsky, & Elsas, 2006). In the alert state, the pupil diameter is largely stable, becoming more unstable with increasing drowsiness due to fluctuations between sympathetic and parasympathetic control (Oken et al., 2006). This instability results in oscillations in pupil size, which can be captured with the pupillary unrest index (PUI) (Lüdtke, Wilhelm, Adler, Schaeffel, & Wilhelm, 1998). We and others have shown that the PUI is sensitive to increased time awake (Maccora, Manousakis, & Anderson, 2018; Regen, Dorn, & Danker-Hopfe, 2013; Wilhelm, Rühle, Widmaier, Lüdtke, & Wilhelm, 1998a), and is predictive of subsequent performance failure (attentional lapses) and increased physiological sleepiness (microsleeps and slow eye movements) (Maccora et al., 2018). Although the PUI metric within the PST is a reliable indicator of drowsiness, and therefore offers the potential for a fitness-for-duty/fitness-to-drive test, the duration of the task (11 min) is too long for a roadside test, and may not be feasible within demanding, time-constrained work environments such as healthcare settings, mining operations, and aircraft flight decks.

This was highlighted previously using the psychomotor vigilance task (PVT), a test of a sustained reaction time. While the full 10-min test duration is optimal for capturing sleepiness-induced performance failure, shortened test durations of as little as 3 min can adequately capture impairments following total sleep deprivation and sleep restriction (Basner & Dinges, 2011; Loh, Lamond, Dorrian, Roach, & Dawson, 2004; Roach, Dawson, & Lamond, 2006). This has subsequently led to the successful implementation of shortened test durations in healthcare settings to monitor alertness across successive night shifts (Ganesan et al., 2019), and the development of tablet-based apps to measure PVT performance in operational settings (e.g. Joggle Research, USA).

While it has been shown previously that the PST task does exhibit a time-on-task effect (like the PVT; Doran, Van Dongen, & Dinges, 2001), both the first and second half increase separately across a night of sleep (Wilhelm, Wilhelm, et al., 1998b). The extent to which the PST can be shortened to make it more operationally practical and feasible remains unknown. We therefore examined whether the PST, a valid and reliable automated test of drowsiness, can be reduced in duration while retaining high accuracy for predicting performance impairment and physiological sleepiness.

## Method

### Participants

Eighteen healthy young adults (10 men, 8 women) aged 18–29 years ( $M = 21.4 \pm 3.2$  years) took part in the study. Participants were non-smokers, consumed less than 300 mg of caffeine per day, and had a body mass index (BMI) within

the healthy range [18–35 ( $M = 23.45$ ,  $SD = 4.31$ )]. They reported a habitual sleep duration of 7–9 h; habitual sleep times between 22:00 and 01:00 and wake times between 06:00 and 09:00; did not nap more than once a week; had no history of medical, psychiatric, or sleep conditions; were free of medications or substances known to affect the central nervous system; did not have any visual impairments, eye conditions, or vision corrected by surgery or corrective lenses; and had not worked shift work or travelled across two time zones in the past 3 months. Female participants were not currently pregnant or using hormonal contraception.

The sample size was based on previous sleep deprivation protocols testing the utility of ocular metrics to predict impairment. Our previous work demonstrated significant results with strong effect sizes ( $Ns = 10–29$ ) (Anderson et al., 2013; Ftouni et al., 2013). With  $N = 18$ , we demonstrate 99% power to detect a medium effect size for linear mixed-model analysis. The present study was approved by Monash University Human Research Ethics Committee. Written informed consent was provided by participants prior to participation in the study, and participants were reimbursed for their time.

### Procedure

To ensure all participants were sleep-satiated prior to admission to the study, all participants maintained a self-selected fixed 9:15 sleep/wake schedule at home for one week. Compliance with this schedule was monitored with Actiwatch-2 activity monitors (Philips Respironics, USA) and time- and date-stamped call-ins at bed and wake time. During this week, participants were required to abstain from the consumption of caffeine, alcohol, nicotine, medication, and recreational drugs, as confirmed by urine toxicology upon arrival at the laboratory. Participants were admitted to a time-isolated, temperature- ( $21\text{ }^{\circ}\text{C} \pm 1\text{ }^{\circ}\text{C}$ ) and light-controlled private room with ensuite, underwent a baseline night of sleep (8 h) scheduled at their habitual bedtime, and woke to a 40-h period of extended wakefulness under modified constant routine conditions. Here, participants were seated upright under dim light conditions ( $< 3\text{ lux}$ ) and provided with hourly isocaloric snacks to equally distribute nutritional intake. Participants were provided 5 min of free movement every hour to stretch and use the bathroom. Alertness testing began 2 h post-wake, and was conducted bi-hourly thereafter. A night of recovery sleep in the laboratory was provided at the end of the 40 h.

### Alertness testing battery

Bi-hourly alertness testing consisted of the pupillographic sleepiness test (PST), followed by the psychomotor vigilance task (PVT), with continuous monitoring of brain activity with electroencephalography (EEG).

Spontaneous oscillations of the pupil were monitored using the F2D2 portable PST (AMTech Pupilknowlogy, Dossenheim, Germany) [see Peters, Grüner, Durst, Hütter, and Wilhelm (2014) for a full device description]. Under conditions of complete darkness, participants fixated on a small light-emitting diode housed within a set of portable goggles, while pupil diameter was measured using an infrared video pupilometer. Test duration was 11 min, pupil diameter was sampled every 40 ms (sampling rate = 25 Hz), and the pupillary unrest index (PUI) was automatically calculated in eight 82.5 s blocks. During the test, participants were asked to open their eyes if the eyes remained shut for more than 5 s—no other conversation was permitted. The test was terminated after a minimum of 5.5 min (four blocks) if the participant was consistently unable to open their eyes long enough for the pupil to be detected.

Vigilant attention was assessed using a 10-min visual PVT (Dinges & Powell, 1985). Participants were required to respond as quickly as possible to an ascending millisecond stopwatch appearing at random intervals (2–10 s). Following a response, the participant's reaction time was displayed on the screen before the next trial commenced. Failure to respond within 10,000 ms sounded an audible tone. Electroencephalogram (EEG: F3, F4, C3, C4, P3, P4) and electrooculogram (EOG) linked to the contralateral mastoids were recorded continuously using Compumedics Profusion 4 software (Compumedics Limited, Melbourne, Australia) and gold cup electrodes. Data was sampled at 512 Hz, with a low-pass filter at 30 Hz, a high-pass filter at 0.3 Hz, and a notch filter at 50 Hz.

## Data analysis

**Data cleaning** PVT responses < 100 ms were removed. Due to non-normal distribution of the data, PVT lapses (responses > 500 ms) were transformed using  $[(\sqrt{n}) + (\sqrt{n+1})]$  (Basner & Dinges, 2009). EEG data during the PVT was visually scored for microsleeps, defined as an intrusion of theta or delta activity > 3 s in the absence of eye blinks. For each PST, the pupillary unrest index (PUI; changes in pupil diameter in mm/min) was automatically calculated using AMTech F2D2 software in eight bins of 82.5 s (Peters et al., 2014). Briefly, PUI is the sum of absolute changes in pupil diameter: data was reduced by calculating the average for periods of 16 consecutive values, and the absolute values of the differences from one 16-value segment to the next are summarised for each 82.5-s bin. The PUI is the normalised value over a 1-min window, which is averaged for each complete 82.5-s bin [see Lüdtke et al. (1998) for full methods]. Every possible test duration was examined by systematically removing bins and manually calculating the mean PUI for the remaining bins (see Table 1). To be included in the calculation of the PUI, 50% of the available blocks had to contain valid data. Those tests that contained

**Table 1** PST test durations and minimum data blocks required for inclusion in data analysis

PST testing blocks	Task Duration	Min. blocks required
8 Blocks	11 min (11 min 0 s)	4
7 Blocks	9.6 min (9 min 38 s)	4
6 Blocks	8.3 min (8 min 15 s)	3
5 Blocks	6.9 min (6 min 53 s)	3
4 Blocks	5.5 min (5 min 30 s)	2
3 Blocks	4.1 min (4 min 8 s)	2
2 Blocks	2.8 min (2 min 45 s)	1
1 Block	1.4 min (1 min 23 s)	1

less than 50% valid data were marked as a 'failed PST'. For all analysis of PUI metrics, failed PSTs were regarded as missing data; however, for the final analysis (pass/fail PST), they were regarded as a 'failed' PST due to inability to complete the test.

**Time course of changes in alertness** For each PST test duration, PUI data was log-transformed, and data from the first 16 h was averaged to form a baseline measurement of rested wake performance (Anderson et al., 2013; Ftouni et al., 2013). Subsequent time points (hours 18–38) were then compared to the baseline value using linear mixed-model analysis to account for inter-individual variability and missing data. Time spent awake was modelled as a fixed factor, and participant was modelled as a random factor. A compound symmetry covariance type was used for all models, as this provided the lowest Schwarz Bayesian criterion (BIC) (Schwarz, 1978). To compare the different duration tests, a linear mixed model was run, with test duration and time spent awake modelled as fixed factors, and participant modelled as a random factor. Here, the main effect of test duration and the interaction with time spent awake were examined. An auto-regressive covariance structure was used. Post hoc pairwise comparisons were conducted within each model as required. A false discovery rate adjustment was applied to control for type I error (Benjamini & Hochberg, 1995). FDR adjusted  $p$  values ( $p_{adj}$ ) are provided for all post hoc tests conducted. All statistical analyses were conducted in SPSS version 24 software (IBM, Armonk, NY).

