Commentary on the Three-Process Model of Alertness and Broader Modeling Issues

Jaques Reifman and Philippa Gander

Reifman J, Gander P. ***Commentary on the three-process model of alertness and broader modeling issues.*** **Aviat Space Environ Med 2004; 75(3, Suppl.):A84–8.**

IN ITS ORIGINAL FORM, the Sleep/Wake Predictor Model was a conceptually simple model that predicted group mean alertness based on the addition of the instantaneous values of three independent processes (9):

1. A sinusoidal circadian process (Process C, sometimes considered as two separately modifiable components, Process C during wake and Process C′ during sleep);
2. A homeostatic process that declines exponentially toward a minimum during wake (Process S) and increases exponentially toward a maximum during sleep (Process S′); and
3. A sleep inertia process, which results in depressed alertness on awakening, and then exponentially improves toward Process S over 3 h (Process W).

Hence, predicted alertness was estimated as the sum of C + S - W with 8 h assumed to provide sufficient time for full recovery of Process S′.

The original version of the model was derived from group mean alertness ratings on a visual analog scale, collected 6 times a day from 12 young volunteers in a 3-wk time-isolation study. For the first 3 d, the participants lived on a 24-h schedule, which was then gradually shortened to 22 h across the remaining 18 d of the study. The circadian and sleep-related components of alertness were derived by comparing the rhythm during entrained conditions and forced internal desynchrony. Regression analyses were then used to compare model predictions with self-reported alertness or fatigue ratings from studies examining normal sleep/waking schedules, total sleep deprivation for 64–75 h, and abnormal sleep/wake schedules in shift work.

Although it is described as a three-process model, Process W is typically omitted from simulations on the basis that the data being modeled do not address functioning during the first 3 h after awakening. In effect, the model is a version of the Borbély two-process model (5). A series of papers published between 1995 and 2000 (1–4,9–11) have explored the model's capacity to predict a range of different outputs, but in each case these involve changes to model parameters and, in some cases, to inputs. This model evolution is summarized in **Table I**.

During this evolution, basic assumptions of the model have been modified over time to improve model fit to the particular data set being modeled. For example, in the simulations of sleep length (3), the previously established set of model parameter values significantly underestimated the duration of daytime sleep episodes. This lead to the proposition that there was circadian modification of the Process S′ during sleep, in contradiction to the basic model assumption that alertness is the simple arithmetic sum S + C of the model outputs. The best-fitting simulations also used a different acrophase of Process C.

To model alertness across successive night shifts (10,11), a series of new assumptions and parameters were introduced. First, it was assumed that the time of waking up sets the phase of Process C. Again, this contradicts the basic model assumption that alertness is the simple arithmetic sum of the outputs of S + C. Second, a "first night compensation effect" was introduced because shift workers rated themselves as progressively more alert than would be predicted over the course of the first night shift. It was also necessary to model a general lowering of alertness on the second night shift, which was proposed to reflect a "cost" of the first night shift effect. Third, a "time on shift" effect was introduced, whereby on night shifts after the first, alertness ratings decreased over the course of each shift before showing a modest improvement at the end of the shift. These new assumptions and parameters are presumably relevant to data sets simulated with earlier formulations of the model, for example, changes in EEG alpha power across the night shift among truck drivers and locomotive engineers. However, the effects of adding these new parameters on model predictions for these other data sets have not been reported.

From the Telemedicine and Advanced Research Center, U.S. Army Medical Research and Materiel Command, Ft. Detrick, MD (J. Reifman); and the Sleep/Wake Research Center, Massey University Wellington Campus, Wellington, New Zealand (P. Gander).

Address reprint requests to: Jaques Reifman, Ph.D., who is a Senior Research Scientist, U.S. Army Medical Research and Materiel Command, MCMR-AT, 504 Scott St., Ft. Detrick, MD 21702-5012; reifman@tatrc.org.

TABLE I. SUMMARY OF THE EVOLUTION OF THE SLEEP-WAKE PREDICTOR MODEL.

