Accident Analysis and Prevention 40 (2008) 396–402

ACCIDENT ANALYSIS & PREVENTION

www.elsevier.com/locate/aap

# Chasing the silver bullet: Measuring driver fatigue using simple and complex tasks

S.D. Baulk [a,b,*], S.N. Biggs [a,c], K.J. Reid [d], C.J. van den Heuvel [a,b,c], D. Dawson [a]

[a] *Centre for Sleep Research, University of South Australia, Level 7, Playford Building, City East Campus, University of South Australia, Adelaide, SA 5000, Australia*

[b] *Adelaide Institute for Sleep Health, Repatriation General Hospital, Daws Road, Daw Park, Adelaide, SA 5041, Australia*

[c] *Discipline of Paediatrics The University of Adelaide Women's and Children's Hospital 72 King William Road North Adelaide SA 5006, Australia*[1]

[d] *Department of Neurology, Northwestern University Feinberg School of Medicine, 710 North Lakeshore Drive, Chicago, IL 60611, USA*

Received 8 December 2006; received in revised form 4 July 2007; accepted 9 July 2007

**Abstract**

Driver fatigue remains a significant cause of motor-vehicle accidents worldwide. New technologies are increasingly utilised to improve road safety, but there are no effective on-road measures for fatigue. While simulated driving tasks are sensitive, and simple performance tasks have been used in industrial fatigue management systems (FMS) to quantify risk, little is known about the relationship *between* such measures. Establishing a simple, on-road measure of fatigue, as a fitness-to-drive tool, is an important issue for road safety and accident prevention, particularly as many fatigue related accidents are preventable. This study aimed to measure fatigue-related performance decrements using a simple task (reaction time – RT) and a complex task (driving simulation), and to determine the potential for a link *between* such measures, thus improving FMS success. Fifteen volunteer participants (7 m, 8 f) aged 22–56 years (mean 33.6 years), underwent 26 h of supervised wakefulness before an 8 h recovery sleep opportunity. Participants were tested using a 30-min interactive driving simulation test, bracketed by a 10-min psychomotor vigilance task (PVT) at 4, 8, 18 and 24 h of wakefulness, and following recovery sleep. Extended wakefulness caused significant decrements in PVT and driving performance. Although these measures are clearly linked, our analyses suggest that driving simulation cannot be replaced by a simple PVT. Further research is needed to closely examine links between performance measures, and to facilitate accurate management of fitness to drive, which requires more complex assessments of performance than RT alone.


*Keywords:* Driving; Fatigue management; Fitness to drive; Performance

## 1. Introduction

### 1.1. Driver fatigue

Despite extensive efforts into research and development of new technologies, driver fatigue remains a major cause of vehicle accidents worldwide. Fatigue plays a role in up to 20% (Horne and Reyner, 1995; Dobbie, 2002) of vehicle accidents, with many tending to be serious or fatal. In Australia, statistics from 1998 indicate that there were 251 fatalities caused explicitly by fatigue-related accidents (16.6% of total road deaths – Dobbie, 2002). It is important to note that, while other road safety issues such as speed and alcohol are increasingly managed with effective and accurate technologies, there are currently no effective comparable on-road measures of driver fatigue. As a result, fatigue has become proportionately more of a problem (from 1990 to 1998, the proportion of fatal crashes involving driver fatigue increased from 14.9% to 18.0% – Dobbie, 2002). Aside from the human misery caused by death and injury, there are significant additional economic costs to be met by governments, industry, and health authorities as a result of these accidents. Thus, establishing a

* Corresponding author at: Centre for Sleep Research, Level 7, Playford Building, University of South Australia, Frome Road, Adelaide, SA 5000, Australia. Tel.: +61 8 8302 6624; fax: +61 8 8302 6623.

*E-mail address:* stuart.baulk@unisa.edu.au (S.D. Baulk).

[1] Present address.

0001-4575/$ – see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.aap.2007.07.008

simple, on-road measure of fatigue, as an additional fitness-to-drive tool, is an important issue for road safety and accident prevention, particularly as many fatigue-related accidents are preventable.

### 1.2. Extended wakefulness and driving impairment: complex measures

It is well established that extended wakefulness leads to increased driving impairment, as measured by both simulated (Philip et al., 2005; Arnedt et al., 2005) and on-road (Philip et al., 2001) driving studies. A recent study using the York Driving Simulator (YDS), a PC-based interactive driving simulator, demonstrated that lane drift, tracking variability, speed deviation, and off-road incidents significantly increased with extended wakefulness (Philip et al., 2005). Research using a high-fidelity, fully interactive driving simulator found that after 60 h of sleep deprivation, mean accident rate rose significantly from 0 to 8 during a 40 min drive and variance in lane position also significantly increased from .05% to 45% (Arnedt et al., 2005). Similarly, Philip et al. (2001) found that sleep restriction significantly increased the risk of inappropriate line crossings by 8.1 times in an on-road study conducted on an open French highway.