**Receiver operating characteristic (ROC) analysis** ROC analysis was conducted using SigmaPlot version 13 (ROC Curves Module; Systat Software, Inc., San Jose, CA) to assess the accuracy of the PUI score at different PST durations in predicting performance impairment, defined here as the number of PVT lapses, and physiological sleepiness, defined as the number of microsleeps during the PVT. Three threshold increases from 'baseline' were defined in order to assess mild–severe impairment: 25%, 50%, and 75%. Thresholds were

calculated according to previous work (Anderson et al., 2013; Chua et al., 2012), and a given time point was classified as either ‘alert’ or ‘drowsy’, depending on whether the impairment criteria (performance or physiological) fell below or above the threshold (Chua et al., 2012). Here, sensitivity is defined as the percentage of ‘drowsy’ time points that were correctly assigned a high PUI score (i.e., the percentage of times the test correctly detected a drowsy individual based on PVT lapses and microsleeps), and specificity is the percentage of times the test correctly detected an alert individual. Positive predictive value (PPV) refers to the percentage of high PUI scores that were genuinely drowsy points (i.e., the percentage of positive PSTs that were also considered impaired on PVT or EEG), and negative predictive value (NPV) refers to the percentage of low PUI scores that were genuinely alert time points. Optimal cut-off values were determined using two criteria: (1) balanced specificity and sensitivity using Youden’s J index (Youden, 1950), and (2) minimum specificity of 85% to reduce the number of false positives, which is important for roadside testing and in line with recommendations for roadside drug testing (Verstraete, 2005). In this analysis, only PSTs with > 50% data were included in the analysis.

**Chi-square analysis of PST pass/fail** Finally, to examine the impact of failing to complete the PST due to ocular occlusion and falling asleep, each PST was dichotomised as a pass/fail. First, we utilised the cut-off scores developed in the ROC curve analysis to predict moderate impairment (performance and physiological) whilst maximising specificity (> 85%), and a ‘failed’ PST was defined as all PSTs where the PUI was above the cut-off score, or there was < 50% valid data due to eye closure. A ‘passed’ PST was any complete PST with a PUI below the cut-off score. Pearson chi-square analysis was conducted to examine the predictive capacity of each PST duration to predict moderate impairment, and sensitivity, specificity, PPV, NPV, and odds ratios were calculated. Second, adjusted cut-off scores were proposed to re-establish maximised specificity (> 85%), and adjusted values reported.

## Results

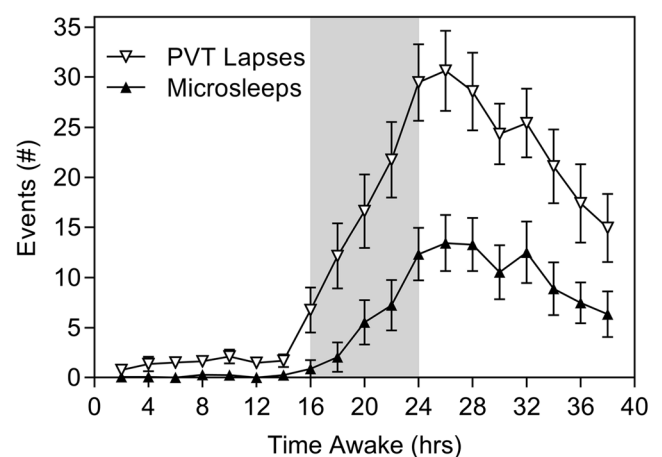
Data were obtained from 18 participants who each completed the 40 h extended wake protocol. Data from one participant was excluded due to binocular miosis in total darkness (pupil diameter < 3 mm; in addition to other data abnormalities). In total, 323 ‘alertness testing’ sessions (PVT + PST) were completed (17 participants x 19 test sessions). Of these, 322 (99.7%) PVT test sessions were included and 322 PST sessions were included (one PST session was lost due to an inability to calibrate the device due to excessive sleepiness;

0.3% data loss). Of the 322 PST sessions, a total of 32 tests were terminated early due to interference from sleepiness-related ocular occlusions, and two tests were terminated early due to the device no longer tracking the pupil—potentially due to poor calibration. EEG was recorded from 321 (99.4%) PVT sessions, with only two excluded due to technical difficulties (0.6% data loss). As per Maccora et al. (2018), PVT lapses [ $F(11,176) = 15.6, p < .0001$ ] and number of microsleeps [ $F(16,174) = 6.69, p < .0001$ ] both showed significant increases from the ‘baseline’ well-rested day. Consistent with PUI, both outcomes showed peak impairment at 26 h post-wake (see Fig. 1).

## Impact of test duration on time course of PUI

Table 2 shows mean PUI and number of data points included at each time point for each test duration. As can be seen, shorter test durations resulted in less data loss, as the most impaired participants who fell asleep part way through the PST were included. PUI increased significantly across time awake for all test durations ( $p < .0001$ ; see Fig. 2). Post hoc comparisons revealed that PUI was higher for all time points (hours 18–38) relative to baseline (hours 2–16) for all test durations ( $p_{adj} < .01$ ) except for the one-block (1.4 min) test duration: here, relative to baseline, the one-block PST exhibited higher PUI values for all time points ( $P_{adj} < .05$ ) except for hour 34 ( $P_{adj} = 0.074$ ).

There was a main effect of test duration on PUI score [ $F(7,352.1) = 13.48, p < .0001$ ], such that compared to the full 11-min test, all test durations  $\leq 5.5$  min resulted in lower PUI scores ( $p_{adj} < .023$ ). Additionally, the 1.4-min duration resulted in lower PUI scores than all test durations ( $p_{adj} < .046$ ), the 2.8-min duration resulted in lower PUI scores than all test durations  $\geq 5.5$  min ( $p_{adj} < .023$ ), and the 4.1-min duration



**Fig. 1** Bi-hourly mean ( $\pm$  standard error) psychomotor vigilance task (PVT) lapses and microsleeps recorded during the 10-min PVT. Grey shaded area represents habitual sleep period

