Executive Summary

This study was conducted with Union Pacific Railroad Train and Engine employees reporting for duty to the San Antonio Kirby Yard from November 3rd through November 8th 2004. During that time, questionnaire assessment of 283 Train and Engine employees (out of a possible 356 who reported for duty) occurred, yieldi

Table of Contents

List of Figures

Background

This project was requested and commissioned by the Federal Railroad Administration (FRA) and was conducted in San Antonio, Texas, to serve as a general assessment of fatigue in the workforce of train and engine employees of the Union Pacific Railroad (UP).

For the purposes of this report, the term 'workforce' pertains to train and engine employees located in the San Antonio area and not to employees associated with other crafts. As a point of reference, the "San Antonio area" refers to employees reporting for duty at the UP Kirby Yard and the South San Antonio Yard to work in the Laredo, Houston, Taylor-Hearne, and Del Rio Pools as well as the Northeast and Southeast Extraboards.

The question of the impact of operator fatigue on railroad safety has been a concern of the National Transportation Safety Board (NTSB) since 1989 (Sherry, 2003). The Association of American Railroads (AAR) began an in-depth study of fatigue issues in its industry in 1992 and the US Government Accounting Office (GAO) issued a report on the fatigue of railroad locomotive Engineers in 1992 that focused attention on the variability of work shift start times (GAO, 1992). The NTSB has urged the Federal Railroad Administration (FRA) to consider changes to the hours of service rules that affect railroad operating employees (Hall, 1998). The recent incident on June 28, 2004, involving the collision of UP freight train MHOTU-23 and BNSF Railway (BNSF) freight train MEAPTUL-126D has also raised questions about the fatigue of locomotive Engineers. According to the public hearing convened by the National Transportation Safety Board (NTSB, 2005) this collision resulted in the death of the UP Conductor, two nearby residents, and the treatment of more than 40 people at local hospitals for the inhalation of chlorine gas. Thirty-five freight cars (19 UP and 16 BNSF) and four UP locomotives derailed, resulting in the release of chlorine, a poisonous gas.

Since some concerns were raised by FRA about the possibility of fatigue in the workforce as a contributing factor to the Macdona accident, the present study was undertaken in an effort to understand the factors affecting the situation in San Antonio. The FRA asked the University of Denver team to conduct a survey of the UP workforce to gather additional information on employee's reports of fatigue.

Fatigue

The issue of fatigue is complicated and subject to considerable misunderstanding. Fatigue has been the subject of a number of scientific investigations and it should be noted that the term fatigue is one that most people can relate to. However, the definition of fatigue, from a scientific standpoint is somewhat less clear. Sherry (2003) noted that in an attempt to understand fatigue, investigators have used several different measures including physiological, behavioral, cognitive, and self-report of mood or subjective

experience. Michielson, De Vries, Van Heck, Van de Vijer, and Sijtsma (2004) suggested that "due to complex interactions between physical and mental elements in task and job demands and consequences of effort, it is difficult to separate" the mental and physical components of fatigue (p. 40). Generally, fatigue in the railroad industry has been taken to mean that an individual suffers a loss of alertness, a loss of mental or cognitive capacity, a reduction in alertness, and a propensity to report feeling sleepy prior to falling asleep.

conditions and an environment adverse to the utilization of such sensitive equipment. Researchers have had more success with the use of actigraphy as a behavioral measure of activity which can be used to infer sleep and wakefulness. These devices, most commonly known as actigraphs, are small wrist-watch size devices that monitor activity and store data for over 60 days. Data from these devices are then available for analysis by standard statistical programs. Actigraph data have been used to obtain reliable and valid measures of sleep and sleep quality. (Sadeh, Alster, Urbach, & Lavie, 1989; Sadeh et al., 1991). The use of actigraph data has been used to differentiate between normal and disturbed sleep-wake patterns of adults, young children, and infants and to assess changes in infant sleep following behavioral interventions. (e.g., Cole, Kripke, Gruen, Mullaney, & Gillin, 1992; Sadeh, Acebo, Seifer, Aytur, & Carskadon, 1995; Sadeh, Hauri, Kripke, & Lavie, 1995; Sadeh, Lavie, Scher, Tirosh, & Epstein, 1991; Sadeh, Sharkey, & Carskadon, 1994). Actigraphy measurements and sleep wake algorithms have also been

requirements. Participants were then asked if they agreed to participate, and if so, were given instructions on how to complete the questionnaires.

