



MONASH University

Accident Research Centre

**THE RELATIONSHIP BETWEEN
DRIVING PERFORMANCE AND THE
JOHNS DROWSINESS SCALE AS
MEASURED BY THE OPTALERT
SYSTEM**

by

Karen Stephan

Simon Hosking

Michael Regan

Ashley Verdoorn

Kristie Young

Narelle Haworth

September, 2006

Report No. 252

MONASH UNIVERSITY ACCIDENT RESEARCH CENTRE
REPORT DOCUMENTATION PAGE

Report No.	Date	ISBN	Pages
252	September 2006	0 7326 2322 7	43

Title and sub-title:

The relationship between driving performance and the Johns Drowsiness Scale as measured by the Optalert system

Author(s): Karen Stephan, Simon Hosking, Michael Regan, Ashley Verdoorn, Kristie Young, Narelle Haworth

Sponsoring Organisation(s): Sleep Diagnostics Pty Ltd

Key Words:

Drowsiness, driving performance, simulation, system evaluation, road safety, fatigue, intelligent transport systems.

Disclaimer

This report is disseminated in the interest of information exchange. The views expressed here are those of the authors, and not necessarily those of Monash University

Reproduction of this page is authorised.

www.monash.edu.au/muarc

Monash University Accident Research Centre,
Building 70, Clayton Campus, Victoria, 3800, Australia.
Telephone: +61 3 9905 4371, Fax: +61 3 9905 4363

Preface

Research Team:

- Dr Michael Regan
- Ms Karen Stephan
- Mr Simon Hosking
- Mr Ashley Verdoorn
- Ms Nicola Fotheringham
- Mr Nebojsa Tomasevic
- Ms Kristie Young

Contents

1	EXECUTIVE SUMMARY	X
1.1	METHOD	XI
1.2	ANALYSES.....	XII
1.3	RESULTS AND DISCUSSION.....	XIII
1.3.1	Lane Excursions	xiv
1.3.2	Comfort of Optalert.....	xvi
1.4	CONCLUSIONS	XVI
2	INTRODUCTION.....	1
2.1	OPTALERT SYSTEM	1
2.2	DRIVING PERFORMANCE MEASURES.....	2
2.2.1	Lateral Placement	2
2.2.2	Steering Control.....	3
2.2.3	Speed and Longitudinal Control.....	3
2.3	AIMS AND HYPOTHESES	4
3	METHOD.....	6
3.1	PARTICIPANTS	6
3.2	APPARATUS	6
3.2.1	Driving simulator.....	6
3.2.2	Optalert drowsiness detection system.....	7
3.2.3	FaceLab	7
3.2.4	Other equipment	7
3.3	DESIGN.....	7
3.4	DRIVING SCENARIOS	7
3.5	VARIABLES MEASURED THROUGHOUT THE DRIVING SCENARIO.....	8
3.5.1	Drowsiness.....	8
3.5.2	Driving performance.....	9
3.6	PROCEDURE.....	9
3.7	ETHICS	10
3.8	DATA ANALYSIS.....	10
4	RESULTS.....	11
4.1	QUESTIONNAIRES	12
4.1.1	Pre-Drive	12
4.2	DRIVING PERFORMANCE	16
4.2.1	Relationship between driving performance and continuous JDS.....	16
4.2.2	Driving performance during periods when the driver was classified as drowsy compared to when periods they were not classified as drowsy	19
5	DISCUSSION.....	37
5.1	LIMITATIONS OF THE STUDY AND FURTHER RESEARCH AND ANALYSIS.....	40
6	CONCLUSIONS.....	41

Figures

FIGURE 1. PARTICIPANTS’ RATINGS OF SUBJECTIVE SLEEPINESS USING THE KAROLINSKA SLEEPINESS SCALE, PRIOR TO THE ALERT AND SLEEP-DEPRIVED EXPERIMENTAL SESSIONS 13

FIGURE 2. PARTICIPANTS’ SCORES ON THE EPWORTH SLEEPINESS SCALE 14

FIGURE 3. PARTICIPANTS’ RATINGS OF THE COMFORT OF THE OPTALERT SYSTEM, AFTER THE ALERT AND SLEEP-DEPRIVED SESSIONS (0=EXTREMELY UNCOMFORTABLE, 10=EXTREMELY COMFORTABLE)..... 15

FIGURE 4. MEAN LATERAL LANE POSITION ACROSS CATEGORIES OF JDS FOR THE ALERT AND SLEEP-DEPRIVED CONDITIONS..... 17

FIGURE 5. STANDARD DEVIATION OF LATERAL LANE POSITION ACROSS CATEGORIES OF JDS FOR THE ALERT AND SLEEP-DEPRIVED CONDITIONS..... 18

FIGURE 6. AVERAGE SPEED ACROSS CATEGORIES OF JDS FOR THE ALERT AND SLEEP-DEPRIVED CONDITIONS..... 18

FIGURE 7. STANDARD DEVIATION OF SPEED ACROSS CATEGORIES OF JDS FOR THE ALERT AND SLEEP-DEPRIVED CONDITIONS..... 19

Tables

TABLE 1: DURATION AND QUALITY OF SLEEP ON THE LAST NIGHT THAT PARTICIPANTS SLEPT, PRIOR TO THE TEST SESSION: MEAN (STANDARD DEVIATION) AND RANGE..... 12

TABLE 2. REPORTED CAFFEINE USE IN THE 24 HOURS PRIOR TO TESTING..... 13

TABLE 3. CATEGORISATION OF JOHNS DROWSINESS SCALE 16

TABLE 4. NUMBER OF OCCASIONS, AND THE PROPORTION OF EVENTS THAT OCCURRED WHEN THE DRIVER WAS CLASSIFIED AS DROWSY BY OPTALERT, WHERE THE ENTIRE CAR WAS OUT OF THE LANE FOR ALERT SESSION, SLEEP-DEPRIVED SESSION AND OVERALL 21

TABLE 5. NUMBER OF OCCASIONS, AND THE PROPORTION OF EVENTS THAT OCCURRED WHEN THE DRIVER WAS CLASSIFIED AS DROWSY BY OPTALERT, WHERE AT LEAST HALF OF THE CAR WAS OUT OF THE LANE FOR ALERT SESSION, SLEEP-DEPRIVED SESSION AND OVERALL 22

TABLE 6. NUMBER OF OCCASIONS, AND THE PROPORTION OF EVENTS THAT OCCURRED WHEN THE DRIVER WAS CLASSIFIED AS DROWSY BY OPTALERT, WHERE ANY PART OF THE CAR WAS OUT OF THE LANE FOR ALERT SESSION, SLEEP-DEPRIVED SESSION AND OVERALL..... 22

TABLE 7. GENERIC 2X2 MATRIX USED FOR THE EVALUATION OF THE OPTALERT SYSTEM. 23

TABLE 8. SENSITIVITY, SPECIFICITY AND PREDICTIVE VALUE OF THE OPTALERT SYSTEM ACCORDING TO ENTIRE CAR LANE EXCURSIONS DURING DISTINCT PREDICTION PERIODS FOR ALERT AND SLEEP-DEPRIVED SESSIONS, POOLED ACROSS PARTICIPANTS. 24

TABLE 9. SENSITIVITY, SPECIFICITY AND PREDICTIVE VALUE OF THE OPTALERT SYSTEM ACCORDING TO LANE EXCURSIONS WHERE AT LEAST HALF OF THE CAR LEFT THE LANE, DURING DISTINCT PREDICTION PERIODS FOR ALERT AND SLEEP-DEPRIVED SESSIONS, POOLED ACROSS PARTICIPANTS..... 25

TABLE 11. MEAN (SD) OF LANE VARIABILITY DURING PERIODS IN WHICH THE DRIVER WAS OR WAS NOT CLASSIFIED AS DROWSY DURING THE PREVIOUS 5 MINUTES 27

TABLE 12. MEAN (SD) OF LANE VARIABILITY DURING PERIODS IN WHICH THE DRIVER WAS OR WAS NOT CLASSIFIED AS DROWSY IN THE PREVIOUS 15 MINUTES 28

TABLE 13. MEAN (SD) OF LANE VARIABILITY DURING PERIODS IN WHICH THE DRIVER WAS OR WAS NOT CLASSIFIED AS DROWSY IN THE PREVIOUS 30 MINUTES 28

TABLE 14. MEAN (SD) AND MEDIAN OF LANE EXCURSION RATES DURING PERIODS IN WHICH THE DRIVER WAS OR WAS NOT CLASSIFIED AS DROWSY IN THE PREVIOUS 5 MINUTES 30

TABLE 15. MEAN (SD) AND MEDIAN OF LANE EXCURSION RATES DURING PERIODS IN WHICH THE DRIVER WAS OR WAS NOT CLASSIFIED AS DROWSY DURING THE PREVIOUS 15 MINUTES..... 30

TABLE 16. MEAN (SD) AND MEDIAN OF LANE EXCURSION RATES DURING PERIODS IN WHICH THE DRIVER WAS OR WAS NOT CLASSIFIED AS DROWSY IN THE PREVIOUS 30 MINUTES 31

TABLE 17. MEAN (SD) AND MEDIAN PROPORTION OF TIME THE CAR WAS OUT OF THE LANE DURING PERIODS IN WHICH THE DRIVER WAS OR WAS NOT CLASSIFIED AS DROWSY IN THE PREVIOUS 5 MINUTES 33

TABLE 18. MEAN (SD) AND MEDIAN PROPORTION OF TIME THE CAR WAS OUT OF THE LANE DURING PERIODS IN WHICH THE DRIVER WAS OR WAS NOT CLASSIFIED AS DROWSY IN THE PREVIOUS 15 MINUTES 34

TABLE 19. MEAN (SD) AND MEDIAN PROPORTION OF TIME THE CAR WAS OUT OF THE LANE DURING PERIODS IN WHICH THE DRIVER WAS OR WAS NOT CLASSIFIED AS DROWSY IN THE PREVIOUS 30 MINUTES.....35

1 EXECUTIVE SUMMARY

Driver drowsiness is a contributing factor in a significant number of single and multiple vehicle crashes. It is estimated that driver drowsiness accounts for around 20 to 30 percent of all traffic crashes in Australia, at a huge cost to society.

A promising countermeasure designed to reduce the incidence of drowsiness-related crashes is a system that can detect drowsiness and issue warnings accordingly. Sleep Diagnostics, Pty Ltd, have developed an in-vehicle drowsiness detection and warning system called Optalert. The Monash University Accident Research Centre (MUARC) was approached by Sleep Diagnostics to conduct an evaluation of the Optalert system (glasses version 5, software version 4.1.5, algorithm version 08075) in terms of its effectiveness in predicting a breakdown in driving performance, using the MUARC advanced driving simulator.

The Optalert system comprises ordinary spectacle frames housing light emitters and sensors that objectively measure eye and eyelid movements. From the many parameters measured, a continuous measure of drowsiness, the Johns Drowsiness Scale (JDS), is derived. The JDS ranges from zero to 10, where an increase in JDS corresponds to an increase in drowsiness.

The Optalert system is designed to emit an auditory warning that informs the driver of approaching drowsiness. Two different warnings are emitted. The optimum JDS level at which the warnings are given has been determined by Sleep Diagnostics in previous research. The first warning is cautionary, and occurs when the JDS reaches a level of 4.5. The second warning is emitted when the JDS reaches 5.0, and indicates that the driver has reached a critical level of drowsiness and should stop driving. Through these warnings, drivers are alerted to take action to safely manage their drowsiness.

Much previous research shows that driving performance degrades as sleepiness increases, particularly drivers' lane-keeping ability, which leads to increased crash risk. The Optalert system seeks to ameliorate this risk by warning drivers that they have reached a level of drowsiness beyond which it would be dangerous to keep driving. Therefore it is of prime interest to determine how actual driving performance changes after a driver's drowsiness level reaches the cautionary and critical JDS levels.

The aim of the current study was to determine whether there is a breakdown in driving performance during the period after which the JDS reaches the cautionary level (4.5) and the critical level (5.0), that is, after the driver is classified as drowsy by the Optalert system. In order to achieve this, the Optalert system warnings were not active within this experiment, that is, the system was only used to measure drowsiness, not to warn the driver of their drowsiness. Based upon previous research using the Optalert system, Sleep Diagnostics defined the period of interest to be 30 minutes after the JDS reached the cautionary or critical levels. The analyses were also repeated for shorter periods, of 15 minutes and 5 minutes after the participant reached the cautionary or critical JDS level. It is important to note that the Optalert system is designed to predict drowsiness well ahead of when a warning would be given, and so, the longer time intervals of 15 minutes and 30 minutes are of greater interest in this study.

Based upon previous research into the effects of drowsiness on driving performance, measures relating to the lateral placement of the vehicle within the lane were the focus of primary interest in this study. Lateral placement and standard deviation of lateral position

(SDLP) were measured continuously throughout the study. Discrete lane excursions were also recorded.

Because previous research has rarely provided an exact definition of a lane excursion, the current study considered three categories of lane excursions. The broadest category included all instances when any part of the car left the lane, whether it was a small part of the car, or the entire car. The second category of lane excursions was a subset of the first, and was defined as those instances when at least half of the vehicle departed from the lane. The third category was a subset of the previous two categories; severe lane excursions were defined as any instance where the entire vehicle departed from the lane of travel, that is, they were true run off-road events. Whilst lane excursions where the entire vehicle leaves the lane are the most dangerous, situations where only a portion of the car departs from the lane can also be dangerous, particularly when parts of the vehicle cross into another lane where another vehicle is travelling.

It was hypothesised that, during periods that drivers were classified as drowsy by the Optalert system, lane excursions would increase in rate (per minute) and duration, and that lane variability would increase relative to periods in which the Optalert system did not classify the driver as drowsy. Speed and speed control were also measured to determine if there was a relationship to drowsiness. In addition, mean time headway while following another vehicle was measured to determine if this varied across periods of drowsiness. Standard test evaluation techniques were used to determine the predictive validity of Optalert.

1.1 METHOD

Twenty-three participants were recruited to take part in the study and 20 participants presented for at least one session. The twenty participants who attended at least one session comprised 15 males and 5 females, with an average age of 23.5 years (standard deviation of 2.9 years). All participants held a full drivers license.

The study was conducted using the Monash University Accident Research Centre's advanced driving simulator. Each participant took part in the experiment on two separate days, one after a night of normal sleep, and one after a night without sleep since 8am the morning before. The order in which participants attended for the sleep deprived and alert days was counterbalanced across participants, to control for order/practice effects.

The simulated driving scenarios consisted of a two-lane rural road, with one lane in each direction. The road had a broken centre line and lateral edge lines, and gravel road edge to provide tactile and audio feedback to participants in the event that they ran off the road. The sessions took approximately 70 minutes to drive when travelling at the signed speed limit.

Driver drowsiness was measured continuously throughout the simulated drives using the Optalert drowsiness detection system. The mean lateral lane position per minute was determined, as was the standard deviation of lateral lane position. Lane excursions were also identified. Other driving performance measures included speed, speed control and mean time headway when following other vehicles.

Upon arrival for their first experimental session, participants completed questionnaires relating to demographic information, simulator sickness symptoms, sleep quality, the Karolinska sleepiness scale and the Epworth sleepiness scale. Participants were instructed

to drive according to the road rules and road signs, to drive as closely as possible to the posted speed limit, to drive straight ahead when they came to cross-roads, and to follow any vehicle in front of them at a safe distance. The data logging systems for the driving simulator and the Optalert driver drowsiness detection system were initiated at the same time to ensure the data could be matched for future analysis and then the participants began the test drive. After their driving session, participants also completed a questionnaire about the comfort of the Optalert system.

After all participants had completed the experimental sessions, the data were parsed and extracted from the raw data files. The data were checked for outliers, and these were further investigated to determine their validity. Where appropriate, data were transformed to enable parametric analyses to be performed.

1.2 ANALYSES

Initial simple descriptive analyses were performed to observe how various driving performance measures differed as drowsiness level (JDS) changed. The driving performance measures considered in this way were mean lateral lane position, standard deviation of lateral lane position, mean speed and standard deviation of speed.

Each of the driving sessions was divided into periods in which the Optalert system classified the driver as too drowsy to drive and not too drowsy to drive. Considering there were two JDS levels of interest (the cautionary level at 4.5 and the critical level at 5.0) and three time periods of interest after the JDS level was reached (5, 15 or 30 minutes), there were six different ways to define a period in which a driver was too drowsy to drive. The sessions were divided into drowsy and non-drowsy periods in these six different ways, and separate analyses performed for each of these six prediction window by JDS level periods. Standard statistical tests were performed to determine if standard deviation of lane position, rate of lane excursions, proportion of total time that lane excursions occurred, and mean time headway when following another vehicle differed between periods when the Optalert system classified the driver as drowsy, compared to periods when the driver was not classified as drowsy.

