Developing Mathematical Models of Neurobehavioral Performance for the "Real World"

Dennis A. Dean, II, Adam Fletcher, Steven R. Hursh and Elizabeth B. Klerman

J Biol Rhythms 2007; 22; 246
DOI: 10.1177/0748730407301376

Developing Mathematical Models of Neurobehavioral Performance for the “Real World”

Dennis A. Dean II,*[,1] Adam Fletcher,[†] Steven R. Hursh,[‡] and Elizabeth B. Klerman*

**Division of Sleep Medicine, Brigham and Women’s Hospital and Harvard Medical School, Boston, MA; [†]Integrated Safety Support, Melbourne, Australia; [‡] Johns Hopkins University School of Medicine and Institutes for Behavior Resources, Baltimore, MD*

Abstract Work-related operations requiring extended wake durations, night, or rotating shifts negatively affect worker neurobehavioral performance and health. These types of work schedules are required in many industries, including the military, transportation, and health care. These industries are increasingly using or considering the use of mathematical models of neurobehavioral performance as a means to predict the neurobehavioral deficits due to these operational demands, to develop interventions that decrease these deficits, and to provide additional information to augment existing decision-making processes. Recent advances in mathematical modeling have allowed its application to real-world problems. Developing application-specific expertise is necessary to successfully apply mathematical models, in part because development of new algorithms and methods linking the models to the applications may be required. During a symposium, “Modeling Human Neurobehavioral Performance II: Towards Operational Readiness,” at the 2006 SIAM-SMB Conference on the Life Sciences, examples of the process of applying mathematical models, including model construction, model validation, or developing model-based interventions, were presented. The specific applications considered included refining a mathematical model of sleep/wake patterns of airline flight crew, validating a mathematical model using railroad operations data, and adapting a mathematical model to develop appropriate countermeasure recommendations based on known constraints. As mathematical models and their associated analytical methods continue to transition into operational settings, such additional development will be required. However, major progress has been made in using mathematical model outputs to inform those individuals making schedule decisions for their workers.

Key words mathematical models, circadian, performance, fatigue, scheduling

The extended, night and/or shift work schedule demands of the 24-h society continue to expand due to military, transportation, security, health care, and economic needs. The most common problems reported by workers on these schedules are disturbed sleep and waketime sleepiness (Akerstedt, 2003; Mallis et al., 2004). From a biological perspective, these complaints are not surprising because these workers are often required to work at times when their physiology promotes sleep (Monk et al., 1985), and they attempt sleep

1. To whom all correspondence should be addressed: Dennis A. Dean II, Division of Sleep Medicine, Brigham and Women’s Hospital, 221 Longwood Avenue, Boston, MA 02115; e-mail: ddean@rics.bwh.harvard.edu.

JOURNAL OF BIOLOGICAL RHYTHMS, Vol. 22 No. 3, June 2007 246-258
DOI: 10.1177/0748730407301376

when their physiology promotes being awake and active. In addition to the mismatched times of sleep/wake and rest/work, if sufficient recovery from the sleep deprivation or restriction does not occur during off-duty times, sleepiness and fatigue can accumulate to dangerous levels. Importantly, self-assessment frequently does not correspond with objective neurobehavioral performance, with individuals reporting greater alertness than their objective performance demonstrates (Van Dongen, 2003). Given this mismatch of subjective and objective assessments, mathematical models can be an important tool in fatigue management and schedule planning. In this article, we discuss the use of mathematical models to provide an objective assessment of performance in 3 different applications: validating models in railroad engineer schedules, predicting performance in transmeridian travel, and optimizing countermeasures to improve performance.

Since the early 1980s, mathematical models of circadian rhythms, sleep and neurobehavioral performance have been used to understand how humans respond to these extended, night, and/or shift work schedules. Initially, models focused on the prediction of sleep and sleepiness in controlled settings such as the laboratory (Borbély, 1982; Kronauer et al., 1982). Later, it was suggested that models had the potential to make operations safer (Belenky et al., 1998). In response, mathematical modeling of neurobehavioral performance was developed from producing tools that are used to understand underlying physiological mechanisms to tools that are applied to operational applications. Many models have been developed; validations of these models have included data from laboratory, simulator, and operational environments (Mallis et al., 2004; Van Dongen, 2004). Based on this work, there has been significant interest and funding from potential users (industry, transportation, military) of these mathematical models to understand and identify physiological mechanisms in specific applications. The potential benefit to the operational community is clear. When appropriately applied, these models could be used to evaluate and compare scheduled shifts, to identify times of fatigue and alertness vulnerabilities within a schedule, and in online monitoring of performance.

The mathematical modeling of neurobehavioral performance included in this article is influenced by the Two Process model of sleep (Achermann, 2004). The model proposes that sleep/wake timing is a function of both sleep (Process S) and circadian (Process C) systems. Process S refers to changes in sleep propensity due to both wake and sleep; it may also be referred to as homeostatic control. The circadian system refers to the signal from the master circadian pacemaker in the suprachiasmatic nucleus of the hypothalamus in mammals (Moore and Eichler, 1972). A limit cycle model is one mathematical modeling approach to describing the circadian system, since it has been shown to capture many of the dynamical features of circadian rhythms. Since ocular light exposure has been shown to have the largest effect on circadian phase (Czeisler et al., 1986), light stimuli are major inputs to some models of the circadian pacemaker. The Two Process model has been extended to predictions of neurobehavioral performance; a nonlinear relationship between Process S and Process C exists among the many outcome measures, including different aspects of performance and mood (Achermann, 2004; Jewett and Kronauer, 1999). Capturing the nonlinear relationship between the sleep/wake and circadian inputs and the neurobehavioral performance and alertness outputs is an important aspect of these mathematical models. A detailed description of mathematical models of performance and fatigue was presented as part of the 2002 Proceedings of the Fatigue and Performance Modeling Workshop (Mallis et al., 2004).

In this article, we present 3 specific applications of mathematical models to illustrate how these models can be used in the "real world." The work was presented during the "Modeling Neurobehavioral Performance II: Towards Operational Readiness" Symposium at the 2006 SIAM-SMB Conference on the Life Sciences. The 3 models demonstrate that, although they have a common aim of objective prediction of neurobehavioral performance in an operational setting, the underlying assumptions within the models are directly influenced by different modeling philosophies, objectives, and specific applications. Development of the Sleep, Activity, Fatigue, and Task Effectiveness (SAFTE) model (Hursh et al., 2004b) was initially motivated by providing objective assessment of military performance in the operational setting and is modeled after the Two Process model. Most recently, SAFTE has been applied to a variety of nonmilitary applications including railroad engineers and accident assessment. The motivating philosophy of the Fatigue Audit Inter*Dyne* (FAID®) model was that the model had to be simple and easy to understand in order to facilitate operational acceptance (Fletcher and Dawson, 1997). The main output of FAID is a fatigue rating. Both Hursh and Fletcher have developed methods for determining the probable timing of workers' chosen sleep episodes given their work hours. Automatic methods for determining sleep episodes from a given work schedule enhance the ability of mathematical models to be applied in the operational setting, since they require less input from individual users or supervisors. The origins of the Kronauer-Jewett Interactive Neurobehavioral Model (Jewett and Kronauer, 1999; Kronauer et al., 1999) was motivated by providing a mathematical basis for understanding the interactions between circadian and sleep rhythms observed in data collected in a research environment. The model outputs include objective and subjective measures of performance and circadian phase. The work presented in this symposium detailed some of the development and additional work required to transition the original models into a format useful for operational settings.