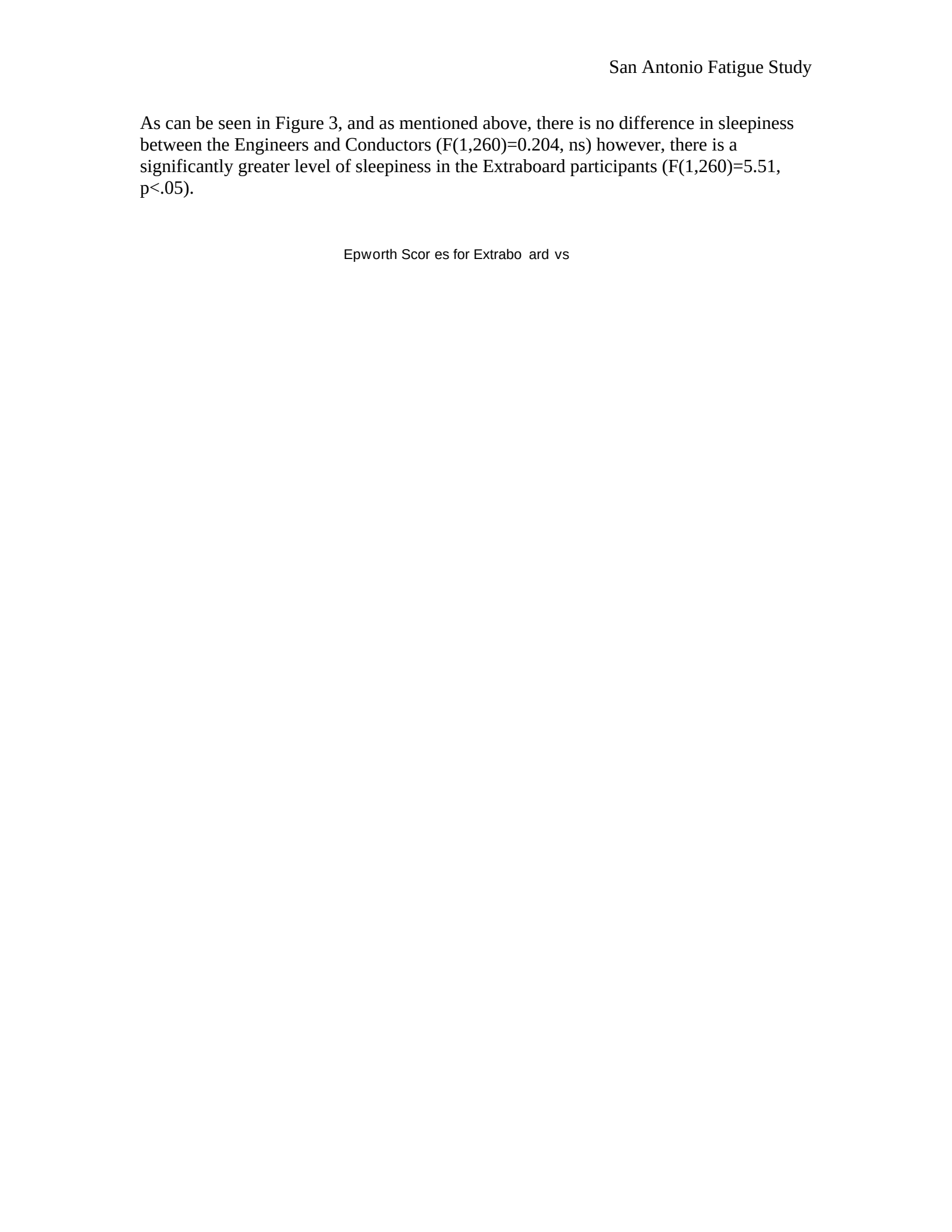
Study Participants

As previously indicated, the participants for the current study comprise the workforce that operates out of the UP Kirby Yard in Sa

a joint labor and management team based on the representativeness of the workload and the geographic distribution of the pools relative to the Kirby Yard. Actigraph participants were chosen to maximize variability in work schedules (one Pool turned on average every 24 hours while the Extraboard could turn every 8 to 10 hours). Participant selection criteria were based on being employed in a specific Pool or Extraboard, planning to work at least the next six weeks, willingness to wear the actigraph daily, and willingness to complete the research questionnaires.