**Table 2** Mean (standard error) PUI score and number of valid tests per time point (bi-hourly) for each PST duration

	11 Min		9.6 Min		8.3 Min		6.9 Min		5.5 Min		4.1 Min		2.8 Min		1.4 Min	
	M (SE)	N	M (SE)	N	M (SE)	N	M (SE)	N	M (SE)	N	M (SE)	N	M (SE)	N	M (SE)	N
Hour 2	7.6 (0.8)	17	7.3 (0.8)	17	7.1 (0.8)	17	6.9 (0.8)	17	6.8 (0.8)	17	6.5 (0.9)	17	5.8 (0.8)	17	5.6 (0.9)	17
Hour 4	8.0 (0.9)	16	7.8 (0.9)	16	7.4 (0.9)	16	7.2 (0.9)	16	6.9 (0.9)	16	6.7 (0.9)	16	6.8 (1.0)	17	6.5 (1.0)	17
Hour 6	7.4 (1.0)	17	7.2 (1.0)	17	7.0 (1.0)	17	6.8 (1.0)	17	6.4 (0.9)	17	6.0 (0.9)	17	5.5 (0.7)	17	5.2 (0.7)	17
Hour 8	6.8 (0.8)	17	6.4 (0.8)	17	6.1 (0.7)	17	5.6 (0.7)	17	5.2 (0.6)	17	4.8 (0.6)	17	4.4 (0.5)	17	4.2 (0.4)	17
Hour 10	7.8 (0.7)	17	7.5 (0.7)	17	7.1 (0.7)	17	6.7 (0.7)	17	6.4 (0.6)	17	5.7 (0.5)	17	5.4 (0.5)	17	5.3 (0.5)	17
Hour 12	6.2 (0.7)	17	6.1 (0.7)	17	5.7 (0.6)	17	5.4 (0.6)	17	5.2 (0.6)	17	4.9 (0.6)	17	4.8 (0.7)	17	5.1 (1.0)	17
Hour 14	6.9 (1.0)	16	6.8 (1.0)	16	6.5 (1.0)	16	5.9 (0.9)	16	5.9 (0.9)	17	5.9 (0.9)	17	5.9 (0.9)	17	6.0 (0.8)	17
Hour 16	7.7 (0.8)	17	7.4 (0.8)	17	7.2 (0.8)	17	7.0 (0.8)	17	6.6 (0.8)	17	6.5 (0.8)	17	6.3 (0.8)	17	6.0 (0.9)	17
Hour 18	9.8 (1.1)	15	9.5 (1.1)	15	9.2 (1.0)	16	8.9 (1.0)	16	8.7 (1.0)	16	8.3 (1.0)	16	8.2 (1.1)	16	7.8 (1.2)	16
Hour 20	10.7 (1.0)	17	10.6 (1.1)	16	10.2 (1.0)	17	9.9 (1.0)	17	9.7 (1.0)	17	9.4 (1.1)	17	9.1 (1.0)	17	8.7 (1.1)	17
Hour 22	13.1 (1.2)	17	12.9 (1.1)	17	12.5 (1.2)	17	12.3 (1.2)	17	12.1 (1.2)	17	11.7 (1.1)	17	11.0 (1.0)	17	9.6 (0.9)	17
Hour 24	15.3 (0.8)	13	15.1 (0.8)	13	14.6 (0.7)	14	14.1 (0.7)	14	13.5 (0.7)	15	12.9 (0.9)	15	12.2 (1.2)	17	10.6 (1.1)	16
Hour 26	15.6 (1.1)	13	15.4 (1.3)	12	15.4 (1.2)	14	15.2 (1.3)	14	14.8 (1.4)	14	14.0 (1.4)	14	12.5 (1.1)	15	11.4 (1.3)	15
Hour 28	14.6 (1.7)	12	13.8 (1.7)	13	15.1 (1.9)	15	14.7 (1.9)	15	13.7 (1.7)	17	12.3 (1.5)	17	11.9 (1.5)	17	10.0 (1.4)	16
Hour 30	13.7 (0.9)	15	13.4 (0.9)	15	13.1 (0.8)	16	12.3 (0.7)	16	11.8 (0.9)	17	10.1 (1.0)	16	9.2 (1.0)	17	8.3 (1.1)	17
Hour 32	13.7 (0.9)	17	13.3 (0.9)	17	12.9 (1.0)	17	12.5 (1.1)	17	11.8 (1.1)	17	10.6 (1.1)	17	9.6 (1.2)	17	8.6 (1.1)	17
Hour 34	12.4 (0.9)	17	11.8 (0.9)	17	11.4 (0.9)	17	11.0 (1.0)	17	10.1 (1.0)	17	8.7 (1.0)	17	8.4 (1.1)	17	7.3 (1.1)	17
Hour 36	12.2 (0.9)	17	11.9 (1.0)	17	11.6 (1.0)	17	10.8 (1.0)	17	10.4 (1.0)	17	9.8 (1.1)	17	8.8 (1.0)	17	7.6 (0.9)	17
Hour 38	11.1 (1.0)	16	10.8 (1.0)	16	10.3 (0.9)	17	9.9 (0.9)	17	9.4 (0.9)	17	9.1 (0.8)	17	8.4 (1.0)	17	7.9 (1.1)	17
F Statistic	11.47*		10.87*		11.98*		10.56*		9.92*		7.79*		6.42*		4.03*	

Note: M = mean; SE = standard error. Maximum number of tests per time point is 17. Hours 2–16 represent ‘baseline’, or normal waking day. Hours 18–38 represent sleep deprivation. \*  $p < .00005$

resulted in lower PUI scores than all test durations  $\geq 6.9$  min ( $p_{adj} < .030$ ). There was no interaction between test duration and time since wake [ $F(77,933.5) = 0.21, P = 1.00$ ].

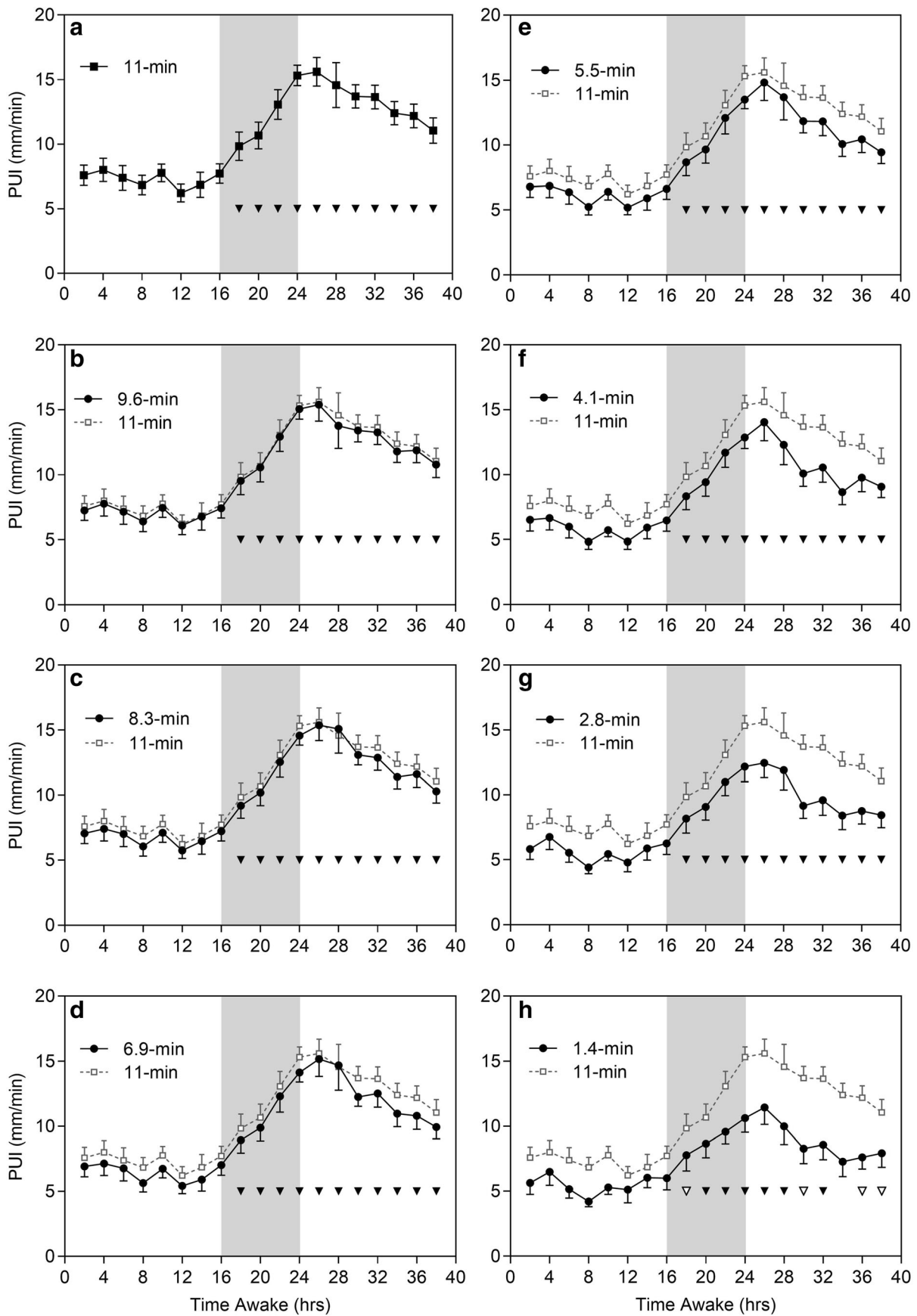
## Impact of test duration on predictive capacity of the PUI

### Ability to predict performance impairment

On average, mild performance impairment was classified as  $10.27 \pm 1.05$  lapses (43% of tests classified as impaired), moderate impairment was classified as  $19.03 \pm 1.83$  lapses (35% of tests classified as impaired), and severe impairment was classified as  $27.78 \pm 2.64$  lapses (22% of tests classified as impaired). At the group level, these were comparable to 16 h, 20–22 h, and 24 h of wakefulness, respectively. All PST durations successfully predicted performance impairment at all impairment levels ( $p < .0001$ ; see Table 3). For mild impairment, the area under the curve (AUC) remained constant at 0.86 for all blocks until the 4.1-min duration task. For moderate and severe impairment, the AUC was relatively stable until shorter test durations (4.1 min and 2.8 min for moderate and severe impairment, respectively). Associated

recommended cut-off scores (using Youden’s J) appeared to be at approximately 9 mm/min, dropping below 8 for shorter test durations (less than 4.1 min when identifying mild impairment, but less than 2.8 min when identifying moderate performance impairment). When maximising specificity (to  $> 85\%$ ), cut-off scores for identifying impairment were higher, at approximately 11 mm/min (see Table 3), and were relatively stable for all test durations longer than 4.1 min (where they dropped below 10 mm/min for identifying mild impairment). Figure 3 shows ROC curves and scatter plots for four different test durations. As seen in panels d–f, optimal cut-off points remained relatively stable until the 2.8 min test duration, although there was a high number of false positives across all test durations (i.e., a PUI score above the cut-off indicating drowsiness, but no performance failure in the alert column).

When specificity was maximised, sensitivity was highest for the 11-min test duration (66.7%) for predicting mild impairment, with the highest values also reported for PPV and NPV (75.9% and 79%, respectively). The difference across all test durations 4.1 min and longer, however, was minimal, with PPV ranging from 75.2% to 75.9% and NPV ranging from 75% to 79%. This was also observed for identifying moderate and severe impairment (see Table 3 and Fig. 3).



