DOT/FRA/ORD – 13/33
Docket No. FRA-2009-0043-0003

CRITERIA AND PROCEDURES FOR VALIDATING BIOMATHEMATICAL MODELS OF HUMAN PERFORMANCE AND FATIGUE:

PROCEDURES FOR ANALYSIS OF WORK SCHEDULES

(Available at http://www.regulations.gov/#!documentDetail;D=FRA-2009-0043-0003)

By Dr. Thomas Raslear

Each railroad covered by 49 CFR 228.407 must perform an analysis of the work schedules of its train employees who are engaged in commuter or intercity rail passenger transportation and identify those schedules that, if worked by such a train employee, may be at risk for a level of fatigue at which safety may be compromised. A level of fatigue at which safety may be compromised, the fatigue threshold, shall be determined by procedures that use a scientifically valid, biomathematical model of human performance and fatigue. This document describes the criteria and procedures that the Federal Railroad Administration (FRA) will use for certifying the scientific validity of a biomathematical model and for determining a fatigue threshold in a biomathematical model. This document also identifies specific such models that are acceptable. Furthermore, it describes the procedures that must be used both to analyze work schedules and to report FRA those schedules that are at risk for exceeding the fatigue threshold.

A. What is a scientifically valid, biomathematical model of human performance and fatigue?

FRA will decide that a model of human performance and fatigue is scientifically valid and may be used for the analysis of work schedules as required by 49 CFR 228.407, if the following two objective criteria are met:

Criterion 1. The model documents sensitivity to circadian, sleep deprivation, sleep recovery, and sleep inertia effects on one or more well-known behavioral or performance-based indicators of fatigue, including vigilance speed, reaction time, lapses, cognitive throughput, alertness, and tendency to fall asleep, similar to that observed in the sleep and fatigue literature as noted below.

Criterion 2. The model is able to demonstrate sensitivity to rail operations human factors accident risk by--

(a) using the Fatigue Accident Validation (FAV) Database to contrast human factor (HF) and nonhuman factor (NHF) accidents (http://www.fra.dot.gov/rpd/policy/1975.shtml);

(b) demonstrating that HF accident risk is a statistically reliable function of modeled performance; and

(c) demonstrating that NHF accident risk is independent of modeled performance.

Criterion 1 requires that a model of human performance and fatigue (model) is demonstrated to be consistent with currently established science in the area of human performance, sleep, and fatigue. The scientific literature in this area has documented that certain patterns of work and/or sleep (model inputs) have known patterns of effects on behavioral or performance-based indicators of fatigue (model outputs). Consequently, model inputs such as—

- the length of time spent working and/or sleeping over long and short periods of time (chronic and acute sleep deprivation or restriction),
- the time of day that work and/or sleep occurs (circadian rhythms), and
- abrupt changes in the time of day that work and/or sleep occurs (phase changes)–

should affect model outputs such as—

- vigilance speed,
- reaction time,
- lapses,
- cognitive throughput,
- alertness, and
- tendency to fall asleep.

In particular, any model should be able to demonstrate that appropriate model inputs result in acute and chronic sleep deprivation/restriction effects, circadian and phase adjustment effects, sleep recovery effects, and sleep inertia effects with regard to one or more of these model outputs. Fidelity to the pattern of time and magnitude of these effects, as documented in the scientific literature, is a requirement of Criterion 1.

There are currently six scientific models that allow work schedules to be evaluated for the effects of fatigue on performance and alertness:

- Sleep, Activity, Fatigue and Task Effectiveness model [SAFTE (Hursh et al., 2004)];
- Fatigue Audit InterDyne model [FAID (Roach, Fletcher, and Dawson, 2004)];
- Three-Process model (Akerstedt, Folkard, and Portin, 2004);
- System for Aircrew Fatigue Evaluation [SAFE (Belyavin and Spencer, 2004)];
- Two-Process model (Achermann, 2004); and
- Circadian Alertness Simulator [CAS (Moore-Ede, Heitmann, Guttkuhn, Trutschel, Aguirre, and Croke, 2004)].

Each of these models has demonstrated sensitivity to circadian, sleep deprivation, sleep recovery, and sleep inertia effects on one or more well-known behavioral or performance-based indicators of fatigue, including reaction time, cognitive throughput, lapses, alertness, and tendency to fall asleep (Balkin, Braun, and Wesensten, 2002; Balkin et al., 2000; Bonnet, 1997; Carskadon and Dement, 1977; Dinges, Orne, and Orne, 1985; Dinges and Powell, 1985; Dinges and Powell, 1989; Folkard and Akerstedt, 1987, 1992; Froberg, 1977; Harrison and Horne, 1996; Jewett, 1997; Jewett and Kronauer, 1999; Lumley, Roehrs, and Zorick, 1986; Mitler, Gujavarty, Sampson, and Bowman, 1982; Monk, 1991; Monk and Embry, 1981; Richardson, Carskadon, and Flagg, 1978; Thorne, Genser, Sing, and Hegge, 1983; Wesensten, Balkin, and Belenky, 1999). These models were recognized as adequate representations of the effects of fatigue on human performance by inclusion in a 2002 workshop cosponsored by the U.S. Department of Transportation on fatigue and performance modeling (Neri, 2004) and are recognized by FRA as satisfying Criterion 1.

