NCIT **National Center for Intermodal Transportation**

A Partnership Between the University of Denver and Mississippi State

RESULTS OF FAST- FAID FATIGUE MODEL CALIBRATION & IDENTICATION OF COUNTERMEASURES FOR HIGH RISK WORK SCHEDULES

Patrick Sherry, PhD
University of Denver

Karen E. Philbrick, PhD.
Mineta Transportation Institute

Prepared
By
Dr. Patrick Sherry

University of Denver
National Center for Intermodal Transportation

October 4 2016

ncit.msstate.edu

Innovators in intermodal transportation education, research, and technology transfer.

Executive Summary

There has been an increased interest in the use of bio mathematical models to understand and predict the impact of extended work hours, exceptional duty rosters, and other work related demands. The present study sought to improve on previously published analyses by analyzing a more representative sample of work schedules typical of the everyday operations of the commuter rail or intercity passenger rail industry. A representative sample of work schedules was obtained that consisted of 101 work schedules in which 61% were morning starts, 36% were afternoon, and 3% were midnight starts. These schedules were then analyzed to produce FAST scores for every 30 minute interval that the employee was working. The results indicate the presence of a highly statistically significant relationship between the two models which supports the assumption that the two models are measuring similar phenomena. Therefore we can assume that the FAID model is also validated and that cutoff scores on the FAID model can also be equated to cutoffs on the previously validated FAST model. The exact score conversion between FAST and FAID is presented using the linear conversion model. IN addition, based on the analysis of work schedules using the FAST model we were able to identify a number of strategic fatigue countermeasures that could be used to mitigate the impact of midnight shift work and reduce the overall risk of accidents considerably.

Introduction

There has been an increased interest in the use of bio mathematical models to understand and predict the impact of extended work hours, exceptional duty rosters, and other work related demands. Bio mathematical models have been developed in various laboratories around the world with the intention of modeling and predicting the physiological and cognitive responses to a variety of different conditions to which the individual has been exposed.

The accuracy of these models for both describing and predicting human behavior was the subject of a conference on fatigue modeling held in Seattle Washington in 2002 and described in a special of issue *Aviation, Space and Environmental Medicine* (Neri, 2004). The most popular and well published models were described and compared using five separate sets of data that were thought to represent common and extreme conditions in the aviation and railroad industry. The conference organizers asked the authors of the models to utilize the prepared data sets and to analyze the data using their models. The models were then compared to determine how well they accounted for the data that they were attempting to model. The results of the conference indicated that none of the models was much better than any of the others in accounting for and predicting human fatigue. In fact the results of the analyses comparing the various models concluded that none of the models was very different from any other. In addition, overall none of the modes was very good at explain or predicting the restricted sleep scenario conditions, the kind of sleep schedule typically faced by people in the rail industry.

Model Calibration

In its November 2010 report, "Procedures *for Validation and Calibration of Human Fatigue Models: The Fatigue Audit InterDyne Tool,*" the Federal Railroad Administration (FRA) described a method for validating fatigue models that involved demonstrating a statistically significant relationship with an already or previously validated model. Previously the FAST model (Hursh, et al., 2004) had been related to an increased risk of human factor caused accidents with scores on the FAST model below 70 (Hursh, Raslear, Kaye, and Fanzone (2006). In the Tabac & Raslear (2010) study, a significant relationship between FAST and FAID was demonstrated and a calibration set as well.

Results of this analysis demonstrated that there was a significant linear relationship between the FAST and the FAID scores and that a biomathematical model was able to be determined. In fact, the published correlation coefficient between FAST and FAID scores was >.90. However, this relationship was based on bin, or ten point interval, means of the FAST scores, comprised of the scores that fell within a ten point range of FAST scores rather than individual pairs of scores. Such an approach reduces the normal variation in the relationship between the independent and dependent variables examined in this analysis. The application of linear regression techniques is typically undertaken with the assumption that the underlying distribution has a moderate amount of variability. By limiting the analyses to bin means the variability is thereby reduced and predictive and explanatory power is reduced considerably. A more robust application of linear regression requires the use of data with more variability.