Table 2. Participants in Pools and Extraboards

As can be seen in Figure 3, and as mentioned above, there is no difference in sleepiness between the Engineers and Conductors ($F(1,260)=0.204$, ns) however, there is a significantly greater level of sleepiness in the Extraboard participants ($F(1,260)=5.51$, $p<.05$).

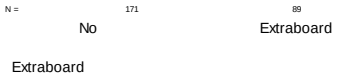


sleepiness compared to a normal population. On a cautionary note, while statistically significant, the practical significance of these findings in this population are not fully known due to the lack of utilization of this instrument with the railroad population.

Non-Extraboard

disturbance and happiness in the general population. A GHQ12 score of 4 or more indicates a high level of psychological distress. Hardy, Shapir, Haynes, & Rick (1999) used the Likert scoring method to validate the GHQ-12 on a sample of 551 National Health Services Staff workers and found that the mean GHQ-12 score was 1.27

Non-Extraboard (Pool) vs. Extraboard	Engineers vs. Conductors



which also indicate there were significant differences between Engineers and Conductors on this measure.

Self Report Hours of Sleep. Respondents were asked to report the number of hours of sleep they had obtained in the last 24. There was a significant difference in the amount of sleep obtained between Extraboard (6.8 hours) and Pool (7.5 hours) ($t=2.01$, $df= 1, 255$, $p<.019$). The distribution was not normally distributed, therefore the Mann-Whitney

Table 4. Group comparisons on se

Craft or Position	N

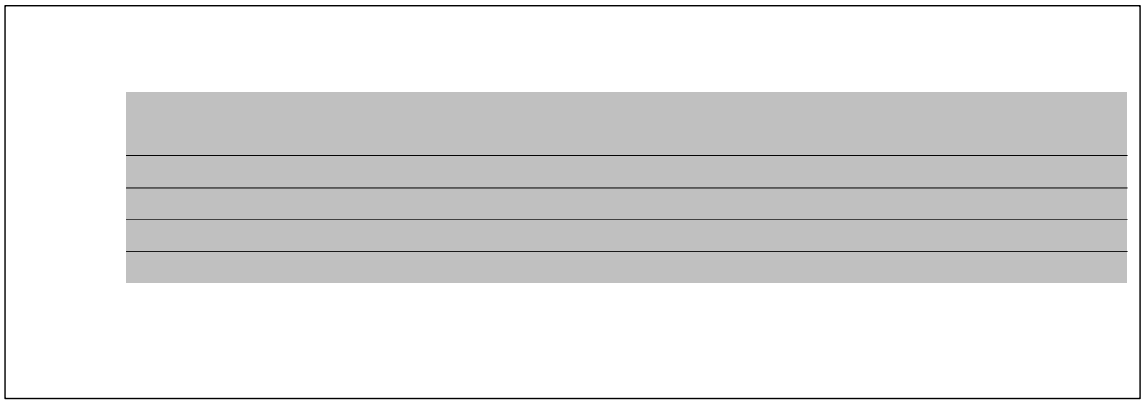
Table 5 presents a similar comparison of Extraboard vs. Pool employees. Significant results are highlighted in yellow.

Table 5. Independent t-tests of selected variables by Assignment (Extraboard vs. Pool).

	Extraboard	N	Mean	Std. Deviation	*t* Test p<	Mann Whitney p<	KS p<
Epworth	No	171	9.4737	4.86683	.020	.03	.29
	Yes	91	10.9890	5.16719			
Pittsburgh Sleep Quality	No	172	15.3663	8.09950	.10	.02	.05

Results of Actigraphy Studies

The use of Actigraphy as a means of determining a person's level of activity as well as sleep is common practice. The present study sought to determine the amount of sleep obtained by a sample of 40 railroad Engineers and Conductors who were asked to wear actigraphs for a total of 30 days. Actigraph results for the entire sample are displayed in Figure 9. Useable data were obtained for only 33 study participants due to missing data, equipment malfunction, and individual's decision to withdraw from the study.



wore the first watch for a 14 day period and then a second for a 16 day period. So, in some cases, the actual data profiled may not be a full thirty days. Note that the standard deviation is 1.87 or a little over one and three quarters hours. Thus, the person is occasionally going with as little as 3 hours of sleep or as much as 6.5 hours of sleep. Overall, however, this individual slept less than 6 hours per night 84.6% of the time and 30% of the time he obtained 4 hours of sleep per night or less.