◀ **Fig. 2** Bi-hourly mean ( $\pm$  standard error) pupillary unrest index (PUI; mm/min) for each PST duration. Panel **a** shows the standard 11-min test duration, and panels **b–h** show each shortened duration, with the 11-min duration as a comparison. Black triangles represent a significant increase from baseline (hours 2–16;  $p_{adj} < .01$ ); white triangles represent a significant increase from baseline ( $p_{adj} < .05$ ). Grey shaded area represents habitual sleep period

### Ability to predict physiological sleepiness

On average, mild physiological sleepiness was classified as  $4.87 \pm 0.75$  microsleeps (29% of tests classified as impaired), moderate impairment was classified as  $9.74 \pm 1.50$  microsleeps (20% of tests classified as impaired), and severe impairment was classified as  $14.60 \pm 2.24$  microsleeps (12% of tests classified as impaired). At the group level, these were equivalent to 16–18 h, 22–24 h, and 26–28 h of wakefulness, respectively. Similar to performance impairment, ROC curve analysis revealed that all PST durations significantly predicted physiological impairment at all impairment levels ( $p < .0001$ ; see Table 4). For mild impairment, AUC remained constant at  $\sim 0.80$  for all test durations 5.5 min and longer. Using Youden's J, recommended cut-off scores remained stable at  $\sim 10$  mm/min until the 4.1-min duration, where they consistently decreased for all levels of impairment. When maximising specificity (to  $> 85\%$ ), cut-off scores were higher. Table 4 presents the sensitivity, specificity, and positive and negative predictive values for all test durations, and Fig. 4 shows ROC curves and scatter plots for four different test durations. As seen in Table 4, sensitivity was highest for the 5.5-min test duration (59%) for predicting mild impairment. Accordingly, the PPV and NPV were also highest (at 61% and 84%, respectively), and were improved relative to the full 11-min duration test (55% and 83%, respectively). This was not observed, however, for moderate and severe impairment, where the 11-min test appeared to perform 'best', although, similar to that described for performance impairment (see above), the difference between all test durations 5.5–11 min was minimal for PPV (moderate impairment, range 47.2–47.9%; severe impairment, range 30.4–31%) and NPV (moderate impairment, range 89.7–92.2%; severe impairment, range 92.5–93.8%). Reduced accuracy across test durations 4.1 min and shorter was consistent across all levels of impairment (see Table 4 and Fig. 4).

### Ability to predict moderate impairment using a pass/fail criteria

Chi-square analysis of the dichotomised PSTs based on the 'maximised specificity' cut-off scores presented in Tables 3 and 4 showed that all PST durations significantly predicted moderate performance impairment and physiological impairment ( $p < .001$ ). As seen in Table 5, using these cut-off scores but including 'failed' PSTs, sensitivity increased slightly and

specificity decreased slightly, with similar odds ratios. Adjustment of cut-off scores to improve specificity back to at least 85% resulted in minor increases in cut-off scores (increases of less than 1.12) and minor changes to sensitivity and specificity.

## Discussion

This paper is the first to systematically investigate the sensitivity of shortened PST durations to sleep loss, as well as the associated predictive capacity to detect performance and physiological impairment at shorter test durations. Taking into consideration various levels of impairment (mild, moderate, and severe) and various types of impairment (behavioural and physiological), our data suggest that a shortened test duration of 5.5 min has the same level of accuracy (and in some cases better accuracy) than the full 11-min test duration, and is thus considered optimal. For operational settings where time is critical, the test may be employed at 4.1 min with relatively little compromise on accuracy, although we do not recommend test durations less than this unless impairment is severe.

The development of a short, objective, and predictive test of drowsiness is critical for early intervention, and optimal for fitness-for duty-tests and fitness-to-drive tests, including roadside testing. The PST is one of the few commercially available tests that allow for a single-time-point assessment of drowsiness, and has been previously shown to predict subsequent performance impairment and physiological sleepiness (Maccora et al., 2018). Demonstrating that the device and PUI metric accurately predict performance impairment and physiological sleepiness at shorter test durations makes this test more operationally practical for field, clinical, and roadside settings, either in experimental research studies or within an operational fatigue risk management strategy.

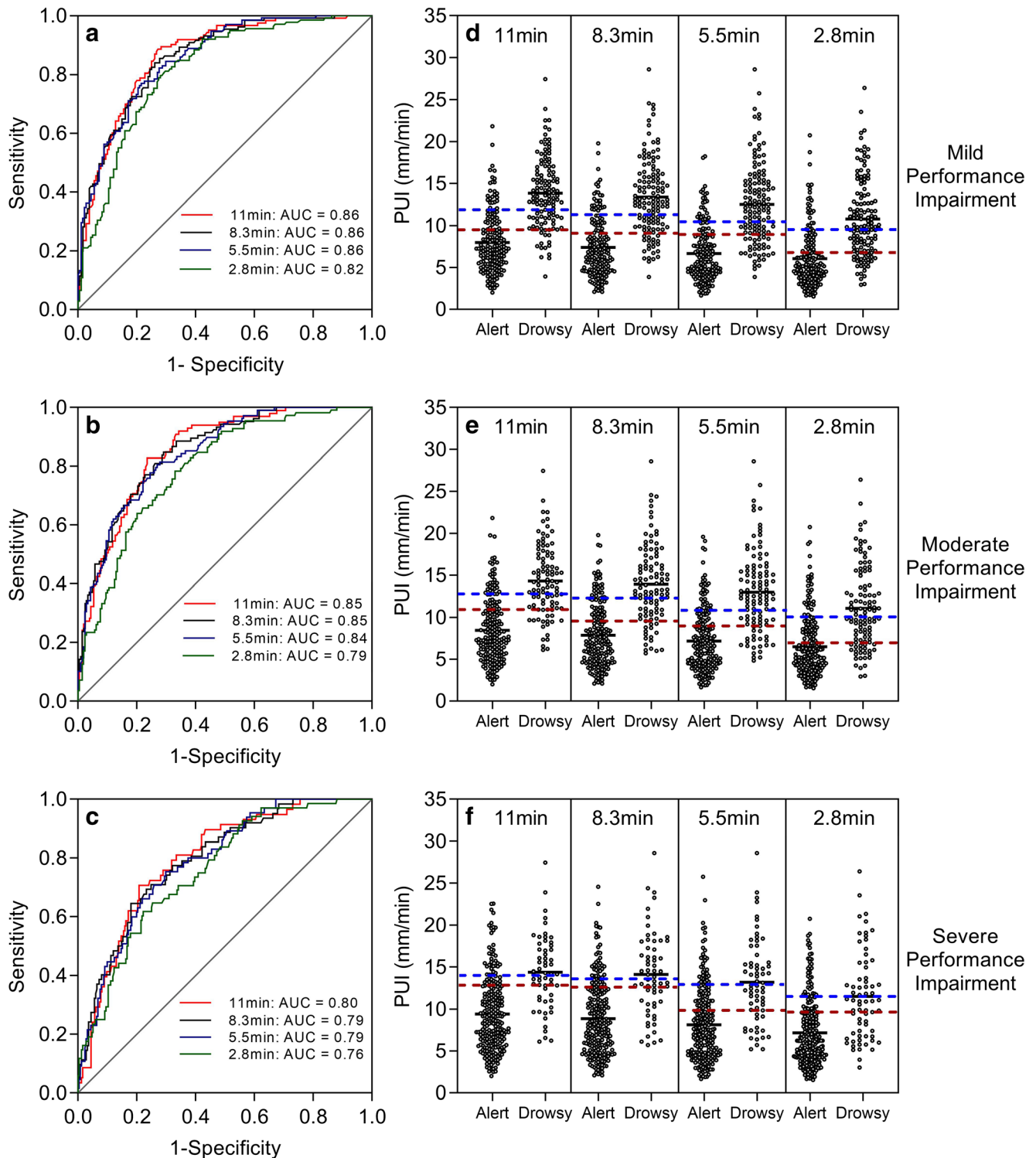
We observed clear time-on-task effects which were consistent with previous data showing a small time-on-task effect of the PST (Wilhelm, Wilhelm, et al., 1998b). Here, we showed a marked reduction in PUI scores for the three shortest PST durations (1.4 min to 4.1 min), as well as a reduction for the 5.5-min duration relative to the 11-min task. This suggests that longer test durations yield higher PUI scores, although no significant interaction with time awake was found. These results highlight the importance of developing individualised cut-offs for each test duration. As we exhibited a clear time-on-task effect of PUI, it is worth noting that a comparison of the predictive capacity of the first 1.4 min (as we have done) with the final 1.4 min of the full 11-min PST would very likely result in different predictive values and different PUI cut-off scores, as the final 1.4 min would likely yield higher mean

**Table 3** Receiver operating characteristic (ROC) curve results for mild, moderate, and severe performance impairment for each PST duration

