U.S. Department of
Transportation

**Federal Railroad
Administration**

Validation of FAST Model Sleep Estimates with Actigraph Measured Sleep in Locomotive Engineers

Office of Railroad
Policy and Development
Washington, DC 20590

788.2-2c

DOT/FRA/ORD-12/05

Final Report
April 2012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE April 2012	3. REPORT TYPE AND DATES COVERED Final Report December 2005 – October 2011

4. TITLE AND SUBTITLE

Validation of FAST Model Sleep Estimates with Actigraph Measured Sleep in Locomotive Engineers

5. FUNDING NUMBERS

DFRA.080088

6. AUTHOR(S)

Judith Gertler,[1] Steven Hursh,[2] Joseph Fanzone,[2] and Thomas Raslear[3]

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

[1]QinetiQ North America, Inc.
Technology Solutions Group
350 Second Avenue
Waltham, MA 02451-1196

[2]Institutes for Behavior Resources
2104 Maryland Avenue
Baltimore, MD 21218

[3]Federal Railroad Administration
1200 New Jersey Avenue SE
Washington, DC 20590

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Department of Transportation
Federal Railroad Administration
Office of Railroad Policy and Development
Washington, DC 20590

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

DOT/FRA/ORD-12/05

11. SUPPLEMENTARY NOTES

Program Manager: Thomas Raslear

12a. DISTRIBUTION/AVAILABILITY STATEMENT

This document is available to the public through the FRA Web site at http://www.fra.dot.gov.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

This report presents the results of a study to validate the AutoSleep sleep prediction algorithm, which is a component of the Fatigue Avoidance Scheduling Tool (FAST). Researchers collected work and sleep data from 41 locomotive engineers by using actigraphy and daily log books and compared these data with AutoSleep predictions developed according to the log-book–recorded work periods. Comparison of the actigraphy data with model predictions on a minute-by-minute basis found an overall agreement between the two 87 percent of the time. Application of Signal Detection Theory to the data indicates that AutoSleep is biased toward underestimating daily sleep. These findings validate the sleep prediction algorithm of FAST and validate its utility for assessing fatigue risk created by typical railroad schedules.

14. SUBJECT TERMS

Fatigue, sleep, actigraphy, locomotive engineer, FAST, Fatigue Avoidance Scheduling Tool

15. NUMBER OF PAGES

29

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT None

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18
298-102

METRIC/ENGLISH CONVERSION FACTORS

ENGLISH TO METRIC

LENGTH (APPROXIMATE)

1 inch (in) = 2.5 centimeters (cm)
1 foot (ft) = 30 centimeters (cm)
1 yard (yd) = 0.9 meter (m)
1 mile (mi) = 1.6 kilometers (km)

AREA (APPROXIMATE)

1 square inch (sq in, in^2) = 6.5 square centimeters (cm^2)
1 square foot (sq ft, ft^2) = 0.09 square meter (m^2)
1 square yard (sq yd, yd^2) = 0.8 square meter (m^2)
1 square mile (sq mi, mi^2) = 2.6 square kilometers (km^2)
1 acre = 0.4 hectare (he) = 4,000 square meters (m^2)

MASS - WEIGHT (APPROXIMATE)

1 ounce (oz) = 28 grams (gm)
1 pound (lb) = 0.45 kilogram (kg)
1 short ton = 2,000 pounds (lb) = 0.9 tonne (t)

VOLUME (APPROXIMATE)

1 teaspoon (tsp) = 5 milliliters (ml)
1 tablespoon (tbsp) = 15 milliliters (ml)
1 fluid ounce (fl oz) = 30 milliliters (ml)
1 cup (c) = 0.24 liter (l)
1 pint (pt) = 0.47 liter (l)
1 quart (qt) = 0.96 liter (l)
1 gallon (gal) = 3.8 liters (l)
1 cubic foot (cu ft, ft^3) = 0.03 cubic meter (m^3)
1 cubic yard (cu yd, yd^3) = 0.76 cubic meter (m^3)

TEMPERATURE (EXACT)

[(x-32)(5/9)] °F = y °C

METRIC TO ENGLISH

LENGTH (APPROXIMATE)

1 millimeter (mm) = 0.04 inch (in)
1 centimeter (cm) = 0.4 inch (in)
1 meter (m) = 3.3 feet (ft)
1 meter (m) = 1.1 yards (yd)
1 kilometer (km) = 0.6 mile (mi)

AREA (APPROXIMATE)

1 square centimeter (cm^2) = 0.16 square inch (sq in, in^2)
1 square meter (m^2) = 1.2 square yards (sq yd, yd^2)
1 square kilometer (km^2) = 0.4 square mile (sq mi, mi^2)
10,000 square meters (m^2) = 1 hectare (ha) = 2.5 acres

MASS - WEIGHT (APPROXIMATE)

1 gram (gm) = 0.036 ounce (oz)
1 kilogram (kg) = 2.2 pounds (lb)
1 tonne (t) = 1,000 kilograms (kg)
= 1.1 short tons

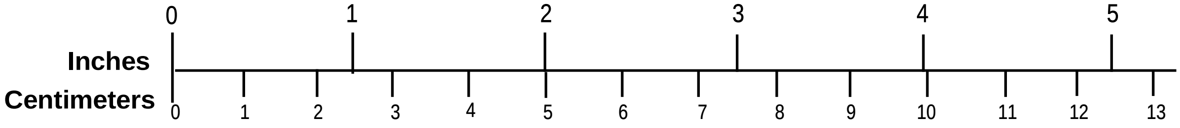
VOLUME (APPROXIMATE)

1 milliliter (ml) = 0.03 fluid ounce (fl oz)
1 liter (l) = 2.1 pints (pt)
1 liter (l) = 1.06 quarts (qt)
1 liter (l) = 0.26 gallon (gal)
1 cubic meter (m^3) = 36 cubic feet (cu ft, ft^3)
1 cubic meter (m^3) = 1.3 cubic yards (cu yd, yd^3)

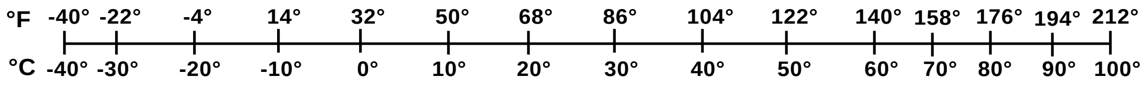
TEMPERATURE (EXACT)

[(9/5) y + 32] °C = x °F

QUICK INCH - CENTIMETER LENGTH CONVERSION



QUICK FAHRENHEIT - CELSIUS TEMPERATURE CONVERSIO



For more exact and or other conversion factors, see NIST Miscellaneous Publication 286, Units of Weights and Measures. Price $2.50 SD Catalog No. C13 10286

Updated 6/17/98

Acknowledgements

This report presents the results of a research study designed to validate the sleep prediction algorithm used by the Fatigue Avoidance Scheduling Tool (FAST). QinetiQ North America (QNA) and the Institutes for Behavior Resources conducted the work for the Federal Railroad Administration (FRA) under contract DTFR53-07-D-00003 with guidance from Mr. Scott Kaye, FRA Office of Safety. The authors worked closely with representatives of the Brotherhood of Locomotive Engineers and Trainmen (BLET) to identify suitable sites for the data collection and to recruit participants. Without the cooperation and interest of the BLET representatives, this work would not have been possible.