HURSH: "MANAGING FATIGUE IN 24/7 OPERATION: FATIGUE MODELING TO IMPROVE RAILROAD SAFETY"

Mathematical models can be useful in managing fatigue and its negative consequences in many industries. To aid in transitioning the theoretical models to operational use, several steps are involved. One involves converting model outputs to usable parameters, such as a fatigue rating scale. A second involves testing the outputs of the model against actual data such as the risk of accidents. For mathematical models of circadian rhythms, sleep, and performance, a conversion of work schedules into predicted sleep times of the worker is also necessary. Finally, providing methods for quick computation of many different possible schedules is important.

Developing a Fatigue Scale and Comparing Model Predictions with Accident Data

The Federal Railroad Administration (FRA) continues to partner with the railroad industry to participate in projects for developing useful tools for managing fatigue in railroad operations. Recently, the FRA has completed the 3rd phase of a research program to validate and calibrate fatigue models for use in predicting and managing fatigue in railroad workers. A fatigue model offers the possibility of objectively assessing and forecasting fatigue so that employees and employers can schedule work and rest to avoid fatigue. A useful fatigue model needs to be calibrated to the demands of a particular job so that the outputs from the model can be related to the risk of meaningful failures of human performance. One important part of calibration of a fatigue model for use as a fatigue management tool is an assessment of whether the tool can predict an increased risk of a human-factors error or risk of having a human factors–caused accident. As part of this assessment, the FRA has sponsored a project in partnership with the 5 Class I rail freight carriers to examine 2-1/2 years of accident data.

For this assessment of the fatigue management tool and its ability to predict increased risks, we examined 30-day work histories of locomotive crews before 400 human-factors accidents and 1000 nonhuman-factors accidents spanning 2-1/2 years. The researchers analyzed all work histories to determine, based solely on the model, the predicted effectiveness of the operators at the time of the accidents. The fatigue model used for these studies—the SAFTE model—was originally developed for the U.S. Army and Air Force by Dr. Steven R. Hursh (Hursh et al., 2004b) and predicts potential fatigue based on an analysis of work schedules. The model has been adopted by the Department of Defense as the standard warfighter fatigue model and has been incorporated into a useful fatigue assessment tool called the Fatigue Avoidance Scheduling Tool (FAST) that uses information about a person's work schedule and typical sleep habits to estimate the amount of sleep that would occur under the work schedule and the effects of that sleep pattern on cognitive effectiveness (Hursh et al., 2004b).

The virtue of having a validated fatigue model, especially if it is calibrated to relate the model to accident risk, is that it could be used by a railroad or other carrier to do a self-assessment of fatigue across their system. For example, freight railroads are organized geographically by terminals where trains are assembled and crews report for work. By evaluating work histories on a terminal-by-terminal basis and using the scores from the model as a metric, the carrier could determine which terminals are experiencing schedules that might be generating fatigue in the crews. Perhaps none of the terminals have a problem or perhaps just a few. With this information, the carrier would be in a position to use this objective assessment to focus fatigue management efforts toward the greatest payoff. Furthermore, after changes are made in operations or crew schedules to reduce schedule-induced fatigue, the carrier could revisit that terminal to assess whether the initiatives have been productive in eliminating or reducing the problem. Therefore, an objective tool could be applied to solve this elusive but safety-related problem. From the industry's perspective, this would be a nonprescriptive, performance-based approach that would not impose a one-size-fits-all regulatory solution. From the public’s perspective, this would have the promise of making a difference for the safety of the industry and the public.

The results of the accident analysis study indicated that a fatigue model could predict an increased risk of human-factors accidents under certain conditions that cause fatigue. SAFTE estimated crew cognitive effectiveness based entirely on work schedule information and opportunities to obtain sleep. Effectiveness is a metric that tracks speed of performance on a simple reaction time test and is strongly related to overall cognitive speed, vigilance, and the probability of lapses (Hursh et al., 2004a). Cognitive effectiveness was interpreted as the inverse of fatigue. The model rates effectiveness on a scale from 0 to 100. There was a reliable linear relationship between crew effectiveness (fatigue) and the risk of a human-factors accident ($r = -0.93$, $P < 0.01$); no such relationship was found for nonhuman-factors accidents ($r = -0.14$, $P > 0.05$). This result satisfied the criteria for model validation. The risk of human-factors accidents was elevated at any effectiveness score below 90 and increased progressively with reduced effectiveness. There was a reliable time-of-day variation in human-factors accidents ($r = 0.71$, $P < 0.05$) but not in nonhuman-factors accidents. At an effectiveness score ≤50, human-factors accidents were 65% more likely than chance. Human-factors accident risk increased reliably when effectiveness was below 70, a value that is the rough equivalent of a 0.08 blood alcohol level or being awake for 21 h following an 8-h sleep period the previous night (Hursh et al., 2004b). Below an effectiveness score of 70, accident cause codes (codes defined by the FRA to indicate the factors that caused the accident, such as passing a stop signal or exceeding authorized speed) were of the sort expected of fatigue, confirming that the relationship between accident risk and effectiveness was meaningful.

Risk was defined as the ratio of the proportion of accidents at a level of effectiveness to the proportion of time spent at work at that level of effectiveness (exposure), based on the 30-day work histories. For schedules with scores below 70, the risk of accidents was elevated relative to chance and greater than the mean risk of a nonhuman-factors accident (Fig. 1A).

For the question “What is the relative accident risk for individuals below a level of X?” the same information can be portrayed as cumulative risk at various criterion levels. Since each point is cumulative and not independent of the prior points, we could not compute correlation coefficients on this chart. Instead, confidence limits on each point were computed relative to chance, and all points that are significantly above or below chance ($P < 0.05$) are included in Figure 1B.