Other, equally adequate models of fatigue and performance may have been developed since 2002 or were inadvertently not included in the 2002 workshop. To satisfy Criterion 1, other models should present evidence to FRA in sufficient detail to prove that the model is consistent with the scientific literature on human performance and sleep/fatigue as noted above.

Criterion 1 is considered a minimal requirement for a model to be qualified for in the purpose of work schedule analysis as required by 49 CFR 228.407. Failure to satisfy Criterion 1 disqualifies that model, and eliminates that model from any further consideration.

Criterion 2 requires that any model of fatigue and performance (including the six models referenced above), to be qualified for use in evaluating work schedules under 49 CFR 228.407, demonstrates sensitivity to the risk of a railroad accident in which an HF is a contributing cause. Satisfying Criterion 2 will entail use of the FAV Database to contrast HF and NHF accidents [http://www.fra.dot.gov/rpd/policy/1975.shtml] (1) to demonstrate that HF accident risk is a statistically reliable function of modeled performance and (2) to demonstrate that NHF accident risk is independent of modeled performance.

FAV serves as the input for models seeking to satisfy Criterion 2. FAV is available as a Microsoft Excel file on the FRA Web site at http://www.fra.dot.gov/rpd/policy/1975.shtml. FAV contains 30-day work histories for locomotive crews prior to involvement in approximately 400 HF accidents and 1,000 NHF accidents. The data represent the work histories of 2,673[1] crewmembers, of whom 732 were involved in HF accidents and 1,944 were involved in NHF accidents.

Models seeking to satisfy Criterion 2 should use the FAV as input to the model to determine a level of performance (model output) at the time of each accident for each employee indicating that the employee was--

- not fatigued (well-rested), or
- mildly fatigued, or
- moderately fatigued, or
- very fatigued, or
- extremely fatigued, or
- severely fatigued.

The performance bin “not fatigued” is determined by the output of the model when sleep occurs or can occur for 8 or more hours, without abrupt phase changes, during the circadian trough between 2200 hours (h) (10 p.m.) and 1000 h (10 a.m.). This is similar to the amount of fatigue produced by the standard 9 a.m.-5 p.m., Monday-Friday work week. The performance bin “severely fatigued” is determined by the output of the model when there is total sleep deprivation for 42.5 consecutive hours. This is similar to the amount of fatigue produced by a permanent night shift schedule with six consecutive 12-h work periods, followed by 1 day off. These two bins are the “anchor” bins for the model. There may be other operational definitions of “not fatigued” and “severely fatigued” that are particularly suited to a model. The use of other operational definitions must be approved by FRA prior to use in validating and calibrating a model.

---

[1] Three employees had two NHF accidents.

The method to be used is described in detail in three FRA reports (Hursh et al., 2006, 2008; Tabak and Raslear, 2010):

STEP A1. The model domain (i.e., the values that the model output can have) is divided into six bins of modeled fatigue levels. One bin is used to capture individuals with severe levels of fatigue. Another bin is used to capture all individuals considered not fatigued (but instead well-rested) by the model. For example, if Model X can have fatigue values that range from 50 to 150, with scores between 50 and 60 representing individuals who are well-rested and with scores greater than 100 representing individuals who are severely fatigued, the bins would be as follows:

Not Fatigued	Mildly Fatigued	Moderately Fatigued	Very Fatigued	Extremely Fatigued	Severely Fatigued
<60	61-70	71-80	81-90	91-100	>100

For each accident in FAV, the amount of time (in hours) that each employee involved in the accident spends in each bin is calculated from the employee's work history. The total time for all employees in each bin is calculated for all accidents. A distribution of the proportion (fraction in decimals or percentages) of employee time as a function of modeled fatigue is constructed for all accidents. This represents the modeled fatigue exposure for all accidents.

