RESEARCH ARTICLE

# Alertness Management in Aviation Operations: Enhancing Performance and Sleep

Mark R. Rosekind, Kevin B. Gregory, and Melissa M. Mallis

Rosekind MR, Gregory KB, Mallis MM. *Alertness management in aviation operations: enhancing performance and sleep.* **Aviat Space Environ Med 2006; 77:1256–65.**

***Introduction:*** Fatigue is an acknowledged safety risk in diverse operational settings. As a result, there has been growing interest in developing and implementing activities to improve alertness, performance, and safety in real-world operations where fatigue is a factor. ***Methods:*** A comprehensive Alertness Management Program (AMP) that included education, alertness strategies, scheduling, and healthy sleep was implemented in a commercial airline. An operational evaluation was conducted with 29 flight crewmembers, first when flying a standard schedule without AMP components (i.e., standard condition) compared with full AMP implementation, which included flying an innovative schedule that incorporated physiological sleep and alertness principles (i.e., intervention condition). The evaluation included objective measures of sleep quantity (actigraphy), psychomotor vigilance task (PVT) performance, and subjective reports of daily activities and sleep. ***Results:*** The results showed that the 3.5-h educational CD improved pre-education test scores from an average 74% correct to a post-education average of 98%. Alertness strategies showed minimal changes, though the daily diary did not allow for refined evaluation of duration, frequency, and timing of use. The intervention condition was associated with significantly more sleep (1 h, 9 min; $p < 0.01$) during the trip period compared with the standard schedule. All performance metrics showed significantly better performance during the intervention condition trip schedule ($p < 0.01$) compared with the standard condition. ***Discussion:*** This first-ever evaluation of a comprehensive AMP showed significantly improved knowledge, support for the use of alertness strategies, and increased sleep and performance during actual operations. The robust and consistent findings support the use of an AMP approach to effectively manage fatigue in operational settings.

**Keywords:** education, alertness strategies, innovative scheduling, actigraph sleep, PVT performance, operational evaluation.

FATIGUE IS ACKNOWLEDGED as a significant safety concern in diverse operational settings, including all modes of transportation, healthcare, public safety, and other 24/7 shift work environments (3,11,14). Efforts have expanded beyond documenting the effects of fatigue to implementing strategies, activities, and programs to minimize known fatigue-related risks and enhance performance, alertness, and safety (7). Given the complexity of fatigue in operational settings, a comprehensive Alertness Management Program (AMP) provides a multi-component approach that addresses different aspects of fatigue (1,13). A comprehensive AMP includes, at a minimum, activities focused on education, alertness strategies, healthy sleep (i.e., addressing sleep disorders), and scheduling.

A U.S. domestic airline engaged us to design and support implementation of a comprehensive AMP. A critical element of the activities included an operational evaluation to determine the effectiveness of the AMP to improve performance, sleep, and other relevant measures during actual flight operations. The education, alertness strategies, and healthy sleep components of the comprehensive program were implemented and distributed on a CD. The scheduling component involved a scientific evaluation of current scheduling practices and policies, development of potential scheduling innovations to address known fatigue factors, and the collection of objective data during actual flight operations that involved the proposed innovative schedules.

All components of the comprehensive AMP were implemented in a group of 29 flight crewmembers, who were intensively evaluated to determine specific outcomes related to education, the use of alertness strategies, and operating innovative flight schedules. Based on the outcomes of this initial evaluation, the airline planned to determine whether to expand implementation of the comprehensive AMP activities throughout the company's flight operations. This paper provides a summary of the AMP component activities and the results of the intensive operational evaluation that included objective performance and sleep measures, and subjective reports of daily activities, sleep, and use of alertness strategies. In this project, data were first collected with the 29 flight crewmembers during current, standard operations without focused education on sleep and alertness issues, without information on alertness strategies and healthy sleep, and when flying regular schedules. All elements of the AMP were then implemented with the 29 flight crewmembers, including education, alertness strategies, healthy sleep, and the operation of innovative flight schedules. The same data and measures were then collected again. The primary goal of this project was to determine if sleep and performance would be significantly enhanced with im-

From Alertness Solutions, Cupertino, CA.

This manuscript was received for review in May 2006. It was accepted for publication in August 2006.

Address reprint requests to: Mark R. Rosekind, Ph.D., Alertness Solutions, 1601 S. DeAnza Blvd., Ste. 200, Cupertino, CA 95014; bealert@alertsol.com.

plementation of a comprehensive AMP compared with standard operations without the AMP.

## METHODS

At the time comprehensive AMP activities were implemented, the airline operated primarily U.S. domestic flights, with some international destinations, in one aircraft type with about 925 pilots. On a daily basis, the airline operated approximately 288 flights to 31 destination cities across the United States and internationally. An introductory DVD describing the AMP was created and distributed to all 925 pilots. The 17-min DVD provided a brief introduction to fatigue as a safety issue in flight operations and included an overview of the planned AMP activities and objectives. Senior leadership and flight operations personnel, pilots, training specialists, and our scientists delivered the information in the DVD. After distribution of the introductory DVD, a web-based background survey was implemented with the entire airline pilot population, and then a subgroup of 29 pilots participated in the full implementation and evaluation of the comprehensive AMP.

### *Web-Based Background Survey*

To explore alertness-related issues among the entire airline pilot population, we conducted a confidential and anonymous web-based survey. An e-mail was sent to all 925 of the airline pilots that invited their participation in the online survey. This background survey included 19 general questions about fatigue and alertness, and 10 basic knowledge questions. At the end of the survey, pilots interested in participating in the AMP component of the project volunteered for the intensive data collection phase.

### *Education*

The comprehensive AMP, which included education, alertness strategies, healthy sleep, and scheduling, was implemented with a group of 29 volunteer pilots. The educational component involved the development of a 3.5-h interactive multimedia CD with all of the educational content, which allowed flexibility and self-pacing for pilots to complete on laptops. The educational CD included a method to determine compliance related to both reviewing material as well as completing quizzes. The six individual educational modules covered: a general introduction to the educational activities; sleep basics; circadian basics; aviation and fatigue; alertness strategies; and healthy sleep. The modules varied in length depending on the content provided. Pilot knowledge acquisition was evaluated with six quizzes: a 20-question quiz administered before and then after completing all of the 6 modules and 5-question quizzes following modules 2–5.

### *Alertness Strategies*

The alertness strategies component was delivered through a 27-slide module included in the educational CD. The alertness strategies module provided information about 10 strategies and guidance on their use as preventive and operational countermeasures (15). Prior to receiving any educational content, the overall use of alertness strategies was examined through questions on the web-based background survey, and their use during actual flight operations was documented through daily diary information collected while operating the standard schedule. After completing the alertness strategies educational component, strategy use during actual flight operations was again documented through daily diary information collected while operating the innovative schedule.

