REPORT OUTLINE

Date: September 2005
ISBN: 1 921168 21 8
Title: Two-up Driving: Fatigue Study
Address: National Transport Commission
Level 15/628 Bourke Street
MELBOURNE VIC 3000
E-mail: ntc@ntc.gov.au
Website: www.ntc.gov.au

Type of report: Final report

Objectives:
The Heavy Vehicle Driver Fatigue (HVDF) review is a key component of the third heavy vehicle reform package. The aim of this on road study was:

To research the fatigue risk implications of two-up driving in the heavy vehicle industry; and

To identify possible criteria from the study to mitigate the risk of heavy vehicle accidents due to driver fatigue. These criteria to be used to develop parameters for two-up driving in the Heavy Vehicle Driver Fatigue (HVDF) reform.

NTC Programs:
Fitness for Duty, Heavy Vehicle Driver Fatigue

Key Milestones:
This report is being released in conjunction with the proposed two-up policy proposal. Public consultation on both draft legislation and revised policies developed for the Heavy Vehicle Driver Fatigue reform will take place later in 2006 upon receipt of the draft legislation.

Abstract:
The practice of two-up driving is used to cover ultra long distances within Australia. Since not enough is known about the practice, the NTC undertook to conduct an on road study of two-up driving to examine the issues surrounding the practice in order to develop clearly defined parameters for its use. Thirty nine drivers were monitored over a period of at least four weeks with an ARRB Fatigue Monitoring Device fitted to their truck. This allowed collection of a continuous stream of data reflecting performance, from which inferences about alertness and fatigue could be drawn.

The general conclusion from the study is that the level of fatigue risk in the two-up heavy vehicle industry is
quite low, with a relatively high and consistent level of driver performance.

**Purpose:**
For information.

**Key words:**
two-up, long distance heavy vehicle transport, alertness, performance, fatigue risk, heavy vehicle driver fatigue, long haul two-up driver teams, bonneted, cabover.

**Enquiries to be addressed to:**
Jeff Potter
Senior Advisor Safety
National Transport Commission
L15/628 Bourke Street
MELBOURNE VIC 3000
FOREWORD

The National Transport Commission (NTC) is a body established under an intergovernmental agreement with a charter to develop, monitor, and maintain uniform or nationally consistent regulatory and operational reforms relating to road transport, rail transport, and inter-modal transport. The NTC is funded jointly by the Australian Government, States and Territories.

Fatigue is one of the main causes of crashes involving heavy vehicle drivers. The Heavy Vehicle Driver Fatigue (HVDF) Review is a key component of the Third Heavy Vehicle Reform Package. The aim of this review is to improve road safety through the implementation of policies and practices addressing the management of fatigue in the road transport supply chain.

Although findings from this on road study relating to fatigue cannot be used as intended, the NTC is making this report available for public comment.

The policy position put forward in two-up policy paper is subject to change. The NTC will be undertaking further consideration of these issues over the coming months and will advance a final draft Heavy Vehicle Driver Fatigue two-up policy proposal in mid 2006, after consultation with transport agencies and industry.

Final Heavy Vehicle Driver Fatigue draft policy proposals will be made available through the NTC website along with the draft legislation and a regulatory impact statement.

Enquiries on the results of this study should be addressed to Mr Jeff Potter, Senior Advisor Safety, Ph 03 9236 5000, email jpotter@ntc.gov.au

Michael Deegan
Acting Chairman
ACKNOWLEDGEMENTS

The study would not have been possible without the support of the participating transport companies and the individual drivers who kindly took part.

We would like to thank the Roads and Traffic Authority (RTA), NSW for the use of their actigraphs during the course of this study.

We would also like to thank ARRB who provided dedicated and capable people to assist with the data collection and the Victorian Institute of Forensic Medicine for the use of their expertise during the course of this study.
SUMMARY

The purpose of the two-up study was to attempt to resolve diverse stakeholder views by investigating the factors contributing to driver fatigue in long haul two-up driving teams and make recommendations for policy based on the findings from this study.

A number of issues were investigated:

- The influence of time of day on driver performance; specifically, whether night time driving was more risky in any sense than daytime driving.
- How driver performance changes across the duration of a trip.
- The effect of time on task and especially whether longer working periods without a reasonable break have a significant detrimental effect on driver performance.
- Whether the length of driving periods is a critical factor in driver performance.
- The amount of rest received by two-up drivers and how that affects driver performance.
- How the nature of the schedules experienced by drivers influences performance.
- The effect of personal factors on fatigue risk.
- The role of different categories of heavy vehicles in influencing performance.

Two-up teams were monitored on ultra long distance trips West to East, South to North and vice-versa to evaluate how two-up teams manage fatigue on these trips and to draw out potential parameters for policy development for two-up.

Thirty-nine drivers were tested over a period of at least 4 weeks with the ARRB Fatigue Monitoring Device fitted to their truck. This allowed collection of a continuous stream of data reflecting performance, from which inferences about alertness and fatigue could be drawn. The general conclusion from the study is that the level of fatigue risk in the two-up heavy vehicle industry is quite low, with a relatively high and consistent level of driver performance. However a number of results may have implications for the way in which two-up driving is conducted.

Firstly, the finding that there appears to be a small but consistent performance decrement during daylight driving may suggest that restrictions on driving after dark could be counterproductive although this may be explained otherwise by environmental factors.

Secondly, there is some evidence that long driving periods (over 9 hours) may have a detrimental effect on driver performance. The 6 hour work/rest cycle adopted by most drivers appears to be a safe one and may be a good model for two-up driving.

There is also evidence that personal factors, especially health factors, are related to driving performance in two-up drivers. As a consequence anything that improves the general health of drivers is a desirable outcome. The amount of rest that drivers obtain when not driving may also be an important influence on performance, suggesting that drivers should be encouraged to use the opportunity to sleep when it arises.

Drug testing of participants in the form of hair follicle testing was undertaken to ensure the fatigue monitoring data collected was not influenced by the use of stimulants. Samples could only be obtained from half of the participating drivers, however three quarters of
these (38% of all participants) showed that stimulants (legal or illicit)\(^1\) had been taken at least once during the data collection period (2 months). While the nature of the test means that it is not possible to determine whether they were affected while driving, the FMD data collected cannot confidently be used to provide a robust scientific basis for determining policy. Instead previous data provided by the heavy vehicle long distance industry and direct consultation with this sector of the industry will form the basis for developing fatigue policy for two-up operations.

It must also be pointed out that the long distance trips were completed safely and successfully by drivers whose hair samples showed no evidence of any stimulant use, indicating that these ultra long distance trips can be undertaken safely if properly managed.

Although there appeared to be indications that the type of vehicle in which two-up driving was conducted may contribute to driver performance, results were inconclusive. The particular cab-over vehicles used in this study did appear to have a detrimental effect on driver performance (however, no significant differences were found in comparisons of the performance data collected) and it should be noted that this analysis only included a small number of vehicles of the same make. Providing advice about the effect of these vehicles on drivers in long haul two-up situations may however be valuable.

Although findings from the study relating to fatigue cannot be used as intended, the NTC is making this report available for public comment, and to ensure that jurisdictions and operators are aware of the need to further investigate the prevalence of both legal and illicit stimulant use in the industry.