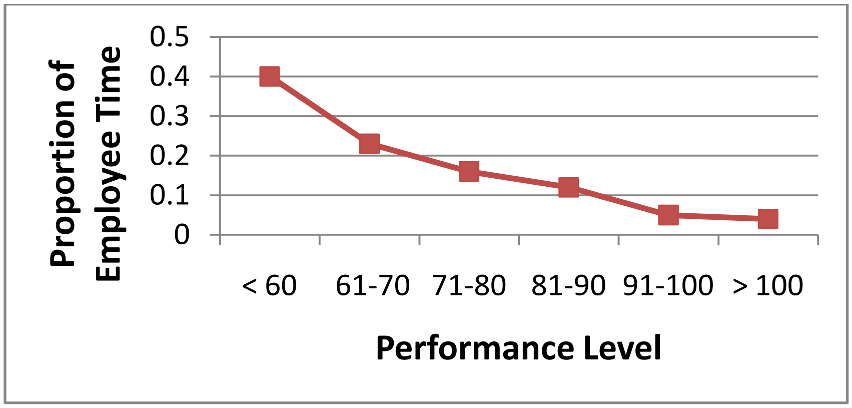
EXAMPLE OF STEP A1: FAV has data for 528,782 h of work for the 2,673 employees. The FAV data is processed by Model X with the following result:

Employee Work Time in Hours as a Function of Modeled Performance Level

Modeled Performance Level						
<60	61-70	71-80	81-90	91-100	>100	Total
211,513	121,620	84,605	63,454	26,439	21,151	528,782

Proportion of Employee (Work) Time as a Function of Modeled Performance Level

Modeled Performance Level					
<60	61-70	71-80	81-90	91-100	>100
0.4	0.23	0.16	0.12	0.05	0.04



STEP A2. For each accident, the modeled fatigue at the time of the accident is calculated for each crewmember. The total number of crewmembers involved in HF and NHF accidents in each of the six contiguous modeled fatigue bins is calculated for the 2,673 crewmembers. A distribution of proportion (or fraction in decimals or percentages) of crewmember accidents as a function of modeled fatigue is constructed for HF and NHF accidents.

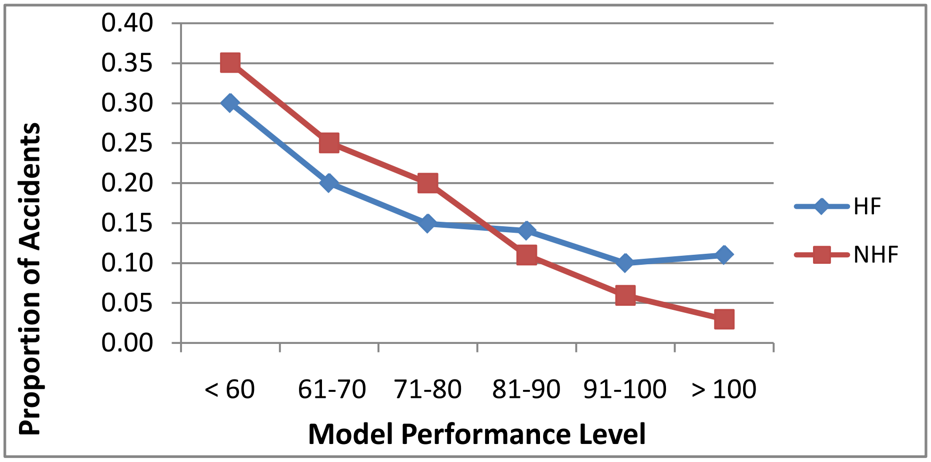
EXAMPLE OF STEP A2: FAV data, processed by Model X produces the following result:

Number of Crewmember Accidents as a Function of Modeled Performance Level

	Modeled Performance Level						
	<60	61-70	71-80	81-90	91-100	>100	Total
HF Accidents	220	146	110	102	73	81	732
NHF Accidents	680	486	389	214	117	58	1,944

Proportion of Accidents as a Function of Modeled Performance Level

	Modeled Performance Level					
	<60	61-70	71-80	81-90	91-100	>100
HF Accidents	0.30	0.20	0.15	0.14	0.10	0.11
NHF Accidents	0.35	0.25	0.20	0.11	0.06	0.03



STEP A3. The Accident Risk Ratio is calculated for HF and NHF accidents for each modeled fatigue level.

Accident Risk Ratio is defined as follows:

$$\frac{\text{Proportion of accidents in fatigue bin } x}{\text{Proportion of employee work time in fatigue bin } x}$$