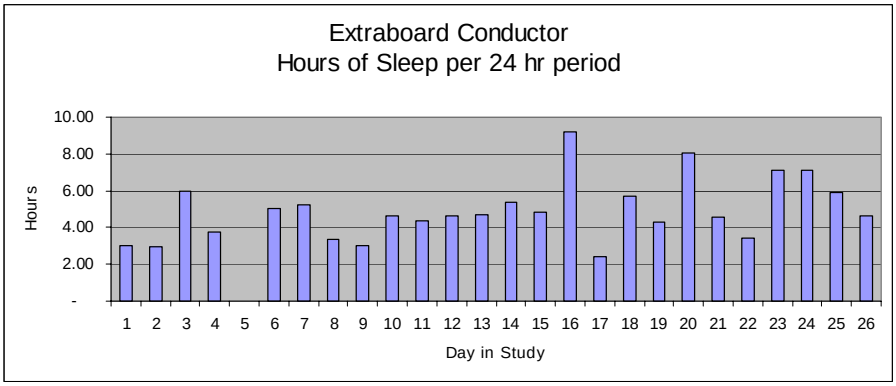
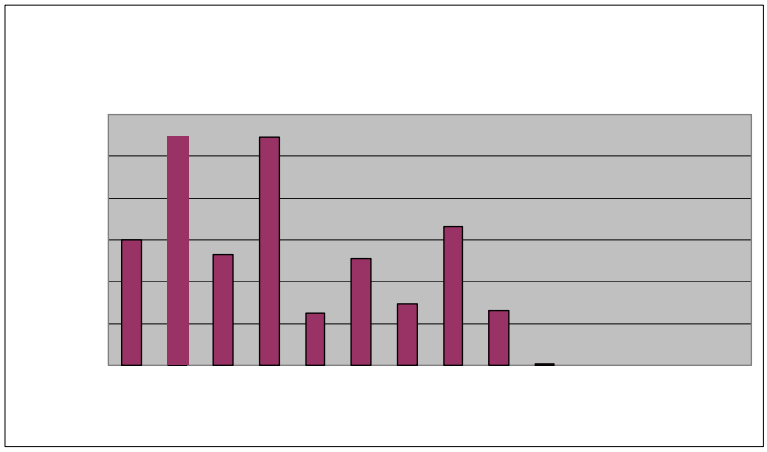
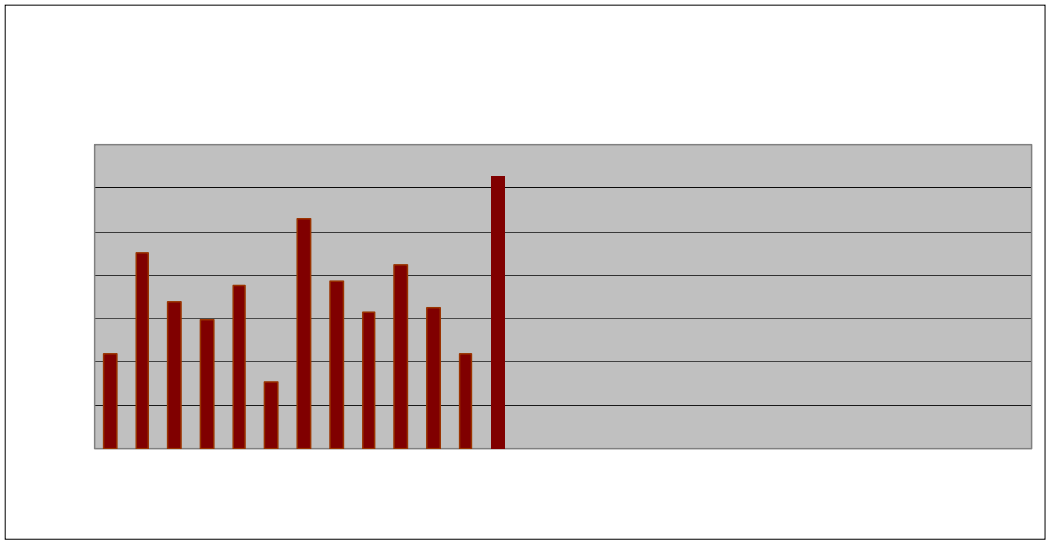


Figure 11. Extraboard Conductor #1 Hours of Sleep per day.