\(^1\) Some of the stimulants detected are present in prescription or over-the-counter medication.
1. INTRODUCTION .................................................................................................................1

2. METHODOLOGY ..............................................................................................................1
   2.1 Subjects .........................................................................................................................2
   2.2 Apparatus ......................................................................................................................2
       2.2.1 ARRB Fatigue Risk Questionnaire ........................................................................2
       2.2.2 ARRB Fatigue Monitoring Device ........................................................................2
       2.2.3 Psychomotor Vigilance Task (PVT) .......................................................................3
       2.2.4 Actigraphs ............................................................................................................4
       2.2.5 Other Sources of Data ............................................................................................4
       2.2.6 Drug Testing .........................................................................................................4
   2.3 Method ..........................................................................................................................4

3. RESULTS ...........................................................................................................................5
   3.1 Interpreting FMD Scores and Blood Alcohol Content (BAC) Equivalence ......................6
   3.2 FMD Results .................................................................................................................6
       3.2.1 Time of Day ...........................................................................................................10
       3.2.2 Trip Effects .........................................................................................................10
       3.2.3 Time on Task .......................................................................................................14
       3.2.4 Schedule Effects ..................................................................................................17
       3.2.5 Personal Variables ...............................................................................................19
       3.2.6 Truck Variables ....................................................................................................22
   3.3 PVT Cross-Validation ....................................................................................................24
   3.4 Drug Testing ................................................................................................................27
   3.5 Sleep Patterns of Drivers .............................................................................................30
   3.6 Fatigue Risk Questionnaire ........................................................................................32

4. FATIGUE ISSUES AND POOR PERFORMANCE ............................................................32

5. SAFE DRIVING PATTERNS ............................................................................................34

6. BIBLIOGRAPHY ...............................................................................................................36
LIST OF FIGURES

Figure 1. Responses per hour and mean response time by time of day– east-west drivers. 6
Figure 2. Responses per hour and mean response time by time of day– north-south drivers 7
Figure 3. Responses per hour and mean response time by time of day – haul truck drivers 8
Figure 4. Average responses by truck number for after-dark driving 8
Figure 5. Average responses by truck number for driving in daylight 9
Figure 6. Average responses by truck number for driving in daylight (6, 7and 13 removed) 9
Figure 7. Average responses by time of day 6, 7 and 13 removed) 10
Figure 8. Mean responses as a function of trip day number 11
Figure 9. Mean responses as a function of time on task period number 12
Figure 10. Average responses by trip day number 13
Figure 11. Average responses by time on task period number 13
Figure 12. FMD scores for all drivers (drivers with the longest period are to the right) 14
Figure 13. FMD scores for all drivers (drivers with the greatest average driving period length are to the right) 14
Figure 14. FMD responses by time of day for long period drivers 15
Figure 15. FMD responses by time of day for short period drivers 15
Figure 16. FMD responses by time on task for long period drivers 16
Figure 17. FMD responses by time on task for short period drivers 16
Figure 18. FMD scores by time of day for long trip schedule drivers 17
Figure 19. FMD scores by time of day for short trip schedule drivers 18
Figure 20. FMD scores by time on task for long trip schedule drivers 18
Figure 21. FMD scores by time on task for short trip schedule drivers 19
Figure 22. Distribution of FRQ scores 20
Figure 23. FMD scores by time of day for low FRQ scorers 20
Figure 24. FMD scores by time of day for high FRQ scorers 21
Figure 25. FMD scores by time on task for low FRQ scorers 21
Figure 26. FMD scores by time on task for high FRQ scorers 22
Figure 27. FMD scores by time of day for cab-over vehicle drivers 22
Figure 28. FMD scores by time of day for bonneted vehicle drivers 23
Figure 29. FMD scores by trip day number for cab-over vehicle drivers 23
Figure 30. FMD scores by trip day number for bonneted vehicle drivers 24
Figure 31. PVT means at each site 25
Figure 32. PVT means at each site (driver 15 data removed) 26
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>33</td>
<td>Mean number of lapses at each site (driver 15 data removed)</td>
<td>26</td>
</tr>
<tr>
<td>34</td>
<td>PVT means by time of day (driver 15 data removed)</td>
<td>27</td>
</tr>
<tr>
<td>35</td>
<td>Number of hair samples containing a stimulant (note that some samples contained more than one drug)</td>
<td>29</td>
</tr>
<tr>
<td>36</td>
<td>Actigraph results and FMD results for operators 7 and 8</td>
<td>30</td>
</tr>
<tr>
<td>37</td>
<td>FC 3+4 responses as a function of rest time through non driving periods</td>
<td>31</td>
</tr>
<tr>
<td>38</td>
<td>FC 3+4 responses as a function of rest time – averaged over journeys</td>
<td>32</td>
</tr>
</tbody>
</table>
1. INTRODUCTION

Two-up, or team driving is an important aspect of long distance transport operations around Australia. With the vast distances between rural centres, it presents many freight forwarders with the opportunity to have their freight delivered much sooner than a solo driver can manage. It benefits truck owners since the vehicle is often utilised for 24-hour periods, rather than being held idle while a safety-conscious driver obtains restorative sleep. As a result an owner will get better usage from the vehicle.

Discussions with large numbers of two-up drivers have emphasised that a successful two-up driving team is similar to a marriage (eg. Hanowski, Wierwille and Dingus, 2003), in that each driver must be comfortable with many aspects of the other’s driving skills, abilities and level of safety on the road. This is said to be the secret between getting, and not getting, the required sleep while the vehicle is in motion. However, if the resting driver does not fully trust the capability and performance of their co-driver, the (apparent) resting driver may not get adequate sleep and will later take control of the vehicle and drive through periods of drowsiness. The replaced driver will then likely find it hard to sleep if he/she knows the state of the driver.

Nevertheless, there are many potential safety benefits in two-up driving, if properly managed, including the potential to obtain relatively long periods of sleep while the other team member is driving and the ability to vary the driving/resting tasks as drivers require (based on two drivers working well together).

A raft of research has been conducted over the past two decades concerning the performance of solo drivers over differing lengths of driving shifts (eg. Williamson, Feyer, Friswell and Finlay-Brown, 2000a, b, c and d). Some research has focussed on two-up driving; however, most of this data has been subjective or anecdotal in nature. Very little objective performance monitoring has occurred. It is therefore problematic for policy makers to stipulate maximum hours of driving and minimum hours of rest and restorative rest breaks for two-up drivers when there are few clearly identified scientific measures of what represents a fatigue risk and what does not.

Under current regulations, two-up drivers can operate a vehicle for 144 hours over 14 days before an extended break away from the vehicle is required. Despite the lack of evidence one way or the other, this is considered too long by fatigue experts and many regulators. But what would be safe limits?

2. METHODOLOGY

A number of trucks were fitted with the ARRB Proactive Fatigue Management System in an effort to capture the performance of drivers as they drove routes such as Perth-Sydney, Perth-Brisbane, Perth-Port Hedland, Toowoomba-Darwin, Sydney-Darwin and Brisbane-Sydney. The two-up drivers involved were instructed to operate as they would normally so that typical patterns of driving could be evaluated. Drivers interacted with the system for a period of approximately four weeks. This allowed enough data to be collected to provide a reliable picture of drivers test performance and to wash out any practice effects.
2.1 Subjects

Subjects were asked by their employers to volunteer for the project. As drivers, they had the option to participate, refuse to participate, or to participate and withdraw at any time without penalty. A total of 39 drivers participated, with a mean age of 44. Thirty-four were male and five were female. Although not formally investigated, informal conversations with drivers over the course of the study suggested that the vast majority were experienced in the two-up situation. This may suggest some degree of self selection.