The authors also thank two individuals at QNA for their significant contributions to this work. Mr. Alex Viale was responsible for the field data collection. Ms. Audrey Murray coded the data from the log books, and Mr. Viale performed the statistical analysis of this data.

Contents

Illustrations

Tables

Executive Summary

The Fatigue Avoidance SchedulingTool (FAST) implements the Sleep, Activity, Fatigue, and Task Effectiveness (SAFTE) biomathematical model of performance and fatigue to generate estimates of performance degradation owing to the individual's level of fatigue. Such models offer opportunities to predict a railroad worker's level of fatigue and resulting performance degradation, based on the individual's work schedule. A key component of FAST is the AutoSleep module, which predicts when the individual working the schedule may sleep. The purpose of the work described in this report was to collect work and sleep data from locomotive engineers by using actigraphy and daily log books and then to validate the AutoSleep sleep predictions according to these data.

Researchers recruited locomotive engineers for participation in the study at three locations: Union Pacific Railroad (UP) in Colton, CA, UP in Roseville, CA, and BNSF Railway in Spokane, WA. A total of 46 individuals participated, and data from 41 was usable for the purposes of the study. Study participants wore the actigraph continuously for 14 days (d). They also kept a daily log in which they recorded commute time, work periods, and personal time as well as sleep periods.

During the 2 weeks of the study, participants averaged 7.7 work periods and the average work period was 9:37 (hours:minutes), with over half exceeding 9:45. A quarter of the work periods exceeded 12 hours (h) because of deadheading, which did not violate Federal Hours of Service limitations in effect at the time of the study. Commute time away from home was slightly shorter than commute time at home. Similarly, call time at an away terminal was shorter than that at the home terminal. Sleep per 24 h, as recorded with actigraphy, averaged 6:58 with a quarter averaging more than 7:26.

Actigraphy data and AutoSleep predictions were compared for the percentage of time that both actigraphy and AutoSleep showed the subject in the same state (sleep or awake) and the mean sleep per subject in 24 h.

AutoSleep requires specification of a normal bedtime, the maximum number of hours of sleep during a workday or a rest day, and the start and end times of a "forbidden zone" when no sleep may occur, typically in the afternoon. Because of the "forbidden zone" and the limitation on daily sleep, perfect agreement between the two sets of data was not possible. With the baseline settings of AutoSleep, model predictions of sleep agreed with the actigraphy data 85 percent of the time, but AutoSleep underestimated sleep in 24 h by 13 min. When the model settings were changed to decrease the "forbidden zone," delay bedtime, and increase maximum rest day sleep, there was 87 percent agreement and AutoSleep overestimated sleep by 6 min. Overall, the sleep patterns and duration estimated by AutoSleep agreed well with those recorded with actigraphy. The divergence is likely due to the forbidden zone and the fact that some participants slept longer than the maximums imposed by AutoSleep. Analysis of the data using signal detection theory confirmed that AutoSleep is slightly biased toward underestimating sleep and provided a rationale for that basis.

The implication of this study is that fatigue assessments associated with train and engine service (T&E) work schedules by using FAST are based on valid expectations of average sleep patterns and, therefore, provide a reasonable estimate of sleep restriction and associated fatigue risk.

These findings further validate the utility of FAST for assessing fatigue risk created by typical railroad work schedules, an important component of a fatigue risk management system.

1. Introduction

1.1 Background

FAST implements the SAFTE biomathematical model of performance and fatigue to generate estimates of performance degradation owing to the individual's level of fatigue. Such models offer opportunities to predict a railroad worker's level of fatigue and resulting performance degradation, based on the individual's work schedule. Performance and fatigue models have two primary components: a biomathematical model of the relationship between sleep and performance capability and a social model of how sleep will be taken by the average person subjected to a specific work schedule that affords certain opportunities to sleep. The second component provides the sleep data that drives the average performance predictions from the first component, and inaccuracies in that component can seriously distort predictions of performance from the biomathematical component. AutoSleep, the sleep prediction component of FAST, was developed using data from a Federal Railroad Administration (FRA)-sponsored study, which in 1992 collected data from 150 railroad engineers (Pollard, 1996). Engineers used a daily diary to record sleep periods, commute time, work time, and time for personal activities. Subsequently, FRA sponsored a series of diary studies of four different groups of railroad employees. The data from these studies were compared with AutoSleep predictions from FAST. That study found for T&E workers 88 percent agreement between log-book–recorded sleep and AutoSleep estimates (Gertler & DiFiore, 2009; FRA, 2011). Because diary data is the participant's estimate of actual sleep, and thus may not be completely accurate, FRA decided to collect sleep and work data from a group of locomotive engineers using actigraphy. The availability of more current and more accurate sleep data will allow for further validation of the applicability of FAST to locomotive engineers.

1.2 Objectives

The purpose of the work described in this report was to collect work and sleep data from locomotive engineers by using actigraphy and daily log books and then to validate the AutoSleep sleep prediction model according to this data.

1.3 Overall Approach

Researchers collected actigraph records of sleep and wake periods along with daily log-book data from 41 locomotive engineers working in freight service. A comparison of sleep prediction results from a biomathematical model with this data provided the means to assess the accuracy and precision of the sleep prediction algorithm. The comparison was done on a minute-by-minute basis with 14 d of data for both the baseline model settings and revised settings that more closely reflected sleep patterns of locomotive engineers found in the prior diary study.

1.4 Scope

Unlike the prior diary study, this data collection and model validation effort did not attempt to characterize the work and sleep patterns of all U.S. locomotive engineers. It was not feasible to use a random sample of locomotive engineers for the collection of actigraph data. Instead, a sample of convenience with volunteer participants was used. Researchers used the data from these volunteers to determine the extent to which the AutoSleep sleep prediction algorithm, developed with data from another group of locomotive engineers, accurately predicted sleep.

1.5 Organization of the Report

Section 2 describes the methodology of the data collection. The analysis of the data is in Section 3, and Section 4 compares the actigraph-recorded sleep to FAST estimated sleep. Section 5 summarizes the overall conclusions of the study.

2. Data Collection

2.1 Methods

Collection of accurate sleep period data requires the use of actigraphy, a technique for recording sleep and wake periods with a device the size of a wrist watch (see Figure 1). The actigraph device must be worn at all times, including sleep, to obtain a valid record of actual sleep periods. Studies have shown that actigraphy is a valid method for assessing sleep durations and sleep/wake activity in healthy adults (Ancoli-Israel et al., 2003).

Figure 1. Actigraph (source: http://www.ambulatory-monitoring.com/products.html)

This study used actigraphy in conjunction with a daily log book for recording commute time, work periods, and personal time as well as sleep periods. Figure 2 contains a sample page from the Locomotive Engineer's Daily Log. Using vertical bars in the appropriate column, the engineer recorded his/her daily activities as sleep, personal time, commute to/from work, or work. Participant engineers recorded the location of their sleep as home or away and provided the actual start and end times for each work period. Through a brief background information form, study participants provided their age, sex, years of experience as a road freight engineer, and call time for home and away-from-home terminals.