Results of the Tabac & Raslear (2010) study determined that a FAID cutoff score of 60 corresponded to a FAST score of 70 following the linear transformation of the FAID score using the parameter weights and constants identified in the study. However, the identification of 60 as the corresponding equivalent to the FAST score may also be the result of unique characteristics of the data set used to generate the linear transformation equation. The data set identified consisted of work schedules of employees involved in either human factor or non human factor caused accidents in the freight industry. Inspection of the data provided by Hursh, Raslear, Kaye, and Fanzone (2006) reveals that most of the accident data provided fell 21:00 and 05:00 hours. Thus, this particular data set might have a slight bias towards lower levels of alertness and higher levels of fatigue. While such a data set is useful in showing the relationship between accident data and fatigue models it is not optimal for calibrating one model to another because the mean of the data set is weighted towards the fatigued end. In this there will likely be a preponderance of scores from both the FAST and the FAID model that would be in the range suggesting a higher risk for fatigue. These scores, due to the law of central tendency, would have the effect of skewing the distribution towards the fatigued end.

Since one goal of these studies is to provide a tool that can apply generally to the passenger rail industry, an alternative methodology would be to use a sample of typical work schedules drawn from the passenger rail industry. Moreover, since the goal is to establish a mathematical relationship between the two models a more robust relationship may be demonstrated by choosing a typical sample of work schedules that represent the likely activities of everyday operations. Thus, the present study sought to analyze a more representative sample of work schedules typical of the everyday operations of the commuter rail or intercity passenger rail industry.

Present Study

Based on the proposed alternative methodology for determining the best calibration of FAST and FAID it was proposed that a representative sample of schedules be analyzed according to the percentage of morning afternoon and midnight schedules. The data submitted suggested that some of the largest railroads had the following percentage breakdown of work schedules.

Percentage of Morning, Afternoon and Nighttime Schedules In Passenger Railroad Operations

On Duty		Off Duty		Metra	SEPTA	MNR	LIRR
3:30 AM	10:00 AM	11:30 AM	10:00 PM	65%	60%	62%	57%
10:00 AM	9:00 PM	1:00 PM	3:00 AM	32%	37%	32%	42%
9:30 PM	3:30 AM	7:00 AM	9:30 AM	2%	2%	6%	1%

Given that this percentage breakdown is consistent for four major commuter railroads a representative sample of work schedules was obtained that consisted of 101 work schedules. In this sample 61% were morning starts, 36% were afternoon, and 3% were midnight starts. These schedules were then analyzed to produce FAST scores for every 30 minute interval that the employee was working. Similarly, *InterDynamics* in Australia, publishers of FAID, analyzed the same data set and prepared a similar of set FAID scores during work periods for every 30 minute interval worked. The data for a typical schedule (e.g. schedule #240) was arranged as follows:

Example of FAST FAID Model Scores			
Date	Time	**FAST**	**FAID**
4/11/2011	14:30	96.26	32
	15:00	96.43	31
	15:30	96.78	31
	16:00	97.26	32
	16:30	97.84	31
	17:00	98.46	29
	17:30	99.06	30
	18:00	99.56	30
	18:30	99.9	30
	19:00	100.01	31
	19:30	99.84	32
	20:00	99.33	34
	20:30	98.45	38
	21:00	97.2	42
	21:30	95.58	44
	22:00	93.63	46
	22:30	91.39	47
	23:00	88.93	48
	23:30	86.32	48
4/12/2011	0:00	83.66	49

The scores for FAST and FAID were arranged in 30 intervals and paired to so that the scores were paired for the same 30 minute interval. These data were then entered into a statistical package and a correlation coefficient was generated. Based on 10,934 FAST-FAID pairs, representing five or six day work schedules, the following statistics were generated.

Descriptive Statistics

	Mean	Std. Deviation	N
FAST	90.63	9.07	10934
FAID	50.07	17.46	10934

There were not an exact number of FAST and FAID scores. The FAID program provides scores at the start, end and for each intervening hour of the work schedule. The FAST program simply calculated the average FAST score for the 30 minute period leading up to the time of day that the work day ended. *InterDynamics* arranged a special run of FAID to produce scores on every half hour of a work schedule and not on the start and end. This enabled half-hourly pairs of FAST and FAID scores to be produced and compared. For the present data set a total of 10795 FAST-FAID pairs were produced and analyzed.