Impairment	PST length	AUC	Youden's J					Maximised specificity (> 85%)						
			Cut-off score	Sens.	Spec.	PPV	NPV	Odds ratio	Cut-off score	Sens.	Spec.	PPV	NPV	Odds ratio
Mild	11-min	0.86 ± 0.02 [0.82–0.91]	9.49	88.62	72.78	68.99	90.34	20.81	11.81	66.67	85.56	75.93	78.97	11.85
	9.6-min	0.86 ± 0.02 [0.82–0.90]	8.96	88.62	72.07	68.55	90.21	20.09	11.82	64.23	85.47	75.24	77.66	10.57
	8.3-min	0.86 ± 0.02 [0.82–0.91]	9.17	83.97	73.89	69.87	86.36	14.69	11.24	64.89	85.00	75.89	76.88	10.47
	6.9-min	0.86 ± 0.02 [0.82–0.91]	8.14	85.50	71.67	68.71	87.16	14.91	11.00	61.07	85.00	74.77	75.00	8.89
	5.5-min	0.86 ± 0.02 [0.82–0.91]	8.96	77.04	78.45	72.73	82.08	12.22	10.48	62.96	85.08	75.89	75.49	9.70
	4.1-min	0.83 ± 0.02 [0.80–0.88]	7.90	78.36	73.48	68.63	82.10	10.03	9.35	65.67	85.08	76.52	77.00	10.91
	2.8-min	0.82 ± 0.02 [0.77–0.86]	6.78	79.71	71.98	68.32	82.39	10.09	9.45	57.25	85.07	74.53	72.43	7.69
	1.4-min	0.77 ± 0.03 [0.72–0.82]	5.19	85.29	57.14	59.59	83.87	7.67	9.80	43.38	85.16	68.60	66.81	4.40
	11-min	0.85 ± 0.02 [0.81–0.90]	10.87	82.83	76.47	63.08	90.17	15.68	12.81	62.63	85.29	67.03	82.46	9.56
Moderate	9.6-min	0.85 ± 0.02 [0.80–0.89]	8.96	90.91	66.01	56.60	93.71	19.42	12.58	63.64	85.22	68.13	82.78	10.27
	8.3-min	0.85 ± 0.02 [0.81–0.89]	9.55	84.76	70.87	59.46	90.12	13.38	12.25	64.76	85.44	69.39	82.63	10.78
	6.9-min	0.85 ± 0.02 [0.81–0.89]	9.49	81.90	72.33	60.14	88.69	11.83	11.94	60.65	85.44	68.09	81.11	9.16
	5.5-min	0.84 ± 0.02 [0.80–0.89]	8.96	80.56	73.08	60.84	87.86	11.24	10.95	65.74	85.10	69.31	82.71	10.80
	4.1-min	0.82 ± 0.02 [0.77–0.87]	9.35	69.44	80.68	65.22	83.50	9.49	10.46	59.26	85.02	67.37	80.00	8.26
	2.8-min	0.79 ± 0.03 [0.74–0.84]	6.98	78.38	66.99	55.48	85.37	7.27	10.07	51.35	85.17	64.77	76.72	6.06
	1.4-min	0.74 ± 0.03 [0.69–0.80]	8.27	57.80	80.86	61.17	78.60	5.79	10.11	43.12	85.65	61.04	74.27	4.52
	11-min	0.80 ± 0.03 [0.74–0.86]	12.81	70.69	79.18	43.96	91.94	8.95	14.08	53.45	85.31	46.27	88.56	6.67
	9.6-min	0.79 ± 0.03 [0.73–0.85]	12.58	68.97	78.28	42.86	91.39	7.96	13.93	50.00	85.25	44.62	87.76	5.78
Severe	8.3-min	0.79 ± 0.03 [0.74–0.85]	12.69	64.52	81.93	47.06	90.27	8.24	13.51	51.61	85.14	46.38	87.60	6.11
	6.9-min	0.79 ± 0.03 [0.74–0.85]	10.89	72.58	74.30	41.28	91.58	7.65	13.18	51.61	85.14	46.38	87.60	6.11
	5.5-min	0.79 ± 0.03 [0.73–0.85]	9.79	75.38	70.12	39.52	91.67	7.19	12.82	49.23	85.26	45.59	86.64	5.43
	4.1-min	0.78 ± 0.03 [0.72–0.84]	9.46	72.31	74.00	41.96	91.13	7.43	12.10	49.23	85.20	46.38	86.59	5.58
	2.8-min	0.76 ± 0.03 [0.70–0.82]	9.63	61.76	77.78	42.86	88.29	5.65	11.56	44.12	85.32	44.78	84.98	4.59
	1.4-min	0.72 ± 0.03 [0.65–0.79]	8.43	60.61	76.19	40.00	88.07	4.92	10.72	40.91	85.32	42.19	84.65	4.02

Note: AUC = area under the curve; Sens. = sensitivity; Spec. = specificity; PPV = positive predictive value; NPV = negative predictive value





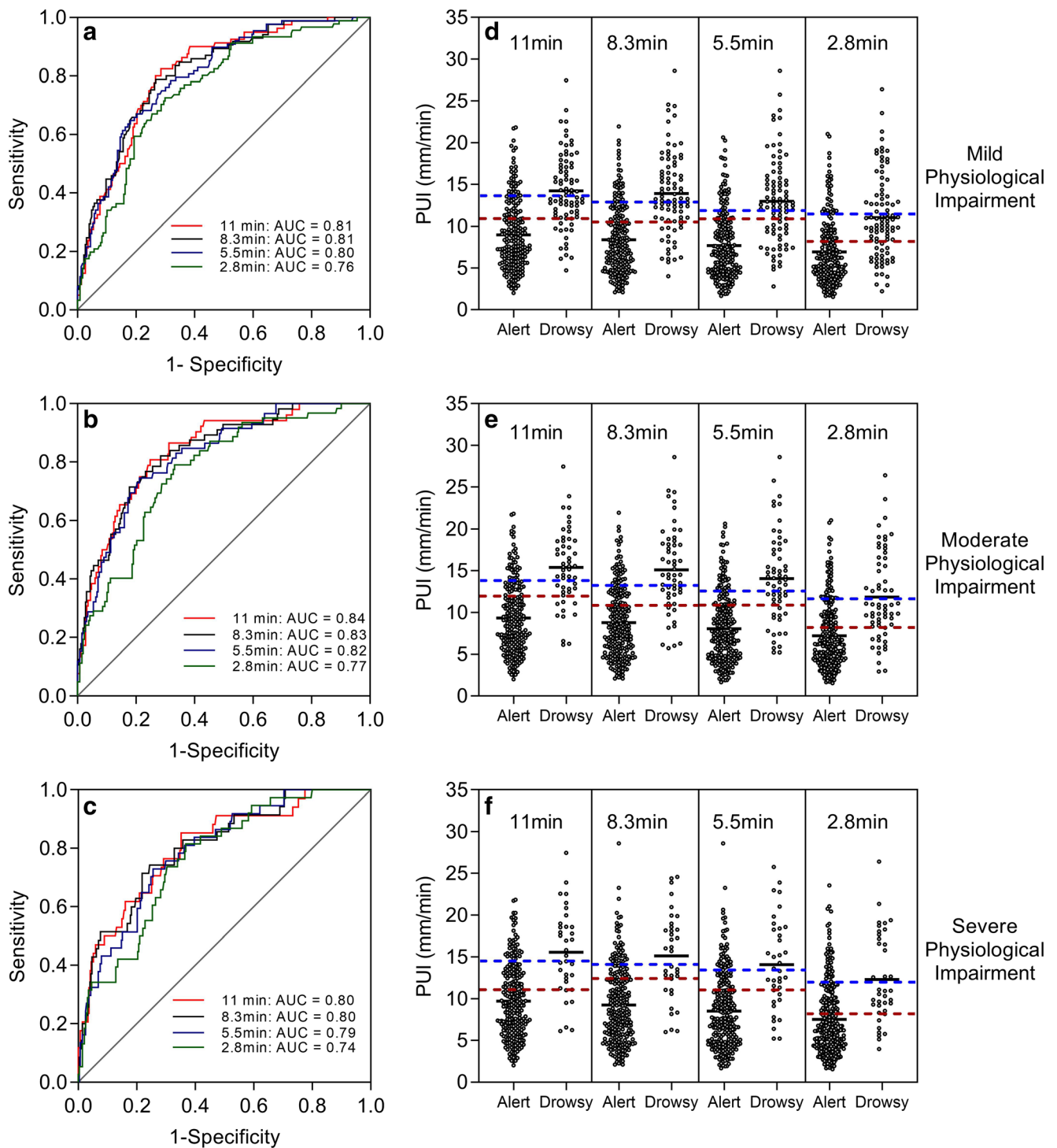
**Fig. 3** Receiver operating characteristic curve analysis for mild (upper panels), moderate (middle panels), and severe (lower panels) performance impairment. Panels a–c show ROC curves for four PST durations, and panels d–f show the mean PUI score for different test durations dichotomised as alert or drowsy based on PVT lapses. Solid bars represent mean PUI for each dichotomisation. Dashed horizontal line

represents optimal cut-off score for each test duration (red = Youden's J; blue = 85% specificity). Alert tests above the dashed line represent false positives (FP); alert tests below the line represent true negatives (TN); drowsy tests above the line represent true positives (TP); and drowsy tests below each line represent false negatives (FN)