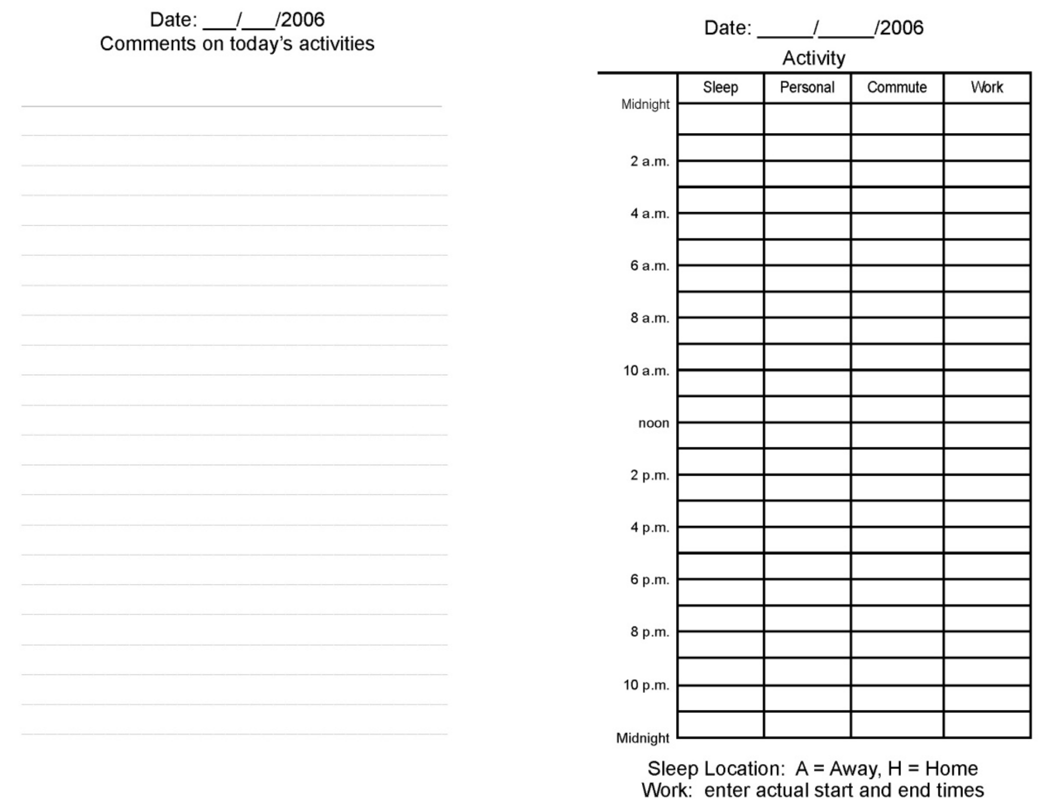
The New England Institutional Review Board approved the study protocol for this effort on February 7, 2006. The National Institutes of Health issued a Certificate of Confidentiality to Foster-Miller[1] on May 30, 2006. Data collection occurred from June through August 2006.

2.2 Participant Recruitment

Researchers recruited for a sample of convenience from engineers at UP locations in Colton, CA, and Roseville, CA, and BNSF Railway in Spokane, WA. Recruitment occurred in partnership with the labor organizations that represent locomotive engineers at these locations.

---

[1] Foster-Miller is now part of QinetiQ North America, Inc.



Date: ___/___/2006
Comments on today's activities

Date: _____/_____/2006

Activity

	Sleep	Personal	Commute	Work
Midnight				
2 a.m.				
4 a.m.				
6 a.m.				
8 a.m.				
10 a.m.				
noon				
2 p.m.				
4 p.m.				
6 p.m.				
8 p.m.				
10 p.m.				
Midnight				

Sleep Location: A = Away, H = Home
Work: enter actual start and end times

Figure 2. Sample Log-Book Page

The only selection criteria for study participants were at least 1 year (yr) of experience as a road freight engineer with untreated sleep disorder.

2.3 Procedure

A researcher was onsite to obtain informed consent from each participant, explain the use and care of the actigraph, and instruct the engineer on how to record his/her activities in the daily log. Two or 3 d after the start of the data collection, the researcher met with each study participant to check that the actigraph was functioning properly and to review the engineer's log-book entries for completeness. At the end of the data collection period, the researcher again met with each participant to collect the study materials and to compensate him/her for participating. Throughout the data collection period, the researcher was available via phone to answer questions and address any problems that might arise. All participants who wore the actigraph for 2 weeks and kept the daily log book for the same period received a $200 gift card to a national retail establishment upon return of the actigraph and daily log. Data collection occurred first in Colton, CA, then Roseville, CA, and finally Spokane, WA.

3. Data Analysis

The researchers used Microsoft Office Excel 2003 and SPSS 13.0 to analyze the log-book data and Action-W 2.5 and Excel 2007 to analyze the actigraphy data. (Ambulatory Monitoring, the vendor for the actigraphs, provides the Action-W software.) Action-W includes a means for the researcher to code the down, or time in bed, periods. The Action-W algorithm identifies sleep that occurs within the down periods. The researchers used the log-book entries to guide coding of the down periods. This Action-W algorithm identifies periods of low activity consistent with sleep during wake periods, but the data used in this study does not include these additional possible sleep periods unless the participant recorded it in the log book.

Sleep recorded by actigraph was compared with sleep predicted from the FAST AutoSleep function by using Microsoft Excel 2007. Both were coded into records including subject identifier, a date/time interval defined by start and end time, the subject's sleep status (awake or asleep), and the source for that status (actigraph or AutoSleep). The two data sets were combined and interleaved to define over 800,000 one-minute intervals during which each subject's status in both data sets could be compared. Four possible combinations of status are shown in Table 1. First, a measure of agreement (proportion correct) was computed consisting of the proportion of time intervals when both sources showed the participant in the same sleep state. This measure consisted of the hits and correct rejections. Then, with the signal detection theory (SDT), this data was analyzed to determine the extent and direction of the bias, or systematic errors, in AutoSleep and the sensitivity of AutoSleep to detect sleep periods.

Table 1. Outcome Matrix

		AutoSleep	
		Yes, asleep	**No, awake**
Actigraph	**Asleep**	Hit	Miss
	Awake	False Alarm	Correct Rejection

3.1 Study Participants

A total of 46 locomotive engineers participated in the study. Table 2 summarizes the distribution of participants by location and the validity of their data. Of the 46 participants, there was usable actigraph data from 41 participants. Three actigraph failures resulted in incomplete or invalid data from three participants. One participant had an allergic reaction to the nickel-coated backing on the actigraph and as a result could not complete the study. A fifth participant removed his actigraph on several nights so his data was not useful to the purposes of the study, which requires a full 14 d of data. The log books for 45 participants were complete. The individual who had the allergic reaction did not complete the log for 14 d, so this summary does not include this participant's information.

A total of 42 men and 3 women had usable log books. Their average age was 46 yr, and they had an average of 13 yr of experience as a locomotive engineer. The median years of experience was 11, indicating that the group was dominated by less experienced people.