Another participant, also an Extraboard Conductor, pictured below, averaged 4.76 hours of sleep with a standard deviation of 2.62. Both of these individuals from the Extraboards would be likely to have a noticeable sleep debt. The participant in Figure 12 paid back a sleep debt on the fourth day of the study, but no evidence of pay back or recuperation is present during the remaining 10 days of the study period. This individual slept less than 6 hours per night 71% of the time and 50% of the time he obtained 4 hours of sleep per night or less.



average number of hours of sleep, this person has a variability of 2.35 hours. Nevertheless, this person appears to have been able to have repaid his/her sleep debt on several occasions. This individual slept less than 6 hours per night 35% of the time and 20% of the time he obtained less than 4.5 hours of sleep per night.



sleep by the same number of consecutive days plus a recovery day. The following list shows the results of this analysis:

Consecutive days with 6 hrs sleep or less	Number of times this set occurs	Percentage
6 days <6hrs	22	18%
5 days <6hrs	9	6%
4 days <6hrs	11	6%
3 days <6hrs	21	10%
2 days <6hrs	37	13%
# of single days < 6hrs	582	
Total Days	875	

The number of times that a set of 6 days, with 6 hours of sleep or less occurred was computed in this analysis. Next, the number of times that a series of six days, plus one day for recovery could have occurred in the 875 days available to the study participants was determined (875 7). These two numbers became the numerator and denominator respectively, yielding the ratio: 22/ (875/7) = .18 or 18% with 6 days or more of 6 hours of sleep or less. The denominator is calculated by taking the total number of days (875) and dividing by 7, which is the number of cons

Houston

The estimated departure time was as close as 7 minutes in one instance, and as far away as 21 hours and 10 minutes in another.
The average estimated departure times ranged from 0:27:30 to 20:44:30.
The estimated departure times were both earlier and later than the times the

Taylor-Hearne

The estimated departure time was as exact in one instance, and as far away as 15

Laredo

The estimated departure time was as exact in one instance, and as far away as 20 hours and 30 minutes in another.

The average estimated departure times ranged from 0:12:00 to 15:45:00.

The estimated departure times were both earlier and later than the times the Engineers actually left.

The estimated departure time was *not* consistently more accurate the later it was checked, indicating that there was little predictability based on the time the line-up was checked.

The total average difference between actual and estimated departure times was 2:32:36.

The frequency of the average time differences appears below:

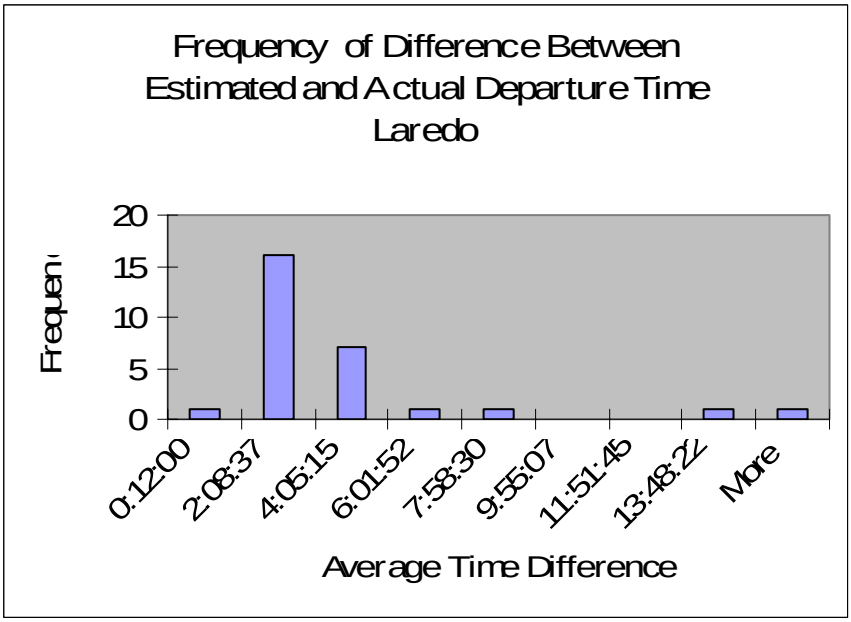


Figure 18. Laredo Line-up estimates.