Correlation Between FAST and FAID

		FAST
FAID	Pearson Correlation	-.729(**)
	Sig. (2-tailed)	.000
	N	10795

** Correlation is significant at the 0.01 level (2-tailed).

The bivariate correlation coefficient that was generated from these paired FAST-FAID scores is shown above. The correlation is statistically significant at beyond the $P<.001$ level and account for 53% of the explained variance. Note that the correlation is negative as would be expected as the FAST scores are higher for lower levels of fatigue while the FAID scores are lower for lower levels of fatigue. The correlation alone indicates the presence of a highly similar relationship between FAST and FAID. There should be no difficulty whatsoever in describing the statistical relationship between the models.

Prediction of FAST Scores

The FRA published a report (Hursh, Raslear, Kaye, and Fanzone, (2006) showing that there is a greater likelihood of human factors caused accidents among persons whose work schedules produce FAST scores below 70. Accordingly, the FRA has accepted FAST as an acceptable method for determining the risk of fatigue in work schedules. Additionally, FRA demonstrated that there was a significant relationship between FAST and FAID scores in its publication (Tabac & Raslear, 2010). Thus, the FRA study suggests that FAST scores below 70 may be sensitive to detecting human factors caused accidents. Other models, like FAID, if they are highly correlated with FAST, can be assumed to show a similar relationship. The goal of a calibration study is to show that the two models are in fact related mathematically. The two data sets obtained were subjected to further analysis using the SPSS Curve Fitting Procedure (SPSS Release 17.0, 2008). This procedure attempts to fit various mathematical equations to the observed data to estimate the underlying relationship. By understanding the underlying relationship and plotting the data we are able to translate the scores of one model or measuring system to another just as we can convert Fahrenheit to Centigrade on a temperature thermometer. Unfortunately, the two models are not measuring exactly the same thing so we expect that there will not be a perfect translation of the two approaches. Nevertheless, as can be seen, with a correlation of -.73 we have a very high degree of confidence that the models are in fact highly correlated. Nunnaly (1978, p245) describes correlations in the .70 neighborhood as being fairly strong but not describing identical tools. The table below shows the results for the analysis of the fit to the data of the various curves estimated.

Predicting FAST from FAID: Model Summary and Parameter Estimates

Dependent Variable: FAST

Equation	Model Summary					Parameter Estimates		
	R Square	F	df1	df2	Sig.	b1	b2	b3
Linear	.531	12239.105	1	10793	.000	-.321		
Quadratic	.556	6758.315	2	10792	.000	-.007	-.003	
Cubic	.559	4560.055	3	10791	.000	-.363	.003	-2.98E-005

The independent variable is FAID.

All of the equations are highly statistically significant in terms of explaining the FAST and FAID scores. The models can be compared to each other by examining the amount of variance accounted for, which is represented in the second column of the table under R Square. This statistic shows how well the model or equation accounts for the observed data. The best model in this case is the Cubic model which accounts for .559 or 56% of the variance, as compared to the others. However, they are all in the same neighborhood and we could not really say at this point that one is highly superior. The cubic is 2.7% better at accounting for the variance and so gets the numerical edge. Most likely the underlying relationship between the models is not linear, but curvilinear. This means that instead of a perfectly straight line, the data are likely arranged in more of a curve with the ends or tails sloping up at either end of the distribution.

As can be seen from the figure below, the purpose of generating the appropriate model is to be able to predict or convert the scores of one model to another. The diagonal line through the center of the darkened section of the graph shows the plot of the relationship between FAST and FAID. The diagonal line is the linear estimate of the relationship between the two models. The vertical red line in the diagram indicates the position of 90 on the FAID scale (which is the current cutoff recommended by the developers of FAID) intersects with the horizontal line from the point of 80 on the FAST axis. The FAID score of 90 corresponds to a score of 80 on the FAST model. The recommended cutoff for FAST is 70. Thus, the present analysis actually identifies a more restrictive or conservative threshold for fatigue than is currently recommended by the FAID authors and than is currently utilized if we accept that the FAST score of 70 as the validated score below which the risk of human factors caused accidents are likely to occur. In other words, by accepting the FAID cutoff of 90 these data suggest that we would obtain a score of 80 on the FAST.