### 1.3. Measuring driver fatigue: "the silver bullet"?

Research has also shown that the risk posed by fatigue and alcohol are comparable in terms of impairment to performance (Dawson and Reid, 1997; Williamson and Feyer, 2000; Fletcher et al., 2003). It was demonstrated by our group that 17 h of sustained wakefulness produced performance impairment on some tasks equivalent to that experienced at a blood–alcohol concentration (BAC) of 0.05 g/dL. A further 7 h (i.e. 24 h in total) of sleep deprivation produced a level of impairment equivalent to that observed in subjects with a BAC of 0.1 g/dL, twice the legal limit in most Western countries (Dawson and Reid, 1997). It is generally accepted that (random) breath testing, the common measure of alcohol intoxication, prevents motor-vehicle accident related deaths, and it is likely that a practical fatigue measure would yield the same result. Not only would simple measures of fatigue have practical implications for on-road use, they would also facilitate objective assessments of fitness to drive. Current medical standards for assessing *fitness to drive* in patients with sleep disorders, such as obstructive sleep apnea (OSA) rely predominately on subjective measures such as the Epworth Sleepiness Scale or structured interviews (Austroads Guidelines, 2003; George et al., 2002), and in some cases the Multiple Sleep Latency Test (George et al., 1996) or Maintenance of Wakefulness Test (Banks et al., 2005) may be used. The most obvious concern with this method is the self-report bias that may occur by those who do not wish for their license to be revoked, irrespective of the danger to themselves and others. There is also increased pressure and responsibility on physicians to identify and report individuals who should not be allowed to continue driving. A tool which could accurately determine fitness to drive is highly desirable and is often referred to (by road transport professionals, researchers and policy makers) as the "silver bullet" for road safety.

### 1.4. Fatigue management systems: simple measures

One of the most common assays of fatigue used in sleep deprivation and performance research is the psychomotor vigilance task (PVT – Dinges and Powell, 1985). The PVT requires responses to a visual stimulus by pressing a response button as soon as the stimulus appears. Research consistently shows that extended wakefulness and cumulative sleep restriction results in an increase in reaction time (the time, measured in milliseconds taken to respond to the stimulus – RT), a decrease in response speed (1/RT), and an increase in lapses (responses > 500 ms – Dinges et al., 1997; Jewett et al., 1999; Rosekind et al., 1994). What makes the PVT a particularly attractive assay of fatigue is that it is simple to perform, and has been shown to have only minor practice effects (Dinges et al., 1997; Jewett et al., 1999; Rosekind et al., 1994). It is also relatively short – whereas driving simulation tests often rely on a longer duration of driving in order to detect fatigue. This test has also been shown to have good test–retest reliability (Kribbs and Dinges, 1994). Mathematical fatigue models, such as those used in fatigue management systems (FMS), have been compared to these simple tests of performance in an attempt to quantify the risk of impairment in the real world (Dawson & Fletcher, 2001; Fletcher & Dawson, 2001; Dorrian et al., 2005). However, while studies have attempted to demonstrate the links between PVT performance and accident risk (van Dongen et al., 2003), performance has not been experimentally compared to more complex tasks such as an interactive driving simulation (i.e. between measures). We have recently measured both types of performance in locomotive engineers (Roach et al., 2001), but did not directly equate performance impairment on the two measures. This would be of benefit to road safety in that methods of fatigue management currently being used and developed for industry are measured against laboratory measures such as the PVT. If we are able to provide a direct link from such a measure to a more realistic simulated driving task, then these systems will have more validity for the domain of driving, and for the assessment of fitness to drive. Therefore, it is the primary aim of this study to examine the potential for a simple task of reaction time (the PVT) to provide this link, by directly comparing performance under conditions of increasing fatigue/sleepiness, using both the PVT and an interactive driving simulation.

## 2. Methods

### 2.1. Design

We used a repeated measures design as part of an extended wakefulness protocol to systematically increase the fatigue levels of the participants and directly compare performance on both PVT and driving simulation tests. 'Wakefulness' refers to the period of time the participants were required to stay awake.

### 2.2. Participants

Sixteen healthy volunteer participants (8 male, 8 female) were recruited for participation. All reported being free of medication, and were within the normal range for body mass index (Heyward and Stolarczyk, 1996: mean BMI = 25.7; S.D. = 5.1). Participants were aged between 22 and 56 years (mean = 33.6 years; S.D. 11.1 years). They had been driving for at least 2 years, and were screened for the absence of sleep disorders using a general health questionnaire which incorporated the Epworth Sleepiness Scale (ESS: Johns, 1991, 1992; Carpenter and Andrykowski, 1998: all participants ≤10; mean 6.53; S.D. 2.89). One volunteer withdrew due to illness at the commencement of sleep deprivation. Therefore, 15 individuals (7 male, 8 female) completed the study. The participants gave written informed consent and were compensated for their participation. The study had approval from the Human Research Ethics Committees of the University of South Australia and the Queen Elizabeth Hospital.