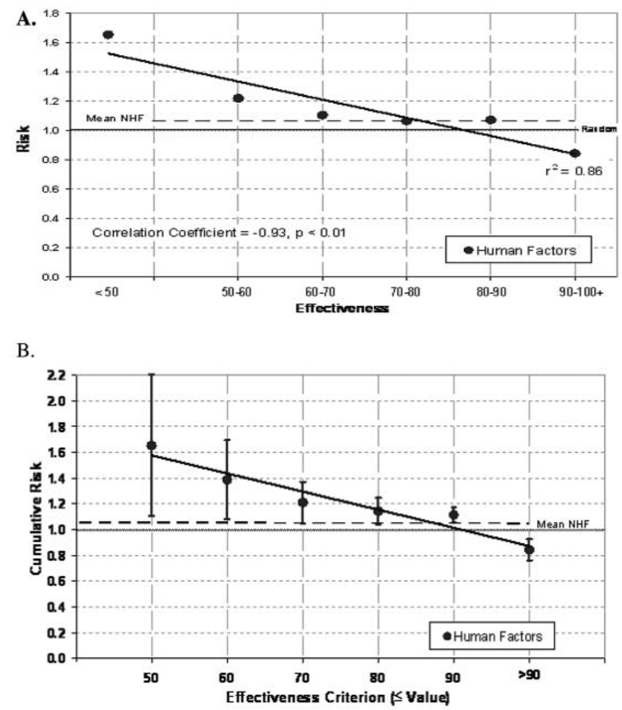

Figure 1. (A) Human-factors accident risk at each level of effectiveness aggregated from 5 railroads for the Sleep, Activity, Fatigue, and Task Effectiveness (SAFTE) model. The solid line represents the correlation. A ratio of 1 is the chance level of risk; less than 1 is reduced risk; greater than 1 is elevated risk. A risk ratio of 1.65 means that accidents were 65% more frequent than chance. (B) Human-factors cumulative accident risk by criterion levels of effectiveness aggregated for 5 railroads for the SAFTE model. The solid line represents a linear relationship between effectiveness criterion score and cumulative risk. The dotted line indicates the nonhuman-factor accident risk.

Predicting Sleep Timing and Managing Fatigue by Evaluating Multiple Possible Schedules

Any work schedule evaluation and design tool requires a method for estimating potential sleep under a work schedule. Dr. Hursh has developed a sleep pattern estimation method called AutoSleep (Hursh et al., 2004a). That sleep estimator was built based on the average performance of engineers reported by Pollard (1996) in a study of work and sleep diaries maintained by a group of about 150 railroad engineers. While the algorithm provided an estimate that closely resembled the average of that aggregate population, the FRA Office of Safety has undertaken a rigorous validation of the sleep estimation algorithm against individual and group data from a range of job classifications in the railroad industry. Two Class I railroads and one of the major railroad unions are cooperating in a study aided by Dr. Hursh to validate and improve the algorithm for estimating sleep under typical work schedules.

To hasten the implementation of corporate fatigue risk management strategies, the FRA has initiated a complementary effort to develop a workforce fatigue risk management tool called the Aggregated Fatigue Risk Management (AFRM) Tool. This tool will use the same methodology incorporated in the work schedule evaluation tool used in the accident analysis study, but will be able to bulk-process work schedule data from an entire workforce and provide standardized reports and metrics to assist a company in evaluating the potential levels of work schedule–induced fatigue that exist at specific work sites or within certain work groups. As a starting point, the tool will use SAFTE, the same mathematical fatigue model validated in the accident analysis study. By incorporating the features of AFRM and SAFTE under a single umbrella, the railroad industry will have the capability of ascertaining the impact of fatigue from both a systemic and individual perspective.

Finally, a work schedule and design tool must be easy to use and informative in the hands of users who are not themselves experts in fatigue analysis. The FRA Office of Safety has undertaken the development of an efficient user interface that will guide the user in the application of the fatigue model for analysis of any work schedule. The initial application of the fatigue analysis "wizard" will be by FRA accident investigators who will apply the tool to assess possible contributions of worker fatigue in railroad accidents.

FLETCHER: "AN APPLICATION TO PREDICT FATIGUE ASSOCIATED WITH HOURS OF WORK, AND A SPECIFIC MODEL TO ESTIMATE SLEEP WITHIN LONG-HAUL AVIATION"

Work in extended- or 24-h environments, and the fatigue that can accumulate, has traditionally been "managed" by limiting the length of shifts and total number of hours worked per week, month, or year (Fletcher and Dawson, 1998; Knauth, 1997). However, this approach fails to acknowledge key factors such as the time of day, or the pattern of work and recovery opportunity (Fletcher and Dawson, 1997; Fletcher and Dawson, 2001; Smiley, 1998; Vespa et al., 1998). In civilian operational settings, accounting for all of the significant factors of influence is difficult, since the only relevant data usually available are the hours each individual has recently worked and the hours they are scheduled to work in the near future. From hours of work data, it is possible to determine 1) the length of shifts and breaks between shifts; 2) the time of day of shifts and breaks between shifts; and, therefore, 3) the pattern of work and nonwork (Dawson and Fletcher, 2001; Fletcher and Dawson, 2001).

Predicting the Effects of Work Hours on Fatigue

Using only hours of work as the input, it has been possible to develop a model to predict the effects of work hours on fatigue and perform validations under laboratory, simulator, and operational conditions (Fletcher et al., 2000; Fletcher et al., 2003; Roach et al., 2001). Free evaluation versions of the Fatigue Audit Inter*Dyne* (FAID) application can be downloaded at www.faidsafe.com. As detailed in the 2002 modeling workshop, this model was the first to use work hours as the primary input. Prior to this, existing 2- and 3-process models required work/rest or wake/sleep profiles as inputs. This requires compliance from workers to record their wake/sleep time regularly. Additions have since been made to a number of these models to enable them to estimate wake/sleep from work/rest (see above).

In using this or any similar application, however, it is critical to explicitly define "hours of work" or other specific inputs. One major reason for this is that in operational settings, it is rarely the model author that enters, analyzes, and interprets data; instead, that task is usually performed by an operational employee. Without guidance, it is impossible for the model user to know what inputs and output interpretations are appropriate. In addition, if model outputs will be compared across various sites within an organization or between organizations, then having the input(s) defined is critical for standardization.

The specific definition of "hours of work" for use with FAID includes all time from the start of a shift to the end of a shift but does not include commute times longer than an hour. Therefore, if operational employees have commutes longer than 1 h, then it is acknowledged that their opportunity for sleep between consecutive shifts is substantially reduced; in this case, FAID might overestimate the recovery value of nonwork periods. In addition, breaks provided within a shift are not included as nonwork time, unless they are substantial enough for recovery sleep to be obtained, as might occur within a split-shift system of a marine pilot. Nominally, it is not considered appropriate to call a break a nonwork period in FAID unless the break is at least 4 h in duration and/or there are specific facilities provided that are conducive to sleep. Finally, FAID has not been validated in contexts that involve multiple time zone transitions (that is, transmeridian travel) as part of work requirements. However, the impact of such transitions is likely to be negligible for 1- or 2-h shifts eastward or westward; this is especially true if an individual returns to the original time zone within 1 or 2 days. Shifts of 3 or more time zones, however, are likely to significantly increase the amount of variance in any actual performance, and FAID will increasingly become inaccurate, especially if the individual remains in the new time zone for more than 1 or 2 days.