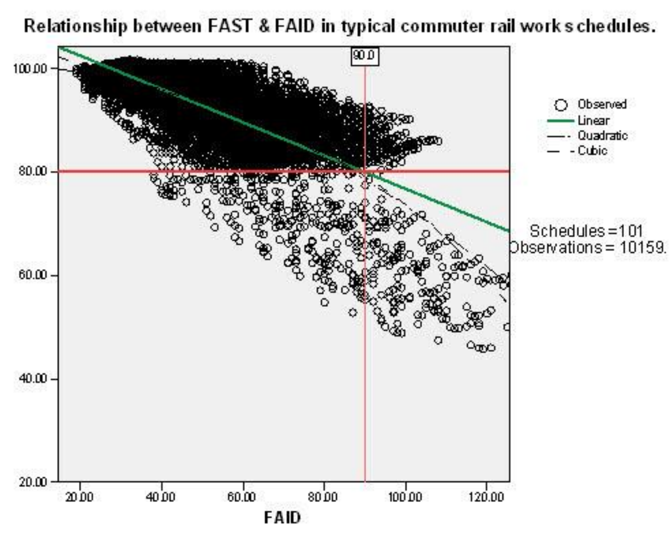


Figure 1. FAST and FAID scores for typical commuter rail operations schedules.

To summarize, work schedules which are above 90 on the FAID scale that would be considered to be at risk for fatigue would also be considered at risk for fatigue on the FAST model as well.

The following histograms present the number of schedule observations that would be identified as at risk for fatigue with varying cutoff scores. For the FAST model, the number of corresponding cases if the fatigue threshold is set at

70 on the FAST model is also 4%. Using the FAID model if the cutoff were set at 90 (as is recommended) the number of observations falling above the threshold is 4.1%.

The differences between the present study and the Tabac & Raslear (2010) study are likely due to differences between the two samples obtained. In the present study, steps were taken to ensure that a representative sample of work schedules was obtained. The present sample consisted of 61% morning, 36% afternoon and 3% midnight shifts. Thus, the present sample is more reflective of actual work practices as opposed to the more atypical schedules that might have been obtained in the earlier validation study sample (Hursh, Raslear, Kaye, and Fanzone (2006) that was based on human factors and non-human factors caused accidents. Since accidents are so rare in the industry it is clear that theirs was an unusual data set. It should also be noted that the validation sample was obtained entirely from freight operations. The present sample is obtained entirely from commuter rail operations. Examining the means for FAST and FAID reported in the Tabac & Raslear (2010) study on page 16, the mean of FAST and FAID is 69 and 59 respectively. In the present study, FAST and FAID means are 90 and 51 respectively. Thus, the present data is probably more representative of normal working hours and times of day.