Analyses using techniques for test evaluation were used to determine the predictive validity of Optalert in predicting periods when discrete lane excursions occurred. In this part of the analysis, the inference was that if the driver was classified as drowsy, then it was predicted that their driving performance would degrade, as measured by lane excursions. To fully assess predictive power, it is necessary to take into account four potential situations:

- A. True positive: the Optalert system classified the driver as drowsy and that poor performance would occur, and a lane excursion did occur;
- B. False positive: the Optalert system classified the driver as drowsy and predicted that poor performance would occur, and a lane excursion did not occur;
- C. False negative: the Optalert system did not classify the driver as drowsy nor predict that poor performance would occur, and a lane excursion did occur; and
- D. True negative: the Optalert system did not classify the driver as drowsy nor predict that poor performance would occur, and a lane excursion did not occur.

These four outcomes can be displayed in a 2x2 matrix as shown below.

		Lane excursion observed during period	
		Yes	No
Driver classified as drowsy by Optalert - Lane excursion predicted	Yes	A: True Positive	B: False Positive
	No	C: False Negative	D: True Negative

From the 2x2 matrix, several measures are derived. First, the sensitivity of the system is given by the proportion of periods in which a lane excursion was observed that were predicted by the system, or, the probability that Optalert predicted poor driving performance would happen in the period, given that one or more lane excursions did happen ($A/A+C$). The specificity of the system is given by the proportion of periods where a lane excursion did not occur and where the Optalert system did not predict poor driving performance, or, the probability that Optalert did not predict poor driving performance in the period, given that no lane excursion occurred ($D/B+D$). In essence, the sensitivity and specificity indicate the proportion of periods in which events (sensitivity) and non-events (specificity) occurred, that were predicted correctly by the system.

Another important task is to determine the predictive value of the system. The positive predictive value (PPV) is defined as the proportion of periods in which the Optalert system predicted that poor driving performance would occur, and one or more lane excursions actually did occur, or the probability that one or more lane excursion occurred, given that the Optalert system predicted it would ($A/A+B$). The negative predictive value (NPV) was defined as the proportion of periods in which the system did not predict poor driving performance, that no lane excursion occurred, or the probability that no lane excursion occurred, given that Optalert did not predict that one would ($D/C+D$). That is, the predictive values indicate the proportion of periods in which the system was correct when it predicted that poor driving performance in terms of lane excursions were (positive predictive value) or were not (negative predictive value) going to occur.

The Optalert system was evaluated by formulating separate 2x2 matrices for cautionary and critical JDS levels, for each prediction window (5, 15 or 30 minutes), each level of lane excursion (i.e., all, half, or severe off-road events), and for each of the alert and sleep-deprived conditions. The driving sessions were divided into periods where the driver was too drowsy to drive and periods where they were not drowsy and each period was then classified into one of the four possible outcomes.

1.3 RESULTS AND DISCUSSION

The simple descriptive analyses showed that as drowsiness increased (as measured by the JDS), the absolute distance of the centre of the car from the centre of the lane (absolute mean lateral lane position) increased, as did the variability in lane position (SDLP). Mean

speed however, did not vary according to drowsiness level, while speed variability seemed to display a more complex non-linear relationship with drowsiness that was not analysed further.

There was a trend for the drivers' lane position to vary more (i.e. for SDLP to increase) when drivers were classified as drowsy by the Optalert system, compared to when they were not drowsy. There was also a consistent trend, for all three classes of lane excursions, for the both the rate of lane excursions and the proportion of time the vehicle was out of the lane to be higher in periods when the driver was classified as drowsy compared with periods where the driver was not classified as drowsy. This was the case for all JDS level/time window combinations. However, not all of these differences were statistically significant. This could possibly be due to the low prevalence of some events, particularly severe lane excursions, and the resulting low statistical power. It was also found that mean time headway when following other vehicles did not differ according to whether the Optalert system classified the driver as drowsy or not, which is worth noting considering that previous research has found that as sleepiness increases, so too does mean time headway.

1.3.1 Lane Excursions

Severe lane excursions where the entire car departed the lane were relatively infrequent, with only seven of the eighteen participants making such excursions, and only five participants doing so more than twice. Fourteen participants made lane excursions where at least half of the car departed the lane. Six of these fourteen participants only made one such excursion, while six of the participants made more than 10 such excursions. All participants made lane excursions where less than half of the car exceeded the lane boundaries.

When considering the predictive validity of the Optalert system in predicting lane excursions, the results consistently showed that for the two more severe classes of lane excursions, the sensitivity and NPV were superior when using the 15 minute prediction window for both the JDS levels of interest. In addition, use of the critical (5.0) level as a cut-off usually improved the specificity and PPV relative to the cautionary (4.5) cut-off. Therefore, the results presented here are those where the driver was classified as drowsy if the JDS had reached the critical level (5.0) in the previous 15 minutes, and was not classified as drowsy if the JDS had not reached 5.0 in the previous 15 minutes. The results for the other JDS level/prediction window combinations are provided in the results section of the report.

The vast majority of occasions when the entire car left the lane (96% in sleep deprived session, 83.3% in alert session) occurred during periods where the Optalert system classified the driver as drowsy. The majority of lane excursions where at least half of the car departed the lane (91.9% in sleep deprived session, 84.6% in alert session) and most of the lane excursions where any part of the car left the lane (79.5% in sleep deprived session, 57.7% in alert session) also occurred when the Optalert system classified the driver as drowsy. For all types of lane excursion, the Optalert system was more successful at predicting lane excursions that occurred during the sleep deprived session than the non-sleep deprived session.

The sensitivity of the Optalert system was better for detecting periods with more dangerous lane excursions, with the greatest sensitivity displayed for periods when entire car lane excursions occurred (sleep deprived session 71.4%, non-sleep deprived session 66.7%),

followed by periods when excursions occurred where at least half the car left the lane (sleep deprived session 56.3%, non-sleep deprived session 42.9%). Sensitivity for detecting periods when lane excursions occurred where any part of the car left the lane was relatively poor (sleep deprived session 44.8%, non-sleep deprived session 35.0%).

The negative predictive value (NPV) of the system followed a similar pattern, with the NPV being highest for periods when entire car lane excursions occurred (sleep deprived session 87.5%, non-sleep deprived session 93.8%), and periods when at least half the car left the lane (sleep deprived 56.3%, non-sleep deprived 75.0%) and very low for periods when excursions occurred where any part of the car left the lane (sleep deprived session 0%, non-sleep deprived session 18.8%). The reason that the NPV was so low for excursions where any part of the car left the lane in the sleep-deprived sessions (0%) was that there were no periods when the car remained entirely within the lane when participants were sleep deprived (that is, there were no false positive or true negatives).

However, for specificity and PPV, the results showed the opposite pattern. Specificity was better for periods when excursions occurred where any part of the car left the lane (sleep deprived session –not able to be calculated, non-sleep deprived session 100%), and lower for periods when at least half of the car departed the lane (sleep deprived session, 69.2%, non-sleep deprived session, 75%) and for periods when entire car lane excursions occurred (sleep deprived session 63.6%, non-sleep deprived session 75%).

PPV was also highest for periods when any part of the car left the lane (sleep deprived session 100%, non-sleep deprived session 100%), lower for periods when at least half of the car left the lane (sleep deprived session 69.2%, non-sleep deprived session 42.9%) and lowest for periods when the entire car left the lane (sleep deprived session 38.5%, non-sleep deprived session 28.6%).

In this sort of evaluation, there is always a trade-off between sensitivity and specificity, and negative and positive predictive values. To determine which are to be valued more highly for a particular system, the consequences of an incorrect “decision”, or prediction, need to be considered. If the decision level cut-off is set too high, then the system will miss many events of interest, and the false negative rate will be high. That is, if the critical JDS level at which the driver is warned to cease driving is set too high, then driving performance may degrade before the driver is classified as too drowsy to drive. This would lead to low sensitivity and negative predictive value (NPV). If on the other hand, the cut-off is set too low, more events will be detected, but the rate of false positives will be high. That is, if the critical JDS level is set too low, then drivers who are not too drowsy to drive will be warned to stop driving. This would lead to low specificity and positive predictive value (PPV). Therefore the consequences of false negatives and false positives must be considered.

In the current situation, the consequence of a false positive (that is, the Optalert system warning the driver that they are too tired to drive when they are not) for on-road driving is that drivers will receive a warning to indicate that they are drowsy when they are not drowsy. If the driver heeds the warning and takes corrective action, such as changing drivers, or stopping for a rest, at worst this will mean it will take longer to complete their journey. The consequences of a false negative on the other hand is that the driver will not receive a warning to indicate they are drowsy when they really are too drowsy to drive, and their driving performance might degrade to the point that they drive out of their lane leading to a collision. Therefore, it was judged that the consequences of false negatives were more dangerous than the consequences of false positives for determining whether a

driver is too drowsy to drive. Thus it is more desirable for the Optalert system to display high sensitivity rather than high specificity. In addition, high negative predictive value is preferred to high positive predictive value.

From the results presented above, the Optalert system showed good sensitivity and very high negative predictive value in predicting lane excursions where the entire car left the road, that is, off-road events. That is, for most periods when an off-road event occurred, the Optalert system correctly predicted the event, by classifying the driver as too drowsy to drive. In addition, for almost all periods in which the Optalert system did not classify the driver as too drowsy to drive, there were no off-road events (severe lane excursions). The sensitivity and NPV were also reasonable for detecting events where at least half of the car left the lane.

The poor sensitivity of the Optalert system in predicting periods when any part of the car left the lane can partly be explained by the extremely high prevalence of these events. Lane excursions where any part of the vehicle left the lane occurred frequently, while periods in which no part of the car ever left the lane were extremely infrequent. As such, the false positive rate is low, simply because of the low rate of non-events. This leads to high specificity and PPV, but low sensitivity and NPV. Alternatively, the prevalence of entire car lane excursions and lane excursions where at least half of the car left the lane was much lower. This means there were fewer false negatives (which is desirable) and relatively more false positives, which are of less concern in this situation. This leads to higher sensitivity and NPV.

There is also another possible explanation for these results. There are many reasons apart from drowsiness that might explain why a driver would cross the boundaries of the lane. Loss of concentration, boredom, inattention or even simply shifting position in the seat might lead to part of the vehicle departing the lane. The Optalert system would not be expected to predict events of this type unless they too were always related to drowsiness. However, the Optalert system did perform well in terms of sensitivity and negative predictive value for the more severe types of lane excursions. It could be suggested that the more severe types of lane excursions are related to drowsiness, while many of the mild lane excursions may have been due to other factors, which could explain Optalert's superior performance in detecting the more severe events.

1.3.2 Comfort of Optalert

Most participants rated the Optalert system as comfortable to wear, with a median rating of 8 out of 10 (where 0 is extremely uncomfortable and 10 is extremely comfortable) and most participants did not believe the Optalert glasses impaired their driving ability. Three participants believed their ability to drive was impaired in both sessions by the Optalert system. Two of these responded that the Optalert frames blocked the visual field, while one felt impaired by the cable dangling over their shoulder. A further three participants believed the Optalert system impaired them during their first driving session with the system. Of these three participants, one attributed the impairment to the weight of the system, one stated that they were not used to wearing glasses, while the remaining participant felt the Optalert system caused blurry vision.

1.4 CONCLUSIONS

The aim of the current study was to determine whether there was a breakdown in driving performance during the period after which a driver reaches a level of drowsiness at which

they would be alerted by the Optalert system that they were becoming drowsy (JDS=4.5) or that they were too drowsy to drive (JDS=5.0).

There was a trend for increased lane variability, increased rate of lane excursions and increased duration of time spent out of the lane when the Optalert system classified the driver as drowsy, compared to periods when the driver was not classified as drowsy, although this trend was not always statistically significant.

The Optalert system performed well in predicting severe lane excursions, where the entire vehicle departed from the lane (run off-road events). The sensitivity of the system for correctly classifying periods when these lane excursions occurred was reasonably high; that is, the Optalert system had classified the driver as too drowsy to drive within the previous 15 minutes for most periods in which the entire vehicle ran off the road. The negative predictive value for run off-road events was very high; that is, for almost all periods in which the Optalert system did not classify the driver as too drowsy to drive within the previous 15 minutes, there were no run off-road events. High sensitivity and high negative predictive value are both very desirable attributes of a system that is designed to warn drivers of drowsiness.

Most participants rated the Optalert system as comfortable to wear and most participants did not believe the Optalert glasses impaired their driving ability.

2 INTRODUCTION

Driver drowsiness is a contributing factor in a significant number of single and multiple vehicle crashes. Unlike the use of blood alcohol content (BAC) to identify the propensity for alcohol-related crashes, there is no simple way to predict the involvement of driver drowsiness in crashes. Instead, subjective criteria are used to determine if drowsiness was a contributing factor, such as if police observed the driver to be asleep or drowsy, and the characteristics of the crash itself (e.g. RTA, 2001). Using such criteria, it is estimated that driver drowsiness accounts for around 20 to 30 percent of all traffic crashes in Australia (e.g. RTA, 2001; VicRoads, 2005) at a huge cost to society.

One promising countermeasure aimed at reducing the incidence of drowsiness-related crashes is the development of systems that are capable of detecting drowsiness and issue warnings accordingly. Various methods have been proposed for measuring the extent of drowsiness, which range from physiological to driving performance-related measures. A number of physiological measures have been examined in the literature and several of these have been found to be sensitive and reliable indicators of driver drowsiness: heart rate and heart rate variability (Riemersma, Sanders, Wildervanck & Gaillard, 1977, cited in Haworth, Triggs & Grey, 1988), brain wave activity (Akerstedt, Kecklund & Knutson, 1991; de Waard & Brookhuis, 1991; Lal & Craig, 2000; 2002), and measurement of head movements. In particular, eyelid closures, slow rolling eye movement and more frequent blinking have all been found to be reliable predictors of the onset of sleep (Wierwille, Wregget, Kirn, Ellsworth, & Fairbanks, 1994).

2.1 OPTALERT SYSTEM

Sleep Diagnostics, Pty Ltd, have developed an in-vehicle drowsiness detection and warning system, called Optalert (Johns, Tucker, Chapman, Michael & Beale 2006). The Monash University Accident Research Centre (MUARC) was approached by Sleep Diagnostics to conduct an evaluation of the Optalert system (glasses version 5, software version 4.1.5, algorithm version 08075) to establish its effectiveness in predicting a breakdown in driving performance, using the MUARC advanced driving simulator. The driving simulator is ideal for research into the relationship between drowsiness and driving performance, as it provides a realistic yet safe method to investigate such issues.

The Optalert system comprises ordinary spectacle frames that house light emitters and sensors that allow the objective measurement of eye and eyelid movements. From the many parameters measured, such as the ratio of the amplitude to the velocity of each movement, the binocular coordination between the movements of each eye, and the duration of all movements, a continuous measure of drowsiness, the Johns Drowsiness Scale (JDS), is derived. The JDS ranges from zero to 10, where an increase in JDS corresponds to an increase in drowsiness.

The Optalert system emits an auditory warning that informs the driver of approaching drowsiness. Two different warnings are emitted. The first warning is cautionary, and occurs when the JDS reaches a level of 4.5. The second warning is emitted when the JDS reaches 5.0, and indicates that the driver has reached a critical level of drowsiness and should stop driving. Through these warnings, drivers are informed to take action to safely manage their drowsiness. A drowsiness-related breakdown in driving performance would be expected to occur in the period after a driver's JDS reaches the cautionary and critical levels.

2.2 DRIVING PERFORMANCE MEASURES

In order to determine the driving performance measures to be used in the present study, it was important to establish the driving performance measures that have previously been demonstrated to be related to drowsiness. However, a review of the previous literature to determine driving performance measures to be used in the current experiment identified an issue of concern. Previous studies have generally used sleep deprivation as the independent measure, and looked at average driving performance across different levels of sleep deprivation. That is, participants attend for several sessions in various states of sleep deprivation, and driving performance is compared across sessions. In these studies, the amount of sleep deprivation is often assumed to be a surrogate measure for drowsiness. However, it cannot simply be assumed that being sleep-deprived is equivalent to being drowsy, or that being non-sleep-deprived is the equivalent of being non-drowsy. Drowsiness changes over time, and can occur when participants are not sleep-deprived as well as sleep-deprived. Equally, some participants may not be drowsy at some point after a night of sleep deprivation. The current study sought to extend on previous studies by measuring the effect of changing drowsiness over time, rather than simply measuring average driving performance across discrete levels of sleep deprivation. It is, however, still important to look to the previous research to determine which driving performance measures may be appropriate to use. Driving performance measures that have been shown to be sensitive to sleep deprivation are lateral position and control, steering control, speed and time-headway control. Each of these is discussed in the following sections.

2.2.1 Lateral Placement

Measures of the lateral position of the vehicle on the road or in its lane of travel have been found to be accurately and reliably related to driver drowsiness (Arnedt, Wilde, Munt & Maclean, 2000; Bittner, Hana, Pousek, Smrcka, Schreib & Vysoky, 2000; Fairclough & Graham, 1999; Peters et al., 1999; Philip et al., 2003). A number of lateral placement measures have been examined in the drowsiness literature. These measures include lane deviations or excursions, run-offs and crashes, standard deviation of lane position, lane position variance, global maximum lane deviation, and the mean square of lane deviation (Bittner et al., 2000). That is, the lateral placement measures that have been shown to be related to driver drowsiness and sleep deprivation are discrete events such as crashes, off-road events and lane excursions, as well as continuous measures of lane variability.