2.2 Apparatus

2.2.1 ARRB Fatigue Risk Questionnaire

The self-report Fatigue Risk Questionnaire (FRQ) obtains information regarding the age, Body Mass Index (BMI), sleep patterns and issues, diet and general exercise level of individuals. The instrument provides an indication of the level of fatigue risk that the individual brings with them into the work environment. Past research with the instrument has shown that drivers with low scores on the FRQ are less likely to become impaired at work than those who have scored high (Mabbott & Lloyd, 2003). Information from the questionnaires will be provided in the relevant sections.

2.2.2 ARRB Fatigue Monitoring Device

The ARRB FMDs are a forced-choice reaction time task measuring human reaction time to an auditory and visual stimulus (for full details see Mabbott, 2002).

In the current study the stimulus box was mounted on the dashboard of trucks. The driver responded by touching one of two pads mounted on the dashboard or alternatively two pads mounted on the steering wheel spokes. This arrangement was designed to enable the measurement of performance while not impeding the driving task.

The stimulus presentations were set to occur only when the truck was travelling above a predefined speed (usually 80 km/h) and did not occur when reversing or when turning movements were difficult (i.e. at low speed when many revolutions of the steering wheel are made). This high cut-off speed was used so that the machine did not produce stimuli when the drivers were operating in heavy traffic, such as on the approaches to large cities like Sydney. Once the critical speed was achieved, the drivers were tested at quasi random intervals varying from at least seven minutes, and up to ten minutes between stimuli. The drivers received no feedback about the speed or accuracy of their responses (unless they took longer than four seconds to respond). All drivers were trained and advised on exactly how the system worked. They were told how to log on and off the system and how to respond appropriately. Specifically, they were told that the system would produce an audible tone every seven to ten minutes accompanied by either a left or a right pointing arrow on the stimulus box. Following a right pointing arrow they should touch the right paddle and following a left pointing arrow they should touch the left paddle.

The fatigue monitoring devices collected data for each driver and the data was downloaded at regular intervals. This was analysed and performance profiles developed for each individual driving team. Performance profiles were also developed for driving schedules or groups of schedules, based on time of day and duration of trips. This was the main tool used to assess whether or not the schedules produce a fatigue risk. The correlation between scores on the Fatigue Risk Questionnaire and the FMD performance profiles was analysed to determine to what extent personal issues influence the fatigue risk in two-up driving.
Validity of the FMD

The validity of a measure can be defined as the extent to which the measure does in fact measure what it purports to measure. There is no one result that can establish the validity of a measure. Rather, validity is established by converging evidence from a number of sources. The FMD is a measure of human vigilance that is sensitive to fatigue. The converging evidence for this comes from a number of sources.

- The FMD taps the kind of cognitive processes known to be sensitive to fatigue effects (e.g. Johnson, 1982).
- The FMD is specifically designed to capture the slow reaction times (lapses) that are most sensitive to fatigue (e.g. Balkin et al., 2000).
- FMD responses are appropriately correlated with eye closures (Mabbott, 2004).
- FMD responses are sensitive to blood alcohol content (Mabbott and Lloyd, 2003).
- FMD responses are appropriately correlated with Psychomotor Vigilance Task (PVT) responses (current study).
- The pattern of FMD responses over a trip are mirrored by the pattern of PVT responses over a trip (current study).
- FMD responses are sensitive to the amount of rest obtained (current study).

In combination, these results provide strong converging evidence for the validity of the FMD as a measure of vigilance that is sensitive to fatigue.

Furthermore, it should be noted that the fact that the pattern of FMD responses across trips in the current study is somewhat unexpected (in the light of existing literature), but nevertheless mirrored by the PVT responses across trips, makes for quite strong evidence that the FMD and PVT are measuring the same thing.

2.2.3 Psychomotor Vigilance Task (PVT)

Human reaction-time (RT) has been used by psychologists and physiologists as an index of motor performance and also alertness. Continuous reaction time performance over a relatively short time period can reveal changes in performance ability caused by fatigue or drugs. Therefore RT performance is also useful as an indicator of fatigue.

The PVT 192 was designed and built by CWE Inc Ardmore PA in consultation with University of Pennsylvania staff Dr Dinges, Dr Kribbs and J Powell. The PVT is sold by Ambulatory Monitoring NY. It is a fully electronic computerised test presentation and data capture system for simple visual or auditory reaction time. The stimulus presentation strategy employed for both modalities is to maintain the stimulus until the subject responds, as opposed to presenting brief fixed length stimuli. When using visual stimuli a LED millisecond counter display acts both as stimulus and performance feedback display. In auditory only mode, there is no feedback. When a stimulus appears the driver presses a button to obtain RT. The test typically last ten minutes.

The device allows the user to vary all major task parameters to suit different applications. All response data is stored in non-volatile Random Access Memory (RAM) with an independent battery backup, and all data may be copied to a computer via the serial port. Nearly every aspect of response behaviour is recorded, such as unstimulated responses and
other inappropriate responses. Analysis software is provided for downloading the data onto a personal computer.

PVT data was collected for a subset of drivers in the study as a way of potentially cross validating the FMD performance measures. PVT measures were taken at the commencement of trips in Perth, at a mid-journey point in Port Augusta, and then either in Brisbane or Sydney. Further PVT tests were conducted on the return journey, in Port Augusta and when the drivers arrived back in Perth. The PVT was not employed as a primary measure of fatigue risk in this study for a number of reasons. Firstly, the logistical issues involved in regular PVT testing throughout the duration of a journey are effectively insurmountable. Secondly, and most importantly, the PVT cannot collect data continuously or ‘in vivo’. In order to complete the PVT the driver must stop driving. It was reasoned that this change of task would be likely to have an alerting effect and thus reduce the possibility of detecting periods of fatigue.

2.2.4 Actigraphs

The actigraphs employed in this study are wristwatch-like devices that measure gross motor activity via an internal accelerometer (Mini Mitter Actiwatch AW-64). In addition they contain an accurate clock and the ability to store data for weeks at a time. These features allow the monitoring of general activity levels across the day. Actigraphy is frequently employed to monitor sleep/wake patterns and sleep quality, as gross movement can be a good indicator of the sleep/wake cycle and sleep quality. Actigraphic measures were included in this study to help flesh out our understanding of the amount and nature of the sleep that is common in a two-up driving situation, to cross validate the performance data, and to lend insight into how the performance of drivers varies as a function of their sleep habits. Ten Actigraphs were utilised. The Actigraphs were worn by ten drivers on the east west run and four drivers on the north-south run to enable comparisons between their performances on different driving schedules.

2.2.5 Other Sources of Data

A brief questionnaire was developed to gather information that may help to provide a better understanding of the influence of the two-up partnership dynamics on performance outcomes. Questions were asked regarding confidence and trust, and their influence on perceived sleep and rest quality.

2.2.6 Drug Testing

Hair samples were collected from drivers at the end of the data collection period and the samples sent to the Victorian Institute for Forensic Medicine for analysis. This method of drug testing was chosen over urine and saliva methods because of its high level of reliability, validity and resistance to manipulation by masking agents. In addition it allowed coverage of the entire data collection window with just one sample. The technique cannot distinguish between drug use whilst driving and drug use at other times. It does, however, provide some information on the types of substances used by drivers during the study period which would have had an effect on the drivers’ alertness if they had been used during driving times.