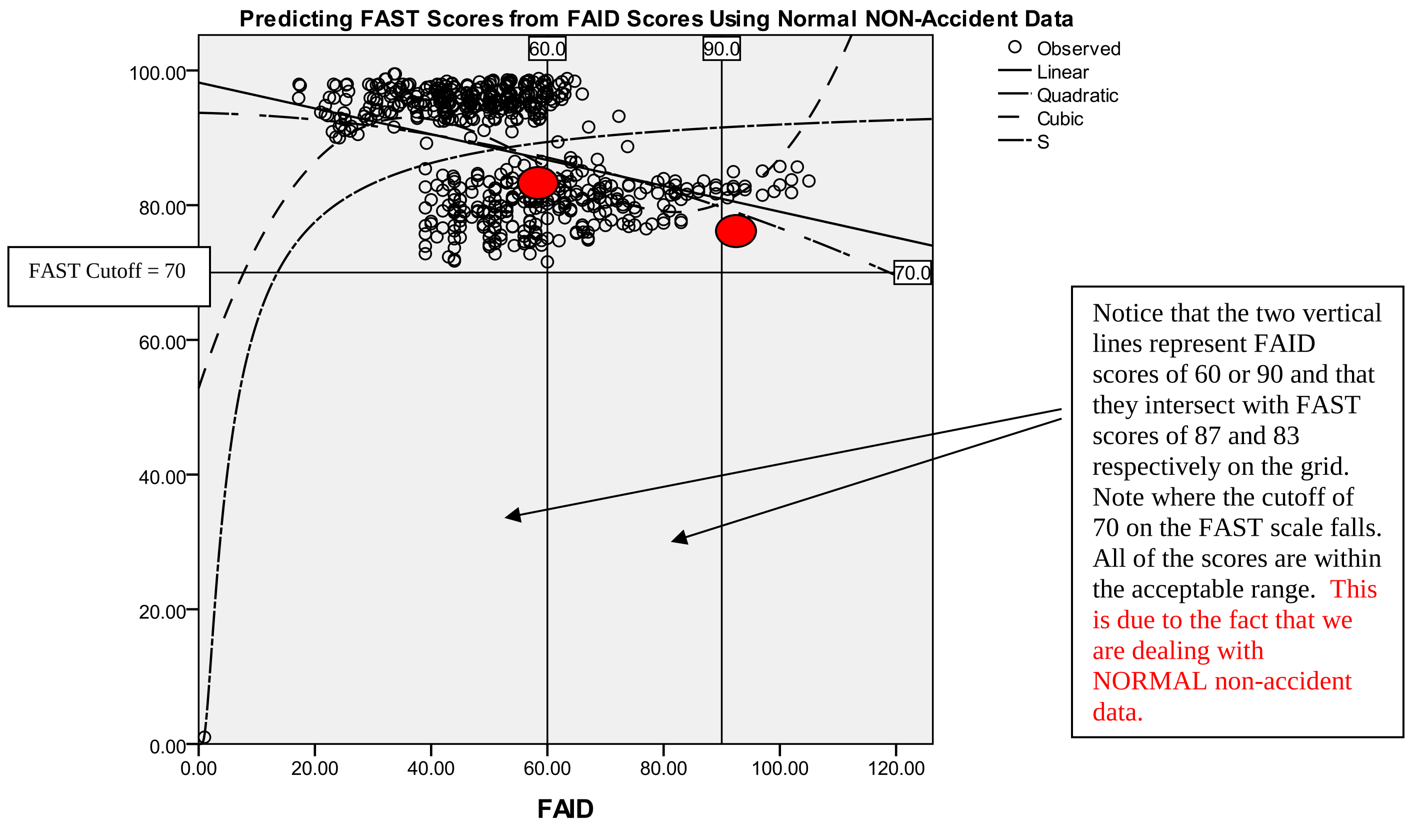


Figure 2. Prediction of FAST Scores from FAID scores for NORMAL working schedules.

Note: Above analysis was based on 577 pairs of observations. Pairing FAST scores with FAID scores was made on an hourly basis.

Table 1. Results of models predicting FAST from FAID scores.

Equation	Model Summary				
	R Square	F	df1	df2	Sig.
Linear	.111	71.516	1	571	.000
Quadratic	.116	37.299	2	570	.000
Cubic	.228	55.895	3	569	.000
S	.721	1472.583	1	571	.000

A second set of analyses, presented in Figures 4 and Table 5, studied a set of so-called normal work schedules that were dealing with employees working essential a 7am to 3pm work schedule without involvement in accidents. Results of the analyses of these schedules reveal a similar pattern in which the scores are highly suggestive of a strong relationship between the two models. Again, suggesting that the two models are tapping into very similar phenomena but with two different metrics.

The results indicate the presence of a highly statistically significant relationship between the two models. In addition, the underlying relationships between FAST and FAID are robust and permit the calculation of scores from one model to the other. By being able to compute FAST scores from the FAID model we can assume that the two models are both measuring similar phenomena. Therefore we can assume that the FAID model is also validated. Cutoff scores on the FAID model can also be equated to cutoffs on FAST. Thus, the present analyses indicate that the two models are highly correlated and both reflect the degree of fatigue in the work schedule. The exact score conversion between FAST and FAID is listed below using the linear conversion model.

$$\textbf{FAST} = (-0.32) * \textbf{FAID} + 108.82$$

FAST	FAID
70.42	120
72.02	115
73.62	110
75.22	105
76.82	100
78.42	95
80.02	90
81.62	85
83.22	80
84.82	75
86.42	70
88.02	65

Based on these results the evidence suggests that have two tools which measure fatigue. Further, the present evidence suggests that there is no reason to set the FAID cutoff lower than 90.

One cautionary note, these scores are estimates of cognitive effectiveness or readiness to perform tasks at a particular point in time given assumptions that the individual has obtained a reasonable amount of sleep prior to doing so. However, these estimates are based on group averages and are not accurate for the estimate of actual individual performance. Variations of work activity, sleeping schedules and opportunities, not to mention individual differences, will all play a significant role in determining actual readiness. The fatigue models are the best estimate of what we might expect in a certain very general situation. Additional research is needed to improve the accuracy of these models in the workplace.