EXAMPLE OF STEP A3: Results for Model X

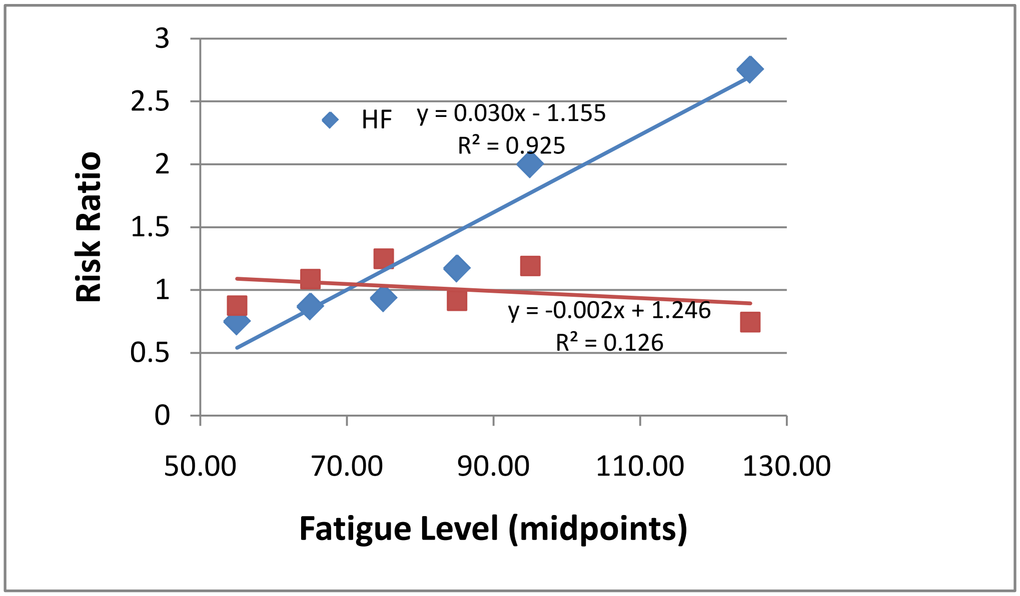
	Modeled Performance Level					
	<60	61-70	71-80	81-90	91-100	>100
Proportion						
HF Accidents	0.30	0.20	0.15	0.14	0.10	0.11
NHF Accidents	0.35	0.25	0.20	0.11	0.06	0.03
WORK TIME	0.40	0.23	0.16	0.12	0.05	0.04
Risk Ratio						
HF Accidents	0.75	0.87	0.93	1.17	2.00	2.75
NHF Accidents	0.88	1.09	1.25	0.92	1.19	0.75

STEP A4. Determine whether the model demonstrates a statistically reliable relationship between the Accident Risk Ratio and modeled fatigue level for HF accidents and whether the model demonstrates that the NHF Accident Risk Ratio is statistically independent of modeled fatigue level.

EXAMPLE OF STEP A4: Correlation coefficients are calculated for (1) the HF Risk Ratio and modeled performance level and (2) the NHF Risk Ratio and modeled performance level. Since the modeled performance is binned, midpoints for each bin should be used.

Performance Level (midpoints)	HF Risk Ratio	NHF Risk Ratio
55	0.75	0.88
65	0.87	1.09
75	0.93	1.25
85	1.17	0.92
95	2.00	1.19
125	2.75	0.75

An Excel spreadsheet can be used to compute the correlations. For HF accidents, the correlation is 0.9621. For NHF accidents, the correlation is -0.3556. Since there are six pairs (n) of modeled performance and risk ratios for each correlation, the number of degrees of freedom for a t-test to determine the statistical significance of each correlation is four (n-2). For 4 degrees of freedom, the absolute value of a correlation must be at least 0.811 to be statistically significant with a probability of 0.05. The figure below shows a scatter plot of the six pairs of HF and NHF data. A least-squares linear regression line is also shown for HF and NHF accidents, along with the equation for each line and the value of the squared correlation coefficient. Consequently, the computations demonstrate that a statistically reliable relationship exists between the Accident Risk Ratio and modeled fatigue level for HF accidents and that no statistically reliable relationship exists between the Accident Risk Ratio and modeled fatigue level for NHF accidents (i.e., Accident Risk Ratio and modeled fatigue level for NHF accidents are statistically independent). This means that Model X satisfies Criterion 2 for use as a scientifically valid model of human performance and fatigue.



Any introductory textbook on statistics will have information on correlation and related t-tests of significance (e.g., Hays, 1963). It should be noted that this example assumes that fatigue level and risk are linearly related. If inspection of the data, plotted as above, does not support a linear relationship, nonlinear correlation and regression methods should be used.

STEP A5. Calculate Cumulative Risk for HF accidents as a function of modeled fatigue criterion level. Calculate mean Cumulative Risk for NHF accidents

EXAMPLE OF STEP A5: Step A5 uses information from Steps A1 and A2 in which the proportion of employee time and proportion of HF and NHF accidents was calculated. Six bins are again arranged as shown below. Note, however, that the bins are now partitioned between the category "not fatigued" (<60) and any level of fatigue (>60). In the table below, 40 percent of the employees' time is "not fatigued," and 60 percent of employees' time has some fatigue ranging from "mildly fatigued" to "severely fatigued." Cumulative Risk is calculated similarly to the Accident Risk Ratio. Cumulative Risk is defined as

$$\frac{\text{Cumulative proportion of accidents in fatigue bin } x}{\text{Cumulative proportion of employee work time in fatigue bin } x}$$

