FATIGUING THE FORCE: USING OPERATIONAL DATA TO IMPROVE THE UNITED STATES AIR FORCE'S MISSION EFFECTIVENESS MODEL

Kelly M. Amaddio
Headquarters Air Mobility Command, Scott Air Force Base, IL
Jennifer P. Howland
Headquarters Air Mobility Command, Scott Air Force Base, IL

Air mobility pilots routinely fly multiple missions spanning several time zones, thereby disrupting their circadian rhythm. As a result, they consistently operate at a sub-optimal performance level. After several fatigue-related accidents, the Air Mobility Command (AMC) Safety Office incorporated the Sleep, Activity, Fatigue, and Task Effectiveness (SAFTE) model into its Aviation Operational Risk Management (AvORM) program to inform aircrew members of their fatigue levels during critical phases of flight. Further analysis indicated that aircrew members experience higher fatigue levels than predicted, which directly reduces flight safety. This study seeks to improve the underlying assumptions within the sleep model to more accurately predict aircrew member performance during critical phases of flight, thereby improving the predictive power of the mission effectiveness model within AvORM. This is the first study to collect operational data from the United States Air Force (USAF) C-17 pilot community using actigraph watches, self-report daily logs, and objective aircraft data to determine the relationship between fatigue and pilot mission effectiveness. Additionally, this study provides policy recommendations to enable aircrew, squadron leadership, and mission planners to mitigate some factors contributing to aircrew fatigue.

In response to numerous Class A mishaps where fatigue was deemed a contributing factor, Air Mobility Command (AMC) has been employing a version of the Sleep, Activity, Fatigue, and Task Effectiveness (SAFTE) model in the Aviation Operational Risk Management (AvORM) program since 2012. Class A mishaps are defined as mishaps resulting in: a total cost of $2 million dollars or more, a fatality or permanent total disability, and/or destruction of an aircraft (AFI 91-204). Schedulers are instructed to use the SAFTE model to plan missions with the goal of keeping the pilot's performance effectiveness level above 70%. Within AvORM, the performance effectiveness graphs indicate expected times for critical phases of flight (e.g. aerial refueling, landing, etc.). Pilots are given a print out of the performance effectiveness graph when starting a mission to aid in situational awareness and plan possible mitigation strategies. While the merits of the SAFTE model are well documented, it has limitations when employed with a unique population such as air mobility pilots.

Development of the SAFTE model was sponsored by the Department of Defense (DoD), and carried out by a collaboration among Walter Reed Army Institute of Research (WRAIR), Air Force Research Laboratory (AFRL) and Science Applications International Corporation (SAIC). The SAFTE model was accepted as the base model for continued DoD development in 2002. The role of the SAFTE model within AvORM is to make predictions of pilot performance effectiveness. Inputs to the SAFTE model within AvORM include scheduled sleep periods (duration and timing) as well as scheduled flight time and duration. The inputs are used to infer an individual's state in terms of his or her sleep reservoir and determine where he or she is within their circadian rhythm. The SAFTE model is a variant of the two-process model proposed originally by Borbely (Borbely, 1982; Daan, 1984). Although other models seek to predict fatigue, the SAFTE model distinguishes itself from others by considering reduced effectiveness due to sleep inertia immediately upon waking and interrupted sleep.

The SAFTE model is highly accurate when predicting the collective effectiveness level for a group of individuals in a laboratory setting over broad time ranges; however, external issues may limit the model's utility in an operational environment. Primarily, the SAFTE model has not been rigorously validated in an operational setting. Tuning the underlying assumptions within the model with field-relevant data will enhance the model's accuracy at planning and mission time.

As depicted in Figure 1, the SAFTE Model employs numerous physiological factors to predict performance effectiveness (Hursh et. al, 2004). The model is very sensitive to the input assumptions. The SAFTE model within AvORM holds the following factors constant: sleep intensity, sleep quality, rate in which the sleep reservoir is replenished and depleted, sleep inertia, and circadian rhythm. While the rate in which the sleep reservoir is replenished and depleted, sleep inertia, and circadian rhythm will not be addressed in this study due to the type of data collected, the sleep intensity and quality are addressed. Replacing those assumptions with measurement-based statistics should improve the accuracy of the predictions and guide planners in minimizing fatigue throughout all phases of the mission.