The major assumptions of the model include the following: 1) Recovery from work-related fatigue by sleeping can be obtained at any time an individual is not working. 2) The actual pattern of recovery or sleep is a subset of the opportunity available and is dictated by time of day and competition from factors such as social pressures. 3) More recent work or nonwork periods weight more heavily on an individual's current fatigue state than time periods further in the past. There is no weighting given to time further back than 7 days or 168 h. 4) Human recovery saturates after a maximum set point (Dawson and Fletcher, 2001; Fletcher and Dawson, 2001).

The primary output from FAID is a fatigue score. The score is determined for each hour of work, based on a weighted valuation of the previous 7-day work and nonwork pattern. The score is presented in tabular or graphical format, and numerous other indicators of work-related fatigue exposure are also provided. These include percentage of compliance with a determined fatigue tolerance threshold, which is ideally determined from the organization's own data. In the absence of an organizationally set threshold, it is possible to interpret the scores against a generic or nominal threshold. The FAID scale ranges from 0 to in excess of 140 points. Validation studies in the field and laboratory have shown that a score of 40 points represents the level of work-related fatigue experienced by someone working Monday to Friday, 0900 h to 1700 h. Any score below 80 is broadly considered to be acceptable. Any score in excess of 100 is considered unacceptable unless fatigue countermeasures are implemented. Any score between 80 and 100 requires further assessment to understand the requirement for countermeasures to minimize errors or accidents. Countermeasures in this context might include the use of checklists or double-checking, additional supervision or communication, or the tactical placement of a break or task rotation.

This scoring breakdown has been independently verified (New South Wales Special Commission, 2003) as an appropriate interpretation of the laboratory, simulator, and field validations (Fletcher et al., 2000; Fletcher et al., 2003; Roach et al., 2001). Given the significance of differences between workplaces and due to factors such as workload, however, it is considered ideal (if not essential) that organizations determine their own FAID tolerance thresholds so they do not require the use of a more generally accepted scoring system. Ideally, factors should be considered in a systematic way within a risk-management process that also considers other work-related factors such as workload, breaks, available countermeasures, degree of training, degree of supervision, and any other identified factors that are likely to impact fatigue. These factors are important and can be accounted for using validated, public domain workload assessment tools coupled with risk-management processes. In cases where risk assessments have been performed, the appropriate FAID scores for a particular set of circumstances can be more precisely determined. In addition, it is worth noting that the application's context of use is best limited to assessments of hours of work across a single workgroup, a range of workgroups or sites, or even an entire organization. FAID is less appropriate—due to its origins and validations—for use to assess fatigue levels of a single individual in a defined set of circumstances.

Predicting Sleep Patterns of International Airline Pilots

More defined sets of circumstances require more specifically parameterized models. One example of a relevant context mentioned above is transmeridian travel. In aviation operations, it is possible to obtain flight schedules for planned hours of work as well as crew flight and duty time hours for hours already worked. In addition, instead of having one set of shift start-and-end times—as would be available in a fixed-time environment—shift times must be represented in other time zones such as the origin and destination of each flight sector. In an operational sense, the added complexity to aviation crew is that not only do sleep/wake opportunities shift in relationship to the work patterns (as with fixed-time environment shift work), but the light-dark cycle also moves depending on the time zone shifts that are being experienced.

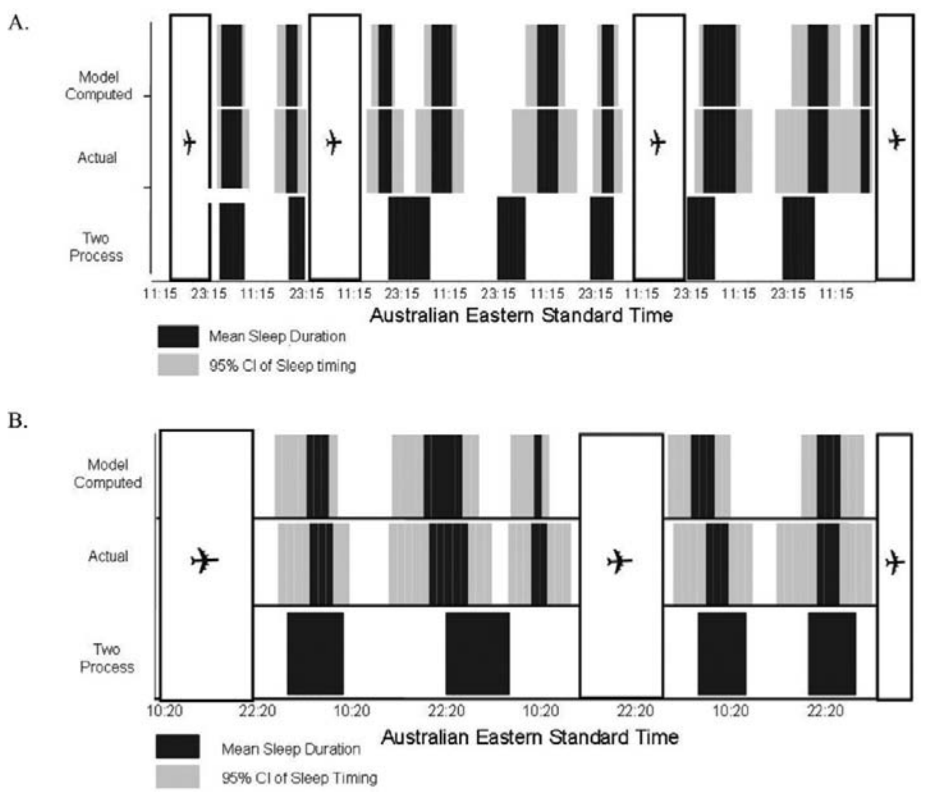

Figure 2. (A) A schematic diagram comparing the Model-Computed, Actual, Two-process sleep data for the Sydney-Bangkok-London-Singapore-Sydney dataset. (B) A schematic diagram comparing the Model-computed, Actual, and Two-process sleep data for the Sydney-Los Angeles-Auckland-Sydney dataset. In both panels, the white blocks indicate duty periods (flights). The black boxes represent sleep duration, and the gray bars represent the 95% confidence interval (CI) of sleep timing (i.e., left gray bars represent 95% CI of sleep onset time, and right gray bars represent 95% CI of sleep duration).

Biologically, this set of circumstances can lead to a situation where individuals have great difficulty in knowing when is the best or worst time to successfully achieve sleep. The decision to sleep is often made using a combination of social and biological factors. At some times, it may be biologically appropriate to sleep but not socially desirable or appropriate, or vice versa.