### 2.3. Procedure

Prior to entering the laboratory participants were asked to provide detailed information about their sleep for one week prior to test sessions using a sleep diary. For each sleep period (including naps), they recorded date and time of sleep onset, the final wake time and the number and length of awakenings during the sleep period. Objective assessments of sleep/wake prior to test sessions were also made using activity monitors and actiware-sleep software (Kushida et al., 2001). Activity monitors (Mini-Mitter Actigraph-L devices) are wrist-worn, and measure activity using a piezo-electric accelerometer with a sensitivity of 0.1 g. An analogue sensor is used to count the number of movements made by the individual wearing the device. The information is collected and stored in 1-min epochs. Actiware™-Sleep software (Cambridge Neurotechnology Ltd.; Mini Mitter Co. Inc.) enables downloading of the raw activity counts. Estimates of sleep derived using these devices and algorithms have been validated against laboratory polysomnography (Kushida et al., 2001; Pilsworth et al., 2001; Stanley et al., 2000) and in shiftwork, and aviation settings (James et al., 2003; Signal et al., 2005). Participants were required to wear the activity monitor on their non-dominant wrist at all times for one week prior to the study, except whilst showering (or in any other situation were the device was likely to be damaged).

Participants arrived at the Centre for Sleep Research laboratory at 19:00 h on a Friday evening, and remained there until 21:30 h on Sunday night. Upon arrival, participants completed a series of training sessions to familiarise them with the testing protocol and eliminate practice effects. A night of baseline sleep (8 h opportunity from 22:00 to 06:00 h) was obtained on the first night, after which 26 h of sleep deprivation without napping commenced. Testing began at 10:00 h, 4 h into extended wakefulness. Subsequent testing occurred at 14:00, 00:00 and 06:00 h (see Fig. 1). Following the period of extended wakefulness, participants were given an 8 h recovery sleep opportunity, from 09:00 to 17:00 h on Sunday. Two hours after waking up from this recovery sleep period, a final test battery was conducted at 19:00 h. Testing bouts were 50 min in length and consisted of a 30-min drive on the YDS, bracketed before and after by a 10-min PVT. Groups of four subjects were tested at each time, alternating between tasks. Subjects were supervised in the laboratory at all times during the test sessions, and were permitted to carry out other quiet activities when not testing (e.g. reading, watching television). At the completion of testing, subjects were sent home in a taxi.

### 2.4. Driving simulator

Driving performance was measured using the York Driving Simulation program (YDS: DriveSim 3.00, York Computer Technologies, Kingston, Ontario, Canada). The program monitors driving impairment on a number of variables (lane drifting, speed deviation, collision status). Lane drifting is the typical manifestation of sleepiness-related driving impairment (Horne and Reyner, 1995). Subjects "drive" using pedals for braking and acceleration, and a standard steering wheel. Studies have found simulated- and real life driving behaviour to correlate quite highly (Lee et al., 2003). Driving performance parameters of speed (kph; Mean kph; S.D. kph), collision status, and road position (%; mean%; S.D.%) for each driving session were automatically detected and logged by the YDS in 1 s intervals. Lane position was denoted as a percentage. A recording of 0% was obtained when the vehicle was off the road to the right, and 100% when it was off the road to the left. Left lane drift was chosen as the measure of impairment for this study as all drivers were instructed to maintain a position in the centre of the left hand lane during the normal course of driving. Due to the need to overtake vehicles and avoid road works by moving into the right hand lane, it was determined that right lane drift did not give a true indication of impairment so was not included in the drift analysis. Left lane drifting incidents were defined as a road position >85%, indicating a crossing of the white markings on the left hand side of the road. As posted speed limit varied (30 and 60 kph), the posted speed limit for each 1 s interval had been previously recorded manually by logging the time of each posted speed limit change throughout the simulation. Speed deviation was then calculated as kph over or under the posted speed limit. Collisions were identified as both collisions into another vehicle or into an object. The YDS logs data 10 times per second,

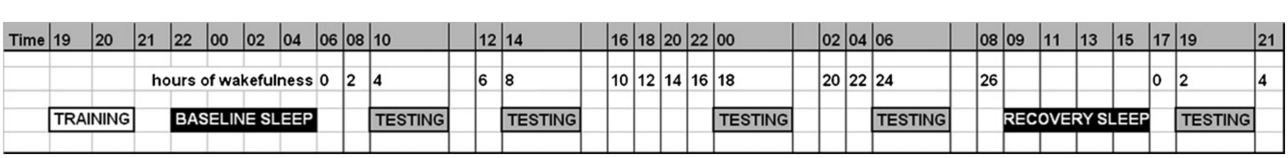


Fig. 1. Study schedule testing sessions are shown by the shaded grey areas.

therefore to facilitate analysis with which to address the aims of the study, a software tool was developed ("*DriverNator*", Quuxo Software, Adelaide, Australia) which enabled the rapid processing of this data into user-configurable periods or epochs of interest. We thus were able to specify variable time intervals for which summary statistics could be rapidly calculated (e.g. 1 s; 30 s; 1 min; 5 min) for each output variable.

### *2.5. Psychomotor vigilance task*

The psychomotor vigilance task (PVT) measures visual reaction time (RT), using a portable device (PVT-192: Ambulatory Monitoring Inc., Ardsley, New York). Each PVT test runs for a period of 10 min. Participants respond to a visual stimulus presented at a variable interval (2,000–10,000 ms) by pressing either the right or left push-button with the thumb of their dominant hand. The LED display shows their RT in milliseconds. Participants are instructed to press the button as soon as they see the numbers appear in the LED window. The number in the display indicates RT in milliseconds, therefore the smaller the number, the quicker the response. Measures extracted from the PVT were mean reaction time (RT), response speed (1/RT), and number of *lapses* (responses >500 ms, i.e. "missed" stimuli). Subjects completed 3 practice trials in training, as research shows that the PVT has a 1–3 trial learning curve (Dinges and Powell, 1985; Dinges et al., 1997).