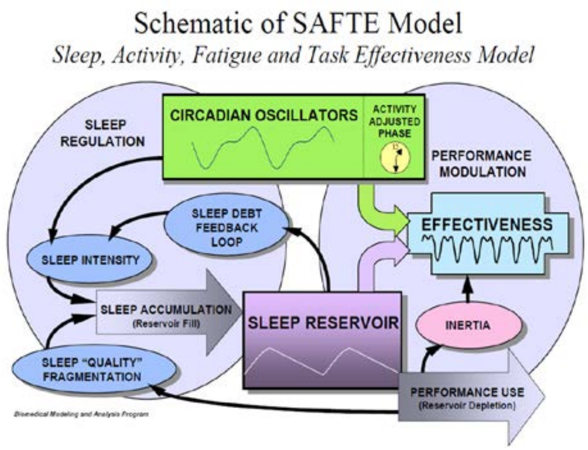

Figure 1. Schematic of the SAFTE Model

Strong assumptions are currently made about the rest-related status of pilots as they begin a mission and about the duration and quality of sleep that crew members obtain during real missions. Within the AvORM program, the SAFTE model assumes that the pilot will start a mission with a sleep reservoir at 90%, start sleeping two hours after landing, sleep for a total of eight hours, experience good quality sleep, experience the same quality sleep with in-flight napping as napping in a bed, and adjust to jet lag at a constant rate of 1.5 hours per day (Hursh, 2004). These factors may be optimized with operational data to reflect a more accurate representation of mobility pilots' operational and fatigue patterns. By using operational data to replace the general population baseline, the SAFTE model will more accurately predict aircrew fatigue and increase safety of flight.

Psychomotor Vigilance Task

This study utilizes the Psychomotor Vigilance Task (PVT) to measure an individual's behavioral alertness. The standard PVT is ten minutes in length; however, operational constraints require this study to use a three-minute PVT. Basner, Mollicone & Dinges found that the 22.7% decrease in effect size garnered by the three-minute verses the ten-minute PVT was an acceptable tradeoff in sensitivity, especially considering the test is 70% shorter in length (2011).

Performance measures on the PVT typically include: lapses of attention, false starts, and response time mean, median, and standard deviation. In this study, we measured performance by mean response time and range. The PVT is best taken in a quiet area free of distractions. However, when the pilots took their PVTs during a mission (whether in flight or on the ground), they were subject to radio calls and other various operational distractions. This led to numerous PVT data points that were outside a person's "normal" range (loosely defined as 100ms to 500ms is a laboratory setting, 100ms to 650ms in an operational setting). Therefore, it was difficult to determine whether a participant's response time increased due to distractions or fatigue. Our method for managing this was to only analyze the response times that were between 100 and 650ms, which comprised 96.8% of the data.

The hypothesis for the study was that mobility pilots are consistently operating at a performance level below the predicted level in the underlying SAFTE model within AvORM. Hence, mobility pilots were more fatigued at critical phases of flight than predicted during the flight planning phase, thereby decreasing the safety of flight. Improving the underlying assumptions within the sleep model would more accurately predict pilot performance during critical phases of flight. It was predicted that sleep duration and quality would be significantly shorter and degraded while on a mission compared to participants' sleep duration and quality at home. A corresponding relationship with the PVT data was expected. Specifically, the mean and range of reaction times should increase throughout a mission and PVTs taken on flying days should be slower than those taken on non-mission days.

While this study addresses the relationship between the predicted performance effectiveness level prior to mission execution and the actual performance effectiveness level, the focus of this paper is the actigraph and PVT analysis.

Method

Participants

Thirty Air Force C-17 pilots stationed at Joint Base Charleston volunteered to participate in this study. All were physically cleared to fly by a flight doctor; therefore nobody had a condition negatively affecting his or her ability to fly safely. Unfortunately, many participants failed to provide demographic information. Of the 30 recruited participants, eleven (nine male and two female) completed all parts of the study so the analysis only includes their data. The mean age of participants was 28.25 years (min 26, max 30). Participants were initially recruited through email and a unit level safety briefing. Interested participants met with the researchers to receive test materials. Pilots unable to attend were directed to the unit safety office to attain test materials.