Both circadian and sleep strategy selection factors have been used to develop mathematical models to predict the sleep patterns of international airline pilots. For one study, data were collected using airline pilots traveling Sydney-Bangkok-London-Singapore-Sydney (SYD-LHR, n = 19, mean layover length of 63.5 ± 1.2 [mean ± SD] h) and Sydney-Los Angeles-Auckland-Sydney (SYD-LAX, n = 16, mean layover length of 35.9 ± 1.1 h). Following this, the 2nd duty period from Los Angeles to Auckland, New Zealand, was conducted. A short layover (23.7 ± 0.9 h) was spent in Auckland, before returning to Melbourne or Brisbane, Australia, followed by a short flight to Sydney. The SYD-LHR data set consisted of 3 separate layovers: 1 in Bangkok while outbound from Sydney (23.8 ± 0.4 h), 1 after landing at London Heathrow (63.5 ± 1.2 h), and 1 in Singapore while inbound to Melbourne (48.3 ± 0.7 h). After a short stop in Melbourne, flight crews then fly to Sydney to complete the pattern. A modeling technique was developed by the authors to predict the timing and duration of sleep during layover periods for each of the SYD-LHR and SYD-LAX data sets. To account for the maximum proportion of variance in the data, and to make optimal use of available data—including the time of arrival into a layover location, and the departure time from that location—the model was constructed to use both prediction (using events from the past to predict events in the future) and retrodiction (using events in the future to predict events occurring in the present). The model was constructed using mixed modeling techniques. The authors have devised a sleep regulation model that splits layover periods into 3 distinct time stages, as there is evidence to suggest that sleep strategies employed by airline pilots change over the course of a layover (Samel and Wegmann, 1988; Samel et al., 1991). The stages are 1) Recovery—lasting from flight arrival until the onset of the 1st local night (2100 h); 2) Personal—lasting from the 1st local night onset until the offset of the final local night period; and 3) Preparation Phase—lasting from the offset of the last local night period until flight departure. Specific results of this study have previously been published (Kandelaars, Fletcher, Eitzen, et al., 2006; Kandelaars, Fletcher, Dorrian, et al., 2006). Although further testing and development are required, it appears that the development of sleep prediction models for estimating airline crew's sleep is plausible.

DEAN: "OPTIMIZING COUNTERMEASURE DESIGN"

Mathematical models have been primarily used to assess activity/rest schedules by predicting alertness and performance outcomes for that schedule. However, mathematical models can also be used to design countermeasures for a given schedule that, without a countermeasure, may result in poor performance or alertness during scheduled working times. We define the countermeasure design problem as the use of mathematical models to automatically or interactively design schedules that meet specific neurobehavioral performance goals given a specific work schedule.

Determination of the appropriate schedule requires both design and evaluation of different schedules. Kostreva et al. (2002) were one of the first to report on the use of a human model of the effect of light on the circadian pacemaker to evaluate alternative shift-work schedules. Other relevant publications (Bagheri et al., 2005; Dean and Klerman, 2005; Dean et al., 2005; Forger and Paydarfar, 2004; Mott et al., 2003) focused on the use of models of the effect of light on determining optimal methods for re-entraining a model of the circadian pacemaker following a shift in sleep-wake schedule. Mott et al. (2003) used model predictive control (described below) to determine the appropriate light intervention to re-entrain the circadian system to a shift in sleep-wake schedule. Model predictive control is a method that converts a mathematical model of a system into a controller for that system. An important point in Mott et al. is that to create the controller, the model was linearized to simplify calculations. Motivated by the work of Mott et al., Bagheri et al. used a nonlinear approach to the problem of determining light interventions (Bagheri et al., 2005). Lastly, Forger (Forger and Paydarfar, 2004) presented the use of the calculus of variations to determine optimal light interventions for sleep-wake schedules. While solving differential equations with boundary value methods as part of the calculus of variations is an elegant and extendable framework for schedule optimization, we found several relevant cases in which the solutions did not converge (D. B. Forger, personal communication, July 2006), restricting its possible use. Whereas Forger considered light level designs that facilitate re-entrainment to a shift in sleep/wake as in Bagheri et al., the work differed in that theoretical approaches to optimal control were presented.

Recently, the concept of a generalized framework for approaching schedule optimization problems was introduced (Dean and Klerman, 2005; Dean et al., 2005). Under this framework, both software and an analysis system are required to approach schedule/countermeasure design. Both software and an analysis system are required since multiple optimal solutions exist for the optimal light countermeasure problem described above (Bagheri et al., 2005). A single solution, however, frequently can be chosen when these multiple solutions are reviewed relative to operational constraints. Consequently, a mechanism with which a user can guide the selection to an optimal solution based on operational constraints is required. Although simple to do in theory, modifying and simulating many possible schedules can be a time-consuming process for the user. Hence, developing a software framework that works in conjunction with analytical methods to generate and compare many possible schedules simplifies and shortens the process for the user.

Features of a Schedule Design System

Through our preliminary work (Dean et al., 2005), we have defined important features and challenges of a schedule design system. These features include 1) a dynamic model that includes the effect of the countermeasure, 2) selecting an objective function for evaluating the system, and 3) methods for optimizing countermeasure features. Each of these features is now discussed.

Dynamic Model Including the Effect of the Countermeasure

An accurate dynamic model of the effect of the countermeasure on the output variable is a prerequisite for optimal countermeasure design. This model should capture the potential nonlinear relationships between the countermeasure and the output variables. Unfortunately, very few dynamic models that include countermeasures directly have been developed. One potential reason is that modelers require experimental data that illustrate the time course of change in the model parameters in response to the countermeasure. This can be an expensive and time-consuming proposition, especially for human data. To be accurate, the models require data both from controlled conditions and from operationally relevant conditions so that they can be amended to be appropriate for both types of conditions.

Selecting an Objective Function

An objective function is the mathematical formulation of the schedule feature that is to be optimized. For example, performance can be optimized across an entire schedule or only during time when critical operations are planned. When choosing an objective function to optimize, both the biological and operational constraints must be met. Most of the work in this area thus far has focused on meeting biological constraints. Published objective functions have included the predicted difference between the circadian state variables (Kostreva et al., 2002), minimizing the amount of light (Forger and Paydarfar, 2004), optimizing the angle of entrainment (Dean et al., 2005), and minimizing resetting time (Bagheri et al., 2005).

In addition to meeting the biological constraints, the operational constraints must be considered. For example, NASA applications require a high level of crew performance during critical time windows, such as launch and space walks. In that application, maximizing performance at certain times is targeted. In applications where napping is scheduled, optimizing the probability of falling asleep may be the target objective function. Therefore, the objective function will be highly dependent on the type of schedules that need to be optimized.