Modeled Performance Level

	<60	>60	>70	>80	>90	>100
Cumulative Proportion						
Employee Time	0.40	0.60	0.37	0.21	0.09	0.04
HF Accidents	0.30	0.70	0.50	0.35	0.21	0.11
NHF Accidents	0.35	0.65	0.40	0.20	0.09	0.03
Cumulative Risk						
HF Accidents	0.75	1.17	1.35	1.67	2.33	2.75
NHF Accidents	0.88	1.08	1.08	0.95	1.00	0.75
NHF Mean			0.96			

STEP A6. Calculate 95 percent confidence intervals for each Cumulative Risk bin, determine Cumulative Risk bin in which HF cumulative risk exceeds both HF Accident Risk Ratio = 1 and the mean NHF risk. This is the fatigue threshold for the model. If the model does not calibrate by this method, proceed to Step A6a.

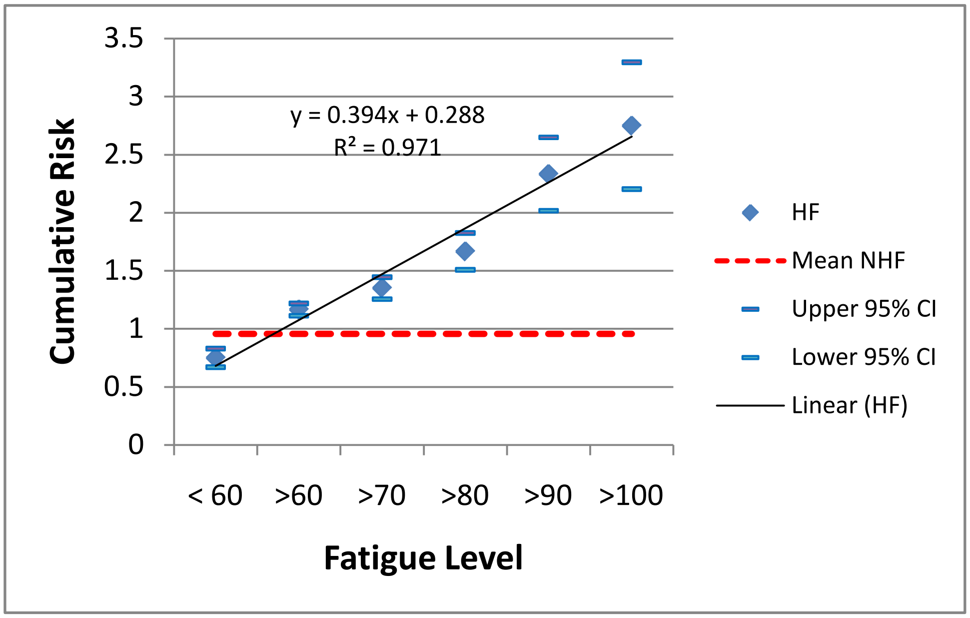
EXAMPLE OF STEP A6: The 95 percent confidence interval (CI) is as follows:

$$CI = \frac{T}{S},$$

where $$T = P \pm 1.96\sqrt{\frac{P\,(1-P)}{N}},$$

P is the proportion of human factor accidents in each bin.
N is the total number of accidents (based on formula 9.26.2, Hays, 1963, p. 291).
S is the cumulative proportion of employee time in each bin.

Cumulative accidents and cumulative employee hours are determined as in STEP 5. The Cumulative Risk for HF accidents, the upper and lower CI, and the mean NHF Cumulative Risk are plotted below for visual inspection:



For Model X, the point at which the HF Cumulative Risk 95 percent CI exceeds both Cumulative Risk =1 and the mean NHF Cumulative Risk is in the bin >60. For Model X, 60 is the fatigue threshold.

It is well known from epidemiology (Lilienfeld and Lilienfeld, 1980) that outcome risk is determined by both magnitude and duration of exposure. In the present case, HF accident risk should therefore be determined by duration of exposure to fatigue as well as fatigue severity. The cumulative proportion of work time spent at a Model X score at or greater than 60 is 0.6 (60 percent). Consequently, there should be less than 60 percent of work time at a Model X fatigue score above 60. On the basis of the same reasoning, calibration results for FAST indicated that there should be less than 20 percent of work time at a FAST score below 70. For FAID, there should be less than 20 percent of work time at a FAID score above 60.

STEP A6a. If a model fails to meet the calibration criteria in STEP A6, calibration may alternately be accomplished by demonstrating a statistically reliable correlation with a model that has been validated and calibrated as described above. The fatigue threshold for the model will be that value that corresponds with the fatigue threshold of the validated and calibrated model by use of a regression or other suitable mathematical equation. The use of STEP A6a is subject to approval by FRA.