### *Healthy Sleep*

The healthy sleep component was delivered as part of a 28-slide module included in the educational CD. The healthy sleep component module provided an introduction to six specific sleep disorders, such as sleep apnea, and included information about their diagnosis and treatment. Tools were included that provided guidance on addressing sleep disorders, and referral resources were identified for follow-up questions and possible evaluation. However, due to medical privacy and confidentiality issues, no data were collected on the number of individuals that pursued further evaluation or the outcomes of those actions.

### *Scheduling*

The scheduling component involved several activities. First, we examined system fatigue factors in current scheduling policies and practices of the airline. Based on this analysis, specific schedules were identified that represented known fatigue factor risks. For example, a common trip schedule that was identified involved an early morning report time, an early transcontinental day flight, and a day sleep period followed by a return night flight with a subsequent repeat of this sequence. **Fig. 1**, top, portrays an actual, standard schedule regularly flown that represents this pattern. The fatigue factors associated with this schedule in-

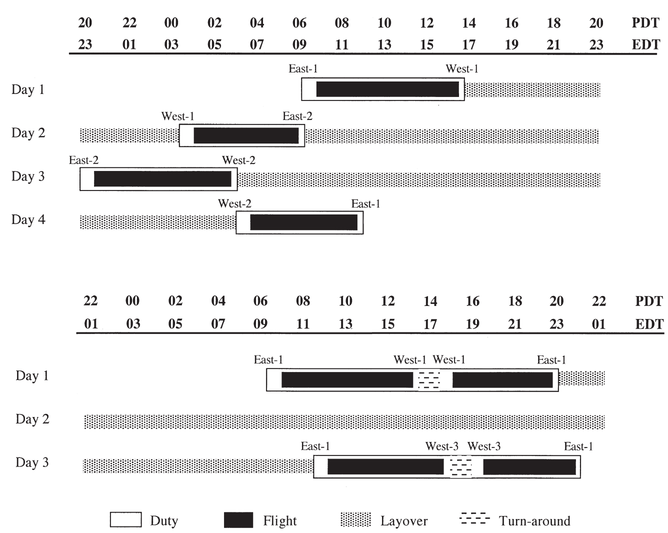


**Fig. 1.** Top: standard condition schedule. Different numbers paired with location labels indicate different cities. For example, East 1 and East 2 are different cities in the eastern time zone. Bottom: intervention condition schedule.

cluded early start times and acute sleep loss, the requirement for multiple day sleep periods and the resultant cumulative sleep debt, and operating multiple flights during a window of circadian low. This pattern resulted from a combination of regulatory requirements (i.e., 8 h scheduled flight time, 9 h minimum rest) and airline route structure (i.e., transcontinental flights).

Second, this specific trip schedule was used as a basis to develop an innovative schedule that would address the identified fatigue factors. Therefore, an innovative schedule was developed that reflected known physiological sleep/alertness principles and included: day flying only, night sleep only, sleep in the home time zone, optimal circadian phase for sleep and duty, and maintaining established duty limits (i.e., continuous hours of wakefulness). The schedule still involved an initial early report time and a potential acute sleep loss and required a longer flight time on a daily basis. This innovative schedule is portrayed in Fig. 1, bottom. A comparison of flight parameters for the standard and innovative schedules is portrayed in **Table I**.

Third, an operational evaluation was conducted after all four components of the AMP were implemented to determine whether the AMP produced the intended performance and sleep benefits. Therefore, to evaluate the overall effects of the AMP, objective data were first collected during actual flight operations on the standard schedule (i.e., standard condition) and prior to receiving any of the education, alertness strategies, or healthy sleep information. Data were collected across 2.5 d pre-trip (baseline), the 4-d trip, and on 3 d post-trip (recovery). The same data were then collected after completing the 3.5-h educational module that included education, alertness strategies, and healthy sleep components and while operating the innovative schedule (i.e., intervention condition). During the intervention condition, data were collected 2.5 d pre-trip (baseline), throughout the 3-d trip, and during 2 post-trip (recovery) days. During the baseline and recovery periods of both the standard and innovative conditions, the crewmembers were free of any company-related duty requirements. The intervention condition provided an evaluation of the comprehensive AMP implemented in the group of 29 flight crewmembers.

Volunteers gave written informed consent for all project activities. All data collected from evaluation of the education component, the use of alertness strategies, and the intensive monitoring with actigraph and personal digital assistant (PDA) were confidential. Although this project was privately funded and thus was not required to conform to the U.S. Code of Federal Regulations regarding human subjects, the protocol met the conditions of the 1964 Declaration of Helsinki. The anonymity of specific individual results was maintained by using randomly assigned ID numbers, and all data are presented in group summary format only.

TABLE I. FLIGHT PARAMETERS BY STANDARD VS. INTERVENTION CONDITIONS.

| | Flight (h) | Duty (h) | Layover (h) |
|---|---|---|---|
| Standard | | | |
| East 1 - West 1 | 6:25 | 7:40 | 10:30 |
| West 1 - East 2 | 4:55 | 5:55 | 13:55 |
| East 2 - West 2 | 5:55 | 6:55 | 23:55 |
| West 2 - East 1 | 5:00 | 6:00 | — |
| Intervention | | | |
| East 1 - West 1 - East 1 | 10:35 | 12:50 | 32:25 |
| East 1 - West 3 - East 1 | 9:30 | 11:45 | — |

Different numbers paired with location labels indicate different cities. For example, East 1 and East 2 are different cities in the eastern time zone.

Data collection during the intensive monitoring phase involved approaches based on NASA methodology (12,16) and included actigraphy, a PDA (in this case a Palm Zire 21) programmed for a 10-min psychomotor vigilance task (PVT), and a subjective daily diary. All participants received training over the phone regarding the use of each device. This training required participants to follow specific instructions and demonstrate successful use of the equipment. The pilots wore the wrist actigraphs on their non-dominant wrists on a 24-h basis throughout the pre-trip (baseline), trip, and post-trip (recovery) periods. The actigraphs were used as a valid and reliable way to monitor sleep/wake cycles and objectively determine sleep quantity and quality (17). The Actiwatch 64 actigraph (Mini-Mitter/Respironics Co., Inc., Bend, OR) was used and pilots were instructed to push the actigraph event marker button to indicate bedtimes, wake times, and when the actigraph was taken on or off. The data from actigraphs were uploaded onto a computer and analyzed to yield such metrics as sleep latency, total sleep time, and sleep efficiency.