**Table 4** Receiver operating characteristic (ROC) curve results for mild, moderate, and severe physiological sleepiness for each PST duration

Impairment	PST length	AUC	Youden's J					Maximised specificity (> 85%)						
			Cut-off Score	Sens.	Spec.	PPV	NPV	Odds ratio	Cut-off Score	Sens.	Spec.	PPV	NPV	Odds ratio
Mild	11 min	0.81 ± 0.03 [0.76–0.86]	10.87	82.50	71.49	51.16	91.86	11.82	13.61	50.00	85.07	54.79	82.46	5.70
	9.6 min	0.81 ± 0.03 [0.75–0.86]	10.34	81.25	70.00	49.23	91.12	9.96	13.21	50.00	85.00	54.79	82.38	5.67
	8.3 min	0.81 ± 0.03 [0.76–0.86]	10.56	78.82	73.21	52.76	90.11	10.17	12.96	54.12	85.27	58.23	83.04	6.83
	6.9 min	0.81 ± 0.03 [0.75–0.86]	8.31	85.88	62.50	46.50	92.11	10.14	12.51	57.65	85.27	59.76	84.14	7.88
	5.5 min	0.80 ± 0.03 [0.75–0.85]	10.84	67.05	79.65	56.19	86.12	7.96	11.94	59.09	85.40	61.18	84.28	8.45
	4.1 min	0.78 ± 0.03 [0.72–0.84]	9.43	70.45	77.33	54.46	87.00	8.00	11.42	53.41	85.33	58.75	82.40	6.67
	2.8 min	0.76 ± 0.03 [0.70–0.81]	8.16	72.53	70.04	48.87	86.41	6.08	11.56	36.26	85.02	49.25	76.89	3.23
	1.4 min	0.70 ± 0.03 [0.64–0.76]	8.27	57.30	77.09	49.51	82.16	4.52	10.60	38.20	85.46	50.75	77.91	3.63
	11 min	0.84 ± 0.03 [0.78–0.90]	11.97	80.77	75.10	40.38	94.92	12.67	13.76	65.38	85.14	47.89	92.17	10.82
Moderate	9.6 min	0.83 ± 0.03 [0.77–0.89]	11.82	78.85	74.60	39.42	94.39	10.95	13.75	57.69	85.08	44.78	90.56	7.78
	8.3 min	0.83 ± 0.03 [0.78–0.89]	10.97	82.14	71.54	38.98	94.76	11.56	13.28	60.71	85.38	47.89	90.76	9.02
	6.9 min	0.83 ± 0.03 [0.78–0.89]	11.00	76.79	74.70	40.19	93.56	9.77	13.06	58.93	85.38	47.14	90.38	8.38
	5.5 min	0.82 ± 0.03 [0.77–0.88]	11.02	74.58	78.82	44.33	93.06	10.67	12.71	57.63	85.10	47.22	89.67	7.77
	4.1 min	0.80 ± 0.03 [0.74–0.87]	9.52	79.66	75.20	42.20	94.09	11.62	12.05	57.63	85.04	47.89	89.63	7.94
	2.8 min	0.77 ± 0.03 [0.71–0.83]	8.16	79.03	66.80	36.09	92.93	7.43	11.72	40.32	85.16	39.68	85.49	3.88
	1.4 min	0.72 ± 0.04 [0.65–0.79]	8.27	63.33	74.61	36.89	89.67	5.08	10.82	38.33	85.16	37.70	85.49	3.57
	11 min	0.80 ± 0.04 [0.73–0.88]	11.02	85.29	64.79	22.95	97.19	10.31	14.47	55.88	85.02	31.03	93.80	6.81
	9.6 min	0.80 ± 0.04 [0.73–0.88]	11.01	81.82	66.29	23.08	96.72	8.85	14.48	51.52	85.39	30.36	93.44	6.21
Severe	8.3 min	0.80 ± 0.04 [0.73–0.88]	12.37	74.29	75.55	27.96	95.83	8.93	14.10	51.43	85.04	30.51	93.20	6.02
	6.9 min	0.80 ± 0.04 [0.73–0.88]	11.00	77.14	70.80	25.23	96.04	8.18	13.87	51.43	85.04	30.51	93.20	6.02
	5.5 min	0.79 ± 0.04 [0.72–0.87]	11.02	72.97	74.37	26.80	95.37	7.54	13.52	48.65	85.20	30.51	92.55	5.45
	4.1 min	0.79 ± 0.04 [0.72–0.87]	9.52	81.08	71.01	26.61	96.55	10.15	12.51	51.35	85.14	31.67	92.89	6.05
	2.8 min	0.77 ± 0.04 [0.69–0.84]	8.16	81.58	63.21	22.56	96.20	7.36	11.92	42.11	85.36	28.07	91.57	4.24
	1.4 min	0.74 ± 0.04 [0.66–0.82]	9.20	63.16	75.18	25.81	93.72	5.19	10.91	44.74	85.25	29.31	91.86	4.68

Note: AUC = area under the curve; Sens. = sensitivity; Spec. = specificity; PPV = positive predictive value; NPV = negative predictive value



**Fig. 4** Receiver operating characteristic curve analysis for mild–severe physiological sleepiness. Panels **a–c** show ROC curves for four PST durations, and panels **d–f** show the mean PUI score for individual tests dichotomised as alert or drowsy based on microsleeps. Solid bars represent mean PUI for each dichotomisation. Dashed horizontal line

represents optimal cut-off score for each test duration (red = Youden's J; blue = 85% specificity). Alert tests above the line represent false positives (FP); alert tests below the line represent true negatives (TN); drowsy tests above the line represent true positives (TP); and drowsy tests below each line represent false negatives (FN)

PUI scores. However, as this lacks operational utility, we have not explored this possibility within this manuscript.

Our study provides optimal cut-off scores and ROC curve results for all test durations (see Tables 3 and 4)

for any future study or operational setting that may wish to employ a shorter test duration. As all test durations significantly predicted performance and physiological impairment above chance (all AUCs above 0.70)—even the 1.4-

**Table 5** Chi-square results for performance failure and physiological impairment using a pass/fail criteria for the PST

Impairment	PST length	Original (max spec.) Cut-off score							Secondary (max spec.) Cut-off score						
		Chi-square	Cut-off	Sens.	Spec.	PPV	NPV	Odds ratio	Chi-square	Cut-Off	Sens.	Spec.	PPV	NPV	Odds ratio
Performance	11 min	78.94*	12.81	66.37	82.86	67.57	82.08	9.54	73.21*	13.43	60.18	85.71	69.39	80.00	9.07
	9.6 min	78.94*	12.58	67.26	82.86	67.86	82.46	9.93	68.47*	13.18	58.41	85.71	68.75	79.30	8.43
	8.3 min	86.71*	12.25	66.37	84.76	70.09	82.41	10.98	83.08*	12.56	63.72	85.71	70.59	81.45	10.54
	6.9 min	86.71*	11.94	63.72	83.81	67.92	81.11	9.09	78.57*	12.03	62.83	85.24	69.61	81.00	9.76
	5.5 min	84.71*	10.95	66.37	84.29	69.44	82.33	10.59	86.16*	11.01	65.49	85.24	70.48	82.11	10.96
	4.1 min	68.11*	10.46	61.06	83.81	66.99	80.00	8.12	70.82*	10.90	59.29	85.71	69.07	76.65	8.74
	2.8 min	49.63*	10.07	52.21	84.76	64.84	76.72	6.08	47.19*	10.25	50.44	85.24	64.77	76.17	5.88
	1.4 min	35.78*	10.11	45.13	85.24	62.20	74.27	4.75	40.21*	10.35	45.13	86.67	64.56	74.59	5.35
Physiological	11 min	74.56*	13.76	71.88	82.49	50.55	92.17	12.04	67.06*	14.46	64.06	85.21	51.90	90.50	10.27
	9.6 min	58.65*	13.75	65.63	82.10	47.73	90.56	8.76	68.13*	14.87	60.94	87.16	54.19	89.96	10.59
	8.3 min	65.95*	13.28	65.63	84.05	50.60	90.76	10.06	65.03*	13.82	62.50	85.60	51.95	90.16	9.91
	6.9 min	62.35*	13.06	64.06	84.05	50.00	90.38	9.39	68.38*	13.57	62.50	86.38	53.33	90.24	10.57
	5.5 min	56.86*	12.71	60.94	84.44	49.37	89.67	8.46	62.71*	13.10	59.38	86.77	52.78	89.56	9.59
	4.1 min	56.86*	12.05	60.94	84.44	49.37	89.67	8.46	56.38*	12.13	59.38	85.21	50.00	89.39	8.42
	2.8 min	22.89*	11.72	42.19	84.82	40.91	85.49	4.08	32.67*	12.56	42.19	88.33	47.37	85.98	5.52
	1.4 min	22.89*	10.82	42.19	84.82	40.91	85.49	4.08	25.80*	10.90	42.19	85.99	42.86	85.66	4.48