A simulator study by Peters and colleagues (1999) examined the effects of sleep deprivation on driving performance. Participants drove the simulator for a period of 40 minutes over four test sessions (one session per day) under varying degrees of sleep deprivation, ranging from 9 hours on day one to 60 hours on day four. Measures of lateral placement, speed and incidence of crashes were recorded. The results revealed that there was a small increase in the number crashes the drivers experienced after moderate sleep deprivation and a marked increase in crashes with progressive sleep deprivation. Lateral placement variance, lane excursions and mean speed also increased with progressive sleep deprivation, particularly after 36 hours without sleep. Lateral placement variance, lane excursions and steering wheel position variance were also correlated with the number of crashes experienced and lateral placement was highly correlated with crashes, accounting for 81% of the variance in crashes.

Research by Arnedt et al. (2000) and Fairclough and Graham (1999) has also revealed similar effects of sleep deprivation on lateral placement control. After 20 hours or one full night of sleep deprivation, drivers exhibited a safety-critical decline in lateral placement

control. In particular, after sleep deprivation, drivers were more likely to drive towards the middle of the road, for their lane position variability to increase and for them to depart from their correct lane and run off the simulated road more frequently. Interestingly, these researchers also found that prolonged sleep deprivation produced a similar magnitude of driving impairment, as did a Blood Alcohol Concentration (BAC) of 0.07 or 0.08%.

A more recent simulator study by Philip and colleagues (2003) has also found that sleep loss is associated with reduced lane position control. Drivers who had been driving for prolonged periods of time with acute sleep loss participated in a 30 minute simulated drive, in which measures of the standard deviation from the centre of the road and the standard deviation of error from the ideal curve of the road were measured. The results revealed that the drivers who had been suffering from sleep loss performed significantly worse than control drivers in terms of the standard deviation of error from the ideal curve, but did not differ from controls on the standard deviation from the centre of the road.

2.2.2 Steering Control

Steering control has been one of the most commonly examined driving performance measures in the drowsiness literature. Research has found that under normal conditions (i.e., no sleep deprivation), drivers make a large number of fine steering adjustments (micro-wheel adjustments) and relatively few large steering adjustments, except when changing lanes. For sleep-deprived drivers on the other hand, research has shown that steering wheel movements become more variable and drivers make fewer micro-wheel adjustments and more large steering adjustments (Fairclough, 1997; Fairclough & Graham, 1999; Salah & Michel, 1994).

Fairclough and Graham (1999) found that steering control, as measured by steering wheel reversal rate per minute (the number of zero-crossings of the wheel), was adversely affected by sleep deprivation. More specifically, they found that the steering wheel reversing rate was lower for the sleep-deprived drivers than control drivers, indicating a reduced level of steering input and correction among these drivers. Research by Skipper, Wierwille and Hardee (1984, cited in Wierwille et al, 1994) has also examined steering control among sleep-deprived drivers and its correlation to eye closure. They found that steering velocity was influenced by sleep deprivation and was highly correlated with eye closure. Fairclough (1997) also found that the magnitude, but not the frequency, of steering change increased in the sleep deprivation condition. The results also revealed that the frequency of steering wheel movements was positively correlated with EEG indicators and subjective measures of drowsiness.

2.2.3 Speed and Longitudinal Control

Measures of speed control, including mean speed and speed maintenance or variability, have not demonstrated consistent results in the literature in terms of their sensitivity to drowsiness and some researchers have suggested not using such measures as indicators of drowsiness (Haworth, et. al, 1988). Earlier research by Mast, Jones and Hiemstra (1966, cited in Wierwille et al, 1994) found that drivers' ability to maintain a constant velocity reduced during the last hour of simulated drives that lasted for four or six hours. However, research by Riemersma, Sanders, Wildervanck and Gaillard (1977, cited in Haworth et al., 1988) failed to find a significant increase in speed variability in drowsy drivers, despite finding a significant effect of drowsiness on measures of steering control. More recent research by Fairclough and Graham (1999) and Arnedt et al. (1999) demonstrated that speed variability was higher for sleep-deprived drivers than for control drivers. The

inconsistencies in these findings may be attributable to differences in the driving tasks or the length of time that the drivers were required to drive for.

Research examining mean speed as an indicator of drowsiness has also revealed inconsistent results. An on-road study by Fairclough (1997) found that drivers in the sleep-deprived condition drove at significantly slower speeds during the test drive than they did in the non sleep-deprived condition. In contrast, Peters and colleagues found that mean speed *increased* with progressive sleep deprivation in lower speed zones (35 miles per hour), but not in the higher zones (55 miles per hour). Differences between these two studies may be due to the fact that in the Fairclough study, drivers were tested on actual roads, whereas participants in the Peters et al. study were tested in a simulator, and therefore may have travelled faster (to perhaps complete the drive faster) because the consequences for doing so were less.

Longitudinal control, as measured by the ability to maintain a safe and constant time headway from the vehicle ahead, is a measure that has also been found to be sensitive to driver drowsiness. Muto and Wierwille (1982) found that the time it took driver's to react to the sudden braking of a lead vehicle increased significantly after driving for 30, 60 and 150 minutes compared to baseline drives. More recently, Fairclough and Graham (1999) found that mean time headway decreased by 0.5 seconds in the last 40 minutes of the test drive compared to the first 40 minutes. They also found that drivers in the sleep-deprived condition maintained longer time headways from the vehicle ahead than did the alcohol-impaired drivers.

2.3 AIMS AND HYPOTHESES

The aim of the current study was to determine whether there is a breakdown in driving performance during the period after which the JDS reaches the cautionary level (4.5) and the critical level (5.0), that is, after the driver is classified as drowsy by the Optalert system. The MUARC driving simulator was used to investigate how driving performance changed during periods in which the Optalert system classified the driver as too drowsy to drive, compared to periods when the driver was not classified as too drowsy to drive. To ensure that a range of drowsiness states was observed throughout the study, participants drove the simulator in both alert and sleep-deprived states. So as not to arouse drowsy participants, the Optalert system warnings were not active within this experiment, that is, the system was only used to measure drowsiness, not to warn the driver of their drowsiness.

Measures relating to the lateral placement of the vehicle within the lane were identified as the driving performance measures of primary interest in this study. These are also intuitively appealing considering that drowsiness-related crashes often involve a vehicle running off the road or into the path of an oncoming vehicle. It was deemed preferable to measure lane excursions rather than simply run-off-road incidents, because crossing the centre line can be just as dangerous as running off the road in terms of road crashes. As well as discrete lane excursions, lateral placement and variability of lateral placement were also measured continuously throughout the study. It was hypothesised that, during periods that drivers were classified as drowsy by the Optalert system, lane excursions would increase in rate (per minute) and duration, and that lane variability would increase, relative to periods in which the Optalert system did not classify the driver as drowsy. In addition, the data were analysed using signal detection theory in order to determine the sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) of the

Optalert system in identifying periods in which lane excursions occurred. A discussion of Signal Detection Theory and definitions of the measures that it provides are provided in more detail in the results section of this report.

Despite previous inconsistent results, speed and speed control were also measured to determine if there was a relationship to drowsiness. In addition, mean time headway while following another vehicle was measured to determine if this varied across periods of drowsiness. It was decided not to include steering variability as a dependent measure in this study, as this is intrinsically related to lateral position and to lane excursions, which were of primary interest in this study.

3 METHOD

3.1 PARTICIPANTS

Twenty-three participants were recruited to take part in the study and 20 participants presented for at least one session. Of the remaining three participants, one withdrew prior to attendance because they were going on holiday, while the other two withdrew for reasons unspecified. The twenty participants who attended at least one session comprised 15 males and 5 females, with an average age of 23.5 years (SD=2.9 years). Eighteen participants were full-time students, while the other two participants were in full-time employment. Four of the full-time students were also employed, one part-time and three full-time. All participants held a full driver's license. The average length of time that participants had held their license was 5.5 years (SD=2.8 years), while the average time spent driving each week was 9.0 hours (SD=5.3 hours). Participants reported predominantly driving in metropolitan and residential environments.

One participant withdrew from the study at the beginning of the first session, while two participants withdrew from the study after completing one test session. The first-mentioned participant experienced symptoms of simulator sickness after only a short time driving the simulator, and for this reason the session was terminated. Of the two participants who did not attend their second test day, one informed the experimenters that they wished to drop out of the study, while the other simply did not attend on their scheduled day. No reasons were given.

Participants were recruited from the MUARC subject database, which includes the names and contact details of people who have expressed interest in participating in research conducted in the driving simulator. The database includes students from the university and members of the general public. Recruitment advertisements were also placed in the Monash University Student Employment Service.

Several exclusion criteria were applied when recruiting participants. Shift workers and people with diagnosed sleeping disorders were excluded due to the possibility of having disrupted circadian rhythms, and the potential to respond differently after a night of sleep deprivation. Sufferers of severe motion sickness or epilepsy were prevented from participating because of the potential for adverse reactions in the driving simulator. People who wear spectacles to drive could not participate, as the Optalert driver drowsiness detection system is currently incompatible with spectacles. People with narrow eyes were also excluded because their eye-shape made it difficult to calibrate the Optalert system. Finally, participants who resided further than 10km, or 15 minutes, from the Accident Research Centre were also excluded, in an effort to contain costs associated with participant transport. Participants were paid \$200 as reimbursement for their time and travel costs.

3.2 APPARATUS

3.2.1 Driving simulator

The study was conducted using the MUARC advanced driving simulator. The simulator consists of a full Holden Calais car body, in front of which is a curved screen with a 180 degree lateral viewing angle from the driver's position. A flat screen is located behind the car for simulation of 60-degree rear view scenes, however this was not used in the current

experiment. Two Silicon Graphics computers power the system, and the images are projected using Barco video projectors. The simulator also includes a 3-dimensional audio system and a partial motion platform. The simulator is operated from a control room, where the experimenter remains for the time that the participant is driving the simulator. There is a two-way communication set-up between the control room and the simulator, and the participant was observed via a television screen. The driving simulator samples a wide variety of driver behavioural data, 30 times per second.

3.2.2 Optalert drowsiness detection system

Driver drowsiness was measured continuously throughout the simulated drives using the Optalert drowsiness detection system. The system consists of regular spectacle frames equipped with light emitters and sensors, which are used to measure eye and eye-lid movements.

3.2.3 FaceLab

The Seeing Machines faceLAB system measured the percent of time the eyelids were closed (PERCLOS) throughout the simulated drives, as well as eye and head movement. This system uses a video-based sensor to record real-time eye and head movements and eyelid position. An infra-red light source was used to ensure that there was enough light in the car cockpit for the faceLAB system to operate correctly in conjunction with the Optalert spectacles.

3.2.4 Other equipment

Images of the drivers' face and the driving scene were recorded using a Panasonic DVR-720-H-s DVD recorder (owned by Sleep Diagnostics Pty Ltd), throughout the driving scenarios.

3.3 DESIGN

Each participant took part in the experiment on two separate days, one after a night of normal sleep, and one after a night with no sleep since 8am the morning before. There were two test drive scenarios, which differed only in the order that events occurred. The order of events was varied to reduce the predictability of event occurrence. The order in which participants attended for the sleep-deprived and alert days, and the order of the different test drives, were counterbalanced across participants to control for order/practice effects.

3.4 DRIVING SCENARIOS

The test drive scenarios consisted of a two-lane rural road, with one lane in each direction. The lane width was 3.5m and there was a broken centre line and lateral edge lines. A gravel road edge extended for 3.5m from the edge of the road, with the last two metres being "rough" gravel, which provided tactile and audio feedback to participants if they ran off the road. This was particularly important to avoid the difficulties that would arise if participants fell asleep and drove off the road, getting lost in the scenario database. It was felt that having the rough gravel road edge was a more realistic option than having the experimenter verbally advise the participant that they must return to the road.

The driving scenario was a loop of approximately 20 minutes, which was driven four times continuously throughout the session. The signed speed limit for most of the scenario was 80km/h, except for a 500 m section of the road that ran through a small town, where the speed limit reduced to 60km/h. The sessions took approximately 70 minutes to drive when travelling at the signed speed limit.

There were 16 events throughout the driving scenario (i.e. four per loop), the order of which was varied between drives. The order of events was somewhat constrained, however, by the inability to change the layout of the database between loops. As such, the location of intersections and towns could not be changed. However, the appearance of other vehicles at intersections, and the behaviour of these vehicles, could vary between loops. Within each loop were two crossroads where the driver had right of way (ROW), one town, and one intersection where the driver did not have ROW. The absence or presence of other vehicles and their behaviour at the crossroads with ROW varied across loops.

The events that occurred throughout the whole driving scenario were:

1. Two cross-roads where the driver had right of way (ROW) with no other vehicle approaching,
2. Two ROW crossroads with an approaching vehicle that stopped at the intersection and gave way to the simulator driver,
3. Two ROW cross-roads at which a vehicle approached the intersection, pulled out in front of the driver and travelled slower than the simulator driver was travelling, thus initiating a car-following episode. The driver was required to maintain a safe following distance until the lead car pulled off the road,
4. Two ROW crossroads at which a vehicle approached the intersection, pulled out in front of the driver and travelled faster than the simulator driver,
5. Four crossroads with a stop sign where the driver did not have ROW, and
6. Four towns, with a reduced speed limit of 60km/h.

For the events in which a vehicle approached the intersection, the direction of approach (from the left or from the right) was varied randomly between events.

3.5 VARIABLES MEASURED THROUGHOUT THE DRIVING SCENARIO

3.5.1 Drowsiness

Drowsiness was measured continuously by the Optalert drowsiness detection system. This system measures a large number of parameters relating to eye and eye-lid movement, from which the Johns Drowsiness Scale (JDS) is derived. The exact parameters measured and the method by which the JDS is derived was not revealed to MUARC. For the analysis to be undertaken, Sleep Diagnostics Pty Ltd provided MUARC with minute-by-minute JDS scores for each participant throughout the experimental sessions. As a second measure of drowsiness, the percent of time the eyelids were closed (PERCLOS) was measured throughout the driving sessions using the FaceLAB system. Participants' eye and head movement were also recorded using FaceLAB.

3.5.2 Driving performance

Driving performance data was extracted from the raw data files as average scores per minute of driving, to enable comparison with the per-minute JDS data. Of primary interest was the ability of participants to keep within their lane. The lane position was measured as the number of metres that the central point of the wheel axle of the vehicle was displaced from the middle of the lane. From this the mean lateral lane position per minute was determined, as was the variability in lateral lane position, in terms of the standard deviation of lateral lane position. Lane excursions were also identified from this data. Three separate definitions of lane excursions were used. The broadest definition was when any part of the car was out of the lane, that is, all lane excursions. A severe lane excursion was defined as when the entire vehicle was out of the lane. A moderate lane excursion was defined as when half or more of the vehicle was out of the lane.

Other driving performance measures included speed control and mean time headway when following other vehicles.

3.6 PROCEDURE

People who expressed interest in participating and met the inclusion and exclusion criteria of the study were informed in detail about the nature and purpose of the research, and provided with an explanatory statement and consent form. Those that still wanted to participate were scheduled for two test days, which were held at least three days apart. The period between test days was chosen so that those participants who took part in the sleep-deprived condition in the first session could obtain restorative sleep before taking part in the alert session.

Participants were given instructions to follow for the day of testing and the previous day, if applicable. For both the alert and sleep-deprived sessions, they were asked not to consume any alcohol, caffeine or stimulants after 10pm the previous night. For the alert session, participants were asked to have a normal sleep the night before testing, of approximately seven to eight hours duration and to wake no later than 8am on the day of testing. For the sleep-deprived session, participants were asked not to have any sleep, including naps, from 8am the previous day¹. For the night *prior to* the night of sleep deprivation, participants were asked to have a normal night's sleep of seven to eight hours duration. The participants were given suggestions for activities to keep them awake, such as playing card games, computer games, going for a walk, and enlisting the help of friends and family. It was suggested that participants refrain from watching television, which can be conducive to sleep. They were informed of the dangers of operating heavy machinery or driving whilst deprived of sleep. Participants attending for their sleep-deprived session were asked to travel to MUARC by taxi, and the taxi-fare was paid for on their arrival.

In general, the alert sessions began between 11:30am to 12pm, whereas the sleep-deprived sessions began at 9:30am. Although it would have been ideal for all sessions to be conducted at the same time of day, constraints on the availability of the simulator precluded this. It also was not possible to counterbalance the times for the alert and sleep-deprived sessions, as the sleep-deprived participants were required to return at 2pm for

¹ Before participating in the sessions, participants were questioned about their sleep behaviour over the previous night. However, since participants were not supervised during the night, these subjective reports could not be verified.

another related experiment. By conducting their first session that day at 9:30am, they were able to have a break in between simulator sessions on that day.