The PVT, a simple, visual reaction time (RT) task, is not dependent on aptitude or skill level and does not have a learning curve. It is sensitive to even small amounts of sleep loss and yields informative metrics on the capacity for sustained attention and vigilance (4–6). PalmPVT software developed at the Walter Reed Army Institute of Research (18) was loaded on the PDA and used to administer the PVT. The conventional 10-min PVTs were completed three times on non-trip days: ~2–3 h after wakeup, ~6 h after wakeup, and ~12 h after wakeup. On trip days, the 10-min PVT was completed in flight within ~60 min after takeoff [after top of climb (TOC)] and within ~60 min prior to landing [before top of descent (TOD)]. These two PVT trials were the only in-flight measures that involved crewmember action. There were four primary outcome measures that were analyzed from the PVT data. They included: 1) mean reaction time (1/RT); 2) duration of lapse domain, which refers to shifts in lapse duration calculated from the reciprocal of the 10% slowest RTs; 3) optimum response speed, which is the average of the reciprocal of the 10% fastest RTs per trial and reflect the best performance an operator is capable of producing; and 4) frequency of lapses, which refers to the number of times the participant fails to respond to the signal or fails to respond in a timely manner (RTs $\geq$ 500 ms). In order to minimize the contribution of lapses greater than 2 s, lapses were transformed using a formula that reduced the proportionality between the SD and mean (5,6).

Subjective daily diary information was recorded just prior to bedtime using the PDA and included the reporting of duty, naps, meals, exercise, alcohol use, caffeine, and other alertness strategies. On awakening, the wake time diary inquired about the previous sleep period (e.g., sleep latency, total sleep time, sleep quality). Using a within-subject design, data from all 29 pilots were collected during a 1-mo period prior to receiving any educational, alertness strategies, or healthy sleep information and while operating the standard schedule (Fig. 1, top). The same data were then collected, also during a 1-mo period, after implementation of the intervention condition, including operation of the innovative schedule portrayed in Fig. 1, bottom. This design required a significant effort on the part of the airline to create and implement the innovative schedule in their regular trip offerings, guarantee that the 29 pilot participants would be assigned to fly both schedules, and organize logistics to have the schedules available in 2 consecutive months. Operational logistics prevented the examination of order effects because it was necessary to operate all of the flight schedules in the intervention condition in 1 mo. Also, to test the full implementation of the comprehensive AMP required that the innovative schedule could not be flown until after receiving and completing the entire educational CD.

The flight schedules flown during both the standard and intervention conditions were revenue-generating flights operated under Federal Aviation Administration (FAA) regulations. The intervention condition flights operated with the usual two primary flight crew and a third augment pilot who did not fly in the cockpit jumpseat, but instead occupied a seat in the passenger compartment in row 1. During operation of the standard and innovative schedules, no pilot (revenue or non-revenue) was allowed to occupy a cockpit jumpseat. This restriction was implemented to minimize any potential disruption of the operational evaluation procedures. Thus, the third pilot was instructed not to interact with the two primary flight crew unless requested. The only interaction among the three crewmembers occurred during the turnaround time (~ 1 h) between the first and second flight of the day. At this time, all three pilots completed a fatigue checklist and discussed operational status regarding the continuation of the flight using the original crew.

## RESULTS

### *Web-Based Background Survey*

A total of 213 pilots (23% response rate) participated in the background survey. The group was 95% men and averaged about 40 yr of age, with 18.4 yr of total flight experience. The pilots reported sleeping an average of 6.8 h on flight days and 7.7 h on non-flight days. Fatigue was identified as "very much" an aviation operations issue by 38% of respondents and moderately by 45%. Night flying (88% of respondents ranked it first or second), long duty days (49%), and early morning report times (38%) were identified as the most commonly reported aspects of flight operations that created the most fatigue. Pilots reported that fatigue most affected concentration (identified by 100% of respondents), vigilance (90%), and decision making (86%). Of the respondents, 86% indicated that they had "nodded off" in the cockpit sometime during their aviation career. Strategies most often used to stay alert during a trip included: cockpit lights (84% of respondents), physical activity/stretching (83%), talking (80%), caffeine (79%), eating/drinking (74%), naps (47%), and oxygen (29%). When asked about knowledge regarding sleep, circadian rhythms, and fatigue, 48% reported "some" and 39% indicated "little." On the 10 knowledge questions, the average number correct was 4.3, with only 6% getting 7 or more correct and 59% getting 4 or less correct.

The following results report the data from the 29 pilots who completed all components of the comprehensive AMP. Due to missing data, results are reported for the maximum available data points.

### *Education*

The pilots involved in full implementation of the comprehensive program completed six quizzes as outlined in the Methods section. The average correct score for the pre-education module quiz was 14.8 (out of 20), or 74% correct (n = 27). The minimum score was 40% and the best score was 95%. Overall, 44% of the pilots scored 80% or better on the pre-quiz.

After completing the 3.5-h educational CD, the average correct score for the post-quiz was 19.6 out of 20 (98% correct). The minimum score was 90% and 67% (18 pilots) had perfect scores. Overall, 100% of pilots scored 80% or better on the post-quiz. Short quizzes after four of the modules showed best knowledge for fatigue basics/strategies (avg. 97% correct), sleep basics (avg. 94% correct), aviation and fatigue (avg. 92% correct), and circadian rhythms (avg. 85% correct). Fatigue basics/strategies had the highest number of perfect scores (85%), while circadian rhythms had the lowest (44%).

### *Alertness Strategies*

Pilots reported use of alertness strategies in the PDA daily diary, which was collected during the standard and intervention conditions. Overall, there were minimal differences in the use of alertness strategies before education and on the standard schedule as compared with after education and flying the innovative schedule (i.e., intervention condition). Specific examples related to caffeine, naps, exercise, and good sleep habits demonstrated these small changes. For caffeine, 96% of pilots reported use during the standard and 93% reported use during the innovative schedules, with the same average daily servings (2.6) and no major changes in consumption comparing baseline to trip to recovery periods. Coffee (~50%) was the most common form of caffeine used, followed by soda (35% on the standard and 41% on the innovative schedule).

For the baseline and recovery periods, 13 pilots reported 31 naps during the standard condition, and 16 pilots reported 29 naps during the intervention condition (the schedules were too different to examine naps during the trip days). On average, pilots reported a 50.8-min nap during the standard condition baseline and recovery periods, while reporting a 57.6-min nap during the intervention condition. The longest duration nap reported during the standard condition was 3 h, while it was 4 h during the intervention condition.

There were 26 pilots who reported exercising in the standard condition (on 48% of all daily entries) and 25 in the intervention condition (on 32% of all daily entries). There were fewer reports of exercise during baseline of the standard schedule (38% of exercise entries) and more during duty (29%) compared with baseline (60% of exercise entries) and duty (16%) of the innovative schedule. Regarding good sleep habits, there were fewer reports of using a regular pre-sleep routine during the standard condition (18% of all entries for items that helped sleep) compared with use during the intervention condition (24%). The most common report of making the room dark and quiet was the same during both conditions (~54%).