Note: \* $p < .001$ ; AUC = area under the curve; Sens. = sensitivity; Spec. = specificity; PPV = positive predictive value; NPV = negative predictive value

min test duration—the test may be adjusted relative to operational needs, and using our cut-off points for decision-making (i.e., employ countermeasure, proceed to rest break, etc.). As the cut-off for accurately detecting impairment changes with decreasing test duration, it is critical that impairment thresholds are modified accordingly, akin to the changing threshold for a PVT lapse at shorter PVT test durations (Basner, Mollicone, & Dinges, 2011). Consistent with the full-duration PST, shorter duration values resulted in better predictive capacity for performance impairment than physiological impairment, and reduced predictive capacity for severe levels of impairment compared to the mild and moderate levels of impairment, with a noticeable reduction in sensitivity. The severe impairment thresholds were equivalent to approximately 24–26 h of sleep deprivation at the group level, which represented peak levels of impairment due to the additive effect of homeostatic sleep pressure and the circadian nadir in alertness (Maccora et al., 2018). Therefore, the reduced sensitivity suggests that the PUI score increases prior to reaching peak impairment and is therefore better at predicting impairment that includes earlier, milder forms of drowsiness (i.e., extended wakefulness of 16–20 h).

Although ocular metrics are highly sensitive to detecting the drowsy state (Cori, Anderson, Shekari Soleimanloo, Jackson, & Howard, 2019), the PST does rely on the eye remaining open to capture an accurate recording of the pupil

and its changing diameter. As drowsiness increases, data loss on the PST increases due to the onset of microsleeps, possibly reducing its efficacy. A shortened test duration of 5.5 min therefore also results in a greater level of data retention. For instance, for the 5.5-min duration, only 2.2% (7/323) of tests were excluded from analysis, compared to 6.2% (20/323) for the full 11-min test duration. This allows for analysis of sleeper individuals who were physically unable to complete the full 11-min test without falling asleep (although operationally this would indicate a ‘failed’ test). It is worth noting that in the real world, an inability to complete the PST would typically be considered a ‘failed’ PST. Analysis of the PST using a pass/fail criterion showed that the predictive efficacy was mostly unaffected, with comparable levels of sensitivity and specificity for detecting moderate impairment. Therefore, we suggest that utilising a 50% data inclusion criteria, marking a PST as either ‘failed’ or ‘impaired’ results in equal predictive value, and could be used in operational or research settings.

The PVT is the most widely used tool for detecting drowsiness, largely due to its sensitivity and lack of learning effects (Balkin et al., 2004; Lim & Dinges, 2008). While considered a reliable candidate for predicting operator fatigue-related deficits, the 10-min duration was considered too long to be acceptable in operational settings, particularly when repeated administration was required (Basner et al., 2011; Dinges & Mallis, 1998). Interestingly, and somewhat consistent with our test duration outcomes, a 5-min task has been consistently

shown to yield similar levels of sensitivity to sleep loss as the standard 10-min PVT (Lamond, Dawson, & Roach, 2005; Lamond et al., 2008; Roach et al., 2006), as has a modified 3-min version (Basner et al., 2011), while test durations of less than 2 min were considered too brief (Loh et al., 2004; Roach et al., 2006).

Our analysis was designed to make the PST more operationally feasible and practical. When evaluating the validity of a drowsiness detection device for fitness for duty or fitness to drive, it is essential to examine not only the capacity to detect drowsiness (i.e., time awake, or during the circadian nadir), but also performance on a different task, and ideally a task that relates most closely to job performance, particularly when job performance represents a serious risk (i.e., driving, ICU monitoring, control room, etc.) (Gilliland & Schlegel, 1993). While we have shown a shortened 5-min PST to accurately detect time awake, performance impairment, and physiological indices of drowsiness, it remains unclear whether the PST accurately predicts real-world performance outcomes such as driving or other operationally relevant outcomes. The portable nature of the F2D2 device, the objective and physiological nature of data recording, and the flexible task duration make the PST an ideal candidate for fitness-for-duty/roadside testing of drowsiness. Future field-based studies are needed, however, to demonstrate its utility in predicting operationally relevant drowsiness-related risk.

**Acknowledgements** The authors thank the study participants, and the technical staff at the Monash Sleep and Circadian Medicine Laboratory—in particular, Rebecca Khan and Mulia Marzuki for assistance in collecting data, and Andrew Laing for sleep scoring. This study was funded by VicRoads. The content is solely the responsibility of the authors and does not necessarily represent the official views of VicRoads.

**Open practices statement** The data used in this manuscript is publicly available at [https://bridges.monash.edu/articles/Shortened\\_PST\\_Prediction\\_-\\_Laboratory\\_Data/12577679](https://bridges.monash.edu/articles/Shortened_PST_Prediction_-_Laboratory_Data/12577679). DOI: 10.26180/5ef6c4f217d2e (Manousakis, Maccora, & Anderson, 2020)

**Author contribution** CA designed the protocol, and JEM and JM collected the data. CA and JEM developed the data analysis plan, JEM and JM completed data cleaning and analysis, and JEM and CA interpreted the data. JEM drafted the manuscript and CA and JM provided edits.

## Compliance with ethical standards

**Conflict of interest** All authors report no conflicts of interest with data as described in this manuscript. In the interest of full disclosure, CA reports receiving a research award/prize from Sanofi-Aventis; research support from VicRoads; speaker fees from Brown Medical School/Rhode Island Hospital, Ausmed, Healthed, and Teva Pharmaceuticals; and contract research funding from Pacific Brands and Rio Tinto Coal Australia; she has served as consultant to the Rail, Bus and Tram Union, National Transport Commission, and the Transport Accident Commission on issues related to fatigue in transportation; and has served as an expert witness on sleep-related motor vehicle crashes. Dr Anderson leads a team within the Cooperative Research Centre for Alertness, Safety and Productivity.

## References

- Anderson, C., Chang, A. M., Sullivan, J. P., Ronda, J. M., & Czeisler, C. A. (2013). Assessment of drowsiness based on ocular parameters detected by infrared reflectance oculography. *Journal of Clinical Sleep Medicine*, 9(9), 907–920. <https://doi.org/10.5664/jcsm.2992>
- Anderson, C., Ftouni, S., Ronda, J. M., Rajaratnam, S. M. W., Czeisler, C. A., & Lockley, S. W. (2018). Self-reported drowsiness and safety outcomes while driving after an extended duration work shift in trainee physicians. *Sleep*, 41(2). <https://doi.org/10.1093/sleep/zsx195>
- Anderson, C., Sullivan, J. P., Flynn-Evans, E. E., Cade, B. E., Czeisler, C. A., & Lockley, S. W. (2012). Deterioration of neurobehavioral performance in resident physicians during repeated exposure to extended duration work shifts. *Sleep*, 35(8), 1137–1146. <https://doi.org/10.5665/sleep.2004>
- Australian Transport Council. (2011). *National Road Safety Strategy, 2011–2020*. Retrieved from
- Balkin, T. J., Bliese, P. D., Belenky, G., Sing, H., Thorne, D. R., Thomas, M., ... Wesensten, N. J. (2004). Comparative utility of instruments for monitoring sleepiness-related performance decrements in the operational environment. *Journal of Sleep Research*, 13(3), 219–227. <https://doi.org/10.1111/j.1365-2869.2004.00407.x>
- Barger, L. K., Cade, B. E., Ayas, N. T., Cronin, J. W., Rosner, B., Speizer, F. E., ... Safety, G. (2005). Extended work shifts and the risk of motor vehicle crashes among interns. *N Engl J Med*, 352(2), 125–134. <https://doi.org/10.1056/NEJMoa041401>
- Basner, M., & Dinges, D. F. (2009). Dubious bargain: trading sleep for Leno and Letterman. *Sleep*, 32(6), 747–752. <https://doi.org/10.1093/sleep/32.6.747>
- Basner, M., & Dinges, D. F. (2011). Maximizing sensitivity of the psychomotor vigilance test (PVT) to sleep loss. *Sleep*, 34(5), 581–591.
- Basner, M., Mollicone, D., & Dinges, D. F. (2011). Validity and sensitivity of a brief psychomotor vigilance test (PVT-B) to total and partial sleep deprivation. *Acta Astronautica*, 69(11–12), 949–959. <https://doi.org/10.1016/j.actaastro.2011.07.015>
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1), 289–300.
- Blanco, M., Biever, W. J., Gallagher, J. P., & Dingus, T. A. (2006). The impact of secondary task cognitive processing demand on driving performance. *Accident Analysis and Prevention*, 38(5), 895–906. <https://doi.org/10.1016/j.aap.2006.02.015>
- Chua, E. C. P., Tan, W. Q., Yeo, S. C., Lau, P., Lee, I., Mien, I. H., ... Gooley, J. J. (2012). Heart rate variability can be used to estimate sleepiness-related decrements in psychomotor vigilance during total sleep deprivation. *Sleep*, 35(3), 325–334. <https://doi.org/10.5665/sleep.1688>
- Connor, J., Norton, R., Ameratunga, S., Robinson, E., Civil, I., Dunn, R., ... Jackson, R. (2002). Driver sleepiness and risk of serious injury to car occupants: Population based case control study. *British Medical Journal*, 324(7346), 1125–1128.
- Cori, J. M., Anderson, C., Shekari Soleimanloo, S., Jackson, M. L., & Howard, M. E. (2019). Narrative review: Do spontaneous eye blink parameters provide a useful assessment of state drowsiness? *Sleep Medicine Reviews*, 45, 95–104. <https://doi.org/10.1016/j.smrv.2019.03.004>
- Crummy, F., Cameron, P. A., Swann, P., Kossmann, T., & Naughton, M. (2008). Prevalence of sleepiness in surviving drivers of motor vehicle collisions. *Internal medicine journal*, 38(10), 769–775.
- Dawson, D., Searle, A. K., & Paterson, J. L. (2014). Look before you (s)leep: Evaluating the use of fatigue detection technologies within a fatigue risk management system for the road transport industry.