Data Set (ref. #)	Measure Simulated	Process C Parameters	Model Output	Best-Fit Regression Model
Objective Sleepiness				
15 male truck drivers (1)	EEG alpha power during night drives	amplitude = 2.5 acrophase = 20:48	1/(S + C)	exponential $r^2 = 0.80$
11 male train drivers (1)	EEG alpha power during night drives	amplitude = 2.5 acrophase = 20:48		exponential $r^2 = 0.94$
8 males in time isolation (1)	EEG alpha power during 28 h without sleep	amplitude = 2.5 acrophase = 19:48		exponential $r^2 = 0.95$
Subjective Sleepiness				
8 males in time isolation (1)	KSS rating during 28 h with 2 h, 4 h, or 8 h sleep.	amplitude = 2.5 acrophase = 20:48	1/(S + C)	linear $r^2 = 0.70$
Objective Sleep Latency				
2 groups of 8 males in time isolation (2)	Sleep latency after wakefulness of 16–40 h.	amplitude = 2.5 acrophase = 20:48	S + C	exponential $r^2 = 0.88$
8 females in time isolation (2)	Sleep latency after wakefulness of 4–12 h	amplitude = 2.5 acrophase = 20:48		linear $r^2 = 0.65$
25 males working a rotating 3-shift system (2)	Sleep latency before or after different shifts	amplitude = 2.5 acrophase = 20:48		N/A errors in both directions, depending on shift
11 male train drivers (2)	Sleep latency before or after different shifts	amplitude = 2.5 acrophase = 20:48		N/A errors in both directions, depending on shift
Objective Sleep Length				
2 groups of 8 males in time isolation (3)	Sleep duration after wakefulness of 16–40 h	amplitude = 0.6 acrophase = 16:48	asleep at bedtime	linear $r^2 = 0.99$
25 males working a rotating 3-shift system (3)	Sleep duration before or after different shifts	amplitude = 0.6 acrophase = 16:48	awake at S′ + C′ = 14.2	linear, intercept correction + 0.1 $r^2 = 0.99$
11 male train drivers (3)	Sleep duration before or after different shifts	amplitude = 0.6 acrophase = 20:48		linear, intercept correction + 0.6 $r^2 = 0.95$
6 male normal day workers in time isolation (3)	Sleep duration after wakefulness of 4 h			
Auditory Vigilance Performance				
12 males in time isolation (4)	Hits on a 30-min task across 64 h of sleep deprivation	amplitude = 2.5 acrophase = 20:48	S + C	linear $r^2 = 0.79$
Alertness Across Shifts				
Parameter setting				
1,114 shift workers on 6 rotating shift systems (four 8-h shift systems, two 12-h shift systems) (10)	Retrospective subjective sleepiness (KSS)	amplitude = 2.5 acrophase = 16:48	S + C	linear $r^2 = 0.34$ to 0.59
		amplitude = 2.5 acrophase = (wakeup + 9 h)		linear $r^2 = 0.63$ to 0.88
Parameter setting				
23 workers on two rotating 12-h shift systems (10)	2-hourly alertness ratings over 28 days	amplitude = 2.5 acrophase = 16:48	S + C	linear $r^2 = 0.23$ to 0.35
		amplitude = 2.5 acrophase = (wakeup + 9 h)		linear $r^2 = 0.40$ to 0.42 $r^2 = 0.40$ to 0.42
		amplitude = 2.5 acrophase = (wakeup + 11 h)		linear $r^2 = 0.47$ to 0.50

TABLE I. CONTINUED.