B. What are the procedures for determining that a schedule exceeds the fatigue threshold?

The procedures for determining that a schedule exceeds the fatigue threshold require the use of a scientifically valid, biomathematical model of human performance and fatigue documented as noted above.
STEP B1. For each unique schedule, the time on duty, time on interim release, and limbo time must be described in a table that shows the time of day spent in each activity, for each day of the week, for one complete cycle of the schedule.

EXAMPLE OF STEP B1: Schedule A has work periods on Sunday, Monday, Wednesday, Thursday, and Friday from 0601 to 1000 and from 1601 to 1900. Schedule A includes a period of interim release from 1001 to 1600. This schedule repeats every week (one complete cycle occurs in a week). The table below shows this schedule using the following symbols:

D – time on duty
R – time on interim release
L – limbo time

	Time of Day																							
DAY	0001 - 0100	0101 - 0200	0201 - 0300	0301 - 0400	0401 - 0500	0501 - 0600	0601 - 0700	0701 - 0800	0801 - 0900	0901 - 1000	1001 - 1100	1101 – 1200	1201 - 1300	1301 - 1400	1401 - 1500	1501 - 1600	1601 - 1700	1701 - 1800	1801 - 1900	1901 - 2000	2001 - 2100	2101 - 2200	2201 - 2300	2301 - 2400
Sun							D	D	D	D	R	R	R	R	R	R	D	D	D					
M							D	D	D	D	R	R	R	R	R	R	D	D	D					
T																								
W							D	D	D	D	R	R	R	R	R	R	D	D	D					
Th							D	D	D	D	R	R	R	R	R	R	D	D	D					
F							D	D	D	D	R	R	R	R	R	R	D	D	D					
Sat																								

STEP B2. Use a validated model to analyze the schedule. Determine the fatigue score for every hour of the schedule marked “D.” Determine the number of work hours in bins as described in Step A1. Calculate the proportion (fraction in decimals or percent) of total work time in bins as described in Step A1.

EXAMPLE OF STEP B2: If Model X, which was validated as in Section A of this Appendix is used, the fatigue scores for specific work periods are as shown below:

	Time of Day																							
DAY	**0001 - 0100**	**0101 - 0200**	**0201 - 0300**	**0301 - 0400**	**0401 – 0500**	**0501 - 0600**	**0601 - 0700**	**0701 - 0800**	**0801 - 0900**	**0901 - 1000**	**1001 - 1100**	**1101 – 1200**	**1201 - 1300**	**1301 - 1400**	**1401 - 1500**	**1501 - 1600**	**1601 - 1700**	**1701 - 1800**	**1801 - 1900**	**1901 - 2000**	**2001 - 2100**	**2101 - 2200**	**2201 - 2300**	**2301 - 2400**
Sun							55	60	65	69	R	R	R	R	R	R	60	65	69					
M							56	62	67	68	R	R	R	R	R	R	62	67	68					
T																								
W							55	60	65	69	R	R	R	R	R	R	60	65	69					
Th							54	57	61	65	R	R	R	R	R	R	64	68	69					
F							58	63	67	69	R	R	R	R	R	R	66	69	71					
Sat																								

The number of work hours in the six bins from Step A1 is shown in the following table:

Employee (work) Time (in hours) as a Function of Modeled Performance Level

Modeled Performance Level						
<60	61-70	71-80	81-90	91-100	>100	Total
6	28	1	0	0	0	35

The proportion of work hours in each bin is shown in the following table:

Proportion of Work Time as a Function of Modeled Performance Level

Modeled Performance Level					
<60	61-70	71-80	81-90	91-100	>100
0.17	0.80	0.03	0	0	0

Step B3. The fatigue threshold for Model X is >60. The cumulative proportion of work time at a score of >60 is 0.8. This schedule is at risk for a level of fatigue at which safety may be compromised. Schedules that exceed the fatigue threshold must be reported, mitigated by one or more elements of a railroad's fatigue management plan, and approved for use by FRA, as required by § 228.407.

Glossary

The **95% Confidence Interval** is the range of estimated values (M) of a parameter that has a probability of 0.95 of containing the true population value (μ) of that parameter. In a normal distribution of sample means, with population mean μ and standard deviation σ_M, the probability is approximately 0.95 that $M - 1.96\ \sigma_M \leq \mu \leq M + 1.96\ \sigma_M$. In other words, the range of values between $M \pm 1.96\ \sigma_M$ is the **95% Confidence Interval** for μ.

Alertness is the state of being awake and watchful, of paying close and continuous attention to events in the immediate environment.

A **Bin** is a range of values of a variable defined by two endpoints. Bins are discrete, nonoverlapping intervals of a variable. For instance if a variable can have values between 11 and 70, bins might be defined as 11-20, 21-30, 31-40, 41-50, 51-60, and 61-70.

Binning refers to the grouping of values of a variable in one of several bins.