Table 2. Summary of Data Obtained from Participant Population

Total number of participants		46
Colton	15	
Roseville	14	
Spokane	17	
Incomplete data		5
Actigraph failures	3	
Allergic reaction	1	
Removed actigraph	1	
Total complete actigraph data		41
Total complete log books		45

3.2 Work Characteristics

Work period characteristics for this group of locomotive engineers are based on data provided in the daily log books for all 45 study participants. In contrast, the actigraphy results are for the 41 participants from whom 14 d of actigraphy data was available. Table 3 summarizes the work period and related call and commute times for work originating at both home and away terminals. Labor agreements for each location specify the call times so there may be some differences in call time by location. For the three study locations, the majority of the participants had at home call times of 2 h and away times of 1.5 h. During the 2-week period of the study, these engineers averaged 7.7 work periods. The work periods averaged 9:37, and half were over 9:45. As Figure 3 illustrates, the distribution of work period lengths is bi-modal, with peaks at the 8–9 h and 11–12 h categories. These peaks may relate to the limitations of the Hours of Service Law. Three-quarters of the work periods were over 8 h. Approximately a quarter of the work periods exceeded 12 h as a result of deadheading time, which did not violate Hours of Service limitations in effect at the time of the study. Commute times when away from home tended to be shorter than those originating at home (see Figure 4).

Table 3. Summary of Work Period Characteristics (h:min)

	Mean	Median
Number of work periods	7.7	8
Work period duration	9:37	9:45
Commute time home	0:48	0:42
Commute time away	0:42	0:38
Call time home	1:45	2:00
Call time away	1:24	1:30

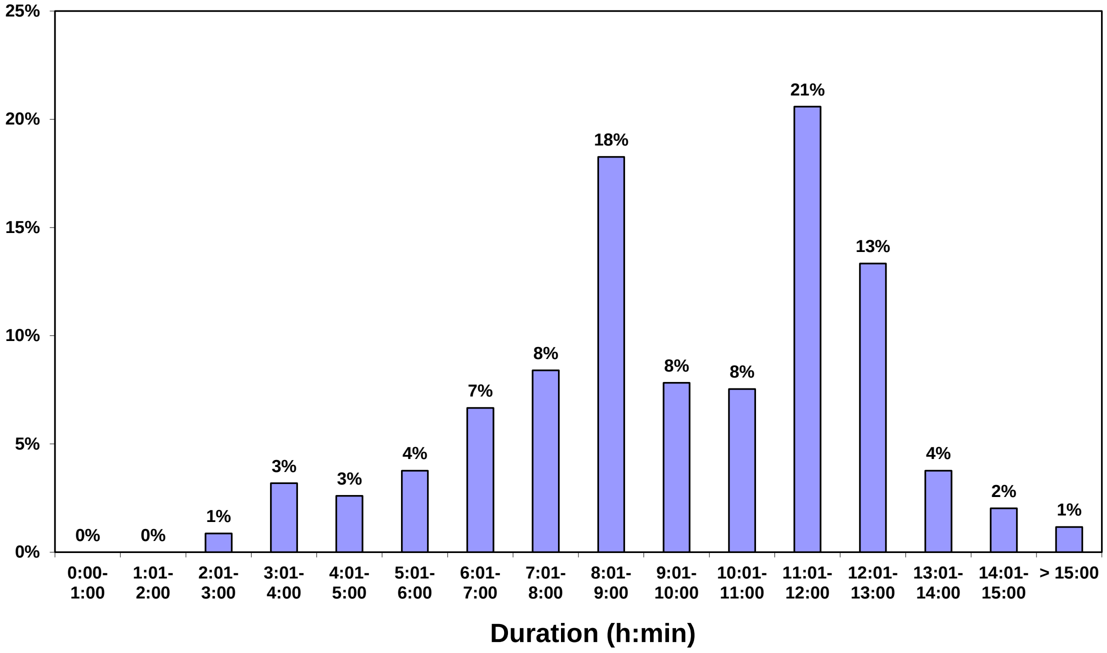


Figure 3. Duration of Work Periods

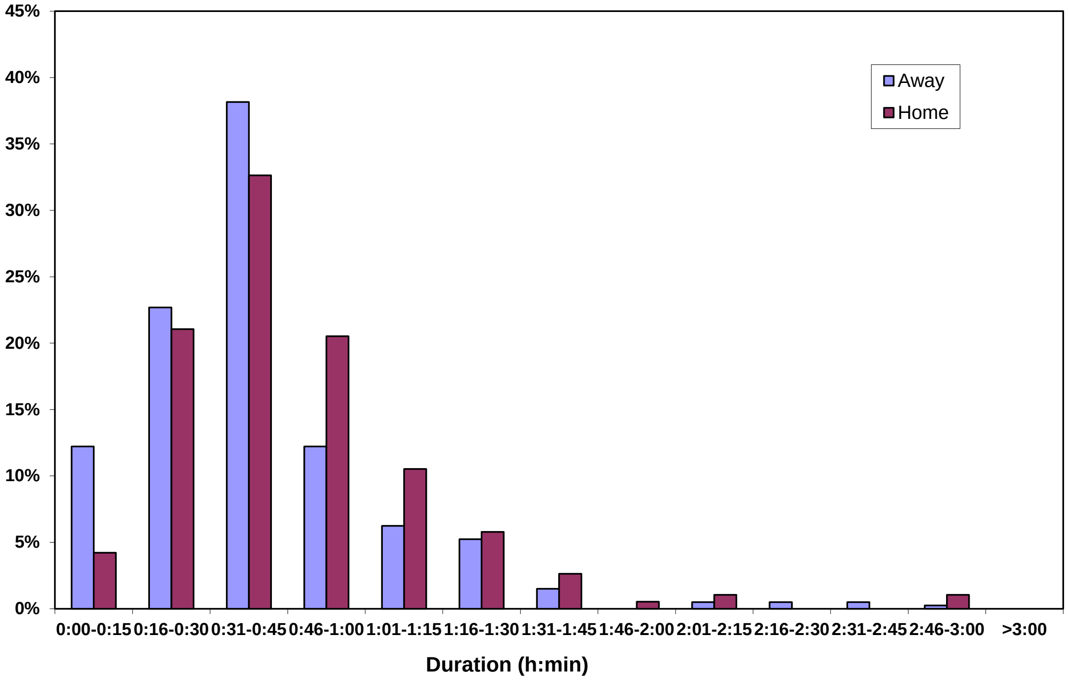


Figure 4. Commute Times

3.3 Sleep Data

Ambulatory Monitoring, the vendor for the actigraphs used in this study, provided the Action-W software for downloading and analyzing the actigraph data. Because AutoSleep is designed to predict sleep while in bed, the down period data from the actigraphs was of interest for the present study. The primary source of sleep data was the actigraphs, although the log books advised the researchers in coding the down periods in the actigraph data.

Instructions to study participants requested that they press the event button on their actigraph when they went to bed and when they woke up. These event markers along with the participant's daily diary assisted the researchers in coding the down periods. The Action-W software identified the beginning and end of each down period and each sleep period within the down period. These sleep periods within the down periods were then compared with the sleep periods predicted by AutoSleep.