Summary of Line-Up Analyses

These analyses indicate that there is substantial variability in the accuracy of the estimated departure times for the four Pools studied. The average difference for the actual and estimated departure times for the Laredo Pool was 2h:32m:36s, for the Taylor-Hearne Pool 3h:15m:34s, for the Del Rio Pool 4h:28m:13s, and for the Houston Pool 4h:36m:18s. Interpreting these differences is speculative at this point. Little comparative data exists for additional analyses. Logic suggests however, that the greater the magnitude of the difference the poorer the prediction. The data from the Del Rio Pool indicates that differences of 7 hours and 54 minutes or more occurred 39% of the time. Data from the Houston Pool indicates that differences of 6 hours and 33 minutes or more

occurred 47% of the time. Data from the Laredo Pool suggests that inaccuracies as large as 4 hours or more occurred 39% of the time. Finally, data from the Taylor-Hearne Pool indicate that inaccuracies of 4:38 hours or more occurred 29% of the time. Clearly, an inaccuracy of more than 5 or 6 hours would make it difficult to plan a days worth of activity. Additional problems occur if these inaccuracies involve shortening the anticipated amount of sleep that could be obtained.

Trip Start Times

Shift start times were looked at for 6 different groups, including both Extraboards and Pools, to see if there were disproportionately more shift starts between the hours of midnight and 5AM. Shift information was collected from 12/17/04 through 2/16/05, for a total of 62 days. If the shift starts were evenly distributed throughout the 24 hour period, 20.83% of the start times would occur between midnight and 5AM.

The lowest percentage of start times between midnight and 5AM occurred on the XE40 Extra Board. There were 1800 total shifts from that board during the 62 days, with 297 of them starting between midnight and 5AM, or 16.50%. This is 4.33% lower than would be expected if the start times were evenly distributed.

The XE30 Extra Board had the next lowest percentage of start times between midnight and 5AM. On this board, there were 1740 total shifts during the 62 days, with 310 of them starting between midnight and 5AM, or 17.82%. This is 3.01% lower than expected if the start times were evenly distributed. .

Of the regular Pools, the Taylor-Hearne Pool, RE46, had the lowest percentage of start times between midnight and 5AM. During the 62 days, there were 1739 total shifts, with 312 of them starting between midnight and 5AM, or 17.94%. This is 2.89% lower than expected if the start times were evenly distributed. .

The Del Rio Pool, RE33, was the only other Pool with a lower percentage of start times between midnight and 5AM than expected. There were 1396 total shifts from that Pool during the 62 days, with 289 of them starting between midnight and 5AM, or 20.70%. This is 0.13% lower than expected if the start times were evenly distributed. .

The Laredo Pool, RE35, had the highest occurrence of shifts starting between midnight and 5AM. During the 62 days, there were 1068 total shifts, with 260 of them starting between midnight and 5AM, or 24.34%. This is 3.51% higher than expected if the start times were evenly distributed. .

Overall, there were 7743 total shifts among all the Pools and Extraboards, with 1468 of them starting between midnight and 5AM, or 18.95%. This is 1.88% lower than expected if the start times were evenly distributed. Therefore, there are not a disproportionately large percentage of shifts starting be

the Laredo Pool had the highest chance of a start time between midnight and 5AM, and the two Extraboards had the least chance of starting during those hours.

Focus Groups

A total of six focus groups were held with employees reporting for work during the week that the investigators were on site. The individuals that were selected to participate in the focus groups were chosen on the basis of convenience so as not to disrupt railroad operations. Three weeks later, individual meetings were held with 10 railroad supervisors at various locations in the San Antonio area and similar questions were asked.

The individuals participating in the focus groups were not identified and no record was kept of their background or years of experience for confidentiality purposes. However, in order to put their comments in context, they were asked to identify their craft. They were also asked to complete the research questionnaire and sign the consent form. Persons who signed the consent form were invited to participate in the focus group.

The format of the focus groups followed the same procedure. Participants were asked five questions. Interviewers took note and listened to their answers. The five questions were:

1. What is your craft?
2. Describe your sleep patterns over the past few weeks.
3. What do you think is the main problem contributing to fatigue /scheduling issues?
4. What needs to be done to change the situation?
5. What are some other factors that might contribute to this problem?

The comments obtained are grouped into

Individuals on some of the boards reported *never* having their boards "rolled" and of having sufficient time to rest and recover. In addition, several individuals indicated that they were not concerned about fatigue as a problem. They reported that if you "focused on work alone" and didn't try to do a lot of other things (e.g., social and family life) that there was sufficient time to obtain rest. This was repeated by several different individuals and indicates that some employees are not concerned about fatigue.