Apparatus

Participants wore an Ambulatory Monitoring Inc. (AMI) Motionlogger watch which features a built-in accelerometer used to record active and sleeping activity during the duration of the testing period. This watch also has a three-minute PVT to assess reaction time, a proxy variable for performance effectiveness. Although participants were also given a daily log to provide information related to their sleep and flight parameters, the analysis of this data is beyond the scope of this paper.

Experimental Procedure

For 30 days, each pilot was instructed to continuously wear the actigraph and complete a series of activities. If not flying that day, they were instructed to complete the three-minute PVT at least three times daily (within 45 minutes after waking and prior to going sleep, and once anytime throughout the day). If flying that day, they were instructed to complete the PVT within 45 minutes after waking, prior to going to sleep, takeoff, and landing. If the participant flew multiple sorties in the day, they were instructed to only complete the PVT within 45 minutes of landing (not on takeoff).

Analysis

Sleep

The data on the actigraph watch was downloaded and cleansed using Ambulatory Monitoring Inc.'s ActionW software. Episodes when the watch was obviously not worn (e.g. taken off to take a shower, prior to the start of study period, etc.) were removed. After the sleep periods were highlighted within the dataset, the ActionW software scored the sleep episodes. The scored sleep was then imported in Fatigue Science's Fatigue Avoidance Scheduling Tool (FAST) to conduct the performance effectiveness analysis.

Eleven participants wore the actigraph watch during 248 total sleep episodes. Six participants flew on multi-day missions that required them to sleep at a hotel. The 155 sleep episodes from these six participants were used for this portion of the analysis. There were 101 bedtime sleep episodes at home ranging from 196 minutes to 612 minutes (μ=410.3, σ=75.5). There were 26 bedtime sleep episodes away from home ranging from 139 minutes to 704 minutes (μ=371.2, σ=130.1). There were 19 nap sleep episodes at home ranging 22 minutes to 126 minutes (μ=76.1, σ=36.9). There were 9 nap sleep episodes away from home ranging from 21 minutes to 114 minutes (μ=52.4, σ=29.0). Duration of sleep episode, quality of sleep (100*sleep minutes/0-0 period), and longest sleep period were analyzed for each sleep episode. The 0-0 period is defined to start when there are twenty minutes of continuous non-movement until the first continuous movement of twenty minutes; hence the period a participant would consider themselves asleep.

The first relationship analyzed was bed time sleep away from home versus at home. An independent samples t-test indicated that bedtime sleep durations away from home were statistically significantly shorter (μ= 371.2, σ = 130.1) than bedtime sleep durations at home (μ= 410.3, σ = 75.5), $t(127) = 1.47$, $p = .08$, $\alpha = .1$. An α of .1 was used due to the highly variable nature of human subjects research. Analysis of variance showed a main effect of sleep location on sleep duration, $F(1, 127) = 3.98$, $p = .048$, $\alpha = .1$. An independent-samples t-test indicated that the length of the longest sleep period away from home were statistically significantly shorter (μ= 151.1, σ = 79.1) than the longest sleep period when sleeping at home (μ= 171, σ = 92.0), $t(127) = 1.84$, $p = .04$, $\alpha = .1$. This suggests that participants were awoken more frequently while sleeping away from home compared to sleeping at home, and therefore not getting as much restorative deep sleep. An independent samples t-test indicated that the quality of sleep away from home was statistically significantly diminished (μ= 93.4, σ = 5.6) than the quality of sleep at home (μ= 95.2, σ= 4.6), $t(127) = 1.82$, $p = .07$, $\alpha = .1$.

The next relationship analyzed was nap time sleep duration away from home and at home. An independent samples t-test indicated that nap duration away from home was significantly shorter (μ= 52.4, σ =29.0) than nap duration at home (μ = 76.1, σ = 36.9) $t(26) = 1.84$, $p = .04$, $\alpha = .05$. The quality of nap sleep was not found to be significantly different away from home and at home.