Data Set (ref. #)	Measure Simulated	Process C Parameters	Model Output	Best-fit Regression Model
Predicting Accident Risk 3 studies of 8-h shift systems (no data on sleep times, modelers "assumed" them) (11)	accident risk on afternoon and night shifts versus morning shift	amplitude = 2.5 acrophase = (wakeup + 9 h)	mean 1/(S + C) across shift	linear $r^2 = 0.89$
	accident risk across 4 night shifts	amplitude = 2.5 acrophase = (wakeup + 9 h)	mean 1/(S + C) across shift + 1st night effect + 2nd night effect + time on task component	linear $r^2 = 0.97$

General Modeling Issues and the Sleep/Wake Predictor Model

Many of the limitations of the current generation of models of neurobehavioral function may have a common explanation. They are all, in one form or another, variations of Borbély's seminal two-process model, which was originally developed solely to explain sleep regulation (5). Borbély's model was not formulated to predict human fatigue and performance, let alone trends in accident risk, which are the outcomes that the current generation of models are trying to predict. Hence, it is unlikely that model enhancements and refinements around Borbély's original model will provide the answers and the new insights needed to address the well-known research gaps in predicting human performance subject to sleep deprivation across the circadian cycle (7,8). Given the common theoretical basis of the current models, it is also not surprising that there is not great variability in model performance—one model performing better in one scenario and worse in another (16).

Following, we discuss four modeling issues: 1) model validation and refinement; 2) predicting group vs. individual performance; 3) model prediction uncertainty; and 4) scaling of model outputs within the context of the Sleep/Wake Predictor Model. However, as articulated above, these issues are prevalent and equally relevant throughout the current generation of sleep and performance models.

Model Validation and Refinement

The Sleep/Wake Predictor Model does not follow a conventional life cycle development protocol of systematic development, validation, and refinement leading to a comprehensive explanation of the underlying mechanisms of sleepiness and alertness. Rather, its evolution appears unsystematic, with ad-hoc adjustments made as necessary, and without reference to formal hypotheses to account for observed model inaccuracies. Whenever the existing model fails to explain new experimental data, new components are added and model parameters are fitted to the new data in order to minimize the differences between "predicted" and observed values. This is potentially a never-ending process, with the model being continually adjusted as new field and laboratory data become available. For example, as discussed above, to enhance model predictive power for the trend in accident risk among rotating shift workers, Folkard et al. (10,11) assumed that the phase of Process C was variable and dependent on the time of waking. Similar ad hoc adjustments are made in other models. For instance, in the SAFTE Model, Johnson et al. (12) modified the reservoir concept to avoid performance over predictions during the recovery phase of individuals who are subject to chronic sleep restriction.

While this approach allows for a model to be tailored to the specific dataset of interest, it provides no insight on the model's predictive power or generalization ability. There is no way of knowing whether the refined model has an improved ability to generalize to new, unanticipated conditions and scenarios. Furthermore, the tailored model may no longer be predictive for data sets simulated with earlier formulations of the model. As pointed out above, it is not clear if the tailoring of the Sleep/Wake Predictor Model to predict alertness across successive night shifts (10,11) is applicable to earlier data sets. Models derived in this fashion cannot be characterized quantitatively, so it cannot be determined, a priori, to which set of scenarios/conditions the models are applicable and to which they are not. Just knowing when a given model should or should not be employed would be extremely useful.

To improve understanding of the underlying neurobehavioral mechanisms, it is more informative to examine how model predictions differ systematically from observations, rather than modifying or adding parameters to improve the fit. In this context, the use of regression models to describe overall fit is unhelpful without consideration of the structure of the residuals. Furthermore, R^2 statistics do not identify at which times of day model predictions are weakest, although this may be of key theoretical and operational significance.

Group Prediction vs. Individual Prediction

Another characteristic of the Sleep/Wake Predictor Model that is common across the other models is the focus on the prediction of group averages, necessarily de-emphasizing individual variation in neurobehav-

ioral functioning. This focus is inconsistent with the recent trend for models to move from being basic research tools to practical field-applicable tools used to optimize work scheduling. Because measures of performance may vary substantially from individual to individual, even if a model is capable of accurately predicting mean group performance, this capability would be of limited benefit without knowing how this translates into predicting the performance of a given individual. In many situations, such as mission-critical operations prevalent in the military environment and in commercial aviation, the emphasis is on the performance of one or two key individuals whose performance could be substantially different from those of a group. It would be extremely valuable if the models could predict individual neurobehavioral function and/or estimate the range of inter-individual variations around group predictions, i.e., estimate the prediction error. However, these capabilities may require a completely new modeling paradigm (6,15) and much greater understanding of sleep and human behavior.