Thus, comments about fatigue were both pro and con. Despite the remarks of some individuals noted above, others indicated that they were comfortable with the situation and not concerned with fatigue. Thus, the impression that the interviewers formed was that the perception of fatigue problems were not necessarily widespread and may reflect individual preferences and differences..

Causes of Fatigue and Scheduling Problems

Many explanations were offered in an effort to explain the current situation. Several people commented on the need for more employees. They acknowledged that the UP had hired a number of people in the last few months but that it still took time to get them trained and ready to work independently. The presence of new hires in the workforce was also described as a source of stress due to the need to supervise the new workers to avoid being injured as a result of mistakes they might make.

Another source of fatigue was thought to be the line-ups. Several individuals commented on the fact that the line-ups were inaccurate and that the inaccuracies prevented them from being able to properly plan their rest periods. The comments were such that the line-ups were not updated in a timely fashion and that they contained trains that did not exist. These were common complaints.

Another theme that emerged from the comments was the notion that management viewed the employees as "robots" who were expected to work long periods of time without time off for families and social matters. One individual indicated that he had worked for 19 days straight and was having trouble getting time off.

Several individuals commented on the fact that fatigue and safety issues were not concerns until the Macdona accident occurred. According to comments that were made the young Conductor that was killed was well-liked and respected and his death was considered a tragedy.

Suggested Remedies

significant and slightly higher elevation on a measure of work related stress as com

Conductor's Extraboard obtained only 4 hours of sleep a day approximately 30% of the time, another participant from the Conductor's Extraboard slept less than 4 hours per night 50% of the time and a Pool Engineer slept less than 6 hours per night 35% of the time and less than 4.5 hours a night 20% of the time. An analysis of the complete actigraph data indicated that 67% of the time study participants obtained less than 6 hours of sleep per 24 hour period. Furthermore, 18% of the participants went 6 days or more in a row with less than 6 hours of sleep per day.

Data provided by FRA suggest that the Engineer involved in the Macdona accident had worked extensively in the days before the accident and may have developed a sleep debt. Recall that the UP crew had gone on duty at San Antonio at 2:45 a.m. and had been on duty about 2 hours and 18 minutes at the time of

quality has been assoc

include the instruction to nap as part of their safety job briefings in order to appropriately guide and direct employees to maintain high levels of alertness. Research has documented the benefits of napping as a means of increasing alertness and cognitive performance following sleep deprivation (Dinges, Whitehouse, Orne, & Orne, 1988; Neri, Oyung, Colletti, Mallis, Tam, & Dinges, 2002). According to UP health and safety staff materials (brochures and videos) concerning napping policies were distributed to supervisors recently. These materials were also sent to the research team. However, further study may be needed on how to increase the likelihood that the napping policy is maintained and used to effectively facilitate alertness and maximize performance.

Study Limitations

This study, like many field studies, has a methodology which, due to the fact that it is not conducted under controlled laboratory conditions, has limitations which prevent generalizations to a wider range of circumstances and conclusions.

One major limitation of the current study is the fact that the data collected for the analyses occurred approximately four months after the Macdona accident. Since that time, changes have been made to the railroad operations in San Antonio. There have also been changes in personnel and in the amount of undisturbed time off between work shifts for train and engine employees. Thus, the present results m

Another limitation of the study is the fact that the amount of sleep reported by the railroad employees both from self report and actigraph data, as well as the level of sleepiness observed on the various self-report measures could be attributable to the presence of un-diagnosed sleep disorders. These conditions could produce similar findings and not be the result of schedule or work patterns.

One other limitation that should be considered when attempting to understand the data and the results of this study concerns the role of individual differences and so-called "outliers." Recent research has suggested that some individuals are more able to tolerate shift work schedules and sleep deprivation than others (Van Dongen, et. al., 2003). These individual differences contribute to make some people more sensitive than others to the effects of changes in sleep schedules and lack of sleep. The present study ma

(r(Tcjsc1)Tj12 0 0 12 299.12 750.343 518.348 Tm(f)Tj12 0 0 12 03 1.27 037.943 608.348 Tm418 Tm(hen3ouo tors)Tj12 0

4. The work schedule data suggest that a large number of employees are working several consecutive work days which may limit the opportunity to recover from sleep debts. While this is certainly needed in some

had. Considerations for the employee's level of alertness throughout the duty period must be considered. A good example of the attempt to incorporate these facts into crew calling are those that were put in place in the original CANALERT project and then modified for more practical application on the Canadian Pacific in Calgary (as discussed in Sherry, 2000).