The sleep analysis used those sleep times for which both actigraph and AutoSleep data existed. Because the start and end of sleep did not generally fall at midnight, the hours of sleep per 24 h were computed as the fraction of time asleep during the 2-week study interval. Sleep per 24 h is shown as both grand mean (based on all time summed across all subjects that the subject was asleep) and the mean of subjects (sum of mean sleep of each subject divided by the number of subjects). Table 4 presents the overall statistics for the actigraphy sleep data.

Table 4. Actigraph Sleep per 24 h (h:min)

Grand Mean	6:57
Mean of Subjects	6:58
Standard Deviation	0:49
Median	6:54
25th percentile	6:28
75th percentile	7:26

Figure 5 displays the distribution of sleep per 24 h for the actigraph data as well as two AutoSleep cases. The bars in Figure 5 represent the predictions of AutoSleep with two different settings of the sleep parameters and the red line is the actigraph data. Overall, the results from the revised AutoSleep settings more closely approximate the distribution of the actigraph results.

Analysis of the data in Figure 5 indicates that the distribution for the baseline AutoSleep settings (blue bars) was statistically different from the actigraphy results, χ^2 (1, N = 41) = 4.89, $p < 0.05$ but for the revised AutoSleep settings (green bars) there was no statistical difference, χ^2 (1, N = 41) = 0.05, $p > 0.05$.

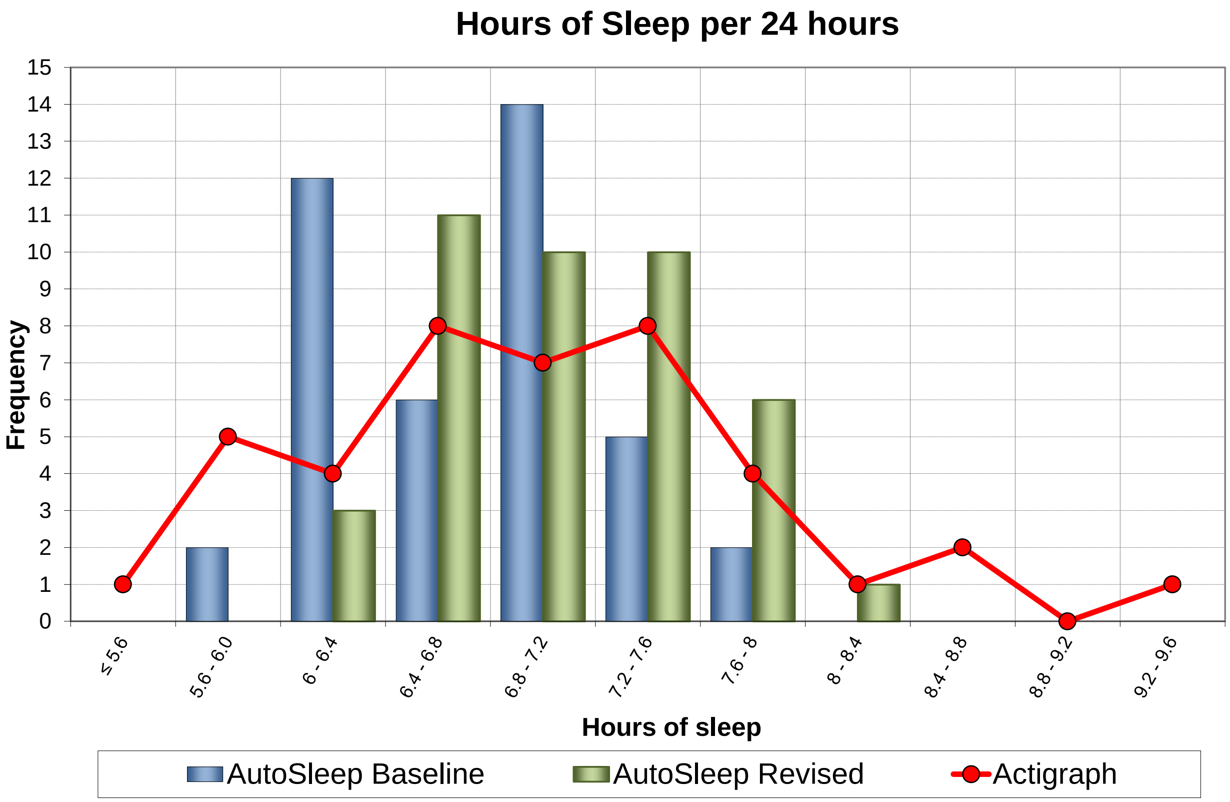


Figure 5. Distribution of Sleep per 24 h across Subjects

4. Comparison of AutoSleep Predictions with Actigraphy Data

Actigraphy data and AutoSleep predictions were compared for the percentage of time that both actigraphy and AutoSleep showed the subject in the same state (sleep or awake) and the mean sleep per subject in 24 h. The AutoSleep predicted sleep state (i.e., awake or asleep) of each participants was compared with the actigraph data on a minute-by-minute basis to determine the proportion of time that the two were in agreement. In addition, the mean sleep in 24 h for AutoSleep was compared with that for the actigraphy data.

AutoSleep requires specification of a normal bedtime, the maximum number of hours of sleep during a workday or a rest day, and the start and end times of a forbidden zone when no sleep is allowed, typically in the afternoon. Prior to this study, the baseline AutoSleep settings assumed a normal bedtime of 10 p.m., a maximum of 8 h sleep during workdays and rest days, and a forbidden zone from 12 noon to 8 p.m. With the FAST model parameters set to these values, estimated sleep patterns agreed with those recorded by actigraphs slightly less than 85 percent of the time, with a standard deviation among subjects of 4.42 percent, but the estimated amount of sleep per 24 h was 13 min less than that recorded by actigraphy. Given the inaccuracy of the baseline AutoSleep settings, a revised set of settings were applied to better predict the actigraph sleep data. The results from the comparison of AutoSleep with locomotive engineer diary data advised the selection of this revised set of AutoSleep settings (Gertler & DiFiore, 2009).

Table 5 contains the settings for the revision and the results for agreement of the FAST estimates with the actigraph data. Table 6 contains the resulting estimates of sleep. The revision has a standard bedtime of 11 p.m., allows more sleep on rest days, up to a maximum of 8.5 h, and has a narrower forbidden zone of 1–7 p.m. (dark green zone in Figure 6). The revised AutoSleep results agreed with actigraphy 87 percent of the time with a smaller standard deviation among subjects of 3.02 percent. The revised settings more accurately predicted the state (sleep vs. awake) of the individual and mean sleep per 24 h slightly exceeded those from actigraphy by 6 min. Figure 6 indicates the distribution of sleep across the 24-hour day aggregated across subjects. The red line is the sleep as recorded by actigraphy. The blue bars show the distribution of sleep according to the baseline settings.

Comparing the distribution of sleep from actigraph records with AutoSleep estimates using the baseline settings shows that the model starts sleep periods too early in the evening and ends them too early in the morning. Hence, alternative settings of the AutoSleep function were tested, referred to as Revised Settings. The distribution for the Revised Settings is closer to that of the actigraphy. The green bars in Figure 6 show the distribution of sleep, based on the Revised Settings. The settings for this case produced a pattern across the day more similar to that recorded by actigraphy. Perfect agreement is not possible with a population-based model. The model does not reflect differences among individual subjects under similar scheduling conditions, so the agreement achieved may be close to optimal given the variance between subjects in preferred bedtime, amount of daily sleep needed, and inclination to sleep in the afternoon during the forbidden zone.