Background on the ASLRRA Industry

In order to gather some information to make some estimates surveys of workforce work schedules were obtained from two companies that manage a number of short line railroads. Two companies, RailAmerica and WATCO provided sufficient data to permit statistical analyses. RailAmerica (RA) provided its entire work history for the 754 employees on its payroll for the month of February 2011.

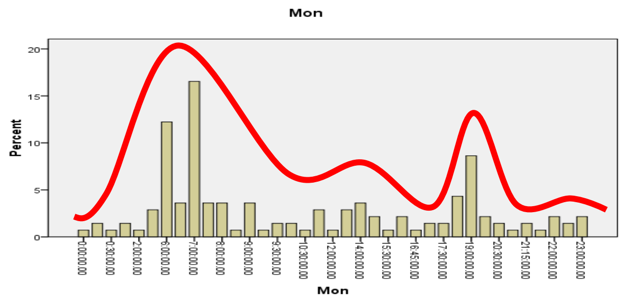


Figure 3. Start times for Short Lines.

The RA employees had 16852 starts or days worked during that period. On the average, the RA employee had 22.35 starts during that time period with an average length of shift equal to 9.48 hours with a standard deviation of 2.33 hours. The maximum hours reported working was 15.89. A little over 15.4 % of the 754 employees reported working a shift over 12 hours during that time period. Data for WATCO companies are not as detailed. Much of the data was recorded by hand. Nevertheless, data was available for 22 different railroads which consisted of work schedules for 204 different work schedules that 384 employees were assigned to (the actual number is uncertain due to missing data). Average shift length was not available for all railroads but it was possible to determine that the majority of the work periods began and ended during daylight hours. Moreover, 75% of the work schedules were 5 days in length, 9% were 6 days in length, and 2% were 7 days in length, the remainder worked 4 days or less.

As can be seen in the following chart, the lowest amounts of sleep are obtained by the persons who work starting at 1am. The data would probably have been a little less favorable if the findings for the 2am stat had been produced by more than two subjects. This average of 7 hours seems high and is probably due to the fact that both of the persons who had this job started regularly at 2am and only worked for six hours. Thus, they had plenty of time to recover.

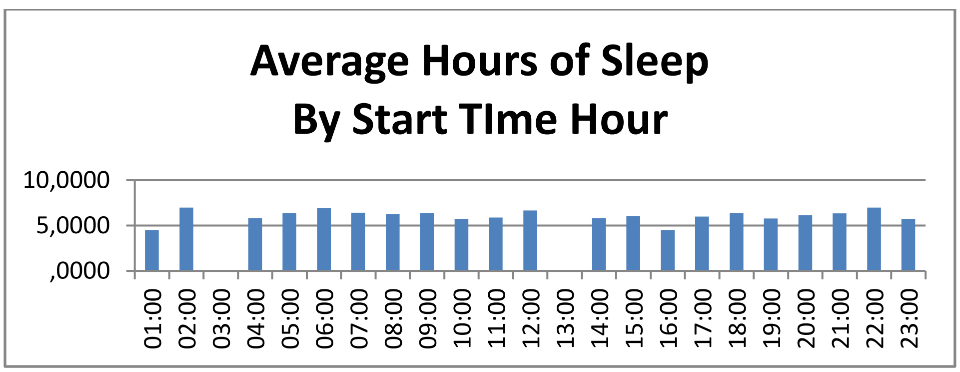


Figure 4. Average Hours of Sleep by Start Hour.

The Epworth mean was 9.0 with a standard deviation of 4.8. Previous research has established a cutoff of 10.0 as the cutoff between normal levels of sleepiness and borderline cases. The clinical cutoff is thought to be 15 or above.

As can be seen from the chart approximately 56.3% of the total number of respondents are below the cutoff. Additionally, a total of 12.5% of the respondents are at or above the clinical cut off score of 15.