*Healthy Sleep*

Information and tools were provided to address sleep disorders, including referral resources for questions and evaluation. However, as previously explained, medical privacy and confidentiality issues, and the sensitivities related to health status, precluded the collection of any data regarding actions taken based on the information and tools provided. Therefore, no data are available about follow-up questions, possible evaluations, or the outcomes of those actions.

*Scheduling*

There were 213 respondents to the web-based background survey, and 87 of these individuals (~41%) volunteered for the intensive monitoring phase of the project. These 87 volunteers were further screened, primarily for operational considerations (e.g., experience with transcontinental flights) and availability, to determine the final group of 29 flight crewmember participants. All volunteer participants were regular line pilots and no leadership pilots or individuals associated with the project in any way were included. The data from the web-based background survey were compared for the 29 pilot participants in the intensive monitoring phase with the overall sample of 213 pilots. The data for the two groups were similar for age (40.6 yr vs. 40.1 yr), gender (100% men vs. 95%), years of flight experience (18.3 vs. 18.4), and total sleep on duty days (6.7 h vs. 6.8 h) and non-duty days (7.7 h vs. 7.7 h). There were no notable differences between the groups, demonstrating that the 29 pilots involved in the intensive monitoring phase were representative of the airline pilot group that responded to the survey.

Overall, the intensive monitoring conducted during the standard and intervention conditions involved a total of 463 actigraph nights collected, 1452 PVTs performed, and 967 daily diaries completed. These data reflected the specific operational requirements of the standard vs. intervention conditions: 261 vs. 202 actigraph nights, 790 vs. 662 PVTs, and 539 vs. 428 diary entries, respectively. The following sleep and performance data are presented for the standard condition that involved no education, alertness strategies, or healthy sleep information and operating a regular, current flight schedule (Fig. 1, top). These data were compared with the intervention condition that involved education, information on alertness strategies and healthy sleep, and operating the innovative flight schedule (Fig. 1, bottom).

One-way within-subjects analysis of variance (ANOVA) was conducted to test for significant differences between time periods (baseline, trip, recovery) by dependent variables (actigraph sleep, daily diary information, PVT performance). Two-way ANOVAs were conducted to test for significant differences between conditions (standard, intervention) by time period and dependent variables. Post hoc comparisons between all pairs of means were conducted using a Tukey test that controls family wise error (8). Data management and analysis were conducted using the JMP (2005, SAS Institute, Cary, NC) software application. Additional analyses were conducted using the program ANOVA, part of the |STAT data analysis package (developed by G. Perlman, http://www.acm.org/~perlman/stat/).

*Sleep Findings*

Summary results for the actigraph sleep data for $TST_{24}$ (total sleep time for the 24-h period that included nap sleep, or any sleep that occurred outside the primary sleep period), $TST_P$ (total sleep time for the primary sleep period), and sleep latency are presented in **Table II**. In the standard condition, there were overall significant differences between the amount of $TST_{24}$ obtained on baseline, trip, and recovery ($p < 0.001$). Pilots slept less during the trip period than during baseline ($p < 0.01$) or recovery ($p < 0.01$). The $TST_{24}$ was greater during recovery than sleep obtained during baseline ($p < 0.01$) or trip ($p < 0.01$). Therefore, reduced sleep obtained during the trip (compared with baseline) was followed by an increased $TST_{24}$ during recovery, indicating recuperation from sleep loss and a cumulative sleep debt.

In the intervention condition, there were no significant differences in $TST_{24}$ obtained across the baseline, trip, and recovery periods. We found a non-statistically significant increase in sleep (~26 min) during the trip period compared with baseline, with a subsequent reduction during recovery to approximately baseline levels. In the intervention condition, pilots obtained a relatively stable sleep amount through baseline, trip, and recovery periods, did not accumulate a significant sleep loss during the trip period, and did not show an increase in sleep amount during recovery days.

A comparison of $TST_{24}$ between standard and intervention conditions is presented in **Fig. 2**. First, it is important to note that there were no significant differences for the amounts of sleep obtained during the baseline period between the two conditions. In the intervention condition, there was a significant increase in total sleep (~1 h, 9 min) obtained during the trip (innovative schedule) period compared with the standard condition ($p < 0.01$). During the recovery period, there was a significant increase in total sleep (~1 h, 1 min) obtained in the standard condition compared with the

TABLE II. ACTIGRAPH SLEEP BY TIME PERIOD AND CONDITION.

| | Baseline | Trip | Recovery | F | p |
|---|---|---|---|---|---|
| Standard | | | | | |
| $TST_{24}$ (h)*†‡ | 6.18 ± 0.62 | 5.32 ± 0.85 | 7.17 ± 0.86 | $F_{(2,48)} = 40.51$ | 0.001 |
| $TST_{P}$ (h)*†‡ | 6.16 ± 0.85 | 4.23 ± 0.71 | 6.81 ± 0.89 | $F_{(2,48)} = 82.85$ | 0.001 |
| Sleep Latency (min) | 8.53 ± 7.52 | 11.42 ± 11.02 | 13.80 ± 15.41 | $F_{(2,48)} = 1.87$ | n.s. |
| Intervention | | | | | |
| $TST_{24}$ (h) | 6.03 ± 0.49 | 6.47 ± 0.69 | 6.15 ± 0.72 | $F_{(2,50)} = 3.12$ | n.s. |
| $TST_{P}$ (h)* | 5.95 ± 0.51 | 6.27 ± 0.71 | 5.81 ± 0.80 | $F_{(2,50)} = 3.53$ | 0.05 |
| Sleep Latency (min) | 11.72 ± 10.17 | 13.54 ± 12.48 | 11.84 ± 12.43 | $F_{(2,50)} = 0.22$ | n.s. |

Values are mean ± SD; n = 25 for the standard condition and n = 26 for the intervention condition; n.s. = not significant.
$TST_{24}$ = Total sleep time for 24-h period (primary sleep and naps).
$TST_{P}$ = Total sleep time for primary sleep period (≥ 2 h).
* indicates significant difference between baseline and trip periods at $p < 0.01$.
† indicates significant difference between trip and recovery periods at $p < 0.01$.
‡ indicates significant difference between baseline and recovery periods at $p < 0.01$.

intervention condition ($p < 0.01$). As indicated in Fig. 2, the significant sleep loss during the trip period of the standard condition was followed by a classic increased recovery sleep period. The intervention condition showed an increase in sleep during the trip period and no increased sleep during the subsequent recovery period.