2.3 Method

Ten trucks operating on the east-west run from a Perth base were fitted with the ARRB Fatigue Monitoring Device to measure the performance of drivers over a four week period.
A further set of six trucks operating on north-south routes to the north-west of Western Australia were also fitted. Unfortunately, due to incorrect wiring, no data was collected from two of these trucks. Additional trucks also fitted with FMDs included one Darwin based truck, operating one round trip per week to Sydney, one truck operating between Brisbane and Sydney, and one truck operating between Toowoomba and Darwin. The east-west trucks were driving one round trip per week while the Perth to north-west trucks were driving two round trips per week.

The Psychomotor Vigilance Task (PVT) was used to capture performance measures to be utilised to cross-validate the measures captured by the ARRB FMDs. PVT testing was conducted for one week and only on the east-west trucks. Tests were performed in Perth, Sydney and Brisbane when the trucks first arrived and when they were nearly ready to depart. In addition, testing was done in Port Augusta at a rest stop, for trucks travelling in both directions.

Ten of the east-west drivers and four of the north-south drivers were fitted with wrist worn actigraphs. This was used to assist in further cross validation of the performance data and to help provide a better understanding of the influence of sleep and rest on the performance of the drivers.

All of the thirty-nine drivers completed the ARRB Fatigue Risk Questionnaire. Of these 39 drivers thirteen had a body mass index over thirty. None of the drivers reported a serious sleep disorder.

3. RESULTS

Over 4400 hours of FMD data was collected. Three main dependent measures were utilised in this study.

1. RT (Reaction Time) is the time (in milliseconds) between the stimulus being presented and the driver responding by hitting a paddle.

2. FC 3+4 (Fatigue Codes 3 and 4) is a measure that combines those responses that are between two and three standard deviations slower than the mean reaction time (Fatigue Code 3) and those responses that are greater than three standard deviations slower than the mean reaction time. This measure gives the number of responses that are in the slowest 2.5% of all responses for a particular driver. It is designed to reflect substantial operational lapses in alertness on the part of drivers. (Note: the mean around which these deviations are defined is calculated over each driver’s entire data set.)

3. RC 2+3 (Response Codes 2 and 3) is a measure that combines those responses where the wrong button is hit (Response Code 2) and those responses where the subject fails to hit the button when stimulated within four seconds (Response Code 3).

Note: in the following figures FC 3+4 and RC 2+3 are read on the left hand scale, RT is read on the right hand scale, Hrs*10 is the number of driving hours multiplied by 10 and is read on the right hand scale.
3.1 Interpreting FMD Scores and Blood Alcohol Content (BAC) Equivalence

The primary measure of performance utilised with the FMD is the FC 3+4 measure. The reason for this is that lapses on this kind of task are thought to be particularly sensitive to fatigue (eg. Balkin et al., 2000; Hartley, personal communication). Such a measure indicates a statistically unusual phenomenon in relation to the subject’s own baseline. This intra-individual comparison is virtually mandatory given the individual variation in response times for speeded tasks. However, one problem with such data is that it is not immediately clear what level of performance decrement is implied by these slow responses and therefore whether this kind of responding represents a risk in any sense. A particularly useful approach to this problem is to relate these slow responses to blood alcohol content. In a recent study by Mabbott and Lloyd (2003) of the effect of BAC on FMD performance while on a driving simulator it was found that FC 3+4 responses began to be produced by operators at a BAC of 0.02. Thus it is reasonable to assume that the FC 3+4 measure is sensitive to decrements in performance equivalent to those that obtain with a BAC of 0.02. As a consequence it is also reasonable to assume that any change in FC 3+4 responses in the current study represents a change in performance equivalent to that which would result from a BAC of 0.02.

3.2 FMD Results

When the mean response times for all drivers were plotted against time of day, it was found that response times during daylight were slower than during the hours of darkness. This was more noticeable for the drivers on the east – west run than for drivers on the north - south run. (see Figures 1 and 2).

![Figure 1. Responses per hour and mean response time by time of day– east-west drivers.](image-url)
Following this discovery, a similar analysis was undertaken of data collected from haul truck drivers at a mine site (CaroSue). The mine site data showed much less variation between daylight and darkness (see Figure 3). This suggests (although does not guarantee) that the effect may not be a circadian rhythm effect. Therefore other explanations were sought. In a haul truck, there is a shield over the cab for protection against accidental dropping of ore onto the cab or windscreen. This shield projects slightly beyond the limits of the cab. As a consequence, it is not as bright inside a haul truck cab as in a typical truck cab in the current study, so the visual stimulus may be generally easier to see. It is possible that the slightly slower reaction times in daylight result from difficulty in seeing which indicator arrow was lit.
The hypothesis that the stimulus may have been difficult to detect under some conditions in some trucks was investigated further by dividing the data into two groups; data collected during the hours of darkness and data collected during daylight hours. The times of sunrise and sunset in Perth at the time of data collection varied between 0600 and 0630 for sunrise and 1800 to 1818 for sunset. For drivers travelling to the eastern states, sunrise would come earlier as the monitoring device operated on Perth time for the whole period. After looking at the response data, daylight was chosen as the hours between 0500 and 1800 inclusive. The daily averages for each truck for darkness and for daylight are plotted in Figures 5 and 6.
Examination of Figure 5 showed that trucks 6, 7 and 13 were very bad in daylight but were quite normal during the hours of darkness. Some trucks are known to have had trouble with the audio stimulus either being turned down too low or not working. For truck 13, the audio stimulus was turned off when the device was examined in the workshop after being retrieved from the truck. Without the audio stimulus, it would have been much more difficult to respond adequately to the visual stimulus. Rather than have this suspect data contaminate the results, daylight data for trucks 6, 7 and 13 was removed from further processing. This reduced the number of hours of data to 4072. The average daily responses by truck number were recalculated and are plotted in Figure 7.
The responses by time of day were similarly recalculated and are plotted in figure 7.

![Average responses by time of day](image)

**Figure 7.** Average responses by time of day 6, 7 and 13 removed

### 3.2.1 Time of Day

Although the responses during daylight are still slower on average than in darkness, the decrement in test performance observed during daylight hours is now broadly comparable with that observed in the mine site environment outlined above. The following analyses are based on this ‘screened’ data. It is worth noting however that even the mine site data shows some evidence of a decrease in test performance during daylight hours, relative to test performance after dark. There is some anecdotal evidence from drivers that this decrement during daylight hours reflects a genuine decrease in alertness. Drivers often mentioned that they became sleepier during the daylight hours than they did after dark. One explanation provided for this was that direct sun shining into the cab and onto the body during the day created a warming that promoted sleepiness and that was not necessarily compensated for by air conditioning. In addition, the daylight driving environment is clearly more demanding (e.g. heavier traffic requiring greater concentration), and this in itself might be expected to be more fatiguing (although it could conceivably have an alerting function also).

### 3.2.2 Trip Effects

In the analyses that follow a trip is defined as all the data collected within an interval bounded by gaps of sufficient length so as to constitute a reasonable break away from the truck. Arguably, ‘a reasonable break’ would allow two consecutive nights sleep away from the truck (Fatigue Expert Group, 2001). Therefore a gap in the FMD data of 48 hours was set as the marker for a trip boundary. Under this scenario, the total trip time exceeded a week on a number of occasions, with the longest being 376 hours, or just over 16 days. Two sets of average responses were calculated, one based on calendar days, and one based on the time on task measured in 24 hour periods starting from the time of the first record.
The first of these allows time of day effects to be examined, while the second is thought to be a better measure of the effects of trip length, particularly as it was common for the trucks on the east-west run to be leaving Perth towards the end of the normal working day. Means of the performance measures as a function of trip day number and time on task period number are plotted in Figures 8 and 9.