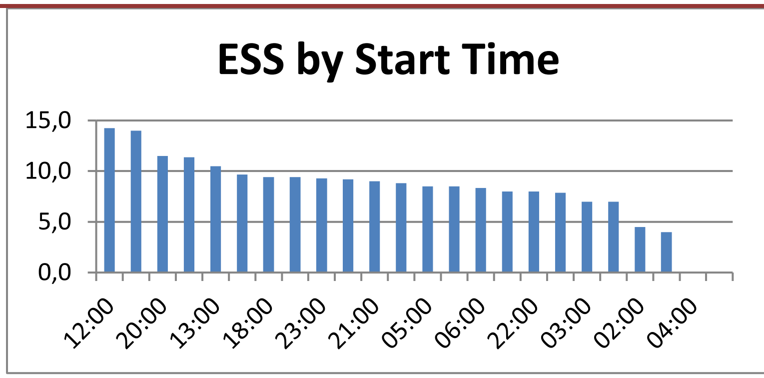


Figure 5. ESS by Start Time.

Interestingly, if we plot the Epworth scores against the start time hour the data reveal an interesting finding. Namely, Epworth scores are higher for those who start work in the late afternoon and early evening. Apparently, the fatigue levels of those persons working the midnight hours are not as pronounced as those from other shifts. Perhaps they have learned to adapt to the demanding conditions of these works schedules.

Conclusion

After reviewing the work schedules and operational demands of the baseline study participants a number of suggested counter measures were reviewed and considered. The operational practicality of these suggestions was reviewed by safety professionals working for the short line rr association. The following countermeasures were considered to be most feasible:

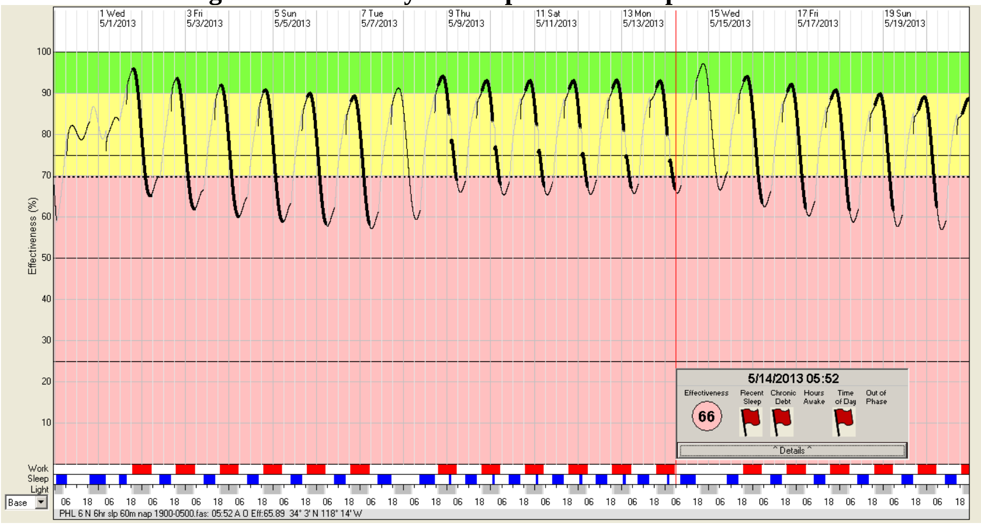
1. Utilization of on-duty naps to offset the negative impact of midnight hours
2. Increase the amount of on-duty supervision
3. Increase the number of breaks
4. Alter the start and end time of work shifts
5. Decrease the number of hours worked
6. Increase the number of employees on the work force

Discussing these options with the FRA and some of the ASLRRA study participants has resulted in a general agreement to go forward and that short-line railroads that want to participate in the waiver will adopt a fatigue countermeasures policy that will include: education, additional supervision, adjustments to hours worked and a napping policy consistent with ones needed to reduce fatigue using the FAST model. In addition, persons engaged in the pilot study will agree to complete the study instruments including the 30 day sleep log needed to document the effectiveness of the proposed countermeasures as well as appropriate pre post baseline data gathering.

Reviewing the countermeasures using the FAST model shows that in most cases a rest period that includes a nap of 30-60 minutes would bring the overall effectiveness levels to nearly within acceptable limits assuming that the participants adhered to proper sleep hygiene prior to and during the time that they worked the six midnight shifts. These results are displayed in Figure 4 in the Appendix. More importantly, a break of 90 minutes sleep time brought the overall effectiveness levels well within the appropriate and recommended cutoffs and guidelines. These results are also displayed in Figure 5 in the Appendix.