11. Further investigation of the impact of work stress and critical incidents on fatigue and alertness needs to be examined. Conversations with employees in focus groups suggest that there was little awareness of the impact of critical incidents on

Glossary of Terms and Acronyms

1. **Boxplot:** A boxplot plots the 25th percentile, the median (the 50th percentile), the 75th percentile, and outlying or extreme values. The length of the box represents the difference between the 25th and 75th percentiles. The horizontal line inside the box represents the median. The "Whiskers" are lines drawn from the ends of the box to the largest and smallest values that are not outliers. The extreme values are cases with the values more than 3 box-lengths from the 75th percentile or 25th percentile. The larger the box, the greater the spread of the data

2. **Electroencephalogram (EEG):** A recording of electrical signals from the brain made by hooking up electrodes to the subjects scalp. EEGs allow researchers to follow electrical impulses across the surface of the brain and observe changes over split seconds of time. In sleep studies, the EEG allows a researcher to determine how stages of sleep change during the night.

3. **Electroocul**

9. **Rapid Eye Movement (REM):** A mentally active period during which dreams occur. REM gives scientists a marker for changes in the brain during sleep.

10. **Stanford Sleepiness Scale (SSS):** Rates an individual's perception of sleepiness during the day on a scale from 1 to 7. A rating of one means the person is fully alert, while a rating of 7 means he or she is struggling to stay awake.

15. Hardy, G. E., Shapiro, D. A., Haynes, C. E., & Rick, J. E. (1999). Validation of the General Health Questionnaire-12: Using a sample of employees from England's health care services. *Psychological Assessment, 11, 159-165.*
16. Harrison, S., Smith, A., & Sykes, R. (2002). Residential rehabilitation courses in the self-directed management of chronic fatigue syndrome: A preliminary evaluation. *Journal of Chronic Fatigue Syndrome, 10(2),* 59-65.
17. Hoddes, E., Zarcone, V., Smythe, H., Phillips, R., & Dement, W.C. (1973). Quantification of sleepiness: A new approach. *Psychophysiology*;10, 431-436.
18. Howell, A.J., Jahrig, J.C., & Powell, R.A. (2004). Sleep quality, sleep propensity, and academic performance. *Perceptual & Motor Skills, 99(2),*525-535.
19. Incalzi, R.A., Marra, C., Salvigni, B.L., Petrone, A., Gemma, A., Selvaggio, D., & Mormile, F. (2004). Does cognitive dysfunction conform to a distinctive pattern in obstructive sleep apnea syndrome. *Journal of Sleep Research, 13(1),* 79-86.
20. Johns, M. W. (1991). A new method for measuring daytime sleepiness: The Eppworth sleepiness scale. *Sleep,1(14),* 540-545.
21. Johns, M.W. (1993). Daytime sleepiness, snoring, and obstructive sleep apnea. The Epworth Sleepiness Scale. *Chest,103*, 30-36.
22. Kilkenny, T.M., Hajjar, A., Zyadeh, N., & Chaftari, P. (1999). The usefulness of the Epworth Sleepiness Scores alone as a preditor of sleep study outcomes. *Ches*t,116 (4), 391.
23. Korszun, A., Young, E.A., Engleberg, N.C., Brucksch, C.B., Greden, J.F., & Crofford, L.A. (2002). Use of actigraphy for morning sleep and activity levels in patients with fibromyalgia and depression. *Journal of Psychosomatic*

47. Thayer, R. E. (1978). Factor analytic and reliability studies on the Activation-Deactivation Adjective Check List. *Psychological Reports, 42, 747-756.*
48. Theorell, T.; Karasek, R. A. (1996). Current issues relating to psychosocial job strain and cardiovascular disease research. *Journal of Occupational Health Psychology, 1, 9-26.*
49. Thorpy M.J. (1992). The clinical use of the multiple sleep latency test: The Standards of Practice Committee of the American Sleep Disorders Association. *Sleep,15(3),* 268-276.
50. Van Dongen, H. P., Maislin, G., Mullington, J. M., & Dinges, D. F. (2003). The cumulative cost of additional wakefulness: Dose-response effects on neurobehavioral functions and sleep physiology from chronic sleep restriction and total sleep deprivation. *Sleep: Journal of Sleep & Sleep Disorders Research,* 26, 117-126.
51. Van Dongen, H. P., Rogers, N.L., & Di