### *2.6. Statistical analyses*

All data were checked for normality prior to analysis. One outlier was identified within the driving data and subsequently eliminated (Tabachnick and Fidell, 1996). PVT data showed a moderate positive skew and were adjusted using the standard method of square root data transformations (Tabachnick and Fidell, 1996). Violations of sphericity were corrected using Huynh-Feldt adjustments; however, degrees of freedom reported in the text are based on the study design. Changes in driving performance and PVT metrics (RT, response speed, lapses) as a result of extended wakefulness were assessed by repeated measures ANOVA. Planned means comparisons were conducted to compare differences where appropriate ($p < 0.05$).

As with standard statistical methodology, correlational analysis was conducted to determine the strength of the relationship between the two measures. However, using a correlation coefficient or regression analysis to compare a new method of measurement against an established one will often show a relationship where none actually exists (Bland and Altman, 1995). We have used a Bland–Altman plot (Bland and Altman, 1986, 1995) to measure the agreement between lane drift and PVT lapses, as two potentially equivalent measures of fatigue. Bland–Altman plots simply graphically illustrate the differences between the two measures (on the *y*-axis) against the mean of both (on the *x*-axis). The mean difference is the estimated bias or the systematic difference between the measures and the standard deviation of the differences indicates the random fluctuations around the mean (Bland and Altman, 1995). The level of concordance between the two measures is determined by calculating the mean and 95% 'limits of agreement' (the mean difference ± 2 S.D.) and then analysing the distribution of the data points. If the distribution shows a significant correlation, the level of agreement between the two measures is low and likely to be affected by random error. If the correlation is not significant, the two measures are in agreement and the new method can be substituted interchangeably for the established one.

## 3. Results

The effects of extended wakefulness on driving parameters are shown in Fig. 2. Subjects had significantly more lane drifting incidents as fatigue increased ($F[3,36] = 11.54$, $p = 0.002$ see Fig. 2A). Planned means comparisons revealed a significant increase in lane drifting between the first (4–6 h), and the last drive (24–26 h) ($t[12] = -4.07$, $p = 0.001$[one-tailed]). There was also a significant decrease in lane drifting after recovery sleep ($t[12] = 4.18$, $p < 0.001$[one-tailed]), almost back to baseline (4–6 h) levels ($t[13] = -1.7$, $p = .056$). Speed deviation varied significantly with extended wakefulness ($F[3,36] = 6.20$, $p = .006$ see Fig. 2A), and means comparisons revealed a significant increase ($t[12] = -3.53$, $p = 0.002$ [one-tailed]). Collision

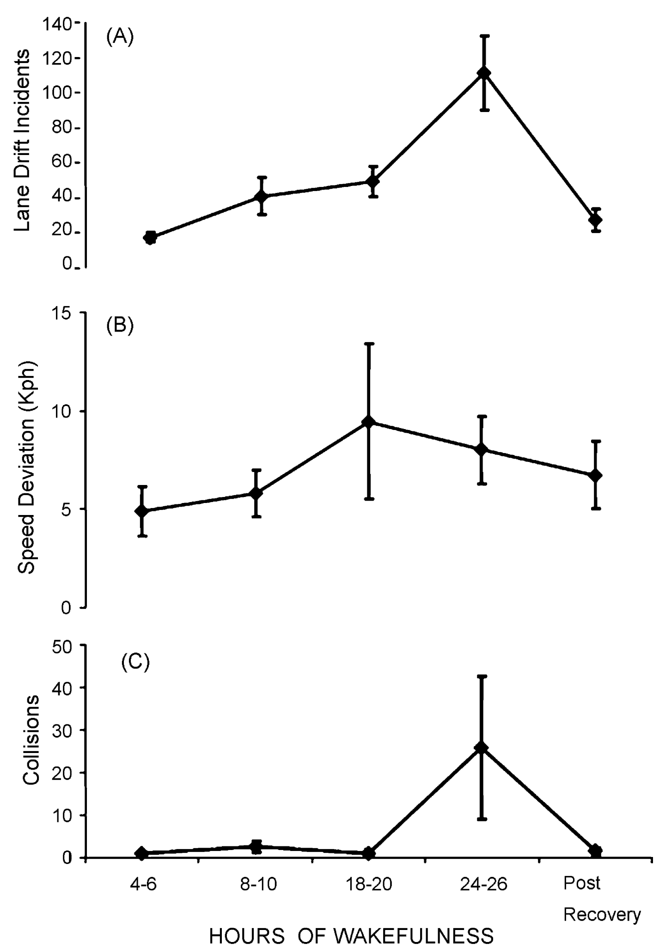


Fig. 2. Mean driving performance data, demonstrating the effects of extended wakefulness on (Panel A) lane drifting; (Panel B) speed deviation and (Panel C) collisions.

status did not show any significant increases with extended wakefulness ($F[3,36] = 1.87$, $p = 0.196$ see Fig. 2C) or subsequent recovery ($t[12] = 1.32$, $p = 0.11$ [one-tailed]).

