Validation of Fatigue Modeling Predictions in Aviation Operations

Kevin Gregory, San Jose State University Research Foundation, kevin.b.gregory@nasa.gov

Siera Martinez, San Jose State University Research Foundation, siera.b.martinez@gmail.com

Erin Flynn-Evans, NASA Ames Research Center, erin.e.flynn-evans@nasa.gov

Problem:

Bio-mathematical fatigue models that predict levels of alertness and performance are one potential tool for use within integrated fatigue risk management approaches. A number of models have been developed that provide predictions based on acute and chronic sleep loss, circadian desynchronization, and sleep inertia. Some are publicly available and gaining traction in settings such as commercial aviation as a means of evaluating flight crew schedules for potential fatigue-related risks. Yet, most models have not been rigorously evaluated and independently validated for the operations to which they are being applied and many users are not fully aware of the limitations in which model results should be interpreted and applied.

Method:

We are comparing the predictions generated from fatigue models to actual alertness and performance data. The evaluation includes five laboratory and field data sets that encompass a wide range of imposed sleep schedules. The model predictions utilize algorithms based on laboratory and field study data including measures of performance, circadian phase, and temperature nadir.

Presented results are from a data set of 44 short-haul commercial aviation pilots using the commercially-available Fatigue Avoidance Scheduling Tool (SAFTE/FAST, version 3.2.0.1T). Data from this pilot group included actigraphy, sleep diary sleep/wake history, and performance measured by the psychomotor vigilance task (PVT), a 5-minute reaction time test completed up to 3 times/day. On duty days, the PVT was completed after waking, in-flight prior to top of descent, and post-duty. Standard outcome metrics from the PVT include mean reaction time (RT), number of lapses (responses > 500 ms) and the inverse mean reaction time (1/RT) (Basner, Dinges, 2011).

For comparison purposes, time points from the model predictions were matched against timing for PVT sessions from the pilot data set.

The pilots worked a fixed-pattern duty schedule with a baseline block (baseline) of five days of short duty hours followed by four days off, five early duty (early) followed by three days off, five daytime starts with many sectors (midday) followed by three days off and then five late duties with finishes that generally ended during the night (late) followed by four days off (Figure 1).