A 2x2 ANOVA with sleep location (home, away) and sleep type (bed, nap) as between-subjects factors revealed a statistically significant main effects of sleep duration, $F(2,155) = 191.9$, $p <.0001$.

Finally, the last relationship analyzed was sleep after local flight days and non-flying days. This indicated whether pilots slept longer after flying compared to a non-flying day. Sleep duration, sleep quality, and longest sleep length after flying a local mission was not statistically different than normal bedtime sleep at home.

PVT

The PVT data was cleansed to remove obvious outliers and analyzed using Microsoft Excel. There were a total of 509 PVTs with 10841 individual button presses (and response times) analyzed. A one-tailed t-test indicated that mean response times for PVTs taken on flying days (μ = 257.74 ms) were significantly faster than those on non-flying days (μ = 265.20), $t(311) = 1.28$, $p = .048$. There was no statistically significant difference between mean PVT trial ranges on flying verses non-flying days.

Mission lengths ranged between one and six days. A positive correlation was found between mission day number and response time, $r(134) = .164$, $p = .028$ (small effect size), but there was no statistically significant correlation found between range of PVT trial scores and mission day. There also was no statistically significant correlation between response time and time of day.

Discussion

Sleep

The operational data collected in this study was incomplete in many aspects. There were three times more home station sleep episodes than mission sleep episodes. This was unavoidable due to the fact that many of the actively participating pilots spent more time at home than away on missions. In particular, the number of nap sleep periods was low which prevents the ability to make any real analytical conclusions.

There are numerous factors contributing to the shorter duration and lower quality of sleep experienced during a mission compared to at home. While policy requires crew rest to be at least twelve hours, there are multiple factors affecting a pilot's ability to get enough sleep during a mission. Transportation to the hotel or crew rest facility may take longer than the 30 minutes currently assumed in the AvORM model, especially when the mission ends after normal operating hours. If a mission ends while it's still daylight outside, pilots will have a harder time falling asleep. The quality of the lodging facilities (sound, light, new bed, etc.), location of the hotel, and mission constraints (aircraft commander being called to put the crew on alert) can also affect sleep duration and quality. Finally, crew rest may simply not be the ideal length to recover from the previous mission and prepare for the next mission. For example, if the crew rest length is too short, the pilots may not be able to fully recover from the previous mission. Conversely, a crew rest period can be too long to get in two sleep periods so pilots are reporting for their next mission with one long sleep period and possibly a short nap.

When a person is sleep deprived, their body will spend subsequent recovery sleep periods in deep sleep (Corsi-Cabrera, 1992; Borbély, 1981). It would have been expected that the pilots would have experienced increased longest sleep periods while on a mission due to the long duty day; however, the opposite was seen. This supports the hypothesis that the sleep environment was not conducive to restorative sleep. Further analysis is needed to determine if the pilots had increased longest sleep periods upon return to home station and for how many subsequent days to return to baseline performance effectiveness levels. Hursh et. al. (2004) found that for some individuals under extreme sleep deprivation, the assumed three days needed to recover was insufficient for performance to return to baseline levels.

Shorter length and lower quality naps when on a mission versus at home was also expected. While mission dependent, flight planners may expect pilots to nap in-flight to keep his or her performance level above 70%. The C-17 aircraft crew rest facilities are not conducive to high quality sleep since they are located underneath the stairs leading to the flight deck. In addition, if an aircraft commander is accompanied by other inexperienced pilots, they may sleep less or have lower quality sleep. It is necessary to note that there were less than half the naptime episodes when on a mission compared to at home. Unexpected though, were the low number of naps at the hotel prior to showing up for a flight while on a mission. When crew rest is approximately 24 hours, it is not conducive for two long sleep periods so many pilots reported that they would have one long sleep period with a nap. Further data collection is required to make any conclusions on this issue.