It should be noted that the gap between trips will include some driving near the start and end of trips when the truck is travelling slower than the speed at which the device activates. After discussion with some of the drivers, the device activation speed was set at 90 kmh for some trucks to prevent the device producing stimuli when travelling in heavy traffic on the approaches to large cities such as Sydney. As a result of this arrangement it is likely that at least one hour of driving is being missed at the start and end of many trips.

Figure 8. Mean responses as a function of trip day number
Figure 9. Mean responses as a function of time on task period number

The last day of trips (defined by a 48 hour break) in Figures 8 and 9 is either day 6, day 13 or day 16. Interestingly, responses are slower at the end of the trip, as indicated by the peaks in fatigue codes 3 and 4 and mean reaction time around those days. This is suggestive of a decrease in alertness and an increase in fatigue towards the end of a trip. However, it is important to note that the absolute level of fatigue risk implied by these results is relatively low. Drivers are responding slower than two standard deviations from their mean response time less than once every two hours. This is quite consistent responding and suggests relatively consistent levels of alertness. Furthermore, drivers’ average reaction times tend to vary by less than half a second over the duration of a trip. While there may be driving situations where a half second period of inattention could be problematic, it would be difficult to argue that such changes could be construed generally as a major risk. In addition, the BAC equivalence of these levels of responding is less than 0.02.

It was calculated that the minimum break that would define a trip duration no longer than seven days was 27 hours (that is, after a 7 day period all drivers had a break of at least 27 hours). Using this definition of what constituted a break between trips, the data was reprocessed and the results plotted in Figures 10 and 11.
As can be seen from Figures 10 and 11 drivers’ test performance is at its best in the middle of a trip (defined as a seven day period bounded by breaks of at least 27 hours) and at its worst towards the end of a trip. Unexpectedly, driver test performance is slightly worse at the beginning of a trip than in the middle of a trip. Thus, while these results are consistent with the idea that fatigue effects build up over time generally, they may also suggest a more complicated picture. Many lifestyle factors may mean that drivers start back at work fatigued. For example, drivers may have been busy all day with normal family activities or may have been involved in loading/unloading activities and hence may not be at their best when they begin a new trip in the evening. This curvilinear relationship between position in the trip and test performance is also evident in Figures 8 and 9.
3.2.3 **Time on Task**

Drivers typically changed positions with their co-driver four times within a 24 hour period. That is, the average driving and rest period was around six hours. However, some teams differed from this pattern quite significantly, with one driver driving up to 18 hours at a stretch.

In order to investigate the effect of time on task on driver alertness FMD results were plotted for drivers in ascending order of maximum and average driving hours (Figures 12 and 13). As can be seen, there is no evidence that drivers who drove longer periods (to the right of the graph) had performance levels that were systematically lower than those drivers who drove shorter periods (to the left of the graph).

![Operator response by increasing maximum drive hours](image)

**Figure 12.** FMD scores for all drivers (drivers with the longest period are to the right)

![Operator responses by increasing average driving hours](image)

**Figure 13.** FMD scores for all drivers (drivers with the greatest average driving period length are to the right)
While the above results suggest that there is no overall relationship between the length of driving periods and performance, it is never the less still possible that drivers who drove longer periods exhibit different time of day effects or trip effects to drivers who drive shorter stints. To this end drivers were divided into two groups (those with long driving stints and those with short driving periods) and the relevant effects plotted in Figures 14, 15, 16 and 17 below. Long driving periods were defined as those 9 hours or longer. Drivers in the short stint group tended to have driving periods that were 6 hours or less.

**Figure 14.** FMD responses by time of day for long period drivers

**Figure 15.** FMD responses by time of day for short period drivers

As can be seen from Figures 14 and 15 there is some evidence that drivers who drove long periods were more likely to exhibit slower responses during daylight hours, suggesting that...
they may be more susceptible to the daytime slowing observed in this data and hence may be less alert than short period drivers during the day. Indeed, there was a statistically significant difference between these two groups in terms of their average daily FC 3+4 responses (p < .05).

The time on task effects depicted in Figures 16 and 17, while they may suggest a difference between long and short stint drivers are difficult to interpret because of the relatively small number of driving hours contributing to the last day of data. It is possible that the long stint drivers actually become more alert on the last leg of the journey (sixth 24
hour period) while the short stint drivers begin to deteriorate. However, this result should be interpreted with caution due to the small sample involved.

3.2.4 Schedule Effects

There were essentially two groups of schedules experienced by drivers in the current study. One involved one round trip per week (the east-west trip and the Sydney-Darwin trip) while the other involved two round trips per week (the Perth-Kununurra trip and the Sydney-Brisbane trip). In order to investigate the effect of driving schedules on driver test performance drivers were grouped into a long trip schedule group and a short trip schedule group and compared. These effects are depicted in Figures 18, 19, 20 and 21 below.

![Long trip - Operator responses by time of day](image)

**Figure 18.** FMD scores by time of day for long trip schedule drivers
While both groups of drivers display the time of day effects discussed previously, there is some evidence that the short schedule drivers may be even more susceptible to these effects, producing slower average reaction times and more critically slow responses during daylight hours. However, there is no significant difference between the two groups in terms of average daily FC 3+4 responses (p > .1).
Figure 21. FMD scores by time on task for short trip schedule drivers

Both groups of drivers tend to display the U shaped relationship between test performance and time on task found generally in this study. Although there may be subtle differences between the two groups, again there is difficulty in making any firm conclusions as the number of driving hours contributing to these plots drops off markedly towards the last 24 hour period.

3.2.5 Personal Variables

The ARRB Fatigue Risk Questionnaire measures a number of areas thought to be relevant to predicting fatigue risk and indeed, as noted above, FMD scores are correlated with fatigue risk scores. There is some research evidence to suggest that personal variables may be some of the most important variables in crash risk. For example (Hanowski et.al.,2003) found that a small percentage of their sample (less than 25%) were responsible for more than 85% of critical incidents (where critical incidents were defined as “an unexpected event resulting in a close call or requiring fast action on the part of a driver to avoid a crash” p.157). Drivers in the current study varied greatly on dimensions such as weight and alcohol consumption. Therefore drivers were divided into two groups on the basis of their FRQ scores, forming a high FRQ group (top third of scorers) and a low FRQ group (bottom third of scorers). Drivers in the middle group were excluded because there appeared to be no natural break point in the distribution of scores around the median (see Figure 22).
Figure 22. Distribution of FRQ scores

These two groups were then compared to determine whether there was any difference in their driving performance as measured by FMD scores. The results are shown in Figures 23, 24, 25 and 26 below.