Using the actigraph data, the cumulative sleep debt (CSD) was calculated for the average amount of sleep lost per day during the trip period (compared with baseline) for the standard and intervention conditions. Overall, there was an average cumulative sleep loss of 1.7 h of sleep per day in the standard condition for the three layover sleep periods. Therefore, the group averaged 5.1 h of total CSD at the end of the trip schedule, with one individual showing a gain of 0.1 h of sleep and another individual having a loss of 12.9 h of sleep over the three layover sleep periods. In the intervention condition, there was an average loss of 0.2 h of sleep per day across two sleep periods. Therefore, the group averaged 0.4 h of total CSD at the end of the trip schedule, with one individual showing a gain of 2.9 h of sleep and another individual having a loss of 3.7 h of sleep over the two trip sleep periods. This represented a significantly greater daily cumulative sleep loss for the standard condition compared with the intervention condition ($p < 0.001$).

Results of subjective flight crew reports of total sleep time and sleep quality ratings entered into the PDA daily diary were also analyzed. Detailed information can be viewed online in Table A*. There were differences in the absolute amounts of reported sleep totals compared with actigraph-determined sleep amounts, a result consistent with subjective vs. objective measure discrepancies typically found (2,9). However, the subjective reports were consistent with the objective actigraph measures in the direction of the findings. For example, in the standard condition, pilots reported less sleep during the trip period compared with baseline and more sleep during recovery than either baseline or trip period ($p < 0.001$). This was consistent with the pattern observed in the actigraph sleep data.

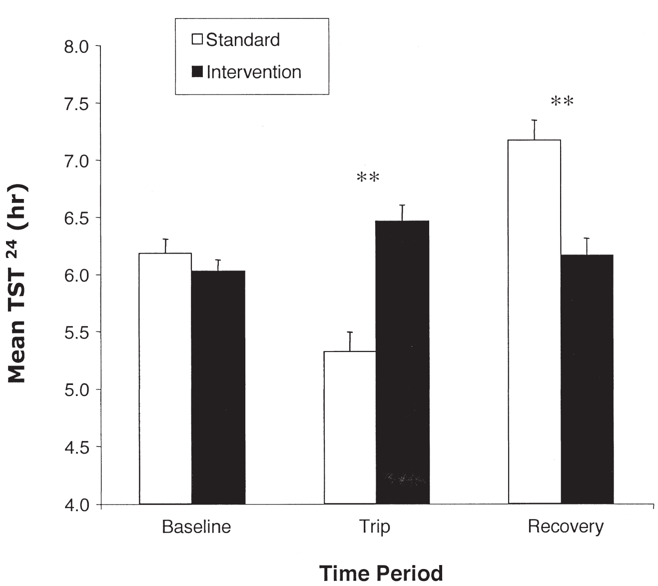


**Fig. 2.** Actigraph mean total sleep time per 24-h period (primary sleep and naps) by time period and condition. ** $p < 0.01$.

*Performance/PVT Findings*

Four performance/PVT metrics were analyzed: mean response speed 1/RT (lower number = worse performance); transformed lapses ($\sqrt{lapses} + \sqrt{lapses + 1}$; lower number = better performance); 1/RT 10% fastest (lower number = worse performance); and 1/RT 10% slowest (lower number = worse performance). Detailed information for these PVT metrics by time period and condition can be viewed online in Table B**. For the standard condition, overall each of the metrics showed significant differences across baseline, trip, and recovery (ranging from $p < 0.05$ to $p < 0.001$). The trip period was associated with worse mean ($p < 0.01$) and slower ($p < 0.05$) reaction times compared with recovery. Transformed lapses ($p < 0.01$) and faster ($p < 0.05$) reaction times were worse during the trip period compared with both baseline and recovery periods. Generally, in the standard condition, performance decreased

* Table A can be found online at http://www.ingentaconnect.com/content/asma/asem.

** Table B can be found online at http://www.ingentaconnect.com/content/asma/asem.

TABLE III. PVT METRICS BY CONDITION AND TIME PERIOD.

| | Standard | Intervention | p |
|---|---|---|---|
| Trip | | | |
| Mean Speed (1/RT) | 3.73 ± 0.70 | 4.31 ± 0.65 | 0.01 |
| Transformed Lapses | 3.77 ± 2.77 | 2.19 ± 1.38 | 0.01 |
| 10% Fastest (1/RT) | 4.86 ± 0.60 | 5.44 ± 0.55 | 0.01 |
| 10% Slowest (1/RT) | 2.53 ± 0.75 | 3.14 ± 0.75 | 0.01 |
| Recovery | | | |
| Mean Speed (1/RT) | 4.01 ± 0.63 | 4.27 ± 0.66 | 0.01 |
| Transformed Lapses | 2.82 ± 1.59 | 2.21 ± 1.44 | n.s. |
| 10% Fastest (1/RT) | 5.16 ± 0.56 | 5.44 ± 0.62 | 0.01 |
| 10% Slowest (1/RT) | 2.76 ± 0.73 | 3.06 ± 0.73 | 0.01 |

Values are mean ± SD; n = 25. PVT = psychomotor vigilance task; n.s. = not significant.
Response speeds are reported as 1/RT × 1000; RT = reaction time.

(worsened) during the trip and improved during the subsequent recovery period.

For the intervention condition, there were overall significant differences across the baseline, trip, and recovery periods for the mean, fastest, and slowest reaction times (ranging from $p < 0.05$ to $p < 0.001$). Performance improved during the trip for both mean and fastest reaction times compared with baseline periods (both $p < 0.01$). Performance also improved during the trip for slowest reaction time compared with baseline ($p < 0.05$). There were no statistically significant differences overall for the transformed lapses across the baseline, trip, and recovery periods. Therefore, generally in the intervention condition, performance showed improvement during the trip period or, at a minimum, maintained consistency across the baseline, trip, and recovery periods.

Comparisons between the standard and intervention conditions showed that at baseline, there were no significant differences among any of the four PVT metrics. However, in the intervention condition, all four PVT metrics showed significantly better performance during the trip (all $p < 0.01$) compared with trip performance in the standard condition (**Table III**). Also, these better performance differences continued into the recovery period, where in the intervention condition, again, three of the PVT metrics analyzed (mean response speed 1/RT, 1/RT 10% fastest, and 1/RT 10% slowest) showed significantly better performance during recovery compared with the standard condition ($p < 0.01$). **Fig. 3** presents the mean response speed 1/RT as an example of the improved performance during the trip and recovery periods for the intervention condition compared with the standard.