The effects of extended wakefulness on PVT parameters are shown in Fig. 3. Repeated measures ANOVA revealed significant differences in mean RT ($F[3,36] = 10.52$, $p = 0.006$ see Fig. 3A). Planned means comparisons showed a significant increase in RT ($t[13] = -3.79$, $p = 0.001$[one-tailed]), which then significantly decreased again after recovery sleep ($t[13] = 3.46$, $p = 0.004$ [one-tailed]). There was no difference between baseline and recovery ($t[14] = -2.06$, $p = 0.059$[two-tailed]). Response speed ([1/mean RT] × 1000) also showed significant differences over extended wakefulness ($F[3,36] = 38.9$, $p < 0.001$ see Fig. 3B). Response speed significantly decreased between the first drive at 1000 h on Saturday and the last drive at 0600 h on Sunday ($t[13] = 8.9$, $p < 0.001$[one-tailed]). After recovery sleep, response speed again significantly increased ($t[13] = -7.27$, $p < 0.001$ [one-tailed]), although not to baseline levels ($t[14] = 1.56$, $p = 0.141$ [two-tailed]). The same pattern was found in PVT lapses, with significant differences found over extended wakefulness ($F[3,36] = 15.39$, $p < 0.001$ see Fig. 3C), and means comparisons showing a significant increase with

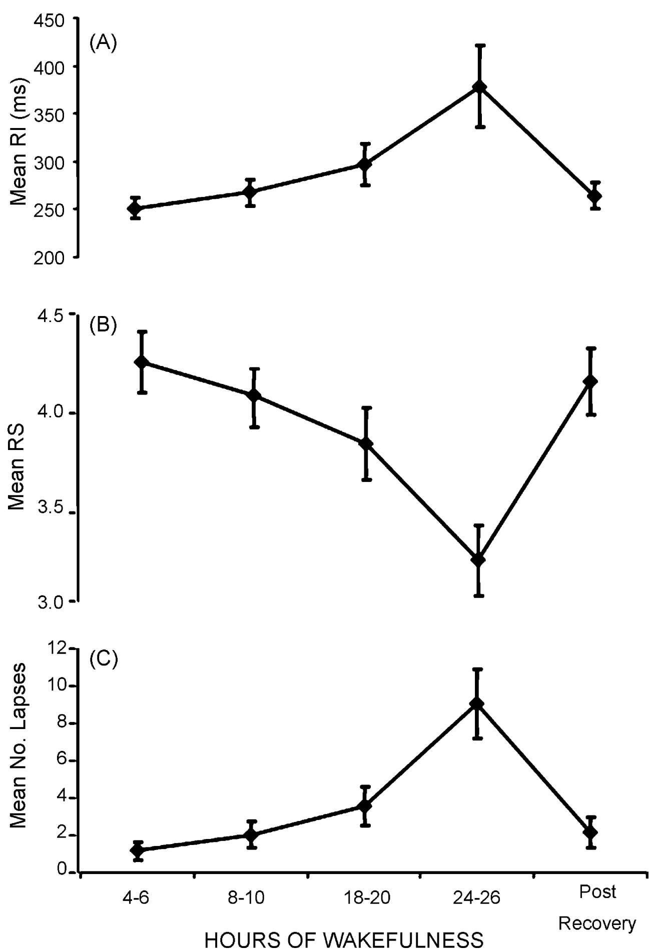


Fig. 3. Mean PVT data, demonstrating the effects of extended wakefulness on (Panel A) reaction time (ms); (Panel B) response speed (1/RT), and (Panel C) number of lapses (RT > 500 ms).

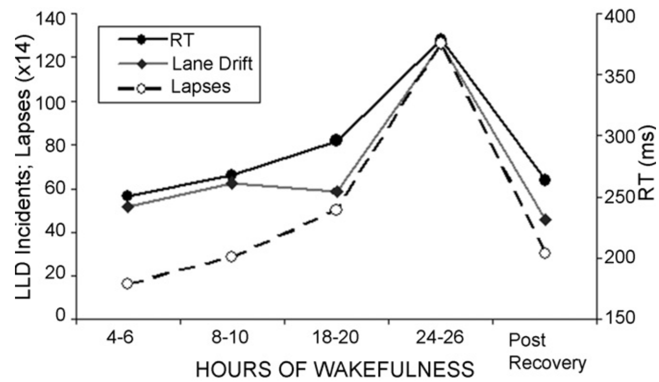


Fig. 4. PVT and driving data: mean lane drift vs. (a) mean reaction time (solid line: $R^2 = 0.96$); and (b) lapses (dotted line: $R^2 = 0.96$).

fatigue ($t[13] = -4.38$, $p < 0.001$[one-tailed]), and decrease after recovery sleep ($t[13] = 4.04$, $p < 0.001$[one-tailed]). However, lapses did not return to baseline levels, and significant differences were found between baseline and recovery ($t[14] = -2.36$, $p = 0.033$[two-tailed]).

To directly compare the simple (PVT) and complex (driving) measures of performance, pairs of variables from the two performance tests were plotted, and $R^2$ values calculated (see Fig. 4). As predicted, left lane drift was significantly correlated to both RT ($R^2 = 0.92$; $p = 0.01$) and lapses ($R^2 = 0.93$; $p < 0.01$).