Methods for Optimizing Countermeasure Interventions

Two ways to approach schedule optimization problems are determining optimal interventions and finding a schedule that meets all the required constraints. Model predictive control and calculus of variations are 2 mathematical methods used to obtain optimal intervention strategies. Model predictive control is a method that converts a mathematical model of a system into a controller for that system. The controller is then used with data measurement from the system to predict future states of the system. A benefit of model predictive control is that the methods are coupled with a vast literature from control theory. Applying the calculus of variation to the schedule optimization problem is a 2-step process. The 1st step is to use a mathematical model to derive an analytical equation to optimize, and the 2nd step is to solve the resulting equations with boundary value methods.

An important theoretical consideration when considering light as a potential countermeasure is type 0 circadian resetting. Type 0 circadian resetting refers to using the predicted singularity in limit cycle models of the circadian pacemaker to obtain shifts as large as 12 h following a correctly timed stimulus. Type 0 resetting has been observed in humans (Jewett et al., 2000) and in simulation studies (Klerman et al., 1996). Developing a method that could be exploited to achieve type 0 circadian resetting has obvious practical advantages.

To further this work of applying models to "real-world" applications, we suggest the following: 1) define a generalized set of schedule optimization problems, such as single "jet lag" sleep/wake shift, rotating shift work, or night-work schedule, that contain both biological and operational constraints; and 2) adapt objective functions to established problems in operational research literature. Once these are complete, the mathematical modeling community can draw on a large body of existing optimization research (Belegundu and Chandrupatla, 1999) to solve these problems.

An Example of Countermeasure Design

We now present an example of using this 3-feature approach. Since our approach to schedule design uses a model of the effect of light on the circadian pacemaker and models of neurobehavioral performance, we first introduce features of these models. As shown in Figure 3A, the Kronauer-Jewett model predicts that during an entrained standard day (8 h sleep, 16 h awake) and moderate light levels, the circadian phase of worst performance (red bars) will maintain a constant relationship to habitual wake and is located during the sleep episode. The model also predicts that performance is high during most of the waking day. The countermeasure schedule design step is to design light interventions so that performance is maximized during the waking day under nonstandard conditions. Due to the dynamic nature of the Kronauer-Jewett models, we can use light pulses to reset the phase of the circadian system. One example of when such an intervention would be desired would be under conditions of a 12-h shift in sleep-wake period where the light level is dim during waking. The model would predict that without intervention, the circadian time of worst performance would occur during the waking day, resulting in a substantial decrease in performance (Fig. 3B).

To redesign the schedule shown in Figure 3B with a countermeasure, a light pulse (for example, 4-h duration and 10,000 lx) was placed on each successive waking day such that entrainment to the new schedule was facilitated. The objective function was to minimize the difference between the predicted baseline phase angle and the phase angle for the shifted schedule. Phase angle is defined as the difference between the circadian minimum and habitual

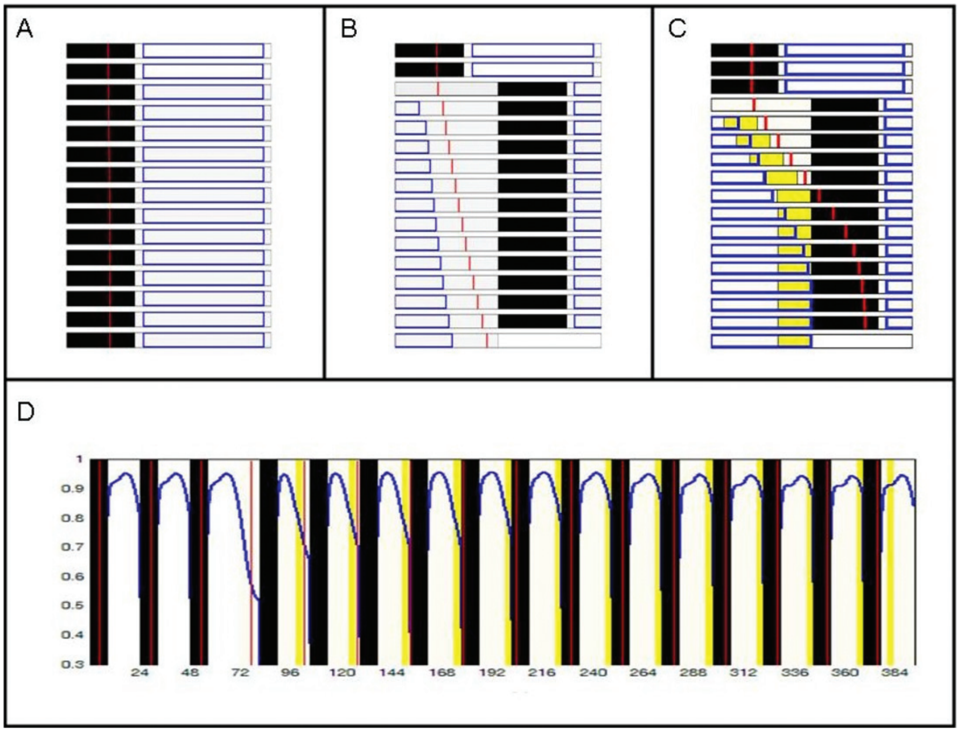

Figure 3. For all panels, raster plots of scheduled sleep (black), predicted circadian phase corresponding to poor performance (red), >90% level of predicted performance (blue), and light countermeasure (yellow). (A) A simulation of 13 days of 8 h of sleep and 16 h of wake is illustrated under entrained conditions. (B) A simulation of 2 days of 8 h of sleep and 16 h of wake, followed by 36 h of wakefulness, which results in a 12-h shift in the sleep period. (C) The schedule in B with an optimally placed countermeasure. (D) Details of the time course of performance over each of the waking days of C.

wake period. From a countermeasure design approach, we selected a discrete form of representation that allows the optimization methodology to separate countermeasure timing from countermeasure strength (Dean et al., 2005). The timing of each pulse was automatically determined with methods that iteratively place pulses of light relative to the predicted circadian phase minimum such that the desired shift in circadian phase is achieved (Dean et al., 2005). As illustrated in Figure 3C, the sample countermeasure intervention results in predicted circadian time of poor performance during sleep and improved performance during the waking day. There is a variation in the time course of the performance over each waking day (Fig. 3D), with higher performance soon after awakening. This pattern of performance is not maintained when the circadian system is not appropriately aligned with the sleep/wake cycle. This mathematical model would predict that the countermeasure increases the portion of the day operating above 90% of maximum performance from approximately 40% to 60%. In contrast, during standard conditions, the portion of the day operating above 90% of maximum performance is 74% of the waking day. Simulation studies suggest that proper scheduling of light pulses can result in substantial improvement of performance during the waking day, especially for the condition of low light levels.

To summarize, we have demonstrated through an example of a shift in sleep-wake that countermeasures can be designed automatically to improve the predicted performance level for a schedule. Once developed, tools that automatically design countermeasures will allow for mathematical models to be used not just to assess a schedule but to recommend an intervention and to design experimental protocols. Our work in schedule design illustrates that developing biomathematical models of performance and alertness is just the first step in making models useful in the operational settings.