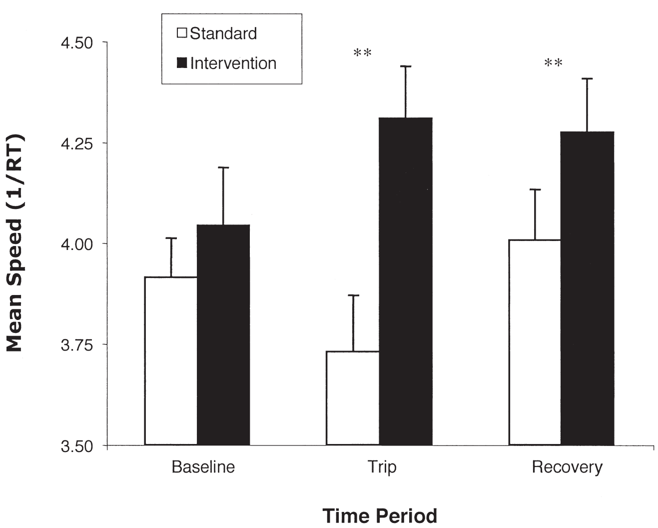


**Fig. 3.** PVT mean speed (1/RT) by time period and condition. ** $p < 0.01$.

TABLE IV. PVT METRICS BY CONDITION AND PHASE OF FLIGHT.

| | Standard | Intervention | p |
|---|---|---|---|
| TOC | | | |
| Mean Speed (1/RT) | 3.84 ± 0.68 | 4.36 ± 0.71 | 0.001 |
| Transformed Lapses | 3.10 ± 2.23 | 2.22 ± 1.64 | 0.01 |
| 10% Fastest (1/RT) | 4.91 ± 0.60 | 5.51 ± 0.60 | 0.001 |
| 10% Slowest (1/RT) | 2.66 ± 0.76 | 3.17 ± 0.82 | 0.001 |
| TOD | | | |
| Mean Speed (1/RT) | 3.64 ± 0.74 | 4.27 ± 0.69 | 0.001 |
| Transformed Lapses | 4.34 ± 3.09 | 2.22 ± 1.35 | 0.001 |
| 10% Fastest (1/RT) | 4.84 ± 0.60 | 5.39 ± 0.61 | 0.001 |
| 10% Slowest (1/RT) | 2.40 ± 0.78 | 3.06 ± 0.80 | 0.001 |

Values are mean ± SD; n = 26. PVT = psychomotor vigilance task; TOC = top of climb; TOD = top of descent. Response speeds are reported as 1/RT × 1000; RT = reaction time.

The comparisons of in-flight performance from after TOC to before TOD between the standard and intervention conditions are presented in **Table IV**. After TOC, in-flight performance was better for all four PVT metrics in the intervention condition compared with the standard condition ($p < 0.01$ for transformed lapses, $p < 0.001$ for the others). Before TOD, in-flight performance is better for all four PVT metrics in the intervention condition compared with the standard condition ($p < 0.001$ for all). An example (transformed lapses) of these differences is portrayed in **Fig. 4**. In the intervention condition, performance was better at the beginning of the flight and maintained throughout the flight compared with the initial worse performance and continued worsening across the flight in the standard condition.

Previously, the physiological sleep/alertness principles were identified that guided the development of the

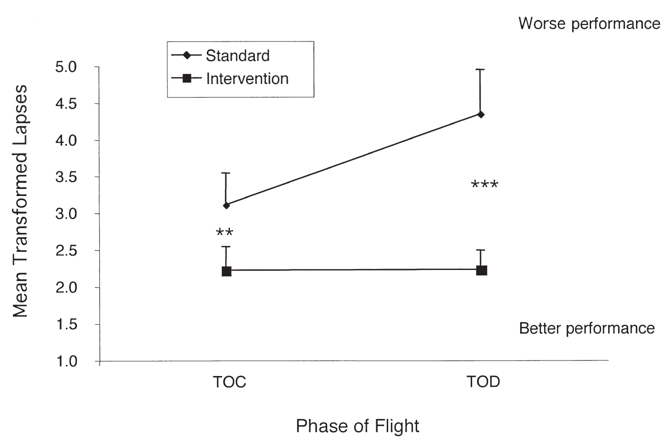


**Fig. 4.** PVT mean transformed lapses by phase of flight and condition. Note that a higher number of lapses indicates worse performance. ** $p < 0.01$; *** $p < 0.001$.

innovative schedule of the intervention condition. These included principles such as day flights only, night sleep periods, and maintaining a regulatory and scientifically based limit on the hours of continuous wakefulness (duty time). However, while the innovative schedule maintained a duty limit significantly less than that allowed under current FAA regulations and consistent with the airline's policies (Table I), it did require an increased flight time. Day 1 of the standard schedule (i.e., standard condition) involved only one westbound daytime transcontinental flight followed by a required daytime sleep period prior to the next day's night flight. Day 1 of the innovative schedule (i.e., intervention condition) involved one westbound daytime transcontinental flight followed by a quick turnaround and a second eastbound daytime/evening transcontinental flight. While an augment pilot flew on each of the eastbound innovative schedule flights, the primary flight crew never deemed it necessary to use the augment pilot for flight duty.

Further analyses examined how daily flight time differences between conditions affected performance. Performance on Day 1 (one transcontinental flight; less than 6 h, 30 min flight time) and Day 2 (one transcontinental flight; less than 5 h flight time) in the standard condition was compared with performance on Day 1 (two transcontinental flights; totaling 10 h, 35 min flight time) and Day 2 (two transcontinental flights; totaling 9 h, 30 min flight time) in the intervention condition. The results of these analyses are presented in **Table V**. The results show that there were no significant performance differences between the standard condition Day 1 and the intervention condition Day 1. However, Day 2 performance in the intervention condition improved in all of the four PVT metrics compared with the standard condition.

## DISCUSSION

Overall, this operational evaluation involved a large group of line pilots, with data collected during actual flight operations, used a within subjects design, involved multiple objective sleep and performance measures and subjective reports, and tested implementation of a comprehensive program that included education, alertness strategies, healthy sleep, and scheduling components. The methodology and approach were based on NASA and other widely used research techniques employed by scientists internationally. The PVT is established as a valid, sensitive measure of vigilance and reaction time, a foundational metric for any operational task and higher cognitive output. The design provided an opportunity for direct comparison between current, standard practices without any AMP components and the full implementation of AMP component activities.