Upon arrival for their first experimental session, participants completed questionnaires relating to demographic information, simulator sickness symptoms, sleep quality, the Karolinska sleepiness scale and the Epworth sleepiness scale. Participants were seated in the driving simulator vehicle, and drove through a short scenario to become familiar with the controls and handling characteristics of the simulator vehicle.

After the familiarisation drive the FaceLAB system was set up to monitor the participant's eye movements, head movements and the percentage of time that the eyelids were closed. The Optalert spectacles were fitted and the data checked for quality. Participants were informed of the possibility of simulator sickness and asked to report any symptoms if they occurred. Participants were instructed to drive according to the road rules and road signs, to drive as closely as possible to the posted speed limit, to drive straight ahead when they came to cross-roads, and to follow any vehicle in front of them at a safe distance. At the beginning of each driving simulation session, the data logging systems for the driving simulator and the Optalert driver drowsiness detection system were initiated at the same time to ensure the data could be matched for future analysis. Once all of this was completed, the participants began the test drive.

While driving the test route, participants were monitored from within the control room for signs of simulator sickness. Their driving was also monitored to ensure that there were no instances where they drove so far off the road that they could not navigate their way back on to the road.

The simulator scenario was complete when the participant had driven four laps of the circuit. A post-drive questionnaire relating to simulator sickness symptoms and the comfort of the Optalert system was also administered. Participants were paid for their participation and, when relevant, given instructions for the next test session. Sleep-deprived participants were reminded of the dangers of driving or operating heavy machinery whilst in that state. At the end of the sleep-deprived session a taxi provided transport for the participant back to their home address.

3.7 ETHICS

The Monash University Standing Committee on Ethics in Research involving Humans (SCERH) granted ethical approval for this study.

3.8 DATA ANALYSIS

After all participants had completed the experimental sessions, the data were parsed and extracted from the raw data files. The data were checked for outliers, and these were further investigated to determine their validity. Where appropriate, data were transformed to enable parametric analyses to be performed.

4 RESULTS

Of the 20 participants who attended for at least one session, complete and analysable sets of data (i.e. two complete sessions of data from both the simulator and the Optalert system) were available for 12 participants, while partial sets of data were obtained for 6 participants. Two participants did not provide any usable data, and none of the data from these two participants is included in any of the analyses (including the surveys), apart from the post-drive questionnaire regarding Optalert comfort.

The six partial data sets comprised the following:

- Two participants provided complete data for the alert session, but no analysable data from the sleep-deprived session. During one of the participant's sleep-deprived session, the steering control in the simulator vehicle malfunctioned, which resulted in premature cessation of the session, and uncertain quality of the data that had been collected to that point. For this reason, the data from this session were not included in the analysis. The other participants did not attend for their sleep-deprived session because they dropped out of the experiment.
- For one participant, complete data were available for the sleep-deprived session, but not for the alert session due to the participants withdrawing from the study.
- Complete data were available for another participant's sleep-deprived session. However, approximately half of the Optalert data for the alert session were missing.
- Another participant provided complete data for the alert session, however only 30 minutes of analysable data was collected for the sleep-deprived session. After 30 minutes in this session, this participant disregarded pre-drive instructions and turned at an intersection. This led to them becoming "lost" in the database, which affected data collection for lane keeping because it was programmed according to the road they were supposed to be travelling on. This session was then terminated.
- The data sets available for a further participant were missing some data. In the alert session, there was no Optalert data from minute 58 until the end of the session at minute 68. In the sleep-deprived session, there was missing data from the simulator from minutes 8 to 13, and missing data from the Optalert system between minutes 13 and 33.

Participants with no usable data:

- One participant displayed erratic behaviour during all sessions, and did not comply with experimental instructions. Although this participant completed both experimental sessions and data were successfully collected, this data were not included in the analyses.
- One participant provided no data, because their session had to be terminated early within the trial due to simulator sickness.

After examination of the data, it became apparent that the data obtained from the FaceLAB system were not of sufficient quality to analyse. Several participants slouched in the driver's seat, or sat with their head resting back against the headrest, particularly in the

sleep-deprived session. When this occurred, the FaceLAB system was unable to collect data that was of appropriate quality for analysis. For other participants, the contrast between their pupil and iris was too low, and the FaceLAB system tracked the frame of the Optalert glasses instead of the participant's eye movements. Even for participants where neither of these situations occurred, the FaceLAB system only achieved a tracking accuracy of between 50% to 60%. For these reasons, the data from the FaceLAB system were not further analysed within this study.

4.1 QUESTIONNAIRES

4.1.1 Pre-Drive

Participants completed questionnaires relating to demographic information, simulator sickness symptoms, sleep quality and duration, and subjective sleepiness (using the Karolinska Sleepiness Scale) prior to both sessions. They were also asked to rate their usual sleep propensity prior to the first test session, using the Epworth Sleepiness Scale.

4.1.1.1 Duration and quality of sleep

Participants were asked to indicate how many hours of sleep they had had the last night that they slept and to rate the quality of their sleep on that night, on a scale of 1 (very poor) to 10 (very good). Four participants in the sleep-deprived session misread the question as asking how many hours of sleep they had the previous night, *not* the last night that they slept. These participants have been excluded from the following table.

Table 1: Duration and quality of sleep on the last night that participants slept, prior to the test session: mean (standard deviation) and range

	Alert session	Sleep-Deprived session
Duration of sleep	8.12 hours (0.99) Range: 7-10, n=17	7.64 hours (1.91) Range: 5-11, n=14
Quality of sleep	7.59 (1.12) Range: 5-10, n=17	7.43 (2.34) Range: 1-10, n=14

A paired t-test was conducted to determine if there was a difference between alert and sleep-deprived sessions in terms of the number of hours of sleep participants obtained on the last night that they slept, and the quality of that sleep. Being a paired analysis, this was restricted to the 13 participants who answered the relevant questions in both sessions. The sleep quality data had to be transformed (squared) to achieve normality, although the untransformed data is reported in Table 1. There was no significant difference between the alert and sleep-deprived sessions in terms of the number of hours of sleep that participants obtained the last night that they slept ($t(12)=1.88$, $p>0.05$), or the quality of that sleep ($t(12)=-0.04$, $p>0.05$).

4.1.1.2 Caffeine consumption

Participants were asked to report if they had consumed any caffeine or alcohol within the previous 24 hours and, if so, what type and in what quantity. Six participants reported consuming caffeine or alcohol in the 24 hours prior to the alert session, while seven reported doing so prior to the sleep-deprived session. Table 2 shows the reported caffeine use of these six participants.

Table 2. Reported caffeine use in the 24 hours prior to testing

	No. participants reporting caffeine or alcohol consumption in the 24 hours prior to testing	Details of alcohol/caffeine consumption
Alert session	6 (of 17)	-1 cup of tea (3 participants) -3 alcoholic drinks at 8pm the night before, and one tea on the day of test -1 Coca-cola the previous night -not reported
Sleep-deprived session	7 (of 17)	-Red bull 12pm the day prior -40 ml Coca-cola -1.5 standard drinks and 1 sustagen -small coke -1 cup (type undefined) -2 glasses of Coca-cola -1 Coca-cola the previous night

4.1.1.3 Subjective sleepiness

Prior to each session, participants were asked to rate how sleepy they felt during the preceding 10 minutes using the Karolinska Sleepiness Scale (KSS). The KSS is a rating scale from 1 (extremely alert) to 10 (extremely sleepy, can't keep awake). As expected, participants reported feeling significantly more sleepy after sleep deprivation than after a full night of sleep ($t(14)=-7.92, p<0.001$) (see Figure 1).

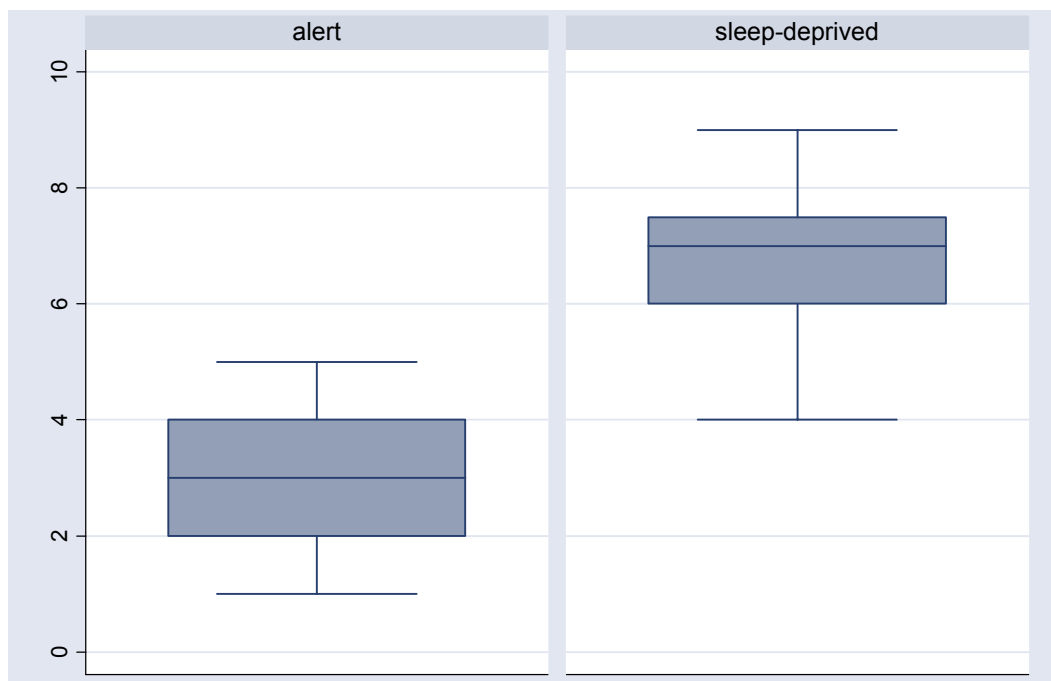


Figure 1. Participants' ratings of subjective sleepiness using the Karolinska Sleepiness Scale, prior to the alert and sleep-deprived experimental sessions

4.1.1.4 General sleep propensity

The Epworth Sleepiness Scale (ESS) was administered to participants prior to the alert session to determine their usual propensity to fall asleep. In the ESS, participants are asked to rate how likely they are to fall asleep in eight situations:

1. sitting and reading,
2. watching TV,
3. sitting inactive in a public place – e.g. a theatre or meeting,
4. as a passenger in a car for an hour without a break,
5. lying down to rest in the afternoon when circumstances permit,
6. sitting and talking to someone,
7. sitting quietly after a lunch without alcohol, and
8. in a car while stopped for a few minutes in the traffic.

The ratings range from zero (would never doze) to 3 (high chance of dozing). The responses for each of the situations are summed to give a total sleep propensity score. The ESS scores for participants in this study ranged from zero to nine, with a median score of 5.5 (see Figure 2). All participants' scores were within the normal range.

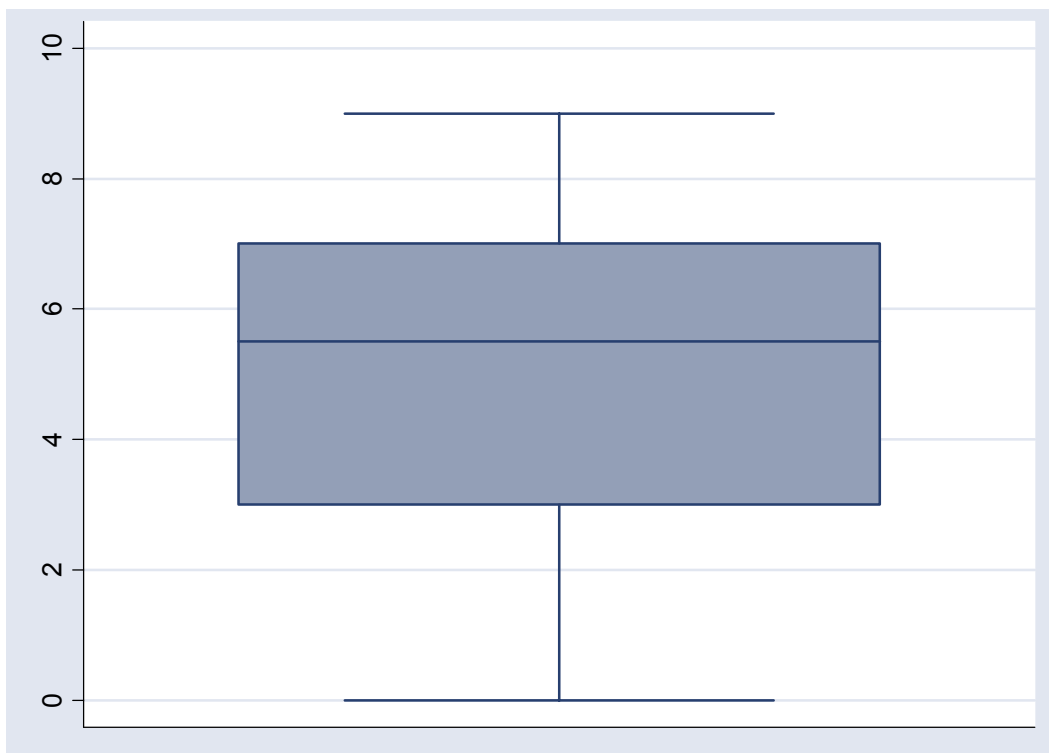


Figure 2. Participants' scores on the Epworth Sleepiness Scale

4.1.1.5 Post-drive

After each session, participants completed a questionnaire regarding the comfort of the Optalert system and its impact on their driving ability. All participants' responses to the post-drive questionnaire are included, except for Subject 18, who suffered from simulator sickness. This participant's session was terminated early, and it was judged that they did not have enough experience with the system to rate how comfortable it was.

Participants were asked to rate the comfort of the Optalert glasses, on a scale from 1 (extremely uncomfortable) to 10 (extremely comfortable). Figure 3 shows that most participants found the Optalert system comfortable to wear, with a median comfort rating of 8 after both the alert session and the sleep-deprived session.

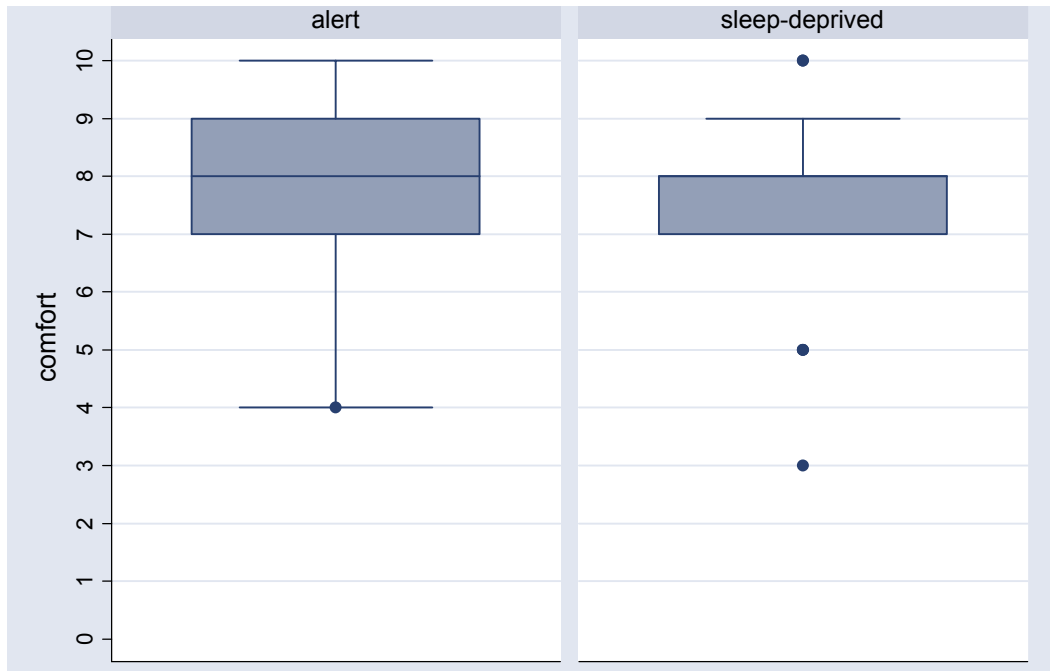


Figure 3. Participants' ratings of the comfort of the Optalert system, after the alert and sleep-deprived sessions (0=extremely uncomfortable, 10=extremely comfortable).

Participants were asked if they believed their ability to drive was impaired by the Optalert glasses during the driving simulator session and, if so, to specify what caused that impairment. The majority of participants (12 of 17 in the alert session and 13 of 17 in the sleep-deprived session) did not believe the Optalert glasses impaired their driving ability. Three participants believed their ability to drive was impaired in both sessions by the Optalert system. Two of these responded that the Optalert frames blocked the visual field, while one felt impaired by the cable dangling over their shoulder.

A further three participants believed the Optalert system impaired them during their first driving session with the system. Of these three participants, one attributed the impairment to the weight of the system, one stated that they were not used to wearing glasses, while the remaining participant felt the Optalert system caused blurry vision.