Figure 23. FMD scores by time of day for low FRQ scorers
As can be seen from Figures 23 and 24 there may be some suggestion that drivers with high FRQ scores are slightly more susceptible to the daytime test performance decrement than are drivers with low FRQ scores. Importantly, high FRQ scorers were also slower in their test responses at all times of the day than were low FRQ scorers, with some evidence that this poor responding also carried over into the night to a much greater extent. It should be noted however that there is no significant difference between the two groups in terms of average daily FC 3+4 responses (p > .1).
3.2.6 Truck Variables

Another factor that may contribute to alertness and fatigue in any driving situation are characteristics of the vehicle. A number of drivers spontaneously mentioned that the ‘cab-over’ vehicles they drove were much less desirable in their eyes than the ‘bonneted’ vehicles. It was thought important to make a comparison between the two kinds of trucks to determine whether the cab-over vehicles used in the study had a greater effect on driver fatigue than the bonneted trucks used in the study. To this end the time of day effects and trip effects described above were investigated for the two groups of trucks separately. In order to hold as many factors constant as possible in making this comparison, trucks of the same make, from within the same transport company, were compared. The FMD scores of drivers of cab-over and bonneted trucks are compared in Figures 26, 27, 28 and 29.
As can be seen from the graphs above, the daytime decrement in FMD performance previously identified is much more marked for those drivers in cab-over vehicles than in bonneted trucks. This is consistent with the drivers’ assertion that the bonneted trucks they drove are more comfortable in a number of respects (especially in their sleeping arrangements) and thus may have the potential to ameliorate vehicle-related performance impairment, such as the daytime test decrement. It should be noted however that there is no significant difference between the two groups in terms of average daily FC 3+4 responses ($p > .1$).
When the two kinds of vehicles are compared on a time on task basis a number of things become apparent. Firstly, that the bonneted trucks tend to be used for the longer journeys; the maximum return journey is six days rather than five days, and secondly, that there is evidence that this sixth day of driving results in a sharp increase in slow responses, on that day. This result should be interpreted with caution however, as there are relatively few data points contributing to this effect.

### 3.3 PVT Cross-Validation

The PVT data was grouped by site into:

- Perth leaving
- Port Augusta east bound
- Sydney arriving and leaving
- Brisbane arriving and leaving
- Port Augusta west bound
- Perth arriving

(note that trucks passed through either Brisbane or Sydney, not both)

The average of all the means for the site were calculated. These are plotted in Figure 31. Eighteen drivers in total were tested with the PVT, all on the east-west route.
In order to cross validate the FMD against PVT scores it was hoped to be able to correlate PVT scores at each location with FMD scores from the immediately prior period. Unfortunately there were too many cases where there was no appropriate overlap between the two kinds of scores for this strategy to be possible. Drivers were not always conscientious in contacting research assistants to let them know their estimated time of arrival, resulting in missed PVT testing sessions. Nevertheless there was a modest correlation between FMD and PVT reaction time means overall ($r=.24$, $p=.1$). The magnitude of this correlation is appropriate given that it incorporates scores averaged over all the other factors that have been shown to affect performance (time of day, time on task etc.).

The variation in the averages of the mean response times between sites is quite small, suggesting that after emerging from the truck, there may be very little difference in the performance ability of the drivers before, during or after a round trip of about five days. The jump in reaction time at the Sydney location appears somewhat anomalous when compared to the FMD data which describes a fairly smooth U shaped curve across this trip. Because of this discrepancy each driver’s data was carefully inspected. It was found that one driver had an unusual number of very long responses in Sydney. This pattern was not evident for any other driver, or indeed for this driver at any other location. It seems likely then that the Sydney data for this driver is not an accurate reflection of their potential. For this reason all the data for this driver was removed and the results replotted in Figure 32 below.
Figure 32. PVT means at each site (driver 15 data removed)

This curve is now quite similar to that obtained from the FMD reaction time data. However, PVT studies of fatigue often employ ‘lapses’ (reaction times slower than 500 ms) as a dependent measure (eg. Balkin et al., 2000; Hartley, personal communication). This kind of measure is somewhat similar to the Fatigue Code measure employed with the FMD. Therefore PVT lapses were also calculated and plotted (Figure 33 below).

Figure 33. Mean number of lapses at each site (driver 15 data removed)

Again these results produce a U shaped curve very similar to that obtained with the FC 3+4 measure in the FMD data.

Thus, these PVT results are consistent with the finding from the FMD data that drivers tend to begin their journey with somewhat slower responses than at a point nearer the middle of the journey and then become somewhat worse again towards the end of the journey. Given that the PVT is generally accepted to be a valid indicator of alertness and fatigue, then the extent to which a change in PVT scores is mirrored by a change in FMD scores is evidence of the validity of the FMD.
The FMD data, in addition to producing the U shaped curve in performance over the course of a trip discussed above, also produced an unexpected time of day effect. For this reason the PVT data was aggregated to investigate time of day effects. This data is plotted in Figure 34 below.

![Figure 34. PVT means by time of day (driver 15 data removed)](image)

While the pattern of these results does not mirror the FMD data in a direct way, it does show the increasing reaction time during daylight hours that characterises the FMD results. In addition these results display a steadily decreasing reaction time function between midnight and 6am, the reverse of the typically claimed pattern for fatigue effects (Wylie et al., 1996). These results reinforce the notion that two-up drivers’ task performance really does decline during daylight hours and that this result with the FMD data is not simply an artefact of the manner in which the FMD stimuli were presented.

It is worth noting that while the PVT result is theoretically important, and useful for purposes of validation, it is difficult to interpret the practical significance of such a result. The difference in reaction time between Perth and Port Augusta is of the order of 20 msec. Such a decrement is unlikely to be important in an actual driving context. However, it must be remembered that PVT reaction time was not measured while driving and therefore is measuring something rather less direct than the ability to react in a driving context. Such interpretive difficulties were another reason for dismissing the PVT as a practical tool in the current study (although the status of PVT reaction time as a proxy for alertness and fatigue is not in question).

### 3.4 Drug Testing

In order to determine the extent to which the use of stimulant drugs, which could temporarily alleviate the effects of fatigue, could be a factor in the performance of participants in this study, hair samples were collected from drivers and analysed by the Victorian Institute of Forensic Medicine. Hair testing was selected for this study because it can be readily and reliably analysed to cover the entire two month study period. It
provides an answer to the question of whether any stimulants which could artificially improve performance have been used, and hence the extent to which the results obtained could be confidently held to reflect the normal human capabilities of drivers who choose to drive in a two-up team. As described below, the finding of a positive result does not indicate that the driver always, or ever, drove while affected by the detected drug, or that the drugs were taken to enhance driving performance.

Drugs, medications and other substances that happen to be in the bloodstream are incorporated into hair as it grows, providing an historical record of that person’s drug consumption. Hair grows at approximately 1cm per month, enabling the analyst to estimate when a drug has been used if the distance from the scalp is measured. The length of hair in the sample can be used to estimate the length of time covered by the analysis. For example, the analysis of a 2cm length of hair taken from the point nearest the scalp will reveal any drug or medication consumed at anytime during the proceeding two months. In contrast, urine testing for stimulants will only reliably cover the preceding 24-48 hours drug use, and saliva 2-6 hours.

The concentration of drug found in a hair sample cannot be interpreted in terms of the quantity or frequency of use, as there is too much individual variation in the amount of drug absorbed in hair. It is also known that the regular washing or perming of hair can reduce the amount of drug present over a long period of time so that hair testing tends to underestimate the amount of drug present in samples covering longer periods of time.

Consequently, while a positive result indicates that a person has used a drug in the time period examined, it does not prove that the drug was used while driving, nor can it be established whether the drug detected is the result of regular use or of a single incident.

Hair samples were successfully collected from 23 drivers, however, due to the length of hair samples, and the gap between the study period and the hair collection, three of these samples did not cover the period in which FMD data was being collected. These three samples have not been included here. Hair samples from 20 of the participants (51% of total study participants) were analysed for the following stimulant drugs:

**Illicit stimulant drugs**

Methamphetamine (also known as speed or crystal meth);

MDMA (Ecstasy); and

PMA (an illicit ‘party drug’ related to ecstasy and methamphetamine).