TABLE V. PVT METRICS BY CONDITION AND TRIP DAY.

| | Standard | Intervention | p |
|---|---|---|---|
| Day 1 | | | |
| Mean Speed (1/RT) | 3.97 ± 0.61 | 4.21 ± 0.73 | n.s. |
| Transformed Lapses | 2.75 ± 1.66 | 2.43 ± 1.78 | n.s. |
| 10% Fastest (1/RT) | 5.08 ± 0.51 | 5.33 ± 0.64 | n.s. |
| 10% Slowest (1/RT) | 2.72 ± 0.74 | 3.02 ± 0.85 | n.s. |
| Day 2 | | | |
| Mean Speed (1/RT) | 3.63 ± 0.78 | 4.42 ± 0.69 | 0.01 |
| Transformed Lapses | 4.10 ± 3.36 | 1.98 ± 1.19 | 0.01 |
| 10% Fastest (1/RT) | 4.78 ± 0.66 | 5.57 ± 0.61 | 0.01 |
| 10% Slowest (1/RT) | 2.44 ± 0.87 | 3.22 ± 0.79 | 0.01 |

Values are mean ± SD; n = 26. PVT = psychomotor vigilance task; RT = reaction time; n.s. = not significant.
Response speeds are reported as 1/RT × 1000.

The consistent and statistically significant findings showed that the comprehensive AMP improved sleep and performance compared with standard operations without the AMP. The intervention condition, compared with the standard condition, was associated with significantly more sleep and better performance during the innovative flight schedule. These findings were consistent across different objective sleep and performance variables, as well as across baseline, trip, and recovery periods. All of the significant findings are consistent with known and established physiological principles related to sleep, circadian rhythms, alertness, and performance. Also, the educational activities were found to be effective and improved pilot knowledge from an average "grade" of C+ to an A. The results obtained from this operational evaluation support the use of a comprehensive AMP, including education, alertness strategies, healthy sleep, and scheduling components to address fatigue and alertness challenges in aviation operations.

However, it should be noted that only one type of innovative schedule in the intervention condition was evaluated and operational and design constraints prevented the examination of order effects. No flight variables or overall flight safety metrics were collected or analyzed. Data regarding use of alertness strategies did not provide a refined basis for better understanding their use (e.g., duration, frequency, and timing) during operations. Also, there is no information about potential long-term changes, for example, in the use of alertness strategies.

The operational approach emphasized by the participant airline necessitated the implementation of a practical and efficient (time and costs) program that was sufficiently comprehensive to address the known complexities associated with fatigue (13). This comprehensive program had to demonstrate significant and operationally relevant outcomes. The program was not intended, or conceived, to address the full range of potential interventions or to determine the relative contributions of any individual program component. These considerations were beyond the operational interests and scope of this project.

The innovative scheduling component that was implemented as part of this AMP presented some challenges. The National Transportation Safety Board (NTSB) continues to recommend that all modes of transportation update the hours-of-service rules to incorporate and reflect current scientific knowledge on sleep, circadian rhythms, and fatigue. This recommendation remains on the NTSB's Most Wanted List (10). However, transportation industry efforts to address the full range of hours-of-service issues and enact regulatory change have occurred only in commercial trucking with recent Federal Motor Carrier Safety Administration actions. This project provides an example of how identified fatigue factors in a specific schedule may be effectively addressed by applying scientific knowledge and may lead to improved sleep and performance during actual operations. By addressing specific fatigue factors, there is an opportunity for a focused intervention and change. The results demonstrated with the AMP, including the innovative schedule, support the need to explore mechanisms that will improve scheduling practices and reflect known physiological principles related to sleep, circadian rhythms, alertness, and performance.

Whenever shown to be useful during operational tests, scientific data should be used to help shape policies and practices. For example, on the background survey, "long duty days" (49%) were reported as the second highest contributor to fatigue in aviation operations. However, this subjective survey finding is inconsistent with the objective data collected during actual operations. During the intervention condition, flight times were longer but conducted within a conservative duty time. This was associated with significantly more sleep on layover and better performance in flight compared with the standard condition. It is critical to note that the innovative schedule in the intervention condition incorporated sleep and circadian physiological principles intended to address the known fatigue factors identified on the standard schedule (specifically day sleep requirements and night flights) and involved only day flights and night sleep. This also highlights the importance of differentiating between flight time (time on task) and duty time (related to continuous hours of wakefulness).

This operational evaluation examined the effects of a comprehensive AMP that included multiple components. This project was not designed to, and does not provide information regarding the individual contributions of the AMP component activities to the overall outcomes that were demonstrated. Therefore, it is not possible to determine the relative contribution of the individual AMP components to the results. Overall, the full AMP component activities were associated with statistically significant and consistent improvements in sleep and performance. The available data do not allow interpretation of how much improvement might be elicited with only partial implementation of the AMP or the use of only one component. Understanding the relative contribution of each of the AMP components would be informative; however, an extensive evaluation that controlled for the different elements of the AMP was beyond the scope, cost, and time available for this specific operational evaluation. To determine the relevant contributions of individual components of an AMP or the potential effectiveness of other interventions, further research will be required.

Based on the significant, consistent, and operationally relevant findings associated with the implementation of the comprehensive AMP, the airline has plans to expand the program. The educational CD will be distributed to the current group of 1300 pilots at the airline, with a 90-d requirement for completion. Educational activities are in development that will extend beyond the pilot group and provide information to mechanic, ground crew, and in-flight groups. Also, an Alertness Scheduling Group is in development that would provide a forum to examine current scheduling policies and practices regarding fatigue factors. The intent is to incorporate physiological principles related to sleep and circadian rhythms that would improve sleep and performance as demonstrated in the innovative schedule implemented in this project.

A review of the scientific literature indicates that this is the first operational evaluation of a comprehensive AMP that included education, alertness strategies, healthy sleep, and scheduling components. Research activities will continue to refine knowledge regarding fatigue in operations and basic findings regarding sleep, circadian rhythms, alertness, and performance. There is also an increasing interest in addressing the known and established safety risks associated with fatigue. Implementing a comprehensive AMP offers one approach to minimizing fatigue-related safety risks and enhancing alertness and performance. The statistically significant and consistent findings demonstrated in this operational evaluation of a comprehensive AMP hold promise for this approach. Data from this operational evaluation are now available to show that a comprehensive AMP can significantly increase knowledge, support the use of alertness strategies, and improve sleep and performance during actual operation of innovative schedules in operational environments.

ACKNOWLEDGMENTS

The authors wish to acknowledge the important contributions of the 29 airline flight crewmember participants in the AMP operational evaluation and the many airline personnel who supported the diverse and complex AMP implementation activities, as well as Colette Staump, Dr. Eddie Jung, and Summer Brandt at Alertness Solutions for their project support. We also thank the anonymous reviewers for their constructive and useful comments and Captain David Neri's independent recommendations, all of which helped to strengthen the manuscript.

Activities related to the operational evaluation and AMP content development were conducted through support provided by the airline. All scientific aspects of this project, including educational content, methods, data analyses, and interpretation were developed and conducted by Alertness Solutions independently from the airline. All authors are Alertness Solutions employees and have no financial interests in the airline.