The **Bin Midpoint** is the middle point of a bin. It is equidistant from the bin endpoints and is often used to represent the whole bin. For example, the midpoint of the bin 11-20 is 15.5.

Circadian refers to events or processes that have a periodicity of approximately 1 day.

Cognitive Throughput is a measure of mental processing speed in humans. In a typical task, such as serial addition and subtraction,[2] the product of speed and accuracy (correct answers per minute) in the task is an index of cognitive throughput.

A **Lapse** is a brief loss of attention, alertness, or wakefulness. Also known as a microsleep, episodes can be as brief as a fraction of a second. In the laboratory, lapses are failures to respond in a reaction time task (see below).

Least-squares Linear Regression is a statistical technique for fitting data to a straight line ($y = mx + b$) such that the sum of the squares of the errors is minimized.

Phase Changes refer to changes in the timing of sleep and wakefulness in circadian rhythms. Phase delays (later wake-up and sleep onset) can be induced by exposure to bright light 2 h before usual bedtime. Phase advances (earlier wake-up and sleep onset) can be induced by exposure to bright light 5 h after usual bedtime. Jet lag (circadian desynchonization) is caused by phase changes because of travel to distant time zones in a short period of time.[3]

---

[2] This task requires a person to add or subtract a value from a previously obtained sum or difference. See Thorne, Genser, Sing, and Hegge, 1983.

[3] See Klein and Wegmann, 1980.

***r*, correlation coefficient** is a statistical measure of the linear ($y = mx + b$) dependence between two variables. The value of r is +1 when there is a perfect positive (increasing) linear relationship. The value of *r* is -1 when there is a perfect negative (decreasing) linear relationship. If *r* is 0, there is no linear relationship between the variables. However, the absence of a linear relationship does not preclude the existence of a nonlinear relationship.

r^2, coefficient of determination is the squared value of the correlation coefficient, *r*. It is the proportion of variance that is accounted for by a linear relationship between the variables.

Reaction Time is the time between the presentation of a sensory stimulus and the response to that stimulus. It is a measure of mental processing speed. In the laboratory, subjects are instructed to respond as fast as possible to the appearance of the stimulus. Generally, a maximum duration to respond is set before a trial is ended. Failure to respond within the maximum duration is considered a lapse (see above). Reaction time should not be confused with response latency, which refers to a response time that is not cued for an "as fast as possible" response.[4]

Sensitivity means that a fatigue model output is readily affected or changed by variables such as sleep deprivation and circadian phase changes.

Sleep Deprivation refers to a person obtaining less than a normal amount of sleep (approximately 8 h in a 24-h period).

Sleep Inertia refers to a lack of alertness and a decline in motor dexterity immediately following an abrupt awakening.

Sleep Recovery refers to a person obtaining more than a normal amount of sleep (approximately 8 h in a 24-h period) following a period of sleep deprivation (see above).

Statistically Reliable has the same meaning as statistically significant (see below).

Statistically Significant in hypothesis testing means that the outcome of a statistical test is unlikely to have occurred by chance. Chance is usually defined as a probability less than 0.05.

Vigilance Speed (see Reaction Time above).

---

[4] See Dinges and Powell, 1985.

References

Achermann, P. (2004). The two-process model of sleep regulation revisited. *Aviation, Space, and Environmental Medicine*, *75,* A75-A83.

Akerstedt, T., Folkard, S., and Portin, C. (2004). Predictions from the three-process model of alertness. *Aviation, Space, and Environmental Medicine*, *75,* A44-A53.

Balkin, T. J., Braun, A. R., and Wesensten, N. J. (2002). The process of awakening: a PET study of regional brain activity patterns mediating the reestablishment of alertness and consciousness. *Brain*, *125*, 2308-2319.

Balkin, T. J, Thorne, D., Sing, H., et al. (2000). *Effects of sleep schedules on commercial driver performance.* Report No. DOT-MC-00-133. Washington, DC: U.S. Department of Transportation, Federal Motor Carrier Safety Administration.

Belyavin, A. J., and Spencer, M. B. (2004). Modeling performance and alertness: The QinetiQ approach. *Aviation, Space, and Environmental Medicine, 75,* A93-A103.

Bonnet, M. H. (1997). Sleep fragmentation as the cause of daytime sleepiness and reduced performance. *Wien Med Wochenschr*, *146*, 332-334.

Carskadon, M., and Dement, W. (1977). Sleep tendency: an objective measure of sleep loss. *Sleep Research, 6,* 200.

Dinges, D. F., Orne, M., and Orne, E. (1985). Sleep depth and other factors associated with performance upon abrupt awakening. *Sleep Research*, *14*, 92.

Dinges, D. F., and Powell, J. W. (1985). Microcomputer analysis of performance on a portable, simple visual RT task during sustained operations. *Behavioral Research Methods, Instrumentation, and Computers*, *17*, 652-665.