One potential approach to model individual performance that would be highly applicable to the military environment is to develop personal models based on individualized on-line measurements of physiologic and non-physiologic variables, such as core body temperature, levels of light exposure, and sleep/wake history (15). This approach may require the coupling of existing parametric models with nonparametric models, such as artificial neural networks, in the development of hybrid models that in addition to being amenable to predicting individual performance could also quantitatively assess the reliability of model predictions through estimation of statistical error bounds. In addition to the intrinsic large data requirements posed in the development of hybrid individualized fatigue and performance models, a related issue is how a personal model handles missing data (for example, sleep times). This is a common occurrence in real-world applications, and the substitution of group data may be inadequate for predicting individual performance.

Estimating Prediction Uncertainties

The applicability of prediction models, be it for predicting group-average or individual performance, is of limited benefit if the model does not also provide a theoretically sound approach for quantitatively estimating the precision of the model predictions for new data for which the outcomes are not known. Thus, in addition to the predicted model value, the model should employ algorithms that simultaneously provide statistical error bounds (such as standard error, confidence intervals, and prediction intervals) to quantitatively determine the bounds within which the predictions may be trusted for a pre-defined coverage probability, e.g., $\pm \epsilon$ with 95% confidence. The algorithms should explicitly allow for the incorporation of the distribution of the data employed in model development (i.e., the "training" data) in the estimation of error bars for new data. Hence, when a prediction is made in a region where the training data are sparse, the error bounds should be wide, indicating less reliability in the resulting prediction. In contrast, in regions where the training data are dense, the bounds should be narrower because the model would have been exposed to similar patterns during model development. This would allow for a quantitative assessment of the accuracy of the model predictions, affording their use when error bounds are within pre-specified range and avoiding their usage when the quality of the prediction deteriorates.

Unfortunately, the nature of the modeling approach employed in the Sleep/Wake Predictor Model as well as the other neurobehavioral function models does not lend itself to the evaluation of prediction uncertainties. Formal statistical methods are not routinely employed in developing the models and, therefore, the uncertainty in inferences based on differential equation models and algebraic models and their sensitivity to model specification and parameter estimation error cannot be evaluated (6). Nevertheless, the model by Moore-Ede et al. (14) attempts to infer lower and upper limits about the model prediction. These limits are obtained by employing a proprietary database, rerunning the model with individuals from the database with different "conditions," and associating the minimum and maximum predicted alertness values for any individual in the database with the lower and upper limits, respectively. Due to a lack of statistical underpinning in this approach, it is unlikely that the suggested limits around the predictions are meaningful. None of the other models presented at the workshop, however, attempted to provide an assessment of the accuracy of the model predictions.

Research is needed to address this fundamental issue and provide statistically based algorithms that make quantitative, objective statements of the reliability of the model predictions. Model reliability cannot be assured if reliability cannot be quantified. This quantification of model reliability might hold the key for enabling the use of neurobehavioral function models for practical, mission-critical applications. Again, achieving this capability might require developing models from scratch based on modeling paradigms that lend themselves to these functionalities.

Scaling of Model Outputs

The outputs of the neurobehavioral functional models are internally scaled into arbitrary numerical units, which are subsequently mapped or normalized to different metrics corresponding to different subjective and objective measures of alertness and performance. Because each model output is normalized to its own arbitrary subjective and objective metric, the models cannot be directly compared with each other and neither can they be directly compared with performance test measurements, such as the Psychomotor Vigilance Task (PVT) and the Karolinska Sleepiness Scale (KSS). It is not clear how these normalizations are performed and validated and what sort of mapping function, e.g., linear, nonlinear, is employed. It is intriguing and disturbing, however, that a model not designed (i.e., normalized) to predict subjective alertness outperforms models tuned to do so in certain applications [see the subjective alertness results of Scenario 2 in Van Dongen

(16)]. The larger issue, though, is whether the outputs provided by the model (e.g., subjective alertness, objective alertness) are appropriate measures of the outcome that the model is trying to predict (e.g., accident risk, functional field performance) (13). This is of particular concern as the Sleep/Wake Predictor Model (and other models) are being proposed to predict accident risk when, in fact, the links between sleep regulation, alertness, and work performance are not well established. Clearly, there is a need to standardize the metrics that models are mapped to and to validate the relationship between the output measures provided by the models and the actual field performance measures of interest.