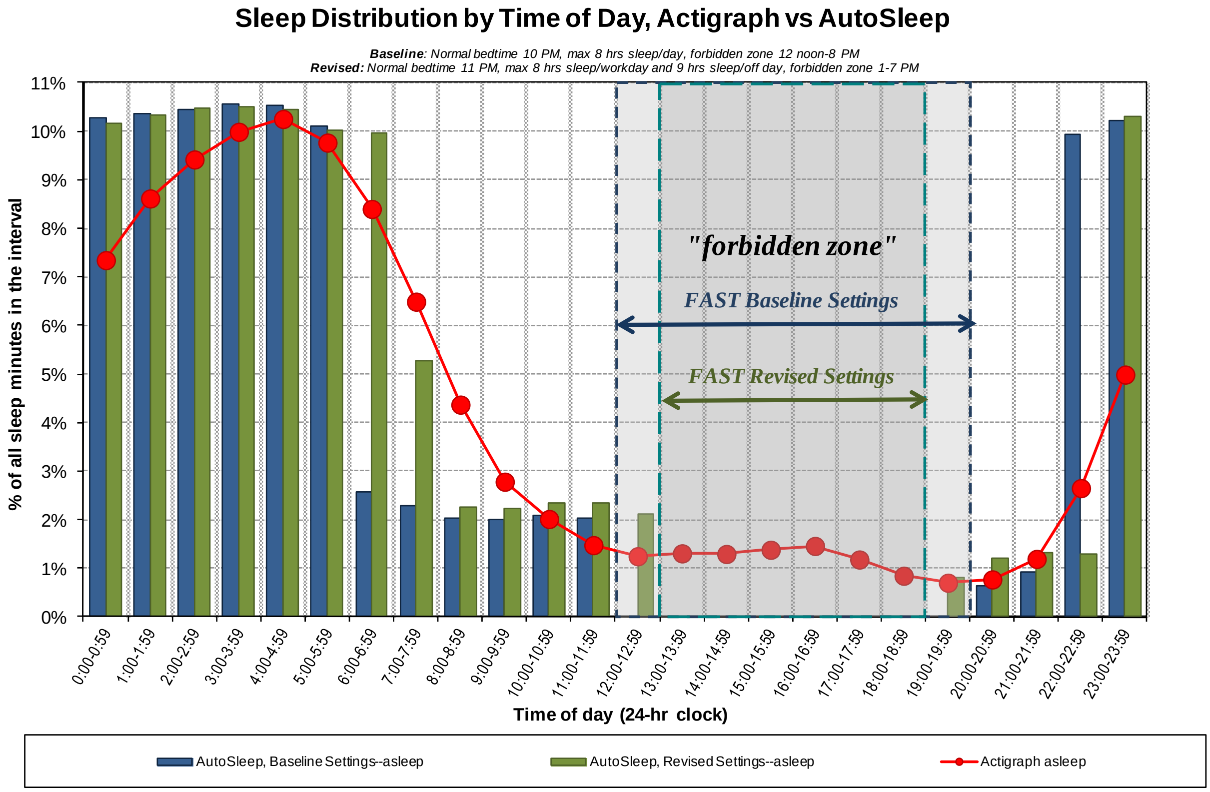


Figure 6. Distribution of Percentage of Sleep as a Function of Time of Day

Table 5. AutoSleep Settings and Percent Agreement with Actigraph Sleep

	AutoSleep Baseline Settings	**AutoSleep Revision Settings**
AutoSleep Settings		
Default bedtime	10 p.m.	11 p.m.
Maximum hours sleep, workday	8	8
Maximum hours sleep, rest day	8	8.5
Forbidden zone start hour	12 p.m.	1 p.m.
Forbidden zone stop hour	8 p.m.	7 p.m.
Results: Percent agreement		
Grand mean	84.65%	86.92%
Mean of subjects	84.65%	86.94%
Standard deviation	4.42%	3.02%
Median	85.27%	87.52%
25th percentile	81.84%	85.46%
75th percentile	87.42%	89.01%

The 95% confidence intervals for each of the estimates of mean sleep are shown in Table 6 and plotted in Figure 7. The confidence intervals for the results from the baseline and revised settings overlap with those for the actigraph data, indicating that they are not statistically different.

Table 6. Accuracy of Sleep Estimates per 24 h Recorded by Actigraph and Estimated by AutoSleep (h:min)

Sleep Metric	Actigraph	AutoSleep Baseline	AutoSleep Revision
Grand mean	6:57	6:45	7:04
Mean of subjects	6:58	6:45	7:04
[95% confidence interval]	[6:43, 7:13]	[6:35, 6:55]	[6:54, 7:14]
Standard deviation	0:49	0:30	0:31
Median	6:54	6:50	7:07
Minimum	5:29	5:54	6:04
Maximum	9:14	7:52	8:00
25th percentile	6:28	6:21	6:34
75th percentile	7:26	7:01	7:25