**Medications**

Phentermine (an appetite suppressing stimulant available only on prescription, but which has been reported as being abused to combat fatigue);

Pseudoephedrine (a mild stimulant which can be used in very high doses to combat fatigue but which is also available in over-the-counter medications such as Sudafed, and various other decongestant and cold medicines);

Ephedrine (a stimulant and decongestant found in some prescription medications and in over the counter herbal preparations containing ephedra or Ma Huang); and

d-Amphetamine (a medication used for attention deficit disorder and narcolepsy, but which is also produced by the body after consumption of methamphetamine. d-Amphetamine was only detected in samples which also contained methamphetamine,
and is likely to have been derived from the natural metabolism of methamphetamine by the body.)

Eight drivers returned a positive result for pseudoephedrine, although in one of these drivers this was the only stimulant found. Of the other seven, two had methamphetamine (only), one had ephedrine, one had phentermine, two had both ephedrine and phentermine and one had ephedrine, phentermine and methamphetamine present. A total of eight drivers provided samples which returned a positive result for methamphetamine. Two tested positive for MDMA (Ecstasy) and four for PMA.

In all, 75% of the samples tested (38% of all 39 drivers in the study) revealed stimulant use both legal and illicit at some time during the study period with approximately one third of samples showing only legal medications. Although this does not prove that these drivers gained chemical assistance to perform to the level shown in this study, it does prevent the results being useful for policy making on safe driving hours since it cannot be said with certainty whether or not some of the performances recorded were only achievable through drug use. It should be pointed out, of course, that drivers who were tested and showed no evidence of drug use (25% of samples tested) undertook these trips safely and successfully.

Figure 35 below depicts the frequency with which particular types of stimulant drugs were detected in the hair sample of drivers tested.

![Bar chart showing the number of hair samples containing a specific stimulant drug](chart.png)

**Figure 35.** Number of hair samples containing a stimulant (note that some samples contained more than one drug)

This Figure simply indicates the frequency with which a positive result was obtained. Although the hair analysis results yield a concentration level, there is no straightforward way to relate this level in the hair sample to behavioural effects or to identify if these drugs were taken during recreational breaks. As a consequence no conclusions can be drawn as to whether these results would have affected any or all of the FMD data for these drivers collected in the study. Nonetheless, these findings raise important issues for further investigation.
The results of testing samples from so small a number of drivers cannot be said in any way to be representative of the industry (or two-up operations as a whole). However they do highlight the need for jurisdictions and industry to be aware of the fact that these substances are being used by people working in the industry. Measures to address the use of stimulants (and other drugs) by drivers are in place or under development in most Australian jurisdictions, and the industry has been active through programs to address worker health and well-being. Further investigation of drug incidence in a wider sample of drivers may provide valuable information to ensure the most effective targeting of these, and other future, measures.

3.5 Sleep Patterns of Drivers

Actigraphs, a watchlike device, are usually worn on either the wrist or the leg and track movement of the limb over time. A choice of sampling intervals between 0.25 secs and 4 mins is available. For this study, the instruments were set to record every 0.5 secs, which provides frequent enough sampling to allow assessment of sleep quality, but also allowed around two weeks worth of data to be collected before memory capacity of the devices was exceeded. However this sampling interval provides too many data points for plotting for visual analysis, so some aggregation of data points was required. Over 99% of data points were in the range of 0 to 400 for the east-west drivers (900 for the north-south drivers) with the remainder being highly variable and greater than 400 indicating very jerky movement. These points were first reset to 400 (900) and then groups of 8 points (corresponding to 4 min intervals) were averaged to provide a new data set. Actigraph data for each truck crew was then plotted with the associated FMD results.

Each graph is for the pair of drivers that made up the crew (in one case a company rotated three drivers per truck to give each driver a week rest. In this case, the three drivers are shown on the one plot). A typical plot is shown in Figure 36.

![Operator response times - truck 4](https://example.com/plot.png)

**Figure 36. Actigraph results and FMD results for operators 7 and 8**

As can be seen from Figure 36 the activity level recorded by the Actigraph is higher when the operator is driving and lower when they are not driving. Given that the movement of the truck will have some effect on the actigraph, it is not straight forward to determine if
the non driver is sleeping or awake. However, it is reasonable to assume that where the actigraph trace approaches the X axis that there is a high likelihood that the wearer is lying in the bunk, at least resting. Conversely, where the actigraph trace does not approach the X axis it is very unlikely that the wearer is lying down and hence unlikely that they are attempting to sleep. For example, the last non driving period for operator eight in the above figure indicates that this person stayed awake during this period, perhaps because they knew that the trip was nearly over.

Usable sets of actigraph data were obtained from eleven drivers who completed between one and four trips while wearing the devices. After identifying the non driving periods for each operator, the proportion of each non driving period in which the actigraph output fell below the criterion for the definition of ‘rest’ was calculated. This ‘rest estimate’ is plotted against the FC 3+4 measure for the trips in which the drivers wore the actigraphs (see Figure 37).

Figure 37. FC 3+4 responses as a function of rest time through non driving periods

Averaging across multiple journeys for each driver yields the plot in Figure 38 below.
As can be seen from this Figure, there is a moderate (but non significant) negative correlation between the amount of rest that drivers enjoy and the number of slow responses that they produce ($r=-.46$), indicating that the less rest that drivers obtain in their non driving periods, the more slow responses they make on the FMD task.

3.6 Fatigue Risk Questionnaire

The mean of the FRQ scores was 74.71 (SD = 21.27) and the scores ranged from a minimum of 44 up to a maximum of 143.

As in previous studies the FRQ scores was positively correlated with FMD reaction time ($r = .37, p < .05$), indicating that drivers who had the most personal fatigue risk factors (e.g. were over weight, had high level of alcohol consumption) were slowest at responding on the FMD task.

4. Fatigue Issues and Poor Performance

The current study was designed to provide objective data on the alertness and fatigue risk associated with long distance heavy vehicle two-up driving in Australia. To this end, a relatively representative sample of two-up driving teams from around the country was studied. Drivers’ alertness while driving was monitored almost continuously over a four week period using the ARRB Fatigue Monitoring Device. The study included teams driving the east-west route between Perth and Sydney, or between Perth and Brisbane; teams driving north-south routes in the east and teams driving north-south routes in the west. However, it should be noted that participation in the study was voluntary for both drivers and transport operators. Indeed, a large number of transport operators and drivers who were approached to participate in the study declined to do so for various reasons. As
a consequence some of the results from the study may not generalise to all drivers and transport operators in Australia.

A growing body of research suggests that time of day is a very important predictor of driving performance, and in fact may be a better predictor in some cases than time on task (eg. see Wylie, Shultz, Miller, Mitler and Mackie, 1996), with midnight to dawn often identified as the period of greatest risk ("Staying alert at the Wheel", Department of Planning and Infrastructure). Interestingly the current results fail to exhibit this pattern, instead showing a general increase in reaction time during daylight hours suggestive of decreased attention to the Fatigue Monitoring Device stimulus. However it is wise to be cautious in interpreting this result because of the possibility of a confound with ambient light resulting in decreased perceptibility of the stimulus during daylight hours. Nevertheless, two results seem to argue against the ‘confound’ explanation. Firstly, the daylight effect varies as a function of Fatigue Risk scores, with drivers who obtain high scores displaying the effect to a greater extent than drivers who obtain low scores. This is important in evaluating the probability of the confound as FRQ scores are unrelated to the in-truck environment in any way. Secondly, like FMD scores, PVT scores also display an elevation during daylight hours. Again, this tends to argue against the ‘confound’ explanation as PVT testing was conducted in a much more consistent lighting environment (indoors with artificial illumination) and without competition from a secondary task (driving).