Thus, the proposed waiver, that allows short line railroads to participate in the pilot study, would, as supported by our preliminary analyses using the FAST model, provide sufficient effectiveness levels if these fatigue counter measures were followed.

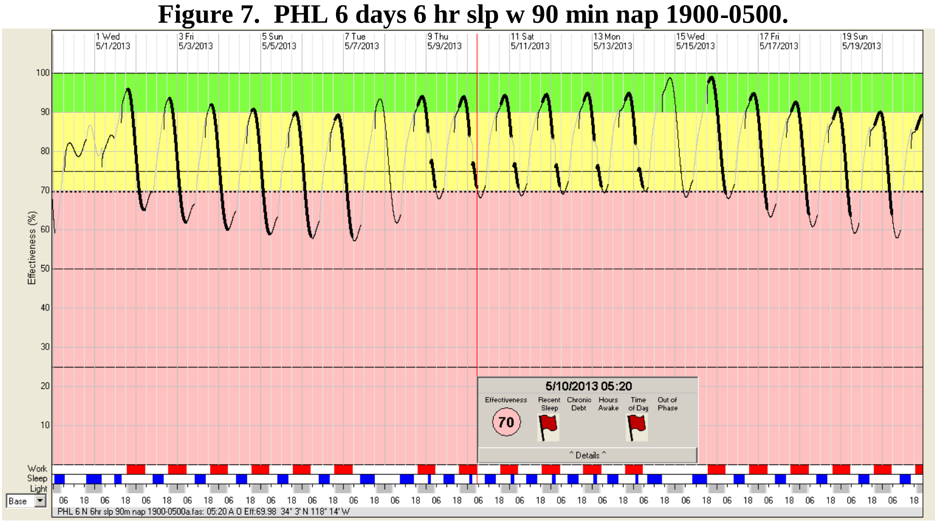
Figure 6. PHL 6 days 6 hr slp w 60 min nap 1900-0500.



PHL 6 N 6hr slp 60m nap 1900-0500.fas Work

Note: Average effectiveness level for days with counter measures implemented is ABOVE the cutoff of 70 and the percent time below criterion is within acceptable range.

Figure 7. PHL 6 days 6 hr slp w 90 min nap 1900-0500.



PHL 6 N 6hr slp 90m nap 1900-0500a-Work

Note: Average effectiveness level for days with counter measures implemented is ABOVE the cutoff of 70 and the percent time below criterion is within acceptable range.

References

Hursh, S.R., Raslear, T.G., Kaye, A.S., & Fanzone, J.F. (2006). *Validation and calibration of a fatigue assessment tool for railroad work schedules, final report* (Report No. DOT/FRA/ORD-08/04). Washington, DC: U.S. Department of Transportation.

Hursh, S.R., Raslear, T.G., Kaye, A.S., & Fanzone, J.F. (2006). *Validation and calibration of a fatigue assessment tool for railroad work schedules, final report* (Report No. DOT/FRA/ORD-08/04). Washington, DC: U.S. Department of Transportation. (http://www.fra.dot.gov/downloads/Research/ord0804.pdf)

Hursh, S.R., Redmond, D.P., Johnson, M.L. Thorne, D.R., Belenky, G., Balkin, T.J., Storm, W.F., Miller, J.C., & Eddy, D.R. (2004). Fatigue models for applied research in war fighting. *Aviation, Space, and Environmental Medicine*, *75*, A44–A53.

Neri, D.F. (2004). Fatigue and performance modeling workshop, June 13–14, 2002. *Aviation, Space, and Environmental Medicine*, *75*, A1–A3.

Nunnally, J. C. (1978). Psychometric theory (2nd ed.). New York: McGraw-Hill.

Roach, G.D., Fletcher, A., & Dawson, D. (2004).A model to predict work-related fatigue based on hours of work. *Aviation, Space, and Environmental Medicine*, *75*, A61–A69.

Tabak, B, & Raslear, T.G. (2010). Procedures for Validation and Calibration of Human Fatigue Models: The Fatigue Audit InterDyne Tool. (Report No. DOT/FRA/ORD-10/14). Washington, DC: U.S. Department of Transportation.