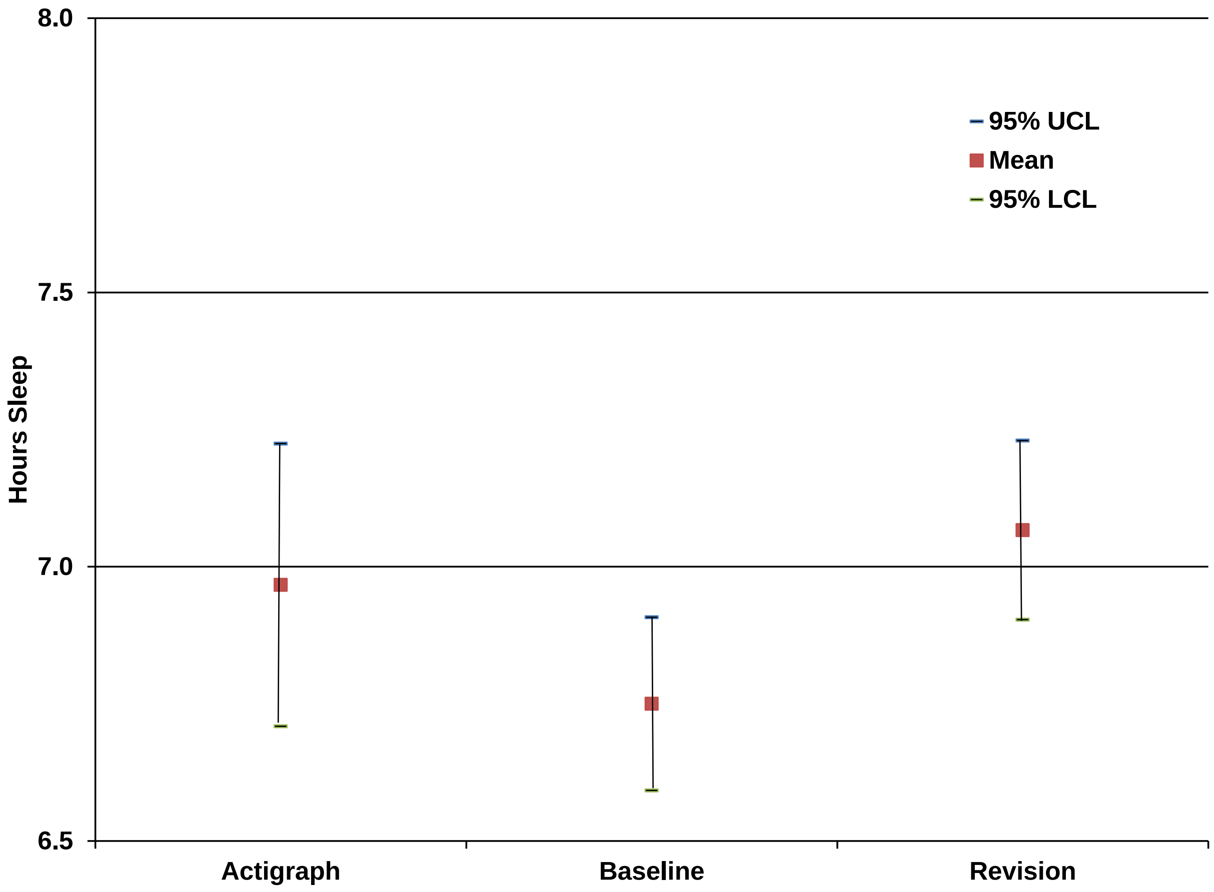


Figure 7. 95% Confidence Intervals for Sleep in 24 h

The frequency data from the AutoSleep/actigraphy comparison was converted to a 2 × 2 probability table representing the four possible outcomes of Table 1 and analyzed by using SDT. (See Appendix A for frequency data used in SDT analyses.) SDT offers a methodology for examining the direction and magnitude of the bias and sensitivity of the AutoSleep estimates. In SDT, bias (β) and sensitivity (*d'*) are independent. This is not the case for other measures such as percent correct in which accuracy is confounded with tendencies to select one alternative over another. Consequently, SDT provides a measure of ability to detect a condition that is not influenced by response bias.

SDT is used in situations in which decisionmaking is uncertain. In the present case, the AutoSleep algorithm is presented with information concerning when a person is working and commuting. On the basis of this information and parameters that are set in the algorithm, a decision is made concerning the sleep/wake status of that person during a time period. The decision is based on a score that has a probability of being associated with a status of being asleep and of being awake. In other words, the probability distribution of algorithm scores for being awake overlaps the probability distribution of algorithm scores for being asleep. This overlap is, in part, due to uncertainty concerning individual characteristics, such as daily bedtime and the amount of sleep needed per day. Consequently, any decision based on the algorithm has two probabilities of being correct (probabilities of Hits [p(Hit), Autosleep says the person is asleep when actigraph says the person is asleep], and Correct Rejections[p(CR), Autosleep says the person is awake when actigraph says the person is awake]) and two probabilities of being incorrect (probabilities of False Alarms [p(FA), Autosleep says the person is asleep when actigraph says the person is awake], and Misses[p(Miss)], Autosleep says the person is awake when actigraph says the person is asleep). The outcome matrix in Table 1 describes this. It should be noted that the actigraph is considered the “state of the world” in this case, and p(Hit) + p(Miss) = 1 because they both are derived from the “asleep” distribution. Likewise, p(FA) + p(CR) = 1 because they are both derived from the “awake” distribution.

The probabilities of hits and false alarms are used to calculate *d’*, a measure of sensitivity that is based on the standardized difference between the means of the “awake” and “asleep” distributions. The algorithm makes status decisions by setting a criterion based on the likelihood of being asleep. This likelihood is determined, in part, by the ratio of the probability of being awake [p(awake)] and the probability of being asleep [p(asleep)]. These two probabilities are called the prior probabilities, and their ratio[p(awake)/p(asleep)] determines bias to respond “awake” versus “asleep.” The setting of the criterion divides the “awake” and “asleep” distributions each into two parts, as Table 1 illustrates. Regardless of where the criterion is placed, *d’* remains the same because it only depends on the difference of the means of the “awake” and “asleep” distributions of AutoSleep. *d’* changes only when the distributions change. This is the sense in which bias and sensitivity are independent in SDT. (Green and Swets (1966) provides a detailed discussion of the principles and application of SDT.)

Table 7 presents the SDT results. The fraction of cases in which both actigraphy and AutoSleep indicated sleep is p(Hit), and the fraction of cases in which the actigraphy indicated that the individual was awake but AutoSleep indicated sleep is p(FA). The p(Hit) and p(FA) values are plotted in the receiver operating characteristic (ROC) curve of Figure 8. Note that the distance of the points from the major diagonal (the line that extends between (0,0) and (1,1)) is equivalent to *d’*, and this is consistent with the numerical values in Table 7. The minor diagonal (the line that extends from (0,1) to (0.5, 0.5)) is the isobias line ($\beta = 1$). Points above the minor diagonal indicate a bias to say “asleep” ($\beta < 1$), whereas points below the line indicate a bias to say “awake” ($\beta > 1$). Distance from the minor diagonal indicates more bias, again consistent with Table 7.

Table 7. SDT Results

Case	p(Hit)	p(FA)	β [95% CI]	p(awake)/ p(asleep)	*d´* [95% CI]
AutoSleep Baseline	0.720	0.102	1.89 [1.87, 1.91]	2.56	1.85 [1.85, 1.86]
AutoSleep Revision	0.782	0.095	1.74 [1.73, 1.76]	2.40	2.09 [2.08, 2.10]