Dinges, D. F., and Powell, J. W. (1989). Sleepiness impairs optimum response capability. *Sleep Research*, *18,* 366.

Folkard, S., and Akerstedt, T. (1987). Towards a model for the prediction of alertness and/or fatigue on different sleep/wake schedules. In A. Oginiski, J. Pokorski, and J. Rutenfranz (Eds.), *Contemporary advances in shiftwork research*. Krakow: Medical Academy. pp. 231-240.

Folkard, S., and Akerstedt, T. (1992). A three process model of the regulation of alertness-sleepiness. In R. Oglivie and R. Broughton (Eds.), *Sleep, arousal and performance problems and promises. A tribute to Bob Wilkinson*. Boston: Birkhauser. pp. 11-26.

Froberg, J. (1977). Twenty-four-hour patterns in human performance, subjective and physiological variables and differences between morning and evening active subjects. *Biological Psychology*, *5*, 119-134.

Harrison, Y., and Horne, J. A. (1996). Long-term extension to sleep—are we really chronically sleep deprived? *Psychophysiology*, *33*, 22-30.

Hays, W. L. (1963). *Statistics for Psychologists*. New York: Holt, Rinehart and Winston.

Hursh, S. R., Raslear, T. G., Kaye, A. S., and Fanzone, J. F. (2006). *Validation and calibration of a fatigue assessment tool for railroad work schedules, summary report* (Report No. DOT/FRA/ORD-06/21). Washington, DC: U.S. Department of Transportation. (http://www.fra.dot.gov/downloads/Research/ord0621.pdf)

Hursh, S. R., Raslear, T. G., Kaye, A. S., and Fanzone, J. F. (2008). *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., and Eddy, D. R. (2004). Fatigue models for applied research in warfighting. *Aviation, Space, and Environmental Medicine, 75,* A44-A53.

Jewett, M. (1997). *Models of circadian and homeostatic regulation of human performance and alertness.* [Dissertation]. Cambridge, MA: Harvard University.

Jewett, M., and Kronauer, R. (1999). Interactive mathematical models of subjective alertness and cognitive throughput in humans. *Journal of Biological Rhythms*, *4*, 588-597.

Klein, K., and Wegmann, H. (1980). The effect of transmeridian and transequitorial air travel on psychological well-being and performance. In L.E. Scheving and F. Halberg (eds.), *Chronobiology: principles and applications to shifts in schedules.* Rockville, MD: Sijthoff and Noordhoff. pp. 339 - 352.

Mitler, M., Gujavarty, S., Sampson, G., and Bowman, C. (1992). Multiple daytime nap approaches to evaluating the sleepy patient. *Sleep*, *5*, 119-127.

Monk, T. (1991). Circadian apects of circadian sleepiness: a behavioral messenger? In T. Monk (Ed.), *Sleep, sleepiness and performance*. Chichester, U.K.: Wiley. pp. 39-63.

Monk, T., and Embry, D. (1981). A field study of circadian rhythms in actual and interpolated task performance. In A. Reinberg, N. Vieux, and P. Andlauer (Eds.), *Night and shift work: Biological and social aspects*. Oxford: Pergamon Press. pp. 473-480.

Moore-Ede, M., Heitmann, A., Guttkuhn, R., Trutschel, U., Aguirre, A., and Croke, D. (2004). Circadian alertness simulator for fatigue risk assessment in transportation: Application to reduce frequency and severity of truck accidents. *Aviation, Space, and Environmental Medicine*, *75*, A107-A118.

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

Lilienfeld, A. M., and Lilienfeld, D. E. (1980). *Foundations of epidemiology*, *2nd ed*. New York: Oxford University Press.

Lumley, M., Roehrs, T., and Zorick, F. et al. (1986). The alerting effects of naps in sleep-deprived subjects. *Psychophysiology*, *23*, 403-408.

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

Richardson, D., Carskadon, M., and Flagg, W. (1978). Excessive daytime sleepiness in man: multiple sleep latency measurement in narcoleptic and control subjects. *Electroencephalography and Clinical Neurophysiology*, *45*, 621-627.

Tabak, B., and 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. (http://www.fra.dot.gov/rpd/downloads/TR_Procedures_or_Validation_and_Calibration_final.pdf)

Thorne, D. R., Genser, S., Sing, H., and Hegge, F. (1983). Plumbing human performance limits during 72 hours of high task load. In *Proceedings of the 24th DRG seminar on the human as a limiting element in military systems*. Toronto: Defense and Civil Institute of Environmental Medicine. pp. 17-40.

Wesensten, N. J., Balkin, T. J., and Belenky, G. (1999). Does sleep fragmentation impact recuperation? A review and reanalysis. *Journal of Sleep Research*, *8*, 237-246.