DISCUSSION

Common to these 3 applications of mathematical models in the operational setting is the development of new methods and algorithms during the application process. New methods required are those for approximating inputs to the model, for associating neurobehavioral performance with risk, and for moving beyond schedule assessment to specific countermeasure intervention recommendations.

A common and major problem in the application outside of the research environment of all mathematical models is estimating sleep-wake information. Approaches to estimating sleep in the field include using a model of sleep dynamics such as those described in Achermann (2004), developing application-specific sleep estimators, and/or collecting sleep-wake data from individuals through sleep logs/diaries or an automated collection system such as activity monitors. A challenge to the community is the development of a sleep predictor model that is valid across a variety of homeostatic and circadian conditions, including chronic sleep restriction. As described above, some application-specific sleep estimators are already in use. Using a mathematical model to predict sleep timing allows for a consistent treatment of sleep estimation across applications and is easily incorporated into scheduling systems. However, the work of Kandelaars, Fletcher, Dorrian, et al. (2006) described above suggests that existing mathematical models for predicting sleep may not contain the necessary inputs for predicting sleep in a challenging sleep environment application such as transmeridian travel. Developing application-specific sleep prediction methods allows for the accurate sleep prediction algorithms at the cost of additional research efforts.

There are advantages and disadvantages to collecting individual sleep/wake data, especially in large field studies. Whereas the individual's data are potentially more accurate, problems arise in the practical handling of data and individual differences in how sleep is recorded. One approach for estimating sleep from individual data is the use of proprietary algorithms for estimating sleep-wake state from actigraphy. Despite the obvious advantages of actigraphy, there are open questions regarding the accuracy of existing sleep-wake algorithms across a wide range of populations (Ancoli-Israel et al., 2003; Stanley, 2003). Furthermore, even if existing actigraphy conversion algorithms are suitable for a particular application, proprietary software often requires for sleep periods to be identified by the inclusion of sleep-wake diary information, thereby continuing to require daily sleep diaries/logs and user cooperation. Another challenge to the community is the development of a publicly validated and automated algorithm for predicting sleep-wake from actigraphy.

An input unique to the Kronauer et al. model is light intensity during the waking day. A possible critique of the simulation example presented above could be that multiple-hour continuous light exposure conditions are not possible in the operational setting. However, there is evidence that using a small number of light intensity values is sufficient for obtaining informative simulation results (Rodriguez et al., 2002) and that continuous light is not required (Gronfier et al., 2004). A sensitivity study of the effect of light on performance predictions could provide needed guidance to the operational community in how to effectively employ the Kronauer model in practice.

Decision makers in the operational setting are usually more concerned with assessing risk than the neurobehavioral performance variables often predicted with existing mathematical models. The work of Hursh represents a long-term effort in developing a quantitative relationship between predicted performance values and risk of adverse advents. Hursh's work represents an achievement of a goal set forth by the modeling community in the 2002 Fatigue and Performance Modeling meeting. The existence of adverse events databases across many disciplines including physicians, health care workers, and shift workers provides an opportunity to develop appropriate statistical methods for associating both group and individual model predictions with risk.

Dean's work suggests a paradigm shift from schedule assessment to schedule design and suggests challenges and opportunities to the mathematical modeling community. The salient challenge is to demonstrate that the schedule optimization methods described in this article can be successfully applied to a realistic cross-section of operational environments. We predict that as schedule design methods become more established, a demand for mathematical models that predict the effect of pharmacological interventions including caffeine, melatonin, and modafinil will be created. The method described above for light countermeasures was designed to be adapted to make recommendations for these other interventions.

There are many open issues regarding the transition of mathematical models into the operational settings. These open issues can be both domain specific (to the individual operational setting) and intrinsic to the modeling process. Although the details of an operational validation process are beyond the scope of this article, one approach to validating operational models is a 2-stage process that ensures the models are first predictive of physiologically relevant measures and a second step that validates an association of those measures with risk. The modeling issues include mathematical definition and optimization methods. The work presented in this article demonstrates that an ongoing relationship between the modeling community and application experts is required to facilitate the transition of these models to operational settings in the real world.

CONCLUSION

The work in this article demonstrates that the fatigue and performance modeling community is actively working with the operational community to use mathematical models to inform decision makers. Integrating the use of mathematical models within the operational setting requires an adaptation process that involves estimating model inputs (e.g., sleep, light), modifying model outputs (e.g., predicting fatigue risk), developing new methods (e.g., countermeasure design, sleep estimation), and tailoring software tools for specific applications. Whereas schedule assessment was the obvious initial goal of each of the coauthors, sustained collaboration within different operational settings required model adaptation and algorithm development that suited the needs of the specific application. Sustained commitment between the modelers and operational personnel is required if mathematical models are to be used effectively and appropriately.

ACKNOWLEDGMENTS

The SIAM-SMB workshop was supported by the AFOSR, especially Dr. Willard Larkin's interest and questions. Dr. Fletcher's work was supported by ARC LPO214128 and ARC DPO0346651. Dr. Hursh's work was supported by the Federal Railroad Administration and the U.S. Department of Transportation. Mr. Dean's and Dr. Klerman's work was supported by NIH K02-HD045459 (EBK), AFOSR FA9550-06-0080, and NSBRI HPF00405.

REFERENCES

Achermann P (2004) The two-process model of sleep regulation revisited. Aviat Space Environ Med 75:A37-A43.

Akerstedt T (2003) Shift work and disturbed sleep/wakefulness. Occup Med 53:89-94.

Ancoli-Israel S, Cole R, Alessi C, Chambers M, Moorcroft W, and Pollak C (2003) The role of actigraphy in the study of sleep and circadian rhythms. Sleep 26:342-392.

Bagheri N, Stelling J, and Doyle J III (2005) Optimal phase-tracking of the nonlinear circadian oscillator. Proc Am Control 5:3225-3240.

Belegundu AD and Chandrupatla TR (1999) *Optimization Concepts and Applications in Engineering*. Upper Saddle River, NJ: Prentice Hall.

Belenky G, Balkin TJ, and Redmond DP (1998) Sustaining performance during continuous operations: The U.S. Army's sleep management system. In *Managing Fatigue in Transportation*, Hartley L, ed, pp 77-86, Oxford, Pergamon.

Borbély AA (1982) A two process model of sleep regulation. Hum Neurobiol 1:195-204.

Czeisler CA, Allan JS, Strogatz SH, Ronda JM, Sanchez R, Rios CD, Freitag WO, Richardson GS, and Kronauer RE (1986) Bright light resets the human circadian pacemaker independent of the timing of the sleep-wake cycle. Science 233(4764):667-671.

Dawson D and Fletcher A (2001) A quantitative model of work-related fatigue: background and definition. Ergonomics 44:144-163.