ACKNOWLEDGMENTS

The authors thank the reviewers for their constructive feedback. Jaques Reifman was supported by the Combat Casualty Care and Military Operational Medicine research programs of the U.S. Army Medical Research and Materiel Command, Ft. Detrick, MD. Philippa Gander was supported by the University of Otago, the Wellington Medical Research Foundation, and the Maurice and Phyllis Paykel Trust.

The opinions or assertions contained herein are the private views of the authors and are not to be construed as official or as reflecting the views of the U.S. Army or the U.S. Department of Defense.

REFERENCES

1. Åkerstedt T, Folkard S. Validation of the S and C components of the three-process model of alertness regulation. Sleep 1995; 18:1–6.
2. Åkerstedt T, Folkard S. Predicting sleep latency from the three-process model of alertness regulation. Psychophysiology 1996; 33:385–9.
3. Åkerstedt T, Folkard S. Predicting duration of sleep from the three-process model of alertness regulation. Occup Environ Med 1996; 53:136–41.
4. Åkerstedt, T, Folkard S. The three-process model of alertness and its extension to performance, sleep latency, and sleep length. Chronobiol Int 1997; 1997:14:115–23.
5. Borbély AA. A two process model of sleep regulation. Hum Neurobiol 1982; 1:195–204.
6. Brown EN, Luithardt H. Statistical model building and model criticism for human circadian data. J Biol Rhythms 1999; 14: 609–16.
7. Dinges DF, Achermann P. Commentary: Future considerations for models of human neurobehavioral function. J Biol Rhythms 1999; 14:598–601.
8. Dinges DF. Critical research issues in development of biomathematical models of fatigue and performance. Aviat Space Environ Med 2004; 75(3, Suppl.):A181–91.
9. Folkhard S, Åkerstedt T. A three-process model of the regulation of alertness-sleepiness. In: Broughton RJ, Ogilvie R, eds. Sleep, Arousal and Performance: Problems and Promises. Boston: Birkhäuser; 1992:11–26.
10. Folkard S, Åkerstedt T, Macdonald I, et al. Beyond the three-process model of alertness: estimating phase, time on shift and successive night effects. J Biol Rhythms 1999; 14:577–87.
11. Folkard S, Åkerstedt T, Macdonald I, et al. Refinement of the three-process model of alertness to account for trend in accident risk. In: Hornberger S, Knauth P, Costa G, Folkard S, eds. Shiftwork in the 21st Century. Frankfurt: Peter Lang; 2000:49–54.
12. Johnson MJ, Belenky G, Redmond DP, et al. Modulating the homeostatic process to predict performance during chronic sleep restriction. Aviat Space Environ Med 2004; 75(3, Suppl.): A141–6.
13. Klerman EB, Jewett ME. Commentary. Model building, quantitative testing, and model comparison. J Biol Rhythms 1999; 14: 621–4.
14. Moore-Ede M, Heitmann A, Croke D, et al. Circadian alertness simulator for fatigue risk assessment in transportation: Application to reduce frequency and severity of truck accidents. Aviat Space Environ Med 2004; 75(3, Suppl.):A107–18.
15. Reifman J. Alternative methods for modeling fatigue and performance. Aviat Space Environ Med 2004; 75(3, Suppl.):A173–80.
16. Van Dogen HPA. Comparison of mathematical model predictions to experimental data of fatigue and performance. Aviat Space Environ Med 2004; 75(3, Suppl.):A15–36.