PVT

While cleaning the data, it was apparent that a possible confounding variable is the location where the individual is completing their PVT. On flying days, pilots are most likely doing their PVTs on the aircraft or around other crew members, possibly causing them to be distracted. When critical radio calls are required, they must respond in a timely manner. However, it is likely that on non-flying days, the pilots are taking the PVT in a quieter area since they have more control over where and when they take it. The data showed that faster response times on flying days verses nonflying days, which was unexpected due to the fatiguing effects of flying and circadian shift. Further analysis is needed to determine the root cause for this finding.

An expected result was the positive correlation between mission day number and response time. As the missions continued in length, the participants responded slower to the PVT stimulus. Long missions usually involve the crossing of multiple time zones and circadian rhythm shifts along with long duty days and short periods of sleep.

Conclusion

This study supports the hypothesis that air mobility pilots are not getting the same sleep duration and sleep quality on a mission than they do at home station. As evidenced by the PVT scores, it appears that there is a compounding effect of fatigue as a mission increases in duration. In the short-term, this may decrease flight safety. In the long-term, Pilcher (1996) found that repetitive sleep deprivation can negatively affect one's health and well-being. Additional data is required to make a more definitive algorithmic conclusion concerning the SAFTE model; however, it is clear that pilots experience shorter and lower quality sleep when on a mission than when at home. Additionally, since naps away from home are shorter and of lower quality, it is advisable that the SAFTE model not count naps as fully restorative sleep.

The authors recommend that the SAFTE model decrease the current eight hour crew rest sleep duration to the mean of six and a half hours of sleep. Next, the sleep quality should be decreased from good quality to poor. With further data collection, the decreased in sleep quality can be more accurately quantified. Finally, the authors recommend that naps not be used as a fatigue mitigation strategy by the flight planners to keep pilot performance effectiveness above 70% since the duration and quality of naps during a mission were of such low quality.

Acknowledgements

We would like to thank the pilots at Joint Base Charleston for participating in this study. The views in this paper do not represent the views of Air Mobility Command or the United States Air Force.

References

AFI 91-204 (2014). Department of the Air Force, Headquarters Air Force Safety Center. http://static.e-publishing.af.mil/production/1/af_se/publication/afi91-204/afi91-204.pdf.

Basner, M., Mollicone, D., & Dinges, D. F. (2011). Validity and sensitivity of a brief psychomotor vigilance test (PVT-B) to total and partial sleep deprivation. *Acta Astronautica,* 69(11-12), 949-959.

Mallis, M.M, Mejdal, S., Nguyen, T.T. & Dinges, D.F. (2004). Summary of the Key Features of Seven Biomathematical Models of Human Fatigue and Performance. *Aviation, Space, and Environmental Medicine* 75:A4-14.

Borbély A.A., 1982. A two process model of sleep regulation. *Human Neurobiology* 1:195-204.

Borbély A.A., Achermann, P. (1999). Sleep homeostasis and models of sleep regulation. *Journal of Biological Rhythms*14:557-568.

Borbély, A. A., Baumann, F., Brandeis, D., Strauch, I., & Lehmann, D. (1981). Sleep deprivation: effect on sleep stages and EEG power density in man. Electroencephalography and clinical neurophysiology, 51(5), 483-493.

Balkin, T., Thorne, D., Sing, H., Thomas, M., Redmond, D., Wesensten, N., Williams, J., Hall, S. & Belenky, G. (2000). Effects of sleep schedules on commercial driver performance. Washington, DC: U.S. Department of Transportation, Federal Motor Carrier Safety Administration; Report No.: DOT-MC-00–133.

Corsi-Cabrera, M., Ramos, J., Arce, C., Guevara, M. A., Ponce-de Leon, M., & Lorenzo, I. (1992). Changes in the waking EEG as a consequence of sleep and sleep deprivation. SLEEP-NEW YORK-, 15, 550-550.

Daan, S., Beersma, D.G.M., & Borbély, A.A. (1984). Timing of human sleep: Recovery process gated by a circadian pacemaker. *American Journal of Physiology* 246:R161-R178.

Hursh, S. R., Redmond, D. P., Johnson, M. L., Thorne, D. R., Belenky, G., Balkin, T. J., & Eddy, D. R. (2004). Fatigue models for applied research in warfighting. *Aviation, space, and environmental medicine*, *75*(3), A44-A53.