4.2 DRIVING PERFORMANCE

4.2.1 Relationship between driving performance and continuous JDS

The driving performance data collected during the simulation sessions were compared to the Optalert drowsiness measurements to investigate the relationship between drowsiness and driving performance, in terms of lateral lane position, lane position variability, speed and speed variability. Because the raw data had already been summarised as per-minute measurements, it was not appropriate to use scatter plots to illustrate the relationship between drowsiness and driving performance. Instead, the continuous drowsiness data were categorised (see Table 3) and line graphs of the average driving performance for each JDS category were plotted. The ten categories of JDS are defined in Table 3.

When inspecting the following graphs, it is important to keep in mind that, although they illustrate the relationship between drowsiness (as measured by the JDS) and driving performance pooled across all participants, they do not account for the fact that not all participants attended for both the alert and sleep-deprived sessions, and not all subjects experienced drowsiness states in the high ranges.

Table 3. Categorisation of Johns Drowsiness Scale

Category	1	2	3	4	5	6	7	8	9
JDS	0.5- 1.4	1.5- 2.4	2.5- 3.4	3.5- 4.4	4.5- 5.4	5.5- 6.4	6.5- 7.4	7.5- 8.4	8.5- 9.5

4.2.1.1 Absolute lateral lane position

Figure 4 shows that, in general, the distance of the centre of the car from the middle of the lane increases as driver drowsiness increases. This relationship appears to vary according to whether the driver is sleep-deprived or not; for JDS categories of 4 and above, there is a trend for the average distance from the middle of the lane to be greater in the sleep-deprived session than the alert session for any given JDS.

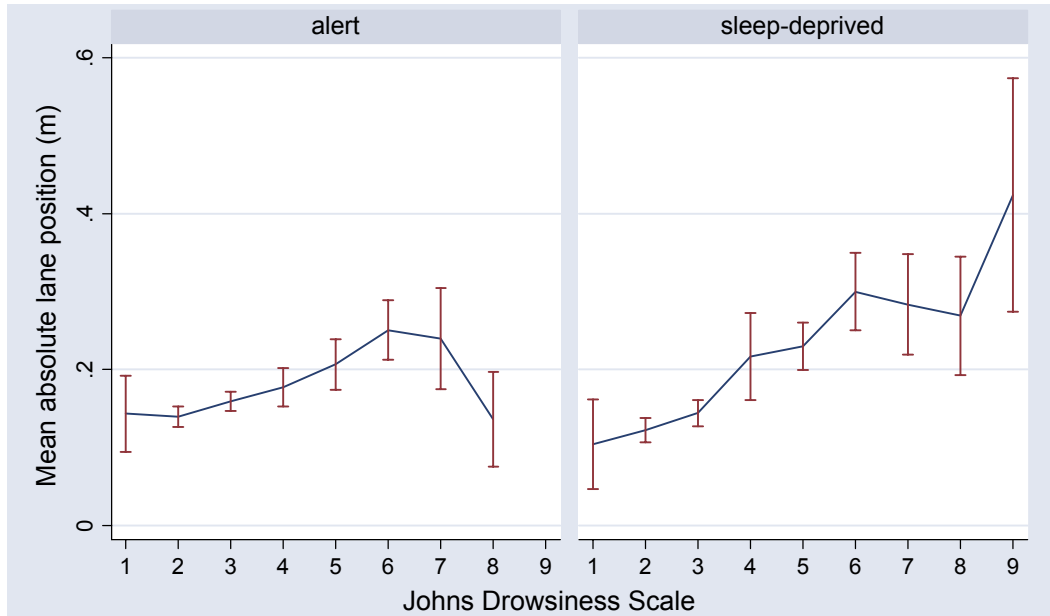


Figure 4. Mean lateral lane position across categories of JDS for the alert and sleep-deprived conditions

4.2.1.2 Variability in lane position

Figure 5 shows that, as drowsiness increases the variability in lane position increases. This relationship appears to vary according to whether the driver is sleep-deprived or not; for JDS categories of 4 and above, there is a trend for the standard deviation of lane position to be greater in the sleep-deprived session than the alert session for any given JDS.

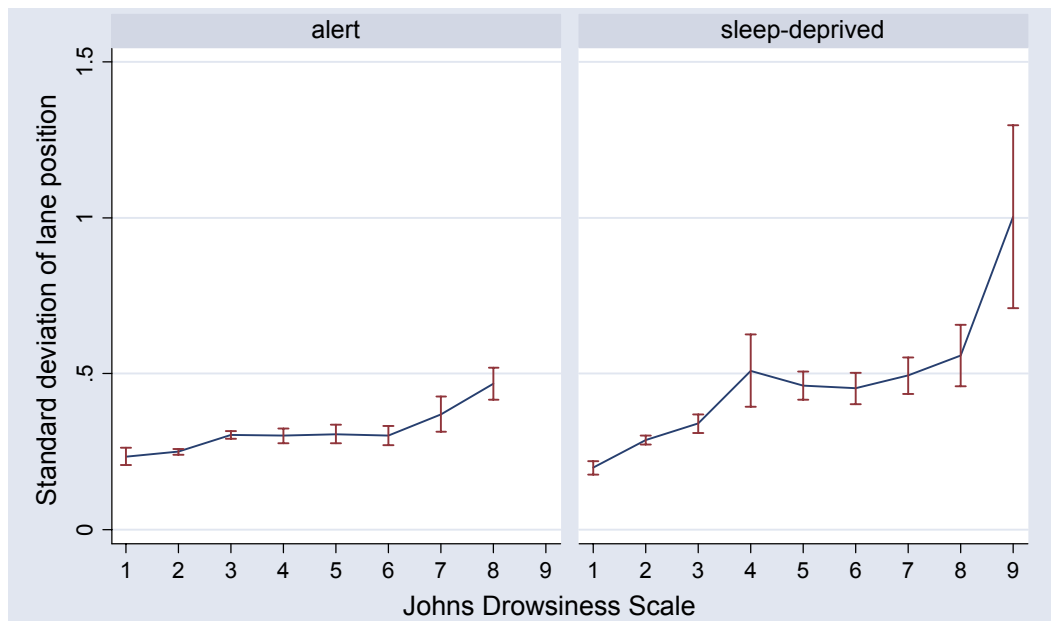


Figure 5. Standard deviation of lateral lane position across categories of JDS for the alert and sleep-deprived conditions

4.2.1.3 Speed

Figure 6 shows that there appears to be little effect of drowsiness on average speed. In addition, average speed does not appear to vary according to sleep status.

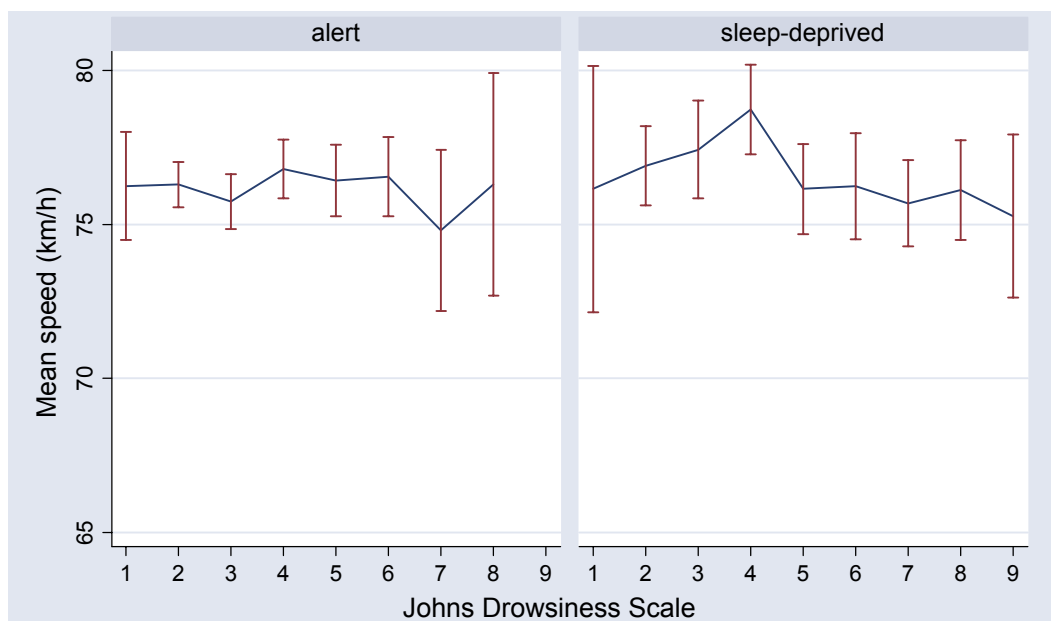


Figure 6. Average speed across categories of JDS for the alert and sleep-deprived conditions

4.2.1.4 Speed variability

There was a trend for speed control to be more variable in the sleep-deprived condition than the alert condition (see Figure 7). Figure 7 shows that there may be an inverse U-

shaped relationship between the JDS and variability of speed control, with speed variability being lower at very low (around 1) and very high (around 8 or 9) JDS scores, and higher throughout the rest of the JDS range.

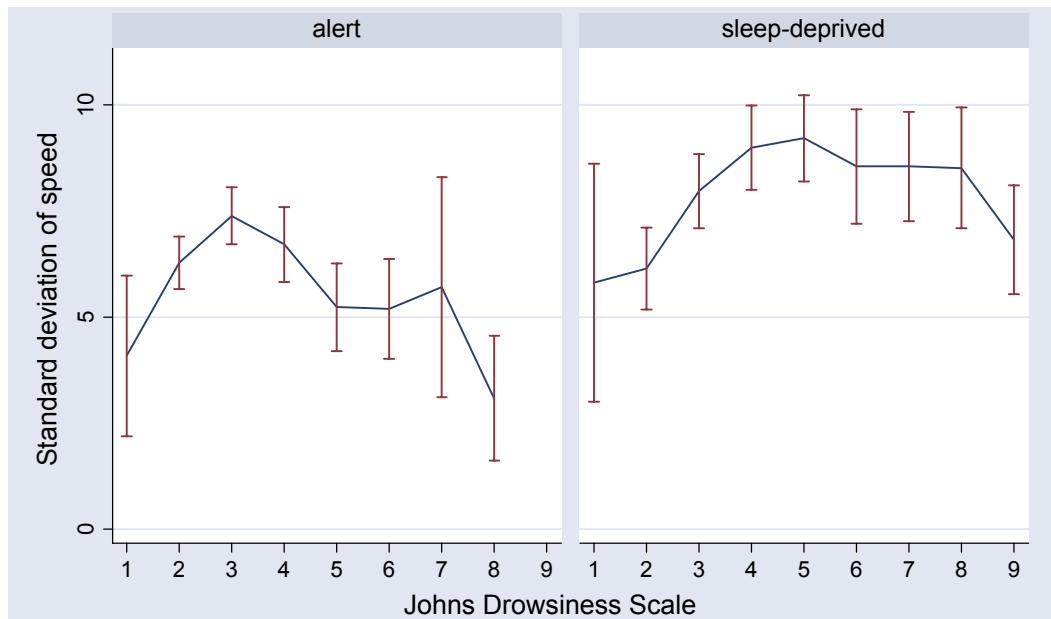


Figure 7. Standard deviation of speed across categories of JDS for the alert and sleep-deprived conditions

4.2.2 Driving performance during periods when the driver was classified as drowsy compared to when periods they were not classified as drowsy

For the main analyses conducted on the data, each of the driving sessions was divided into periods in which the Optalert system had classified the driver as too drowsy to drive and periods in which the driver was not classified as drowsy. Then it was determined whether or not driving performance differed in periods when the drivers were drowsy to periods when the drivers were not drowsy. Several criteria were used to identify periods in which a driver was classified as drowsy.

Previous research with the Optalert system (conducted by Sleep Diagnostics) has identified two JDS levels of interest. The first occurs when the JDS reaches 4.5, which is the level at which the driver would be cautioned that they are becoming drowsy. The second occurs when the JDS reaches 5.0, which is the level at which the driver would be instructed that they are too drowsy to drive.

A performance breakdown due to drowsiness would be expected to occur during a specified period after either of the levels of interest has been reached. Sleep Diagnostics have specified 30 minutes as the period of interest, i.e. the “prediction window” in which drowsiness is expected to affect driving performance. Two other prediction windows were also investigated, the durations of which were 15 minutes and 5 minutes.

Thus, there were six different ways to define a period in which drowsiness would be expected to affect driving performance:

- either 5, 15 or 30 minutes after the cautionary JDS level (4.5) was reached; and

- either 5, 15 or 30 minutes after the critical JDS level (5.0) was reached.

The sessions were divided into drowsy and non-drowsy periods in these six different ways, and separate analyses performed for each of the six prediction windows by JDS level periods. The analyses detailed below are restricted to times during each session where data from both the driving simulator and the Optalert system were available.

The proportion of total lane excursions predicted by the Optalert system, and the sensitivity, specificity, and positive and negative predictive value of the JDS in correctly identifying periods in which a lane excursion occurred is presented.

Additionally, standard statistical tests were used to determine if driving performance, in terms of variability of lateral lane position, the rate (per minute) of lane excursions, the average duration per period that the vehicle was travelling out of the lane, and the mean time headway when following another vehicle, varied between periods in which the driver was classified as drowsy compared to periods in which the driver was not classified as drowsy. It was decided not to further investigate average lateral position, average speed or standard deviation of speed. This was because, from previous research and from the continuous data presented previously, it was felt that lane position variability is a better measure to relate to drowsiness than average lane position. Furthermore, average speed did not appear to vary with drowsiness, while the speed variability displayed a relationship with drowsiness that would not be sensible to categorise in this way.

4.2.2.1 Frequency of lane excursions & proportion predicted by Optalert system

Three definitions of lane excursions were used. The first (all events) was when any part of the car departed from the lane. The second (half events) was when at least half of the car departed from the lane. The third (severe events) was defined as when the entire car departed the lane.

Initial descriptive analyses of the data revealed that all participants made lane excursions, however the majority of these involved less than half of the car exceeding the lane boundaries. Fourteen of the eighteen participants made lane excursions where at least half of the car departed the lane. Six of these fourteen participants only made one such excursion, while six of the participants made more than 10 such excursions. Lane excursions where the entire car departed the lane were less frequent, with only seven participants making such excursions, and only five participants doing so more than twice.

The total number of each type of lane excursion in each experimental session was determined for each participant. Table 4 displays the number of lane excursions in which the entire car was out of the lane, for each session and overall, pooled across participants. Table 5 displays the same data for lane excursions where at least half of the car was out of the lane, while Table 6 shows lane excursions where any part of the car left the lane. In addition, the proportion of lane excursions that occurred during periods where the driver was classified as drowsy was determined, and is displayed in each of the tables.

Inspection of the tables reveals what appears to be an odd result. The number of events appears to differ according to the JDS level/time window combination used for prediction, even though all of the data were from the same driving simulation sessions. This is not actually the case – the number of events was constant, however, the way that the sessions were divided into prediction periods meant that, for some combinations, some of the data had to be disregarded. Only prediction periods which had associated driving performance

data for the whole specified Optalert prediction window of interest were used. For example, if the JDS reached the cautionary level 10 minutes prior to the end of the driving simulation session, the following ten minutes of data were disregarded from the 30 minute and 15 minute prediction windows. However, they were still used for the 5 minute prediction windows. This was to ensure that driving performance for the whole prediction period could be observed, otherwise it may have been claimed that the Optalert system did not predict an event within 30 minutes, even though there was only 10 minutes of observable driving behaviour. By disregarding non-complete prediction periods, the false positive rate (i.e. periods where the Optalert system classified the driver was drowsy and that a lane excursion would occur and no such event occurred) was not artificially inflated.

Table 4 shows that the vast majority (87-97%) of occasions when the entire car left the lane occurred during periods where the Optalert system classified the driver as drowsy. A comparison of the percentage of predicted events in the sleep-deprived and alert sessions reveals that the Optalert system predicted a greater proportion of the lane excursions that occurred when the participants were sleep-deprived (83%-98%) than the lane excursions that occurred when participants were not sleep-deprived (50%-86%).

Table 4. Number of occasions, and the proportion of events that occurred when the driver was classified as drowsy by Optalert, where the entire car was out of the lane for alert session, sleep-deprived session and overall

Prediction period: JDS level and time window		Excursions where the entire car exceeded the lane boundaries					
JDS Level	Time window	Total number	% when classified drowsy	No. in sleep-deprived session	% when classified drowsy	No in alert session	% when classified drowsy
4.5	5 min	110	87.3	103	88.3	7	71.4
4.5	15 min	104	97.1	98	98.0	6	83.3
4.5	30 min	115	97.4	108	98.1	7	85.7
5.0	5 min	112	80.4	104	82.7	8	50.0
5.0	15 min	105	95.2	99	96.0	6	83.3
5.0	30 min	101	95.0	94	95.7	7	85.7

Comparing across the different JDS level/time window prediction combinations, choosing a JDS level of 4.5 as the cut-off to divide the time into periods when the driver is drowsy leads to a greater proportion of entire car lane departures being predicted than the choice of a JDS level of 5.0 for the cut-off. Regarding the most effective time-window for prediction, 30 minute time windows had a slight edge over 15 minute windows in terms of the proportion of events that occurred when the driver was classified as drowsy, with a bigger gap between the 15 minute windows and the 5 minute windows.