To determine if this was a true association, a Bland–Altman plot was constructed for paired comparisons between number of lapses and lane drifting. The plot of the difference between lapses and lane drift showed a bias of 43.79 (95% CI, 14.7–72.9 see Fig. 5). The lower limit of agreement was −61.28 (95% CI, −257.55–134.99). The upper limit of agreement was 148.86 (95% CI, −47.41–345.13). The square of the difference between the two performance scores was tested for association with the mean incident score using regression analysis and was found to be statistically significant ($p < 0.001$), indicating a systematic error within the measures. That is, the level of concordance between the two measures deteriorated with extended wakefulness within the limits of the study protocol.

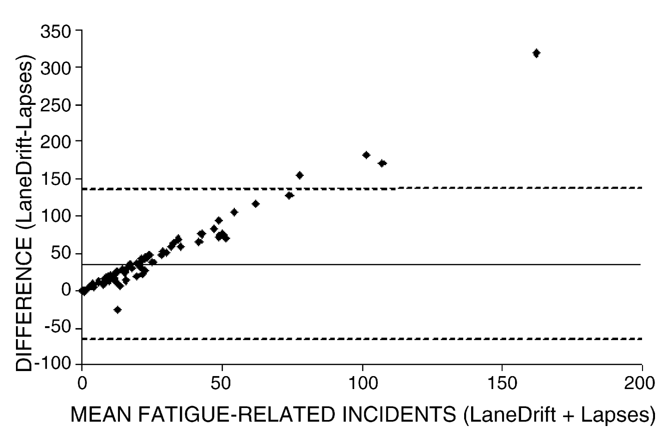


Fig. 5. Bland–Altman plot showing paired comparisons between lapses and left lane drift. The solid line represents the mean (43.79) and the dotted lines represent the limits of agreement. The positive regression trend indicates an increase in the difference between the measures as the mean impairment also increases.

## 4. Discussion

Consistent with previous studies, extended wakefulness resulted in performance impairment in all measures. Driving performance measures (lane drifting, speed deviation and collision status) showed some variation with increasing wakefulness. As expected, lane drift was the most sensitive measure of fatigue, showing the most significant impairment over extended wakefulness and the least between subject variability (Philip et al., 2003). Speed deviation also showed significant impairment with extended wakefulness, but due to between subject variability, especially during the third driving period, this result was weaker. As can be seen from Fig. 2B, this time period also shows the greatest impairment in speed deviation. While this seems counter-intuitive, when circadian factors are accounted for, the result is not surprising. The third drive occurred between the hours of 12–2 a.m., corresponding with the onset of the circadian nadir in performance (Zee and Turek, 1999). It is likely that at this time, due to the circadian driven reduction in mental alertness, some participants sacrificed monitoring their speed in favour of attempting to avoid accidents. The fact that collision status remained low during this same time period supports this theory. Collision status did not show any significant increase over hours of wakefulness. As shown in previous studies (Philip et al., 2003; Otmani et al., 2005), collision status is the least sensitive of all driving measures. One possible reason for this is that it is also the most tangible impairment measure to the drivers. That is, most subjects may be highly sensitive to having an accident thus pay more attention to avoiding it even in simulated situations. However, collision status showed the greatest between subject variability, with some participants crashing excessively and others not at all. Again, this is not an unrealistic result as not only will subjects vary in their resilience to fatigue, but they will also vary in their experience of computerised tasks. Participants with experience playing computer games are generally more comfortable with a computer generated driving simulator. Therefore, it is important to ensure that practice effects are adequately addressed within study design, using extensive opportunity for subjects to practice simulated driving.

Statistical analyses using correlational methods showed a clear agreement between lane drifting and PVT reaction time. However, as described by Bland and Altman (1995), a significant correlation between measures does not always indicate an agreement between them. In this study, the Bland–Altman analysis demonstrated a systematic error within the measures. This is denoted by the positive regression between the limits of agreement on the Bland–Altman plot (Fig. 5). As the subjects became increasingly impaired, the difference between the measures increased. If a genuine association existed between the measures, the regression line would be horizontal and the correlation zero, as the variances between the measures would remain equal with increasing impairment. Thus, this result indicates that a simple measure of RT is not indicative of impairment on a more complex task such as driving. It is likely that this error is caused by the large variation in driving between subjects. With increased wakefulness, subjects showed large differences in the magnitude of impairment on the driving simulator, which was not seen in the PVT data. The level of impairment in the PVT measures remained relatively equal across subjects. It is important to note however, that both measures showed similar patterns of impairment over the period of extended wakefulness. That is, both measures showed that impairment steadily increased for the first 20 h, after which there was a sharp increase in impairment at 24–26 h, and a dramatic decrease after recovery sleep. This suggests that, although simple RT is not effective at determining driving impairment on its own, it may be essential as a component of a test battery which could be used to examine fitness to drive. Further research is required to examine this supposition.