- Sleep Medicine Reviews*, 18(2), 141-152. <https://doi.org/10.1016/j.smr.2013.03.003>
- Dinges, D. F., & Mallis, M. M. (1998). *Managing fatigue by drowsiness detection: can technological promises be realized?*. Retrieved from Pergamon.
- Dinges, D. F., & Powell, J. W. (1985). Microcomputer analyses of performance on a portable, simple visual RT task during sustained operations. *Behavior Research Methods*, 17(6), 652-655. <https://doi.org/10.3758/bf03200977>
- Doran, S. M., Van Dongen, H. P. A., & Dinges, D. F. (2001). Sustained attention performance during sleep deprivation: Evidence of state instability. *Archives Italiennes de Biologie*, 139(3), 253-267.
- Ftouni, S., Rahman, S. A., Crowley, K. E., Anderson, C., Rajaratnam, S. M. W., & Lockley, S. W. (2013). Temporal dynamics of ocular indicators of sleepiness across sleep restriction. *Journal of Biological Rhythms*, 28(6), 412-424. <https://doi.org/10.1177/0748730413512257>
- Ftouni, S., Sletten, T. L., Howard, M. E., Anderson, C., LennÉ, M. G., Lockley, S. W., & Rajaratnam, S. M. W. (2012). Objective and subjective measures of sleepiness, and their associations with on-road driving events in shift workers. *Journal of Sleep Research*, no-no. <https://doi.org/10.1111/j.1365-2869.2012.01038.x>
- Ganesan, S., Magee, M., Stone, J. E., Mulhall, M. D., Collins, A., Howard, M. E., ... Sletten, T. L. (2019). The Impact of Shift Work on Sleep, Alertness and Performance in Healthcare Workers. *Scientific Reports*, 9(1). <https://doi.org/10.1038/s41598-019-40914-x>
- Gilliland, K., & Schlegel, R. E. (1993). *Readiness to perform: A critical analysis of the concept and current practices*. Retrieved from
- Lamond, N., Dawson, D., & Roach, G. D. (2005). Fatigue assessment in the field: validation of a hand-held electronic psychomotor vigilance task. *Aviat Space Environ Med*, 76(5), 486-489.
- Lamond, N., Jay, S. M., Dorrian, J., Ferguson, S. A., Roach, G. D., & Dawson, D. (2008). The sensitivity of a palm-based psychomotor vigilance task to severe sleep loss. *Behav Res Methods*, 40(1), 347-352. <https://doi.org/10.3758/bm.40.1.347>
- Lee, M. L., Howard, M. E., Horrey, W. J., Liang, Y., Anderson, C., Shreeve, M. S., ... Czeisler, C. A. (2016). High risk of near-crash driving events following night-shift work. *Proc Natl Acad Sci U S A*, 113(1), 176-181. <https://doi.org/10.1073/pnas.1510383112>
- Lim, J., & Dinges, D. F. (2008). Sleep deprivation and vigilant attention. *Annals of the New York Academy of Sciences*, 1129, 305-322. <https://doi.org/10.1196/annals.1417.002>
- Loh, S., Lamond, N., Dorrian, J., Roach, G. D., & Dawson, D. (2004). The validity of psychomotor vigilance tasks of less than 10-minute duration. *Behavior Research Methods*, 36(2), 339-346. <https://doi.org/10.3758/bf03195580>
- Lüdtke, H., Wilhelm, B., Adler, M., Schaeffel, F., & Wilhelm, H. (1998). Mathematical procedures in data recording and processing of pupillary fatigue waves. *Vision Research*, 38(19), 2889-2896. [https://doi.org/10.1016/S0042-6989\(98\)00081-9](https://doi.org/10.1016/S0042-6989(98)00081-9)
- Maccora, J., Manousakis, J. E., & Anderson, C. (2018). Pupillary instability as an accurate, objective marker of alertness failure and performance impairment. *Journal of Sleep Research*, 28(2), e12739. <https://doi.org/10.1111/jsr.12739>
- Manousakis, J. E., Maccora, J., & Anderson, C. (2020). *Shortened PST Prediction - Laboratory Data*. Retrieved from: [https://bridges.monash.edu/articles/Shortened\\_PST\\_Prediction\\_-\\_Laboratory\\_Data/12577679](https://bridges.monash.edu/articles/Shortened_PST_Prediction_-_Laboratory_Data/12577679)
- Mulhall, M. D., Sletten, T. L., Magee, M., Stone, J. E., Ganesan, S., Collins, A., ... Rajaratnam, S. M. W. (2019). Sleepiness and driving events in shift workers: the impact of circadian and homeostatic factors. *Sleep*, 42(6). <https://doi.org/10.1093/sleep/zsz074>
- Oken, B. S., Salinsky, M. C., & Elsas, S. M. (2006). Vigilance, alertness, or sustained attention: physiological basis and measurement. *Clinical Neurophysiology*, 117(9), 1885-1901. <https://doi.org/10.1016/j.clinph.2006.01.017>
- Peters, T., Grüner, C., Durst, W., Hütter, C., & Wilhelm, B. (2014). Sleepiness in professional truck drivers measured with an objective alertness test during routine traffic controls. *International archives of occupational and environmental health*, 87(8), 881-888. <https://doi.org/10.1007/s00420-014-0929-6>
- Regen, F., Dorn, H., & Danker-Hopfe, H. (2013). Association between pupillary unrest index and waking electroencephalogram activity in sleep-deprived healthy adults. *Sleep Medicine*, 14(9), 902-912. <https://doi.org/10.1016/j.sleep.2013.02.003>
- Roach, G. D., Dawson, D., & Lamond, N. (2006). Can a shorter psychomotor vigilance task be used as a reasonable substitute for the ten-minute psychomotor vigilance task? *Chronobiol Int*, 23(6), 1379-1387. <https://doi.org/10.1080/07420520601067931>
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2), 461-464.
- Stevenson, M. R., Elkington, J., Sharwood, L., Meuleners, L., Ivers, R., Boufous, S., ... Wong, K. (2014). The role of sleepiness, sleep disorders, and the work environment on heavy-vehicle crashes in 2 Australian states. *Am J Epidemiol*, 179(5), 594-601. <https://doi.org/10.1093/aje/kwt305>
- Tefft, B. C. (2018). Acute sleep deprivation and culpable motor vehicle crash involvement. *Sleep*, 41(10). <https://doi.org/10.1093/sleep/zsy144>
- Verstraete, A. G. (2005). Oral fluid testing for driving under the influence of drugs: history, recent progress and remaining challenges. *Forensic Sci Int*, 150(2-3), 143-150. <https://doi.org/10.1016/j.forsciint.2004.11.023>
- Wilhelm, B., Rühle, K. H., Widmaier, D., Lüdtke, H., & Wilhelm, H. (1998a). Objective measurement of severity and therapeutic effect in obstructive sleep apnea syndrome by the pupillographic sleepiness test. *Somnologie*, 2(2), 51-57. <https://doi.org/10.1007/s11818-998-0008-x>
- Wilhelm, B., Wilhelm, H., Lüdtke, H., Streicher, P., & Adler, M. (1998b). Pupillographic assessment of sleepiness in sleep-deprived healthy subjects. *Sleep*, 21(3), 258-265.
- Wolkow, A. P., Rajaratnam, S. M. W., Anderson, C., Howard, M. E., & Mansfield, D. (2019). Recommendations for current and future countermeasures against sleep disorders and sleep loss to improve road safety in Australia. *Intern Med J*, 49(9), 1181-1184. <https://doi.org/10.1111/imj.14423>
- Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer*, 3(1), 32-35. [https://doi.org/10.1002/1097-0142\(1950\)3:1<32::AID-CNCR2820030106>3.0.CO;2-3](https://doi.org/10.1002/1097-0142(1950)3:1<32::AID-CNCR2820030106>3.0.CO;2-3)