Table 5 reveals that, similar to entire car lane excursions, the majority (77%-96%) of lane excursions where at least half of the car departed the lane occurred when the driver was classified as drowsy by the Optalert system. Again, a greater proportion of the events that occurred during the sleep-deprived session happened when the driver was classified as drowsy (79%-97%) compared with the alert session (57%-86%).

Similar to the more severe lane excursions, choosing a cut-off level of 4.5 for classifying the driver as drowsy led to prediction of a greater proportion of events where at least half the car left the lane than if a cut-off level of 5.0 was chosen. Again, choosing a prediction time window of 30 minutes or 15 minutes was more effective than a time window of 5 minutes.

Table 5. Number of occasions, and the proportion of events that occurred when the driver was classified as drowsy by Optalert, where at least half of the car was out of the lane for alert session, sleep-deprived session and overall

Prediction period: JDS level and time window		Excursions where at least 1/2 of the car exceeded lane boundaries					
JDS Level	Time window	Total number	% when classified drowsy	No. in sleep-deprived session	% when classified drowsy	No in alert session	% when classified drowsy
4.5	5 min	281	88.3	252	90.1	29	72.4
4.5	15 min	268	95.5	242	96.7	26	84.6
4.5	30 min	292	95.9	264	97.0	28	85.7
5.0	5 min	291	76.6	261	78.9	30	56.7
5.0	15 min	272	91.2	246	91.9	26	84.6
5.0	30 min	267	92.5	239	93.3	28	85.7

Table 6. Number of occasions, and the proportion of events that occurred when the driver was classified as drowsy by Optalert, where any part of the car was out of the lane for alert session, sleep-deprived session and overall

Prediction period: JDS level and time window		Excursions where any part of the car exceeded lane boundaries					
JDS Level	Time window	Total number	% when classified drowsy	No. in sleep-deprived session	% when classified drowsy	No in alert session	% when classified drowsy
4.5	5 min	2092	72.3	1439	81.7	653	51.8
4.5	15 min	1981	78.4	1377	86.4	607	60.1
4.5	30 min	1999	79.5	1411	87.3	588	60.7
5.0	5 min	2107	63.7	1453	71.2	654	47.1
5.0	15 min	2020	73.0	1125	79.5	603	57.7
5.0	30 min	1912	72.4	1324	79.4	588	56.6

The Optalert system had classified the driver as drowsy for most of the events where any part of the car left the lane (64%-80%), although not as successfully as for the two more severe classes of lane excursions discussed previously (see Table 6). The proportion of lane excursions that occurred when the driver was classified as drowsy was higher for the sleep-deprived session (71%-87%) compared with the alert session (47%-61%). Similar to the other lane excursion types, choosing a cut-off level of 4.5 for classifying a driver as drowsy led to prediction of a greater proportion of events where the entire car left the lane than if a cut-off level of 5.0 was chosen. Again, choosing a prediction time window of 30 minutes or 15 minutes was more effective than a time window of 5 minutes.

4.2.2.2 Evaluation of Optalert using techniques for test evaluation

In order to fully evaluate the Optalert system, it is necessary to know more than simply the proportion of lane excursions that occurred that happened when the Optalert system classified the driver as drowsy. In this part of the analysis, the inference was that if the driver was classified as drowsy, then it was predicted that their driving performance would degrade, as measured by lane excursions.

- A. To fully assess predictive power, it is necessary to take into account four potential situations: True positive: the Optalert system classified the driver as drowsy and that poor performance would occur, and a lane excursion did occur;
- B. False positive: the Optalert system classified the driver as drowsy and predicted that poor performance would occur, and a lane excursion did not occur;
- C. False negative: the Optalert system did not classify the driver as drowsy nor predict that poor performance would occur, and a lane excursion did occur; and
- D. True negative: the Optalert system did not classify the driver as drowsy nor predict that poor performance would occur, and a lane excursion did not occur.

These four outcomes can be displayed in a 2x2 matrix as shown in Table 7.

Table 7. Generic 2x2 matrix used for the evaluation of the Optalert system.

		Lane excursion observed during period	
		Yes	No
Driver classified as drowsy by Optalert - Lane excursion predicted	Yes	A: True Positive	B: False Positive
	No	C: False Negative	D: True Negative

From the 2x2 matrix, several measures were derived. Firstly, the sensitivity of the system was defined as the proportion of periods in which a lane excursion was observed that were predicted by the system, or, the probability that Optalert predicted a lane excursion would happen in the period, given that one or more lane excursions did happen ($A/A+C$). The specificity of the system was defined as the proportion of periods where a lane excursion did not occur and where the Optalert system did not predict an event would occur, or, the probability that Optalert did not predict a lane excursion in the period, given that no lane excursion occurred ($D/B+D$). Specificity is also equal to 1-false positive rate. In essence, the sensitivity and specificity indicate the proportion of periods in which events (sensitivity) and non-events (specificity) occurred, that were predicted correctly by the system.

Another method of analysis was to determine the predictive value of the system. The positive predictive value (PPV) was defined as the proportion of periods in which the Optalert system predicted that an event would occur, and one or more event actually did

occur, or the probability that one or more lane excursion occurred, given that the Optalert system predicted it would (A/A+B). The negative predictive value (NPV) was defined as the proportion of periods in which the system did not predict an event would occur, that no event occurred, or the probability that no lane excursion occurred, given that Optalert did not predict that one would (D/C+D). That is, the predictive values indicate the proportion of periods in which the system was correct when it predicted that an event was (positive predictive value) or was not (negative predictive value) going to occur.

The Optalert system was evaluated by formulating separate 2x2 matrices for cautionary and critical JDS levels, for each prediction window (5, 15 or 30 minutes), each level of lane excursion (i.e., all, half, or severe off-road events), and for each of the alert and sleep-deprived conditions. Thus, there were 36 2x2 matrices generated for this part of the analysis.

The driving performance data for each participant were inspected for every distinct prediction period within their driving sessions to determine whether or not a lane excursion occurred. If, during a period in which a driver was classified as drowsy, one or more lane excursions occurred, this was recorded as one true positive. If no lane excursion occurred, this was recorded as one false positive. Alternatively, if during a period in which a driver was not classified as drowsy, one or more lane excursions occurred, this was designated a false negative. If no lane excursion occurred, it was designated a true negative. These results were then pooled across all participants for the analysis of the sensitivity, specificity, and positive and negative predictive value of the Optalert system. It is important to understand that the units of analysis for each subject were the prediction periods, and whether or not an event occurred during that period. Therefore, the outcome is binary; either an event occurred or it did not, and the number of events per period was irrelevant.

4.2.2.2.1 Signal Detection Theory analysis of lane excursions where the entire car departed the lane

The sensitivity, specificity and predictive value of the Optalert system according to entire car lane excursions during distinct prediction periods for alert and sleep-deprived sessions, pooled across participants are shown in Table 8, and summarised below.

Table 8. Sensitivity, Specificity and predictive value of the Optalert system according to entire car lane excursions during distinct prediction periods for alert and sleep-deprived sessions, pooled across participants.

Prediction period: JDS level and time window		Excursions where the entire car exceeded lane boundaries							
		Alert sessions				Sleep-deprived sessions			
JDS Level	Time window	Sensitivity	Specificity	PPV	NPV	Sensitivity	Specificity	PPV	NPV
4.5	5 min	60.0	71.0	25.0	91.7	70.0	57.6	33.3	86.4
4.5	15 min	66.7	70.8	22.2	94.4	83.3	60.9	35.7	93.3
4.5	30 min	50.0	71.4	14.3	93.8	85.7	63.2	46.2	92.3
5.0	5 min	50.0	70.0	25.0	87.5	61.5	63.0	44.4	77.3
5.0	15 min	66.7	75.0	28.6	93.8	71.4	63.6	38.5	87.5
5.0	30 min	50.0	77.8	20.0	93.3	66.7	68.8	44.4	84.6

Sensitivity: In the alert sessions, of the periods in which one or more entire car lane excursions occurred, the JDS classified the driver as drowsy during a reasonable proportion of these periods – that is, the sensitivity of the JDS algorithm was reasonable (50%-67%). The sensitivity in the sleep-deprived sessions was better than for the alert sessions (62%-86%).

Specificity: The specificity of the JDS algorithm was good. In the alert sessions, the specificity ranged between 70% to 78%. The specificity was also reasonable for the sleep-deprived sessions, and ranged between 58% to 69%. This means that for most of the periods in which an entire lane excursion did not occur, the Optalert did not classify the driver as drowsy. This translates to a reasonably low false positive rate (1-specificity).

Positive predictive value: Of the periods where the Optalert system classified the driver as drowsy, the proportion in which one or more entire lane excursion did occur was quite low. The PPV in the alert sessions ranged from 14% to 29%. For the sleep-deprived session the PPV was 33%-46%.

Negative predictive value: For the majority of the periods during which the Optalert system did not classify the driver as drowsy, no entire car lane excursions occurred. That is, the NPV was high (88%-94%) in the alert sessions. NPV was also good for the sleep-deprived sessions, although not as high as for the alert sessions (77%-93%).

4.2.2.2.2 Signal Detection Theory analysis of lane excursions where at least half of the car left the lane

Table 9. Sensitivity, Specificity and predictive value of the Optalert system according to lane excursions where at least half of the car left the lane, during distinct prediction periods for alert and sleep-deprived sessions, pooled across participants.

Prediction period: JDS level and time window		Excursions where at least half of the car exceeded lane boundaries							
		Alert sessions				Sleep-deprived sessions			
JDS Level	Time window	Sensitivity	Specificity	PPV	NPV	Sensitivity	Specificity	PPV	NPV
4.5	5 min	40.0	69.2	33.3	75.0	70.6	65.4	57.1	77.3
4.5	15 min	42.9	70.0	33.3	77.8	75.0	70.6	64.3	80.0
4.5	30 min	33.3	70.6	28.6	75.0	75.0	71.4	69.2	76.9
5.0	5 min	36.4	68.0	33.3	70.8	45.8	56.3	61.1	40.9
5.0	15 min	42.9	75.0	42.9	75.0	56.3	69.2	69.2	56.3
5.0	30 min	33.3	78.6	40.0	73.3	50.0	70.0	66.7	53.8

Table 9 shows the sensitivity, specificity and predictive value of the Optalert system according to lane excursions where at least half of the car left the lane, during distinct prediction periods for alert and sleep-deprived sessions, pooled across participants. The results of these analyses are summarised below.

Sensitivity: The sensitivity of the Optalert system for lane excursions where at least half of the car left the lane was worse than the sensitivity of the system for the more severe category of lane excursions discussed previously. The sensitivity ranged from 33% to 43% in the alert session and from 46% to 75% in the sleep-deprived session.

Specificity: The specificity of the Optalert system was reasonable for lane excursions where at least half of the car left the lane. The specificity was 68% to 79% in the alert sessions and 56% to 71% in the sleep-deprived sessions.

Positive predictive value: The positive predictive value of the Optalert system was low during the alert sessions and ranged between 29% to 43%. The PPV was improved during the sleep-deprived sessions, and ranged from 57% to 69%.

Negative predictive value: The negative predictive value of the Optalert system was reasonable in the alert sessions, and ranged from 71% to 78%. For the sleep-deprived sessions, the NPV ranged from 41% to 80%.

4.2.2.2.3 Signal Detection Theory Analysis of lane excursions where any part of the car left the lane.

Table 10. Sensitivity, Specificity and predictive value of the Optalert system according to lane excursions where any part of the car left the lane, during distinct prediction periods for alert and sleep-deprived sessions, pooled across participants.

Prediction period: JDS level and time window		Excursions where any part of the car exceeded lane boundaries							
		Alert sessions				Sleep-deprived sessions			
JDS Level	Time window	Sensitivity	Specificity	PPV	NPV	Sensitivity	Specificity	PPV	NPV
4.5	5 min	34.4	75.0	91.7	12.5	52.5	100	100	13.6
4.5	15 min	39.1	100	100	14.3	50.0	100	100	6.67
4.5	30 min	35.0	100	100	18.8	54.2	100	100	15.4
5.0	5 min	33.3	66.7	83.3	16.7	45.0	-	100	0
5.0	15 min	35.0	100	100	18.8	44.8	-	100	0
5.0	30 min	27.8	100	100	13.3	40.9	-	100	0

The sensitivity, specificity and predictive value of the Optalert system according to lane excursions where any part of the car left the lane, during distinct prediction periods for alert and sleep-deprived sessions, pooled across participants, are shown in Table 10.

Sensitivity: Only approximately one third of the periods where lane excursions when any part of the car left the lane occurred during periods in which Optalert classified the driver as drowsy. The sensitivity ranged between 28% to 39% in the alert sessions. In the sleep-deprived sessions, the sensitivity ranged between 41 to 54%.

Specificity: The specificity of the Optalert system for events where any part of the car left the lane was quite high when it could be calculated. That is, when no lane excursion occurred, the Optalert system had not classified the driver as drowsy for the majority of these periods. The specificity in the alert session was high, ranging between 67% to 100%. For the sleep-deprived session, the specificity could not be calculated when the JDS level for prediction was 5.0, because there were no periods in which no lane excursion occurred. Specificity could however be calculated for the sleep-deprived session for each prediction period where the JDS level of interest was 4.5, and for these instances, the specificity was 100%.

Positive predictive value: The PPV of the Optalert system for events where any part of the car left the lane was high. PPV ranged from 83% to 100% in the alert sessions, and was 100% for all JDS level/time windows in the sleep-deprived session.

Negative predictive value: The NPV (i.e., the proportion of times that the Optalert system did not classify the driver as drowsy and no lane excursion occurred) was very low for events where any part of the car left the lane. This means that for most periods where the Optalert system did not classify the driver as drowsy, there were instances in which some part of the car left the lane. NPV ranged from 13% to 19% in the alert sessions, and from zero to 15% in the sleep-deprived sessions.

4.2.2.3 Lane position variability

The variability in lateral lane position (as measured by the standard deviation of lane position) was compared for time periods in which the driver was classified as drowsy by the Optalert system compared to when the driver was not classified as drowsy. Reciprocal transformation was used to achieve normality of the data in order for two-way repeated measures analyses of variance to be performed, although only untransformed means and standard deviations are reported below. The repeated factors were Drowsiness (i.e. whether or not the driver was classified as drowsy by Optalert) and Sleep (i.e. whether the participant was alert or sleep-deprived during the session). Sleep status was included as an independent factor to investigate whether sleep status differentially affected the effect of drowsiness on lane variability; that is, whether there was an interaction between sleep status and drowsiness. For each of the six different prediction periods, there was no significant interaction between drowsiness and sleep status. Due to this lack of interaction and because our interest lies in the performance of the Optalert system, only the main effect of drowsiness (i.e., periods when the driver was classified as drowsy compared to periods when the driver was not classified as drowsy) on lane position variability is reported below.

4.2.2.3.1 Five minute prediction window

The lane variability for periods when the JDS had reached the cautionary level (4.5) or the critical level (5.0) sometime in the preceding 5 minutes was not significantly different to the lane variability during periods when the JDS had not reached the cautionary ($F(1,14)=3.50, p>0.05$) or critical levels ($F(1,13)=3.71, p>0.05$) in the preceding 5 minutes. Although there was no significant difference in lane variability between periods when the driver was classified as drowsy and periods when the driver was not classified as drowsy, it can be seen in Table 11 below that there was a trend for lane variability to be higher when the driver was classified as drowsy.

Table 11. Mean (sd) of lane variability during periods in which the driver was or was not classified as drowsy during the previous 5 minutes

Was the JDS level reached within previous 5 minutes?	JDS level	
	Caution (4.5)	Critical (5.0)
Yes	0.77 (0.45)	0.88 (0.60)
No	0.65 (0.87)	0.61 (0.66)

4.2.2.3.2 15 minute prediction window

The lane variability for periods when the JDS had reached the cautionary level (4.5) or the critical level (5.0) sometime in the preceding 15 minutes was not significantly different to the lane variability during periods when the JDS had not reached the cautionary ($F(1,)=$, $p>0.05$) or critical levels ($F(1,12)=3.16$, $p>0.05$) in the preceding 15 minutes. Again however, Table 12 shows that there was a trend for lane variability to be higher during periods when the driver was classified as drowsy compared to periods when they were not.