Dean DA II, Forger DB, and Klerman EB (2005) Designing optimal light intervention schedules for experimental and operational settings. Sleep Abstr Suppl 28:A69-A69.

Dean DA II and Klerman EB (2005) Using domain specific information to design optimal circadian adjustment schedules. Computational Physiology: From Genome to Physiome Symposium, San Diego, CA.

Fletcher A and Dawson D (1997) A predictive model of work-related fatigue based on hours of work. J Occup Health Safety Aust N Z 13:471-485.

Fletcher A and Dawson D (1998) A work-related fatigue model based on hours-of-work. In *Managing Fatigue in Transportation*, Hartley L, ed, pp 189-208, Oxford, Pergamon.

Fletcher A and Dawson D (2001) A quantitative model of work-related fatigue: empirical evaluations. Ergonomics 44(5):475-488.

Fletcher A, Lamond N, van den Heuvel C, and Dawson D (2003) Prediction of performance during sleep deprivation and alcohol intoxication by a quantitative model of work-related fatigue. Sleep Res Online 5(2):67-75.

Fletcher A, Roach GD, Lamond N, and Dawson D (2000) Laboratory based validations of a work-related fatigue model based on hours of work. In *Shiftwork in the 21st Century: Challenges for Research and Practice*, Hornberger S, Knauth P, Costa G, Folkard S, eds, pp 93-98, Frankfurt am Main, Peter Lang.

Forger DB and Paydarfar D (2004) Starting, stopping, and resetting biological oscillators: in search of optimum perturbations. J Theor Biol 230(4):521-532.

Gronfier C, Wright KP, Kronauer RE, Jewett ME, and Czeisler CA (2004) Efficacy of a single sequence of intermittent bright light pulses for delaying circadian phase in humans. Am J Physiol Endocrinol Metab 287(1)E174-E181.

Hursh SR, Balkin TJ, Miller JC, and Eddy DR (2004a) The fatigue avoidance scheduling tool: modeling to minimize the effects of fatigue on cognitive performance. SAE Trans 113:111-119.

Hursh SR, Redmond DP, Johnson ML, Thorne DR, Belenky G, Balkin TJ, Storm WF, Miller JC, and Eddy DR (2004b) Fatigue models for applied research in warfighting. Aviat Space Environ Med 75:A44-A53.

Jewett ME, Khalsa SBS, Kronauer RE, and Czeisler CA (2000) Light-induced suppression of endogenous circadian amplitude in humans. Nature 350:59-62.

Jewett ME and Kronauer RE (1999) Interactive mathematical models of subjective alertness and cognitive throughput in humans. J Biol Rhythms 14(6):588-597.

Kandelaars KJ, Fletcher A, Dorrian J, Baulk SD, and Dawson D (2006) Predicting the timing and duration of sleep in an operational setting using social factors. Chronobiol Int 23(6):1265-1276.

Kandelaars KJ, Fletcher A, Eitzen GE, Roach GD, and Dawson D (2006) Layover sleep prediction for cockpit crews during transmeridian flight patterns. Aviat Space Environ Med 77(2):145-150.

Klerman EB, Dijk DJ, Kronauer RE, and Czeisler CA (1996) Simulations of light effects on the human circadian pacemaker: implications for assessment of intrinsic period. Am J Physiol 270:271-282.

Knauth P (1997) Changing schedules: shiftwork. Chronobiol Int 14(2):159-171.

Kostreva M, McNelis E, and Clemens E (2002) Using a circadian rhythms model to evaluate shift schedules. Ergonomics 45:739-763.

Kronauer RE, Czeisler CA, Pilato SF, Moore-Ede MC, and Weitzman ED (1982) Mathematical model of the human circadian system with two interacting oscillators. Am J Physiol 242:R3-R17.

Kronauer RE, Forger DB, and Jewett ME (1999) Quantifying human circadian pacemaker response to brief, extended, and repeated light stimuli over the photopic range. J Biol Rhythms 14:500-515.

Mallis MM, Mejdal S, Nguyen TT, and Dinges DF (2004) Summary of the key features of seven biomathematical models of human fatigue and performance. Aviat Space Environ Med 75:A4-A14.

Monk TH, Fookson JE, Kream J, Moline ML, and Pollak CP (1985) Circadian factors during sustained performance, background and methodology. Behav Res Methods Instruments Comp 17:19-26.

Moore RY and Eichler VB (1972) Loss of a circadian adrenal corticosterone rhythm following suprachiasmatic lesions in the rat. Brain Res 42(1):201-206.

Mott C, Mollicone D, van Wollen M, and Huzmezan M (2003) Modifying the human circadian pacemaker using model based predictive control. Proceedings of the 2003 American Control Conference, June 4-6, 2003.

New South Wales Special Commission into the Waterfall Rail Accident (2003) *Special Commission of Inquiry into the Waterfall Rail Accident New South Wales* [electronic resource]. Sydney, Australia: The Enquiry:e.law, 6304-6401.

Pollard JK (1996) Locomotive engineer's activity diary. Final Report DOT/FRA/RRP 96/02. Washington, DC: U.S. Department of Transportation, Federal Railroad Administration.

Roach GD, Dorrian J, Fletcher A, and Dawson D (2001) Comparing the effects of fatigue and alcohol intoxication on locomotive engineers, performance in a rail simulator. J Hum Ergol (Tokyo) 30(1-2):125-130.

Rodriguez DM, Oyung RL, Barger LK, Mallis MM, Jewett ME (2002) Flight deck light exposure of pilots during long-haul trips between the United States and Japan. Sleep Abstr Suppl 25:A420-A421.

Samel A and Wegmann HM (1988) Sleep and circadian rhythms of an airline pilot operating on the polar route: a case study. Aviat Space Environ Med 59(5):443-447.

Samel A, Wegmann HM, Summa W, and Naumann M (1991) Sleep patterns in aircrew operating on the polar route between Germany and east Asia. Aviat Space Environ Med 62(7):661-669.

Smiley A (1998) Fatigue management: lessons from research. In *Managing Fatigue in Transportation*, Hartley L, ed, pp 1-23, Oxford, Pergamon.

Stanley N (2003) Actigraphy in human psychopharmacology: a review. Hum Psychopharmacol 18:39-49.

Van Dongen HPA (2003) The cumulative cost of additional wakefulness: dose-response effects on neurobehavioral functions and sleep physiology from chronic sleep restriction and total sleep deprivation. Sleep 26(2):117-126.

Van Dongen HPA (2004) Comparison of mathematical model predictions to experimental data of fatigue and performance. Aviat Space Environ Med 75:A15-A36.

Vespa S, Wylie D, Mitler M, and Shultz T (1998) Study of commercial vehicle driver rest periods and recovery of performance in an operational environment. In *Managing Fatigue in Transportation*, Hartley L, ed, pp 119-165, Oxford, Pergamon.