As expected, all measures returned to, or near baseline after an 8 h recovery sleep. Consistent with recent studies, 8 h of continuous sleep after 26 h of extended wakefulness appears sufficient for the body and brain to return to normal functioning. Interestingly, some performance measures were better after recovery sleep than at baseline. There are two possible reasons for this. Firstly, this difference may represent an end-of-test effect. Secondly, and most importantly, the baseline and recovery testing sessions were conducted at different times of day, which would alter the underlying levels of alertness due to the human circadian rhythm. While we have compared data on driving/PVT performance for baseline and recovery, it is important to note that this comparison may be confounded by the circadian (i.e. time of day) element. Although both test sessions occurred within the protocol following an 8 h sleep opportunity, the baseline session occurred at 10:00 h, while the recovery session took place at 19:00–20:00 h. Clearly, the difference in circadian alertness at this point is not only notable, but also vulnerable to individual differences (e.g. the older subjects may have been less alert than the younger subjects in the recovery session).

While this sample was sufficiently powerful to show statistically significant differences, larger subject numbers are ultimately required to allow broader application of the findings. This type of analysis can be repeated in larger studies and may generate results which have potential to assist in determining fitness to drive in at-risk populations such as those suffering from sleep disorders.

While measures of simple (PVT) and complex (simulated driving) performance are clearly linked and may correlate highly, our results using the Bland–Altman technique show for the first time that there is a systematic error which prevents a true mathematically predictive relationship. Thus, while simple tests of performance such as the PVT are undoubtedly useful in the measurement of fatigue, they are not able to replace more complex tests such as driving simulation. We suggest that further research is required to identify potential *combinations* of tools which may facilitate adequate management of fitness to drive, rather than a single, "silver bullet" measure. Such combinations will naturally require being restricted to the laboratory or research environment, unless more compact and brief versions can de developed in the future to allow roadside testing, for example.

## Acknowledgements

The authors gratefully acknowledge the support of the Australian Transport Safety Bureau (ATSB). The authors wish to thank Alison Bagnall, Ryan Higgins, James Swift, Sophie Walenczykiewicz and Dayna Griffin for help with recruitment and data collection. We thank Michael Gratton of Quuxo Software for software programming, and Martin York for assistance with the YDS. We would also like to thank Jill Dorrian and Katie Kandelaars for help with data analysis, and Sally Ferguson for reading the manuscript.

## References

Arnedt, J.T., Geddes, A.C., MacLean, A.W., 2005. Comparative sensitivity of a simulated driving task on self-report, physiological, and other performance measures during prolonged wakefulness. J. Psychol. Res. 58, 61–71.

Austroads Guidelines, 2003. Assessing fitness to drive for commercial and private vehicle drivers, third ed., September 2003: ISBN 0 85588 507 6.

Banks, S., Catcheside, P., Lack, L.C., Grunstein, R.R., McEvoy, R.D., 2005. The Maintenance of Wakefulness Test and driving simulator performance. Sleep 28 (11), 1381–1385.

Bland, J.M., Altman, D.G., 1986. Statistical methods for assessing agreement between two methods of clinical measurement. The Lancet i, 307–310.

Bland, J.M., Altman, D.G., 1995. Comparing methods of measurement: why plotting difference against standard method is misleading. The Lancet 346, 1085–1087.

Carpenter, J.S., Andrykowski, M.A., 1998. Psychometric evaluation of the Pittsburgh sleep quality index. J. Psychosom. Res. 45, 5–13.

Dawson, D., Fletcher, A., 2001. A quantitative model of work-related fatigue: background and definition. Ergonomics 44 (2), 144–163.

Dawson, D., Reid, K., 1997. Fatigue, alcohol and performance impairment. Nature 388 (6639), 235.

Dinges, D.F., Powell, J.W., 1985. Microcomputer analyses of performance on a portable, simple visual RT task during sustained operations. Behav. Res. Methods Instrum. Comput. 22 (1), 69–78.

Dinges, D.F., Pack, F., Williams, K., Gillen, K.A., Powell, J.W., Ott, G.E., Aptowicz, C., Pack, A.I., 1997. Cumulative sleepiness, mood disturbance, and psychomotor vigilance performance decrements during a week of sleep restricted to 4–5 h per night. Sleep 20 (4), 267–277.

Dobbie, K., 2002. Fatigue related crashes: an analysis of fatigue related crashes on Australian roads using an operational definition of fatigue. Australian Transport Safety Bureau: 30.

Dorrian, J., Rogers, N.L., Dinges, D.F., 2005. Psychomotor vigilance performance: a neurocognitive assay sensitive to sleep loss. In: Kushida, C. (Ed.), Sleep Deprivation: Clinical Issues, Pharmacology and Sleep Loss Effects. Marcel Dekker Inc., New York, pp. 39–70.

Fletcher, A., Dawson, D., 2001. A quantitative model of work-related fatigue: empirical evaluations. Ergonomics 44 (5), 475–488.

Fletcher, A., Lamond, N., van den Heuvel, C., Dawson, D., 2003. Prediction of performance during sleep deprivation and alcohol intoxication by a quantitative model of work-related fatigue. Sleep Res. Online 5 (2), 67–75.

George, C.F., Boudreau, A.C., Smiley, A., 1996. Simulated driving performance in patients with obstructive sleep apnea. Am. J. Respir. Crit. Care Med. 154 (1), 175–181.