Table 12. Mean (sd) of lane variability during periods in which the driver was or was not classified as drowsy in the previous 15 minutes

Was the JDS level reached within previous 15 minutes?	JDS level	
	Caution (4.5)	Critical (5.0)
Yes	0.61 (0.25)	0.66 (0.28)
No	0.40 (0.13)	0.45 (0.37)

4.2.2.3.3 30 minute prediction window

The lane variability for periods when the JDS had reached the cautionary level (4.5) sometime in the preceding 30 minutes was significantly higher than the lane variability during periods when the JDS had not reached the cautionary level in the preceding 30 minutes ($F(1,12)=8.07$, $p<0.02$). There was however, no significant difference in the lane variability for periods when the JDS had reached the critical level (5.0) in the preceding 30 minutes compared to periods in which this had not occurred ($F(1,8)=4.78$, $p>0.05$).

That is, the lane variability was worse in the 30 minutes after the JDS had reached the cautionary level compared to periods when the cautionary level had not been reached. However, lane variability was not significantly worse in the 30 minutes after the JDS had reached the critical level, compared to periods when the critical level had not been reached. Although not significant, the trend for increased lane variability after the critical level had been reached (i.e. when the driver was classified as too drowsy to drive) was still evident (see Table 13).

Table 13. Mean (sd) of lane variability during periods in which the driver was or was not classified as drowsy in the previous 30 minutes

Was the JDS level reached within previous 30 minutes?	JDS level	
	Caution (4.5)	Critical (5.0)
Yes	0.63 (0.41)	0.67 (0.49)
No	0.37 (0.11)	0.37 (0.09)

* significant differences between periods shown in bold type

4.2.2.4 *Rate of lane excursions*

The rate of each type of lane excursion (per minute) was determined for the periods in which the driver was classified as drowsy compared to periods when the driver was not classified as drowsy. The event rate was used rather than the absolute number of events because the periods were not of a constant duration, and so the rate of events per minute was a better comparator than the absolute number of events during the period. It was not possible to transform the data to achieve normality, so the non-parametric Wilcoxon Signed Ranks test was used. Because this statistical test is not capable of testing a full 2x2 design, data for each drowsiness period was pooled across both experimental sessions

(alert and sleep-deprived). This was justified because the previous analysis of lane variability found no interaction between sleep state and drowsiness, and so it was assumed that there would be no such interaction for the lane excursions either. Thus, only the main effect of drowsiness was investigated.

4.2.2.4.1 5 minute prediction window

Table 14 shows the mean (with standard deviations) and median of lane excursion rates during periods in which the driver was or was not classified as drowsy in the previous 5 minutes.

All lane excursions: The rate of lane excursions where any part of the car left the lane was significantly higher for periods when the JDS had reached the cautionary level (4.5) or the critical level (5.0) sometime in the preceding 5 minutes, than the rate of lane excursions during periods when the JDS had not reached the cautionary ($z = -3.24, p < .002$) or critical levels ($z = -3.30, p = 0.001$) in the preceding 5 minutes. That is, driving performance, in terms of the rate of events where any part of the car left the lane, was worse during periods in which the driver was classified as drowsy compared to period when the driver was not classified as drowsy.

Half-car lane excursions: The rate of lane excursions where at least half of the car was out of the lane was significantly higher for periods when the JDS had reached the cautionary level (4.5) or the critical level (5.0) sometime in the preceding 5 minutes, than the rate of half car lane excursions during periods when the JDS had not reached the cautionary ($z = -2.19, p < .03$) or critical levels ($z = -2.01, p < 0.05$) in the preceding 5 minutes. That is, driving performance, in terms of the rate of events where at least half of the car left the lane, was worse during periods in which the driver was classified as drowsy compared to when they were not classified as drowsy.

Severe lane excursions: The rate of lane excursions where the entire car departed the lane was not significantly different for periods when the JDS had reached the cautionary level or critical level sometime in the preceding 5 minutes compared to periods when the JDS had not reached the cautionary level ($z = -1.78, p > 0.05$) or critical level ($z = -1.73, p > 0.05$) in the preceding five minutes. However there was a trend for worse performance when the driver was classified as drowsy.

Table 14. Mean (sd) and median of lane excursion rates during periods in which the driver was or was not classified as drowsy in the previous 5 minutes

Was the JDS level reached within previous 5 minutes?		JDS level					
		Caution (4.5)			Critical (5.0)		
		All events	Half events	Severe events	All events	Half events	Severe events
Yes	Mean	1.39	0.22	0.08	1.45	0.25	0.10
	(sd)	(1.14)	(0.35)	(0.13)	(1.20)	(0.39)	(0.17)
	Median	0.88	0.04	0	0.97	0.04	0
No	Mean	0.53	0.04	0.01	0.59	0.06	0.02
	(sd)	(0.61)	(0.07)	(0.03)	(0.70)	(0.13)	(0.05)
	Median	0.27	0.01	0	0.34	0.01	0

* significant differences between periods shown in bold type

4.2.2.4.2 15 minute prediction window

Table 15. Mean (sd) and median of lane excursion rates during periods in which the driver was or was not classified as drowsy during the previous 15 minutes

Was the JDS level reached within previous 15 minutes?		JDS level					
		Caution (4.5)			Critical (5.0)		
		All events	Half events	Severe events	All events	Half events	Severe events
Yes	Mean	1.39	0.23	0.09	1.48	0.25	0.10
	(sd)	(1.20)	(0.36)	(0.14)	(1.21)	(0.38)	(0.16)
	Median	0.67	0.02	0	0.92	0.04	0
No	Mean	0.36	0.01	0.002	0.41	0.02	0.003
	(sd)	(0.30)	(0.02)	(0.007)	(0.30)	(0.04)	(0.01)
	Median	0.26	0	0	0.31	0.01	0

* significant differences between periods shown in bold type

Table 15 shows the mean (with standard deviations) and median of lane excursion rates during periods in which the driver was or was not classified as drowsy during the previous 15 minutes.

All lane excursions: The rate of lane excursions where any part of the car was out of the lane was significantly higher for periods when the JDS had reached the cautionary level (4.5) or the critical level (5.0) sometime in the preceding 15 minutes, than the rate of lane excursions during periods when the JDS had not reached the cautionary ($z = -3.30$, $p < .002$) or critical levels ($z = -3.18$, $p < 0.002$) in the preceding 15 minutes. That is, driving

performance, in terms of the rate of events where any part of the car left the lane, was worse during periods in which the driver was classified as drowsy compared to periods in which they were not classified as drowsy.

Half-car lane excursions: The rate of lane excursions where at least half of the car was out of the lane was not significantly different for periods when the JDS had reached the cautionary level or the critical level sometime in the preceding 15 minutes compared to periods when the JDS had not reached the cautionary level ($z = -1.88, p > 0.05$) or the critical level ($z = -1.75, p > 0.05$) in the preceding 15 minutes. That is, there was no significant difference in the rate of lane excursions where the car was at least half out of the lane between periods when the driver was classified as drowsy compared to when they were not, however there was a trend for these lane excursions to be more frequent when the driver was drowsy.

Severe lane excursions: The rate of lane excursions where all of the car left the lane was significantly higher for periods when the JDS had reached the cautionary level (4.5) or the critical level (5.0) sometime in the preceding 15 minutes, than the rate of lane excursions during periods when the JDS had not reached the cautionary ($z = -1.99, p < 0.05$) or critical levels ($z = -1.49, p < 0.05$) in the preceding 15 minutes. That is, driving performance, in terms of the rate of events where the entire car left the lane, was worse during periods in which the driver was classified as drowsy compared to periods when they were not classified as drowsy.

4.2.2.4.3 30 minute prediction window

Table 16. Mean (sd) and median of lane excursion rates during periods in which the driver was or was not classified as drowsy in the previous 30 minutes

Was the JDS level reached within previous 30 minutes?	JDS level						
	Caution (4.5)			Critical (5.0)			
	All events	Half events	Severe events	All events	Half events	Severe events	
Yes	Mean	1.40	0.28	0.11	1.45	0.30	0.11
	(sd)	(1.29)	(0.41)	(0.18)	(1.40)	(0.45)	(0.18)
	Median	0.68	0.04	0	0.81	0.11	0
No	Mean	0.36	0.01	0.002	0.40	0.01	0.003
	(sd)	(0.30)	(0.02)	(0.007)	(0.30)	(0.03)	(0.01)
	Median	0.28	0	0	0.28	0.01	0

* significant differences between periods shown in bold type

The mean (with standard deviations) and median of lane excursion rates during periods in which the driver was or was not classified as drowsy within the previous 30 minutes are shown in Table 16.

All lane excursions: The rate of lane excursions where any part of the car departed the lane was significantly higher for periods when the JDS had reached the cautionary level (4.5) or the critical level (5.0) sometime in the preceding 30 minutes, than the rate of lane

excursions during periods when the JDS had not reached the cautionary ($z = -3.18$, $p < .002$) or critical levels ($z = -2.67$, $p < 0.01$) in the preceding 30 minutes. That is, driving performance, in terms of the rate of events where any part of the car left the lane, was worse during periods in which the driver was classified as drowsy compared to periods in which the driver was not classified as drowsy.

Half-car lane excursions: There was no significant difference between the rate of lane excursions where at least half of the car was out of the lane for periods when the JDS had reached the cautionary level or the critical sometime in the preceding 30 minutes compared to periods when the JDS had not reached the cautionary level ($z = -1.82$, $p > 0.05$) or the critical level ($z = -1.67$, $p > 0.05$) in the preceding 30 minutes. However, there was a trend for lane excursions to be more frequent when the driver was classified as drowsy.

Severe lane excursions: The rate of lane excursions where all of the car was out of the lane was not significantly different for periods when the JDS had reached the critical level sometime in the preceding 30 minutes compared to periods when the JDS had not reached the critical level ($z = -1.72$, $p > 0.05$). However, there was a trend for worse performance when it was expected. There was, however, a significantly higher rate of lane excursions in the 30 minute period after the JDS reached the cautionary level compared to periods when the JDS had not reached the cautionary level in the preceding 30 minutes ($z = -2.22$, $p < 0.03$).

4.2.2.5 Proportion of total time the vehicle was out of the lane

The proportion of time in each prediction period that the vehicle departed the lane, for each of the three definitions of lane excursion, was determined for the periods in which the driver was classified as drowsy compared to periods in which they were not classified as drowsy. It was not possible to transform the data to achieve normality, so the non-parametric Wilcoxon Signed Ranks test was used. Because this statistical test is not capable of testing a full 2x2 design, data for each drowsiness period was pooled across both experimental sessions (alert and sleep-deprived). This method was justified because the analysis of lane variability found no interaction between sleep state and drowsiness, and so it was assumed that there would be no such interaction for the lane excursions either. Thus, only the main effect of drowsiness was investigated.

4.2.2.5.1 5 minute window

The mean (with standard deviations) and median proportion of time the car was out of the lane during periods in which the driver was or was not classified as drowsy in the previous 5 minutes are shown in Table 17.

All lane excursions: The proportion of time that any part of the car departed the lane was significantly higher for periods when the JDS had reached the cautionary level (4.5) or the critical level (5.0) sometime in the preceding 5 minutes, than for periods when the JDS had not reached the cautionary ($z = -2.73$, $p < .01$) or critical levels ($z = -2.48$, $p < 0.02$) in the preceding 5 minutes. That is, driving performance, in terms of the proportion of time that any part of the car departed the lane, was worse during periods in which the driver was classified as drowsy compared to periods when the driver was not classified as drowsy.

Half-car lane excursions: There was no significant difference between the proportion of time that at least half of the car departed the lane for periods when the JDS had reached the cautionary level or the critical level sometime in the preceding 5 minutes compared to

periods when the JDS had not reached the cautionary level ($z = -1.28, p > 0.05$) or the critical level ($z = -1.26, p > 0.05$) in the preceding 5 minutes. However, there was a trend for the proportion of time that at least half of the car had departed the lane to be higher when the driver was classified as drowsy.

Severe lane excursions: There was no significant difference between the proportion of time that the entire car departed the lane for periods when the JDS had reached the cautionary level or the critical level sometime in the preceding 5 minutes compared to periods when the JDS had not reached the cautionary level ($z = -0.94, p > 0.05$) or the critical level ($z = -1.02, p > 0.05$) in the preceding 5 minutes. However, there was a trend for the proportion of time that the entire car departed the lane to be higher when the driver was classified as drowsy.

Table 17. Mean (sd) and median proportion of time the car was out of the lane during periods in which the driver was or was not classified as drowsy in the previous 5 minutes

Was the JDS level reached within previous 5 minutes?		JDS level					
		Caution (4.5)			Critical (5.0)		
		All events	Half events	Severe events	All events	Half events	Severe events
Yes	Mean	0.06	0.01	0.005	0.07	0.02	0.006
	(sd)	(0.07)	(0.02)	(0.009)	(0.07)	(0.04)	(0.01)
	Median	0.03	0.0006	0	0.04	0.0007	0
No	Mean	0.02	0.003	0.001	0.02	0.003	0.001
	(sd)	(0.04)	(0.01)	(0.003)	(0.04)	(0.009)	(0.003)
	Median	0.006	0.0001	0	0.008	0.0002	0

* significant differences between periods shown in bold type

4.2.2.5.2 15 minute window

Table 18 shows the mean (with standard deviations) and median proportion of time the car was out of the lane during periods in which the driver was or was not classified as drowsy in the previous 15 minutes.

All lane excursions: The proportion of time that any part of the car had departed the lane was significantly higher for periods when the JDS had reached the cautionary level (4.5) or the critical level (5.0) sometime in the preceding 15 minutes, than during periods when the JDS had not reached the cautionary ($z = -3.30, p < .002$) or critical levels ($z = -3.18, p < 0.002$) in the preceding 15 minutes. That is, driving performance, in terms of the proportion of time that any part of the car had departed the lane, was worse during periods in which the driver was classified as drowsy compared to when they were not classified as drowsy.

Half-car lane excursions: There was no significant difference between the proportion of time that at least half of the car had departed the lane for periods when the JDS had reached the cautionary level or the critical level sometime in the preceding 15 minutes

compared to periods when the JDS had not reached the cautionary level ($z = -1.70$, $p > 0.05$) or the critical level ($z = -1.61$, $p > 0.05$) in the preceding 15 minutes. However, there was a trend for the proportion of time that at least half of the car had departed the lane to be higher when the driver was classified as drowsy.

Severe lane excursions: The proportion of time that the entire car had departed the lane was not significantly different for periods when the JDS had reached the critical level sometime in the preceding 15 minutes compared to periods when the JDS had not reached the critical level ($z = -1.49$, $p > 0.05$). However, there was a trend for worse performance when the driver was classified as drowsy. There was a significantly higher proportion of time that the entire car had departed the lane in the 15 minute period after the JDS reached the cautionary level compared to periods when the JDS had not reached the cautionary level in the preceding 15 minutes ($z = -1.99$, $p < 0.05$).

Table 18. Mean (sd) and median proportion of time the car was out of the lane during periods in which the driver was or was not classified as drowsy in the previous 15 minutes

Was the JDS level reached within previous 15 minutes?		JDS level					
		Caution (4.5)			Critical (5.0)		
		All events	Half events	Severe events	All events	Half events	Severe events
Yes	Mean	0.06	0.01	0.005	0.07	0.01	0.006
	(sd)	(0.07)	(0.02)	(0.009)	(0.07)	(0.02)	(0.01)
	Median	0.02	0.0004	0	0.04	0.0007	0
No	Mean	0.009	0.0003	0.00004	0.01	0.0004	0.00007
	(sd)	(0.01)	(0.0006)	(0.0001)	(0.01)	(0.001)	(0.0002)
	Median	0.006	0	0	0.008	0.0002	0

* significant differences between periods shown in bold type

4.2.2.5.3 30 minute window

The mean (with standard deviations) and median proportion of time the car was out of the lane during periods in which the driver was or was or not classified as drowsy during the previous 30 minutes are shown in Table 19.

All lane excursions: The proportion of time where any part of the car had departed the lane was significantly higher for periods when the JDS had reached the cautionary level (4.5) or the critical level (5.0) sometime in the preceding 30 minutes, than for periods when the JDS had not reached the cautionary ($z = -3.18$, $p < .002$) or critical levels ($z = -2.67$, $p < 0.01$) in the preceding 30 minutes. That is, driving performance, in terms of the proportion of time that any part of the car departed the lane, was worse during periods in which the driver was classified as drowsy compared to periods in which they were not.

Half-car lane excursions: There was no significant difference between the proportion of time that at least half of the car departed the lane for periods when the JDS had reached the cautionary level or the critical level sometime in the preceding 30 minutes compared to

periods when the JDS had not reached the cautionary level ($z = -1.89$, $p > 0.05$) or the critical level ($z = -1.55$, $p > 0.05$) in the preceding 30 minutes. However, there was a trend for the proportion of time that at least half of the car had departed the lane to be higher when the driver was classified as drowsy.

Severe lane excursions: The proportion of time that the entire car had departed the lane was not significantly different for periods when the JDS had reached the critical level sometime in the preceding 30 minutes compared to periods when the JDS had not reached the critical level ($z = -1.72$, $p > 0.05$). However, there was a trend for worse performance when it was expected. There was a significantly higher proportion of time that the entire car had departed the lane in the 30 minute period after the JDS reached the cautionary level compared to periods when the JDS had not reached the cautionary level in the preceding 30 minutes ($z = -2.22$, $p < 0.03$).