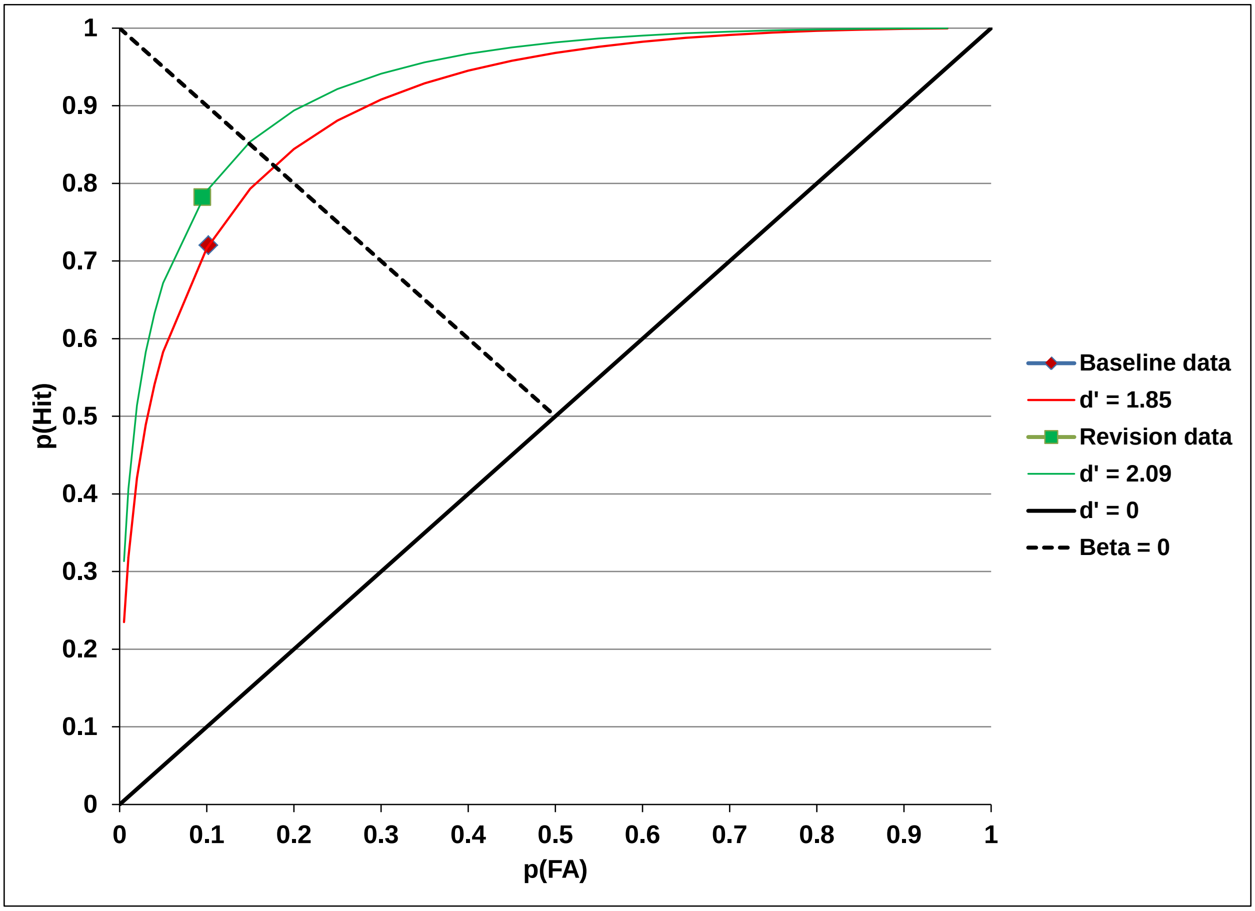


Figure 8. p(Hit) versus p(False Alarm) ROC Curve

A β > 1 indicates a bias for AutoSleep to decide "awake" when actigraphy indicates "asleep." In both the baseline and revision cases, β was close to 2. This result is consistent with mean sleep in 24 h, as predicted by AutoSleep, being less than the actigraphy results. When the AutoSleep parameters were changed to reduce the afternoon/forbidden zone and increase allowable sleep on rest days (revision), p(Hit) increased with only a negligible decrease in p(FA). At the same time, the bias (β) toward predicting awake state was reduced from 1.89 to 1.74, and *d´* increased. As noted above, *d'* only changes when changes occur in the distributions of "awake" and "asleep" from AutoSleep. Changing the forbidden zone, allowable sleep on rest days, and bedtime would logically change the "awake" and "asleep" distributions of AutoSleep scores to be more consistent with actigraphy. Moreover, the revision of AutoSleep parameters also changes the ratio p(awake)/p(asleep) toward slightly more sleep than with the baseline settings. This is shown in Table 7. Given that the great overall probability of awake than sleep in any given day and that the ratio is about 2:1 in favor to awake, one would expect that bias (β) would be close to 2. Furthermore, when the AutoSleep settings were changed in the revision to increase maximum rest day sleep, this should reduce the ratio p(awake)/p(asleep) and, likewise, reduce bias for awake. Both of these outcomes were confirmed in Table 7.

5. Conclusions

With data from 41 locomotive engineers, AutoSleep, the sleep estimation algorithm in FAST, provided accurate predictions of actigraph-recorded sleep with baseline settings but modifications to the default bedtime and the maximum allowed sleep per day on rest days resulted in a more accurate prediction. With the revised AutoSleep settings, overall congruence between actigraph sleep and estimated sleep was 87 percent, and the AutoSleep predicted sleep per day exceeded actigraph sleep by 6 min. The estimates of mean sleep in 24 h for both the baseline and revised AutoSleep parameters were not statistically different from the actigraph results.

Complete agreement with the actual sleep patterns of all 41 participants is impossible to achieve with a model that has a single set of parameter settings. There are actually two sources of errors, and the settings of the model only reduce one of those error types. One source of error is the result of inaccurate prediction of the average subject. An example of this kind of error is the setting for the width of the forbidden zone. The revisions of the baseline model reduced this error by allowing sleep later into the afternoon and earlier in the evening across subjects. The second source of error is differences between subjects, even when experiencing the same set of work conditions. An example of this kind of error is the setting of the typical bedtime. Some subjects prefer to retire early, around 10 p.m., whereas others prefer a bedtime of 11 p.m. or later. Any model with a single setting for bedtime will be inaccurate for those subjects with a different preferred bedtime, unless the model was individualized for individual preferences. For the average person, however, the sleep estimates closely reflect both the pattern and amounts of sleep measured by actigraphy, providing a sound basis for predicting performance and fatigue.

The implication of this study is that fatigue assessments associated with T&E work schedules using FAST are based on valid expectations of average sleep patterns and, therefore, provide a reasonable estimate of sleep restriction and associated fatigue risk. These findings further validate the utility of FAST for assessing fatigue risk created by typical railroad work schedules, an important component of a fatigue risk management system.

6. References

Ancoli-Israel, S., Cole, R., Alessi, C., Chambers, M., Moorcroft, W., & Pollack, C. (2003). The role of actrigraphy in the study of sleep and circadian rhythms. *SLEEP*, *26*(3), 342-392.

Federal Railroad Administration. (2011). *Measurement and estimation of sleep in railroad worker employees* (RR11-02). Retrieved August 12, 2011, from http://www.fra.dot.gov/rpd/downloads/RR_Measurement_and_estimation_of_Sleep_final_1102.pdf.

Gertler, J., & DiFiore, A. (2009). *Work schedules and sleep patterns of railroad train and engine service workers* (Report No. DOT/FRA/ORD-09/22). Washington, DC: Federal Railroad Administration. (Available at http://www.fra.dot.gov/downloads/Research/ord0902.pdf).

Green, D. C., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Pollard, J. K. (1996). *Locomotive engineer's activity diary* (Report No. DOT/FRA/RRP-96/02). Washington, DC: Federal Railroad Administration. (Available at http://www.fra.dot.gov/downloads/Research/rrp9602.pdf).

Appendix A.
Frequency Counts

AutoSleep Baseline

		AutoSleep Yes, Asleep	No, Awake	Total
Actigraph	Asleep	171,737	66,707	238,444
	Awake	59,568	524,544	584,112
	Total			822,556

AutoSleep Revision

		AutoSleep Yes, Asleep	No, Awake	Total
Actigraph	Asleep	186,379	52,065	238,444
	Awake	55,561	528,551	584,112
	Total			822,556

Abbreviations and Acronyms

BLET	Brotherhood of Locomotive Engineers and Trainmen
d	day(s)
FAST	Fatigue Avoidance Scheduling Tool
FRA	Federal Railroad Administration
h	hour(s)
min	minute(s)
QNA	QinetiQ North America
ROC	receiver operating characteristic
SAFTE	Sleep, activity, Fatigue, and Task Effectiveness
T&E	train and engine service
UP	Union Pacific Railroad
yr	year