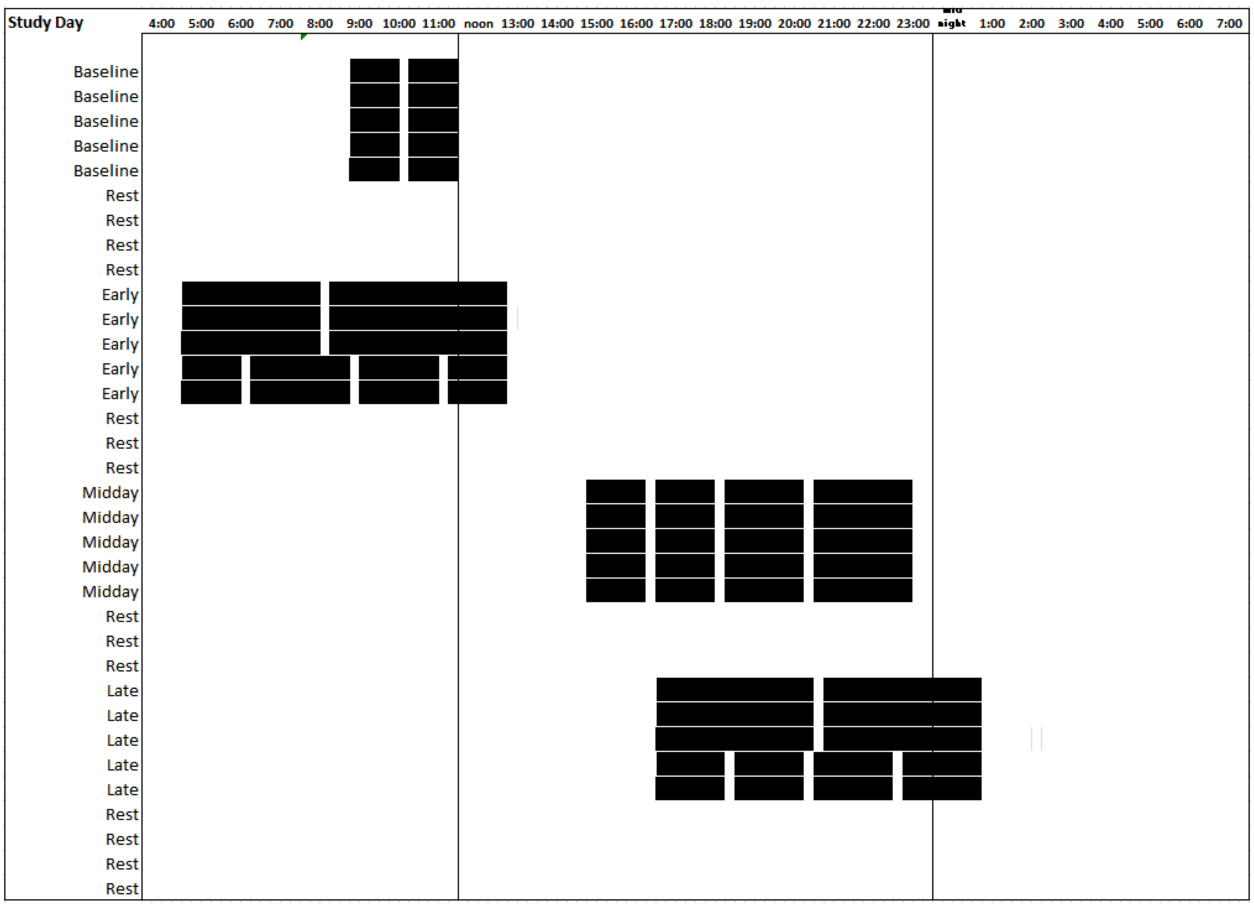

Figure 1. Pilot Schedule of Flight Duty. Rest=Day Off; Baseline = short sectors, short duty, variable start time; Early = early departure, 2-4 sectors; Midday = midday departure, heavy workload of generally 4 sectors; Late =late arrival, 2-4 sectors including some long flights

Sleep diary data was used as input to the model. As not all sleep information available for each subject was contiguous, in addition to running the data in its raw form, a more continuous dataset for each subject was created to include single-day imputations of sleep period timing. Imputations were taken by averaging the sleep or wake times from surrounding days of the same duty block. Complete contiguous data was available for one-third of the subjects while those with missing data averaged less than 3 gaps in their data sets. Runs were also completed with and without self-identified nap information included.

Primary output from SAFTE/FAST is presented in terms of changes in cognitive effectiveness, expressed as a percent of well-rested baseline performance such that a value of 100 is a predicted result of performing at an equivalent level as to when a well-rested state. The first three days of output were used to establish a baseline for the model and were not considered during the analysis of the model output.

Results:

Comparing the raw data mean effectiveness score against the mean 1/RT allows for a visual assessment in which each measure is oriented in the same manner, with higher values representing better performance (Figure 2).

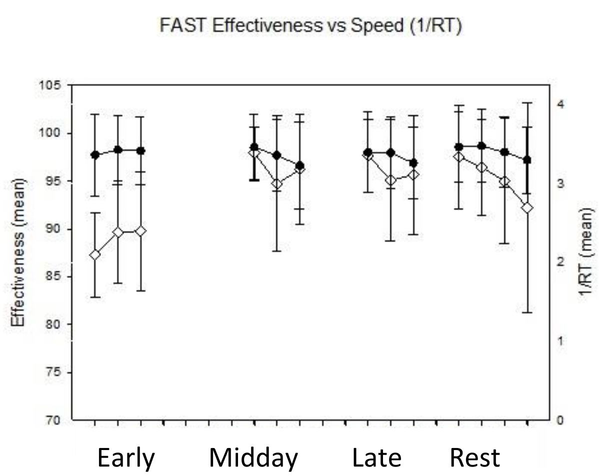

Figure 2. Mean effective and 1/RT values calculated for daily time bins (0300-0600, 0600-1200, 1200-1800, 1800-2400, 2400-0300) for early, midday, late and rest schedule periods. Open diamonds = model predictions, filled circles = response speed.

In Figure 3, SAFTE/FAST results based on both raw and imputed values are compared against the mean 1/RT.

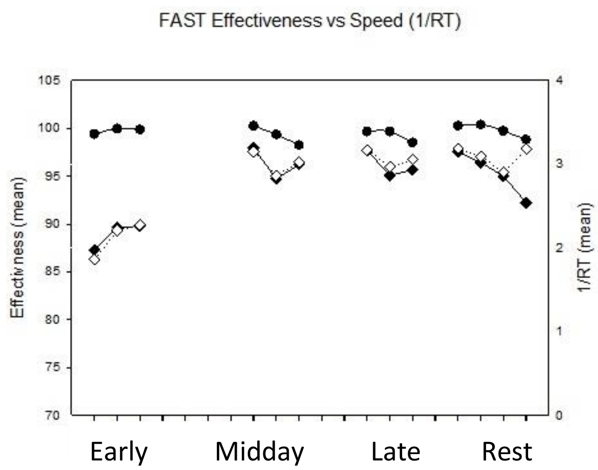

Figure 3. Mean effectiveness score from raw and imputed data plotted against 1/RT values calculated for daily time bins (0300-0600, 0600-1200, 1200-1800, 1800-2400, 2400-0300) for early, midday, late and rest schedule periods. Filled diamonds = model predictions (raw), open diamonds = model predictions (imputed), filled circles = response speed.

Analytical techniques that are being further explored include non-linear mixed models to better assess the time-related effects of predicted and actual circadian phase.

Discussion:

Technology-based tools such as models are envisioned as potentially powerful mechanisms for managing fatigue in complex work environments. Aviation operations can challenge flight crew with early morning starts and late nights, limited opportunities for rest, time zone changes and workload stresses. For these tools to provide appropriate guidance they must be able to accurately model the interaction of these physiological factors in such settings.

While based on similar physiological principles, different models offer unique attributes that may provide better application to different operational scenarios. In our current evaluation, the SAFTE/FAST model predicted some aspects of the studied aviation operation well, and not so well for other aspects. We anticipate that this general finding will stay consistent through the complete evaluation of all models and all data sets.

Summary:

Our evaluation will provide researchers and safety management personnel with a comprehensive understanding of the capabilities and limitations of fatigue management modeling tools.

Reference:

Basner M; Dinges DF. Maximizing sensitivity of the psychomotor vigilance test (PVT) to sleep loss. *SLEEP* 2011;34(5):581-591