Table 19. Mean (sd) and median proportion of time the car was out of the lane during periods in which the driver was or was not classified as drowsy in the previous 30 minutes

Was the JDS level reached within previous 30 minutes?	JDS level						
	Caution (4.5)			Critical (5.0)			
	All events	Half events	Severe events	All events	Half events	Severe events	
Yes	Mean	0.07	0.01	0.007	0.07	0.02	0.008
	(sd)	(0.08)	(0.02)	(0.01)	(0.08)	(0.03)	(0.01)
	Median	0.02	0.001	0	0.03	0.003	0
No	Mean	0.01	0.0003	0.00004	0.01	0.0004	0.00008
	(sd)	(0.01)	(0.0006)	(0.0001)	(0.01)	(0.0009)	(0.0002)
	Median	0.006	0	0	0.007	0.0002	0

* significant differences between periods shown in bold type

4.2.2.6 Mean time headway

There were two events during each simulated drive in which the participant was required to follow a vehicle that turned into the road ahead of them and travelled at a slower speed. The mean time headway during these car-following episodes was recorded, and analyses were performed to determine if the mean time headways that participants chose were different between periods when they were classified as drowsy and when they were not classified as drowsy. Sleep status (alert or sleep-deprived) and event number (first or second) were also included as factors.

Parametric analyses revealed no effect of drowsiness on the mean time headway between the participant's vehicle and the vehicle in front ($p > 0.05$). In addition, sleep status did not affect mean time headway ($p > 0.05$). The only significant main effect was for event number: participants chose a significantly greater time headway when following the lead vehicle for the second time the event occurred within the driving scenario, compared to the time headway when following the lead vehicle the first time the event occurred ($p < 0.05$).

These results were consistent for all of the JDS level/time window prediction combinations.

5 DISCUSSION

Lane excursions are an outcome of particular interest in this study, due to the increased risk of collision with another vehicle or roadside objects when the vehicle leaves the lane in which it is supposed to be travelling. It seems reasonable to assume that the risk of a collision is greater as the proportion of the car that leaves the lane of travel increases. Severe lane excursions, where the entire vehicle departed from the lane, were relatively infrequent within this study, with only seven of the eighteen participants making such excursions, and only five participants doing so more than twice. Incidents where at least some part of the car left the lane, however, were common, with all participants making such lane excursions during their driving sessions.

In terms of the lane excursions that occurred, the vast majority of occasions when more than half of the car departed the lane of travel occurred during periods when the driver was classified as drowsy by Optalert (including lane excursions where the entire car departed the lane). Optalert also classified drivers as drowsy when most lane excursions where any part of the car left the lane occurred, although this proportion was not as high as for the more severe types of lane excursions. A greater proportion of lane excursions that occurred in the sleep deprived session occurred when the driver was classified as drowsy compared to the alert session. However, for most JDS level/prediction window combinations the proportion of lane excursions that occurred when the driver was classified as drowsy in the alert session was still reasonable. Regarding the choice of JDS value to use as a cut-off for classifying the driver as drowsy, and hence predicting lane excursions, a greater proportion of excursions was predicted when a cut-off value of 4.5 was used to define periods of impaired performance due to drowsiness, compared to the cut-off value of 5.0. In addition, using a prediction period of 30 minutes after the cut-off value was reached led to prediction of a higher proportion of events, followed closely by a 15 minute period, while a 5 minute period led to a lower proportion of events that were predicted.

It is not particularly surprising that the lower JDS cut-off of 4.5, and longer time windows, led to prediction of a greater proportion of events that occurred. In essence, using a lower cut-off and a longer time window simply means that a greater proportion of time within each driving session was classified as a period in which the driver was drowsy and a performance decrement would be expected. This in turn leads to a higher probability that any event that occurred lay within the time frame that a performance decrement would be expected. However, by choosing a low cut-off and a longer prediction period, it would be expected that not only would the detection rate increase, but so would the number of times that the driver was classified as drowsy and no such lane excursion occurred, i.e. the false positive rate.

Further analysis using techniques for test evaluation took into account all four possible outcomes (true positives, true negatives, false positives and false negatives) rather than true positives only (i.e., proportion of correct predictions only). The method that was used involved dividing the driving sessions into periods where a performance decrement was expected (because the driver was drowsy) and periods when it was not (because the driver was not drowsy), and each period was then classified into one of the four possible outcomes. This means that multiple lane excursions that occurred during a period were treated the same as a single lane excursion, and both were classed as a positive lane excursion within that period.

In any evaluation of a system using this method, a decision has to be made as to the consequences of an incorrect “decision”. The performance of a system is never perfect and

there is always a trade-off between sensitivity and specificity, and negative and positive predictive values. If the decision level cut-off is set too high, then the system will miss many events of interest, and the false negative rate will be high. That is, if the critical JDS level at which the driver is warned to cease driving is set too high, then driving performance may degrade before the driver is classified as too drowsy to drive. This will lead to low sensitivity and NPV. If on the other hand, the cut-off is set too low, more events will be detected, but the rate of false positives will be high. That is, if the critical JDS level is set too low, then drivers who are not too drowsy to drive will be warned to stop driving. This will lead to low specificity and PPV. Therefore the consequences of false negatives and false positives must be considered.

In the current situation, the consequence of a false positive (that is, the Optalert system warning the driver that they are too tired to drive when they are not) for on-road driving is that drivers will receive a warning to indicate that they are drowsy when they are not drowsy. If the driver heeds the warning and takes corrective action, such as changing drivers, or stopping for a rest, at worst this will mean it will take longer to complete their journey. The consequences of a false negative on the other hand are that the driver will not receive a warning to indicate they are drowsy when they really are too drowsy to drive, and their driving performance might degrade to the point that they drive out of their lane leading to a collision. Therefore, it was judged that the consequences of false negatives were more dangerous than the consequences of false positives for determining whether a driver is too drowsy to drive. Thus it is more desirable for the Optalert system to display high sensitivity rather than high specificity. In addition, high negative predictive value is preferred to high positive predictive value.

The Optalert system was more sensitive for detecting more severe lane excursions, with the greatest sensitivity displayed for entire car lane excursions, followed by excursions where at least half the car left the lane. That is, for most periods when an off-road event occurred or when at least half of the car left the lane, the Optalert system correctly predicted the event, by classifying the driver as too drowsy to drive. Sensitivity for detecting lane excursions where any part of the car left the lane was relatively poor. The negative predictive value of the system followed a similar pattern, with the NPV being highest for entire car lane excursions and very low for excursions where any part of the car left the lane. That is, for almost all periods in which the Optalert system did not classify the driver as too drowsy to drive, there were no off-road events, or events where at least half of the car left the lane. However, for specificity and positive predictive value, the results were the opposite; specificity and PPV were better for excursions where any part of the car left the lane and poor for entire car lane excursions.

In terms of the best combination of JDS level/prediction window to use, the results consistently showed that, for the two more severe classes of lane excursions, the sensitivity and NPV were superior when using the 15 minute prediction window for both the JDS levels of interest. In addition, use of the critical (5.0) level as a cut-off usually improved the specificity and PPV relative to the cautionary (4.5) cut-off. The differing performance of the Optalert system in predicting different classes of lane excursions can partly be explained by the relative prevalence of these lane excursions. As mentioned previously, lane excursions where any part of the vehicle left the lane occurred frequently throughout the whole driving session, and so the number of false negatives could be expected to be high. On the other hand “non-events”; that is, periods in which lane excursions of this type did not occur were extremely infrequent. As such, the false positive rate was low because of the low rate of non-events. This led to high specificity and PPV, but low sensitivity and NPV. Alternatively, the prevalence of entire car lane excursions and lane excursions where

at least half of the car left the lane was much lower. This means there were fewer false negatives (which is desirable) and relatively more false positives, which are of less concern in this situation. This leads to high sensitivity and NPV.

There is also another possible explanation for these results. There are many reasons apart from drowsiness that may explain why a driver might cross the boundaries of the lane. Loss of concentration, boredom, inattention or even simply shifting position in the seat might lead to part of the vehicle departing the lane. The Optalert system would not be expected to predict events of this type unless they too were always related to drowsiness. However, the Optalert system did perform well in terms of sensitivity and negative predictive value for the more severe types of lane excursions. It could be suggested that the more severe types of lane excursions are related to drowsiness, while many of the mild lane excursions may not have been, explaining Optalert's superior performance in detecting the more severe events.

It is also worth noting the difference in the proportion of the total events that occurred that was correctly predicted by the Optalert system, and the sensitivity of the system in detecting whether an event occurred during a distinct prediction period. For example, 80% to 97% of the total number of entire car lane excursions was predicted by the Optalert system. However, the Optalert system detected only 50% to 67% of the number of periods in which one or more entire lane excursions occurred. This discrepancy is due to the fact that most participants' multiple severe lane excursions occurred during the one prediction period, for example one participant had 15 entire car lane excursions during one period where they were classified as drowsy and poor performance was predicted. In the sensitivity analysis, this would have been counted as one true positive, whereas in the crude analysis, this was included as 15 correctly classified events.

There was a consistent trend, for all three classes of lane excursions, for both the rate of lane excursions, and the proportion of time the vehicle was out of the lane to be higher in periods where the driver was classified as drowsy compared to periods when the driver was not classified as drowsy. This was the case for all JDS level/time window combinations. However, not all of these differences were significant. For lane excursions where any part of the car left the lane, the rate and the duration of time out of the lane was always significantly different according to whether the driver was classified as drowsy or not. However this was not the case for the more severe types of lane excursions. This is surprising, given that so many of the more severe lane excursions occurred during periods when the driver was classified as drowsy. There is a possibility that this is an issue of statistical power. As noted previously, the trend for impaired performance was evident when the driver was drowsy, however, the differences were only consistently significant for the lane excursions where any part of the car left the lane. These lane excursions happened frequently, and to every participant. However, lane excursions where half or more of the car departed from the lane were less frequent, and did not happen to every participant. There may not have been enough statistical power to detect an effect.

Average speed did not appear to be related to drowsiness level, and speed control displayed a complex relationship that was not investigated further. Mean time headway when following other vehicles did not differ according to whether the Optalert system predicted poor performance or not, which worth noting considering that previous has research has found that as sleepiness increases, so too does mean time headway.

5.1 LIMITATIONS OF THE STUDY AND FURTHER RESEARCH AND ANALYSIS

Not all participants attended for both the alert and sleep-deprived sessions, and hence not all data-sets were complete. In addition, not all participants experienced both drowsy and non-drowsy periods during each session; for example, one participant was drowsy for the whole time during both sessions, whereas another participant was not classified as drowsy during either the alert or sleep-deprived sessions. This makes comparative analysis of drowsy and non-drowsy periods within subjects difficult. In future studies of this kind, all experimental sessions could be extended until every driver experiences both drowsy and non-drowsy periods. However, this could lead to experimental sessions being quite prolonged, and different participants would need to drive for different lengths of time. These inconsistencies could cause problems with the analyses.

The results presented in the current report focussed mainly on the comparison of driving performance between periods in which the Optalert system classified the driver as drowsy compared to periods when the driver was not classified as drowsy. This necessitated prior specification of JDS cut-off levels and time windows for prediction. The relationship between the continuous JDS and driving performance, in terms of lane position and variability and speed control was only investigated descriptively within this study. Empirically derived JDS cut-off values and time windows for prediction of driving performance could be determined from the continuous driving performance data; however, statistical modelling of this data is a complex task. Complex time series analyses could be performed, but the authors are unaware of a technique that could calculate cross correlations with lags while taking into account the within subjects design, incorporating the different sessions (alert and sleep-deprived), as well as taking into account the missing data and sessions. It is possible that this analysis could be done separately for each participant and for each session, but valuable information may be lost with this approach. Another potential analysis method involves Fourier transformation, however this would also need to be analysed separately for each subject and each session. However a successful analysis of this type could lead to empirically derived JDS levels and time windows for the prediction of poor driving performance, which could be validated in future research.

6 CONCLUSIONS

The aim of the current study was to determine whether there is a breakdown in driving performance during the period after which a driver reaches a level of drowsiness at which they would be alerted by the Optalert system that they were becoming drowsy (JDS=4.5) or that they were too drowsy to drive (JDS=5.0).

There is a trend for increased lane variability, increased rate of lane excursions and increased duration of time spent out of the lane when the Optalert system classifies the driver as drowsy, compared to periods when the driver is not classified as drowsy, although this trend was not always statistically significant.

The Optalert system performed well in predicting severe lane excursions, where the entire vehicle departs from the lane (run off-road events). The sensitivity of the system for correctly classifying periods when these lane excursions occurred was reasonably high; that is, the Optalert system had classified the driver as too drowsy to drive within the previous 15 minutes for most periods in which the entire vehicle ran off the road. The negative predictive value for run off-road events was very high; that is, for almost all periods in which the Optalert system did not classify the driver as too drowsy to drive within the previous 15 minutes, there were no run off-road events. High sensitivity and high negative predictive value are both very desirable attributes of a system that is designed to warn drivers of drowsiness.

Most participants rated the Optalert system as comfortable to wear and most participants did not believe the Optalert glasses impaired their driving ability.

7 REFERENCES

- Akerstedt, T., Kecklund, G., & Knutsson, A. (1991). Manifest sleepiness and spectral content of the EEG during shift work. *Sleep, 14*, 221-225.
- Arnedt, J.T., Wilde, G.J.S., Munt, P.W., & Maclean, A.W. (2000). Simulated driving performance following prolonged wakefulness and alcohol consumption: separate and combined contributions to impairment. *Journal of Sleep Research, 9*, 233-241.
- Bittner, R., Hana, K., Pousek, L., Smrcka, P., Schreib, P., & Vysoky, P. (2000). Detecting drowsiness states of a car driver. In R.W. Brause, & E. Hanisch (Eds.), *Medical Data Analysis, Lecture Notes in Computer Science*. Frankfurt: Springer.
- De Waard, D., & Brookhuis, K.A. (1991). Assessing driver status: a demonstration experiment on the road. *Accident Analysis and Prevention, 23*, 297-307.
- Fairclough, S.H. (1997). Monitoring driver drowsiness via driving performance. In I. Noy (Ed.), *Ergonomics and safety of intelligent driver interfaces* (pp. 363-379). New Jersey: Lawrence Erlbaum Associates.
- Fairclough, S.H., & Graham, R. (1999). Impairment of driving performance caused by sleep deprivation or alcohol: A comparative study. *Human Factors, 41*, 118-128.
- Haworth, N.L., Triggs, T.J., & Grey, E.M. (1988). *Driver fatigue: Concepts, measurement and crash countermeasures*. (CR 72). Federal Office of Road Safety, Canberra, ACT.
- Johns, M.W., Tucker, A.J., Chapman, R.J., Michael, N.J. & Beale, C.A. (2006). A new scale of drowsiness based on multiple characteristics of blinks: The Johns Drowsiness Scale. *Sleep, 29*:A365
- Lal, S.K., & Craig, A. (2000). Driver drowsiness: psychophysiological effects. *Fourth International Conference on Drowsiness and Transportation*, Fremantle, Western Australia.
- Lal, S.K., & Craig, A. (2002). Driver drowsiness: Electroencephalography and physiological assessment. *Psychophysiology, 39*, 313-321.
- Muto, W. H. and Wierwille, W. W. (1982). The effects of repeated emergency response trials on performance during extended-duration simulated driving. *Human Factors, 24*, 693-698.
- Peters, R.D., Wagner, E., Alicandri, E., Fox, J., Thomas, M., Thorne, D.R., Sing, H., & Balwinski, S.M. (1999). Effects of partial and total sleep deprivation on driving performance. *Public Roads, 62*.
- Philip, P., Taillard, J., Klein, E., Sagaspe, P., Charles, A., Davies, W.L., Guilleminault, C., & Bioulac, B. (2003). Effect of drowsiness on performance measured by a driving simulator in automobile drivers. *Journal of Psychosomatic Research, 55*, 197-200.

RTA (2001). Driver fatigue: Problem definition and countermeasure summary. Roads Traffic Authority of NSW, Downloaded from the world wide web: http://www.rta.nsw.gov.au/roadsafety/downloads/driverfatigue_pdcn.pdf

Salah, K., & Michel, V. (1994). Drowsiness of the driver: EEG and vehicle parameters interaction. *14th International Conference on Enhanced Safety of Vehicles*, Munich, Germany.

VicRoads (2005). Fatigue and road safety. Downloaded from <http://www.vicroads.vic.gov.au/>

Wierwille, W.W. (1994). Overview of research on driver drowsiness definition and driver drowsiness detection. *14th International Conference on Enhanced Safety of Vehicles*, Munich, Germany.

Wierwille, W.W., Wregget, S.S, Kim, C.L., Ellsworth, L.A., & Fairbanks, R.J. (1994). *Research on vehicle-based driver status/performance monitoring: Development, validation and refinement of algorithms for detection of driver drowsiness*. (DOT HS 808 247). National Highway Traffic Safety Administration, Washington, DC.