REFERENCES

1. Bagian TM, Rosekind MR. Human factors in aerospace systems design and operations. In: DeHart RL, Davis JR, eds. Fundamentals of aerospace medicine. Philadelphia: Lippincott, Williams & Wilkins; 2002:516–37.
2. Baker FC, Maloney S, Driver HS. A comparison of subjective estimates of sleep with objective polysomnographic data in healthy men and women. J Psychosom Res 1999; 47:335–41.
3. Barger LK, Cade BE, Ayas NT, et al. Extended work shifts and the risk of motor vehicle crashes among interns. N Engl J Med 2005; 352:125–34.
4. Belenky G, Wesensten NJ, Thorne DR, et al. Patterns of performance degradation and restoration during sleep restriction and subsequent recovery: a sleep dose-response study. J Sleep Res 2003; 12:1–12.
5. Dinges DF, Kribbs NB. Performing while sleepy: effects of experimentally-induced sleepiness. In: Monk TH, ed. Sleep, sleepiness and performance. West Sussex, UK: John Wiley; 1991:97–128.
6. Dinges DF, Powell JW. Microcomputer analyses of performance

on a portable, simple visual RT task during sustained operations. Behav Res Methods Instrum Comput 1985; 17:652–5.
7. International Conference on Fatigue Management in Transportation Operations Proceedings; 2005 Sept. 11–15; Seattle, WA. Washington, DC: U.S. Dept. of Transportation Federal Motor Carrier Safety Administration; 2005.
8. Keppel G. Design and analysis: a researcher's handbook, 3rd ed. Englewood Cliffs, NJ: Prentice-Hall; 1991:173–5.
9. Lewis SA. Subjective estimates of sleep: an EEG evaluation. Br J Psychol 1969; 60:203–8.
10. National Transportation Safety Board. NTSB most wanted transportation safety improvements 2006. Retrieved February 2, 2006, from: http://www.ntsb.gov/Recs/brochures/MostWanted_2005_06.pdf.
11. National Transportation Safety Board and NASA Ames Research Center. Fatigue Symposium Proceedings; November 1–2, 1995; Tyson's Corner, VA. Washington, DC: National Transportation Safety Board; 1995. NTSB/RP-95–02.
12. Neri DF, Oyung RL, Colletti LM, et al. Controlled breaks as a fatigue countermeasure on the flight deck. Aviat Space Environ Med 2003; 73:654–64.
13. Rosekind MR. Managing work schedules: an alertness and safety perspective. In: Kryger MA, Roth T, Dement WC, eds. Principles and practice of sleep medicine. Philadelphia: Elsevier Saunders; 2005a:680–90.
14. Rosekind MR. Underestimating the societal costs of impaired alertness: safety, health and productivity risks. Sleep Med 2005b; 6(Suppl. 1):S21–3.
15. Rosekind MR, Gander PH, Connell LJ, Co EL. Crew factors in flight operations X: alertness management in flight operations education module. Moffett Field, CA: NASA; 2001. Technical Memorandum 2001–211385.
16. Rosekind MR, Graeber RC, Dinges DF, et al. Crew factors in flight operations IX: effects of planned cockpit rest on crew performance and alertness in long-haul operations. Moffett Field, CA: NASA; 1994a. Technical Memorandum 108839.
17. Sadeh A, Hauri P, Kripke D, Lavie P. The role of actigraphy in the evaluation of sleep disorders. Sleep 1995; 18:288–302.
18. Thorne DR, Johnson DE, Redmond DP, et al. The Walter Reed palm-held psychomotor vigilance test. Behav Res Methods 2005; 37:111–8.

TABLE A. DAILY DIARY SLEEP BY TIME PERIOD AND CONDITION.

| | Baseline | Trip | Recovery | F | p |
|---|---|---|---|---|---|
| Standard | | | | | |
| TSTp (h)*†‡ | 6.63 ± 1.07 | 5.39 ± 1.29 | 7.57 ± 1.04 | $F_{(2,50)} = 27.89$ | 0.001 |
| Sleep Quality Rating | 3.49 ± 0.55 | 3.47 ± 0.70 | 3.80 ± 0.68 | $F_{(2,50)} = 2.13$ | n.s. |
| Intervention | | | | | |
| TSTp (h)*† | 6.73 ± 0.86 | 7.42 ± 0.96 | 6.76 ± 1.05 | $F_{(2,50)} = 5.11$ | 0.01 |
| Sleep Quality Rating | 3.68 ± 0.54 | 3.65 ± 0.58 | 3.66 ± 0.68 | $F_{(2,50)} = 0.03$ | n.s. |

Values are mean ± SD; n = 26; n.s. = not significant.
$TST_P$ = Total sleep time for primary sleep period (≥ 2 h).
Sleep Quality Rating were as follows: (1) = "very poor"; (3) = "fair"; and (5) = "very good."
* indicates significant difference between baseline and trip periods at $p < 0.01$.
† indicates significant difference between trip and recovery periods at $p < 0.01$.
‡ indicates significant difference between baseline and recovery periods at $p < 0.01$.

TABLE B. PVT METRICS BY TIME PERIOD AND CONDITION.

| | Baseline | Trip | Recovery | p |
|---|---|---|---|---|
| Standard | | | | |
| Mean Speed (1/RT)† | 3.90 ± 0.48 | 3.71 ± 0.69 | 3.98 ± 0.63 | 0.001 |
| Transformed Lapses*† | 2.84 ± 1.40 | 3.81 ± 2.73 | 2.91 ± 1.63 | 0.05 |
| 10% Fastest (1/RT)*† | 5.06 ± 0.41 | 4.85 ± 0.59 | 5.14 ± 0.57 | 0.001 |
| 10% Slowest (1/RT)† | 2.63 ± 0.55 | 2.51 ± 0.74 | 2.73 ± 0.73 | 0.05 |
| Intervention | | | | |
| Mean Speed (1/RT)*‡ | 4.04 ± 0.72 | 4.31 ± 0.65 | 4.27 ± 0.66 | 0.001 |
| Transformed Lapses | 2.59 ± 1.48 | 2.19 ± 1.38 | 2.21 ± 1.44 | n.s. |
| 10% Fastest (1/RT)*‡ | 5.15 ± 0.64 | 5.44 ± 0.55 | 5.44 ± 0.62 | 0.001 |
| 10% Slowest (1/RT)§ | 2.89 ± 0.82 | 3.14 ± 0.75 | 3.06 ± 0.73 | 0.05 |

Values are mean ± SD; n = 26 for the standard condition and n = 25 for the intervention condition.
Response speeds are reported as 1/RT × 1000; RT = reaction time; n.s. = not significant.
* indicates significant difference between baseline and trip phases at $p < 0.01$ and § at $p < 0.05$.
† indicates significant difference between trip and recovery phases at $p < 0.01$.
‡ indicates significant difference between baseline and recovery phases at $p < 0.01$.