George, C.F., Findley, L.J., Hack, M.A., McEvoy, R.D., 2002. Across-country viewpoints on sleepiness during driving. Am. J. Respir. Crit. Care Med. 165 (6), 746–749.

Heyward, V.H., Stolarczyk, L.M., 1996. Applied Body Composition Assessment. Human Kinetics, Leeds.

Horne, J.A., Reyner, L.A., 1995. Sleep related vehicle accidents. Brit. Med. J. 310, 565–567.

James, F.O., Komourian, J., Morin, C., Boivin, D.B., 2003. Correlation of actigraphic measures of sleep quality with nightcap and polysomnography in diurnal sleep of night shift workers. Sleep 26, A401–A402, Abstract 1011. R.

Jewett, M.E., Dijk, D.J., Kronauer, R.E., Dinges, D.F., 1999. Dose-response relationship between sleep duration and human psychomotor vigilance and subjective alertness. Sleep 22 (2), 171–179.

Johns, M.W., 1991. A new method for measuring daytime sleepiness: the Epworth Sleepiness Scale. Sleep 14 (6), 540–546.

Johns, M.W., 1992. Reliability and factor analysis of the Epworth Sleepiness Scale. Sleep 15 (4), 376–381.

Kribbs, N.B., Dinges, D.F., 1994. Vigilance decrement and sleepiness. In: Harsh, J.R., Ogilvie, R.D. (Eds.), Sleep onset mechanisms. American Psychological Association, Washington DC, pp. 113–125.

Lee, H.C., Cameron, D., Lee, A.H., 2003. Assessing the driving performance of older adult drivers: on-road versus simulated driving. Accid. Anal. Preven. 35, 797–803.

Kushida, C.A., Chang, A., Gadkary, C., Guilleminault, C., Carrillo, O., Dement, W.C., 2001. Comparison of actigraphic, polysomnographic, and subjective assessment of sleep parameters in sleep-disordered patients. Sleep Med. 2 (5), 389–396.

Otmani, S., Pebayle, T., Roge, J., Muzet, A., 2005. Effect of driving duration and partial sleep deprivation on subsequent alertness and performance of car drivers. Physiol Behav. 84 (5), 715–724.

Philip, P., Vervialle, F., Le Breton, P., Taillard, T., Horne, J.A., 2001. Fatigue, alcohol, and serious road crashes in France: factorial study of national data. Brit. Med. J. 322, 829–830.

Philip, P., Taillard, J., Klein, E., Sagaspe, P., Charles, A., Davies, W., Guilleminault, C., Bioulac, B., 2003. Effect of fatigue on performance measured by a driving simulator in automobile drivers. J. Psychosom. Res. 55, 197–200.

Philip, P., Sagaspe, P., Moore, N., Taillard, J., Charles, A., Guilleminault, C., Bioulac, B., 2005. Fatigue, sleep restriction and driving performance. Acc. Anal. Prev. 37, 473–478.

Pilsworth, S.N., King, M.A., Shneerson, J.M., Smith, I.E., 2001. A comparison between measurements of sleep efficiency and sleep latency measured by polysomnography and wrist actigraphy. Sleep 24; Supplement, Abstract #171R, p. A106.

Roach, G.D., Dorrian, J., Fletcher, A., Dawson, D., 2001. Comparing the effects of fatigue and alcohol consumption on locomotive engineer's performance in a rail simulator. J. Hum. Ergol. 30 (1–2), 125–130.

Rosekind, M.R., Graeber, R.C., Dinges, D.F., Connell, L.J., Rountree, M.S., Spinweber, C.L., Gillen, K.A., 1994. Crew factors in flight operations IX: effects of planned cockpit rest on crew performance and alertness in long-haul operations. (NASA Technical Memorandum No. 108839). Moffett Field, CA: National Aeronautics and Space Administration.

Signal, T.L., Gale, J., Gander, P.H., 2005. Sleep measurement in flight crew: comparing actigraphic and subjective estimates to polysomnography. Aviat. Space Environ. Med. 76 (11), 1058–1063.

Stanley, N., Dorling, M.C., Dawson, J., Hindmarch, I., 2000. The accuracy of mini-motionlogger and actiwatch in the identification of sleep as compared to sleep EEG. Sleep, 23 (Supplement, Abstract#1536N), A386.

Tabachnick, B.G., Fidell, L.S., 1996. Using multivariate statistics, third ed. Harper Collins, New York.

van Dongen, H.P., Maislin, G., Mullington, J.M., Dinges, D.F., 2003. The cumulative cost of additional wakefulness: dose-response effects on neurobehavioural functions and sleep physiology from chronic sleep restriction and total sleep deprivation. Sleep 26 (2), 117–126.

Williamson, A.M., Feyer, A.-M., 2000. Moderate sleep deprivation produces impairments in cognitive and motor performance equivalent to legally prescribed levels of alcohol intoxication. Occup. Environ. Med. 57 (10), 649–655.

Zee, P.C., Turek, F.W., 1999. Introduction to sleep and circadian rhythms. In: Turek, F.W., Zee, P.C. (Eds.), Regulation of Sleep and Circadian Rhythms: 1–17. Marcel Dekker Inc., New York.