If one accepts that the daylight effect is real then the question arises as to the cause. Firstly it is likely that traffic is heavier during daylight hours, perhaps resulting in drivers allocating more attention to negotiating traffic and hence having less ‘spare’ attention for responding to the stimulus. If correct, this in itself is a factor that could increase the risk of crashes during daylight hours. It is also possible that drivers are actually less alert during daylight driving than they are after dark. A number of drivers in the study spontaneously mentioned the problems with drowsiness caused by the sun warming the body directly during the day. This may be a more significant problem in a country like Australia with a very warm climate. Whatever the cause of this decreased performance during the day it seems likely that that drivers in this study were less vigilant during daylight hours and thus it is possible that they may have been more fatigued. The contrast between this finding and the typical finding of decreased night time performance may be explicable in terms of the unique character of two-up driving. For example, it may be that two-up drivers drive for shorter stints and sleep more than do solo long haul drivers. It is also possible that experienced two-up drivers are less susceptible to time of day effects because they have the opportunity, and have learned, to modify their sleep-wake cycle successfully.

While previous studies may suggest that time on task is not necessarily a good predictor of driver performance, the current study shows a very distinct relationship between a drivers progression through a trip and FMD performance. Specifically, they appear to exhibit a curvilinear relationship. That is, drivers’ performance gradually improves over the course of a trip and then declines again as the end of the trip approaches. A methodology that simply took a snapshot of performance at the start and finish of a trip may fail to detect any change, as drivers are not much worse at the end of the trip than they are at the start of a trip. However, the FMD methodology clearly shows that drivers are not at their best at either the beginning, or at the end of a trip. This suggests that there may well be a decrease in alertness towards the end of the trip, but that there is also relatively impaired alertness at the beginning of a trip. It is likely that this effect is the result of what drivers do on their days off or if engaged in loading/unloading or other heavy physical activities immediately prior to a trip. It should be noted however that the change in performance that is observed,
while consistent, is of a magnitude that may render it of relatively minor practical significance; the average change in the number of slow responses by drivers over the course of a trip being in the order of around one every four hours.

The most common driving/rest pattern in the study was six hours of driving followed by a six hour break, although there were some drivers who varied from this pattern markedly. There was some evidence from this study that drivers who drove longer stints exhibited a greater fatigue risk than those who drove shorter stints. In particular, drivers who drove longer stints (nine hours or more) seemed to be more susceptible to the daytime reduced performance observed for two-up drivers generally. In addition, the magnitude of the effect is worth noting. Long stint drivers were approximately twice as likely to exhibit slow responses during the day as short stint drivers, with up to one critical response per hour compared to one every two hours for short stint drivers. It may be that drivers who drove longer stints were attempting to conform to a ‘normal’ work/sleep pattern and as such were not benefiting from the advantages that may accrue from the shorter rest/work cycle that is possible in two-up driving. Furthermore, a moderate negative correlation was detected between FMD performance and amount of rest, suggesting that drivers who took advantage of the opportunity to rest in their break were less impaired in their performance across the journey.

While some previous research has suggested that individual driver characteristics may play an important role in driver performance (e.g. Hanowski et al., 2003), the current study failed to detect such factors to any great degree. However, there was some suggestion that drivers who possess a number of characteristics deemed to contribute to a fatigue risk did exhibit a relatively greater daytime slowing in response times than drivers possessing fewer of these risk factors. In addition reaction speed to the FMD task overall was positively correlated with fatigue risk scores. This suggests that, to the extent that drivers possess fatigue risk characteristics, their alertness may be compromised and their fatigue risk elevated.

It is apparent from the current study that two-up driving does not necessarily employ the best (e.g. most comfortable) vehicles for the job. A number of drivers complained about the discomfort associated with long haul driving in the cab-over vehicles available to them, claiming that they were rougher and more difficult to sleep in. The objective performance data is consistent with this claim, with drivers of the cab-over vehicles appearing to be more susceptible to the general daytime performance decrement. Interestingly this effect holds even though the drivers of cab-over vehicles tend to drive trips that are around a day shorter than drivers of bonneted trucks. The cab-over vehicles used in this study, in addition to providing a rougher ride, have less space in the bunk and less room in the cab generally than the bonneted vehicles used. Both of these factors are likely to compromise sleep and rest quality and hence result in reduced driver performance.

5. SAFE DRIVING PATTERNS

A number of conclusions can be drawn from the current study that may be useful and informative for evaluating and designing the driving patterns of two-up heavy vehicles to aid in the reduction of fatigue related crashes.

- The idea that night driving is more risky than day driving is not supported by this study. If anything drivers appear to perform more poorly during the day than after dark although the results are not statistically significant. A number of drivers in this
study volunteered the information that they preferred driving at night and the data
tend to support the wisdom of this preference. As a consequence any regulatory
measures that have the effect of restricting the amount of night driving by two-up
drivers may be counterproductive. That would be likely to result in an increase the
amount of day time driving and hence increase the amount of driving where
performance is less than optimum and through a higher density of traffic.

- There is some evidence that drivers who drive for long stints before handing over to
  their partner show greater decrements in performance than those who drive shorter
  stints before handing over. This may suggest that drivers should aim for a six hour
  work/rest cycle.

- Although there is evidence that drivers’ performance becomes poorer over the
duration of a trip, it also tends to start off relatively poorly, being best in the middle
of a trip. This finding, previously reported by Feyer et. al. (1997) has a number of
implications. Most obviously it suggests that what drivers do on their days off
(whether the break is 48 hours or 27 hours) may be just as important as on the job
factors in affecting performance. Any measures that improve the performance of
drivers at the start of a journey are likely to carry over for the rest of the trip and
result in a benefit to performance across the whole trip. Thus some kind of fatigue
education and training may be one way of improving the performance ‘start point’.

- It appears to be the case that drivers who take the opportunity to rest in between
driving stints are able to perform better. This may suggest that drivers should be
discouraged from spending too much time playing games and watching movies etc.
during breaks, rather being encouraged to rest and perhaps taught techniques to
facilitate rest, relaxation and sleep in the truck environment.

- There is some suggestion from the data in this study that drivers on short trip
schedules may be slightly more susceptible to the daytime performance decrement
observed for drivers generally, although this is a rather small effect. It may be that
drivers on short trip schedules have more constraints on how they manage their
work/rest cycles and therefore are less able to optimise these cycles to reduce
performance decrements. Perhaps lengthening some short trip schedules slightly or
rethinking schedules to reduce the interruption of rest periods by loading stops and
the like may help in this regard.

- The oft-cited complaint by drivers that the cab-over trucks they drove were less
comfortable than the bonneted ones, and unsuitable for long haul work does seem
to be supported by this study. As such it probably makes sense to recommend that
only the most comfortable vehicles be employed for long haul two-up driving.

- As in previous studies, a number of personal factors such as weight, alcohol
consumption and diet are seen to predict overall driver performance. This suggests
that education and training focussing on the importance of health issues for long
haul truck drivers may be valuable.

- Finally, the fact that 38% of participants tested positive for some form of stimulant
effectively compromised at least one third of the data collected during the course of
the study and with no details of stimulant usage of a further 62% cast some doubt
on the remaining data collected and the interpretation of that data.
6. BIBLIOGRAPHY


