Aircrew Actual vs. Prescriptive Sleep Schedules and Resulting Fatigue Estimates

Megan B. Morris; Bella Z. Veksler; Michael A. Krusmark; Alex R. Gaines; Helen L. Jantscher; Glenn Gunzelmann

BACKGROUND: Fatigue is an insidious and costly occurrence in the aviation community, commonly a consequence of insufficient sleep. Some organizations use scheduling tools to generate prescriptive sleep schedules to help aircrew manage their fatigue. It is important to examine whether aircrew follow these prescriptive schedules, especially in very dynamic environments. The current study compares aircrew sleep during missions to prescriptive sleep schedules generated by a mission scheduling tool.

METHODS: Participating in the study were 44 volunteers (M_{age}= 28.23, SD_{age}= 4.23; $Proportion_{male}$= 77.27%) from a C-17 mobility squadron providing 25 instances of sleep and mission data (80 flights total). Aircrew wore actigraph watches to measure sleep during missions and prescriptive sleep schedules were collected. Actual and prescriptive sleep was compared with calculated performance effectiveness values per minute across mission flights.

RESULTS: Prescriptive schedules generally overestimated effectiveness during missions relative to estimated actual sleep, potentially causing shifts in effectiveness to ranges of increased risk requiring elevated fatigue mitigation efforts. Actual and prescriptive effectiveness estimates tended to increasingly diverge over the course of missions, which magnifies differences on longer missions.

DISCUSSION: The current study suggests that aircrew sleep during missions often does not align with prescriptive sleep schedules generated by mission planning software, resulting in effectiveness estimates that are generally lower than predicted. This might discourage aircrew from using mission effectiveness graphs as a fatigue mitigation tool. Additionally, because fatigue estimates factor into overall operational risk management processes, these schedules might underestimate risks to safety, performance, and health.

KEYWORDS: aviation, scheduling tool, biomathematical model.

Morris MB, Veksler BZ, Krusmark MA, Gaines AR, Jantscher HL, Gunzelmann G. *Aircrew actual vs. prescriptive sleep schedules and resulting fatigue estimates*. Aerosp Med Hum Perform. 2021; 92(10):806–814.

Aircrew fatigue is an insidious and costly occurrence in the aviation community.[4,7] A common contributor to aircrew fatigue is insufficient sleep during operations.[12,13] Operation tempo, time zone changes, and operations spanning the entirety of the day, among other factors, can cause aircrew sleep issues.[13,14] As a result, numerous organizations have implemented duty hour and crew rest policies to help promote adequate sleep,[5,8] and some have started using scheduling tools to create prescriptive sleep schedules for aircrew to follow during operations.[10]

Air Mobility Command (AMC) uses one of these sleep scheduling tools as part of their general risk management program, Aviation Operational Risk Management (AvORM). The tool is based on FlyAwake (flyawake.org) and the Fatigue Avoidance Scheduling Tool® (FAST),[10] which use a biomathematical fatigue model, the Sleep, Activity, Fatigue, and Task Effectiveness™ (SAFTE)[11] model, to create optimized sleep schedules that maximize performance effectiveness predictions based upon mission parameters entered by planners. The tool generates a mission effectiveness (ME) graph, which includes these prescriptive sleep schedules along with a resulting performance effectiveness curve and flight timing information throughout the mission (see **Fig. 1**).

Aircrew members are commonly given the ME graph before a mission and, if the crew has access to an authorized

From the Air Force Research Laboratory, Wright-Patterson AFB, OH, USA; TiER1 Performance Solutions, Covington, KY, USA; L3Harris Technologies, Dayton, OH, USA; the U.S. Air Force Academy, CO, USA; and the Air Force Joint Test Program Office, Nellis AFB, NV, USA.

This manuscript was received for review in November 2020. It was accepted for publication in July 2021.

Address correspondence to: Dr. Megan Morris, 2620 Q St., Bldg. 852, Wright-Patterson AFB, OH 45433-7955, USA; megan.morris.3@us.af.mil.

DOI: https://doi.org/10.3357/AMHP.5820.2021

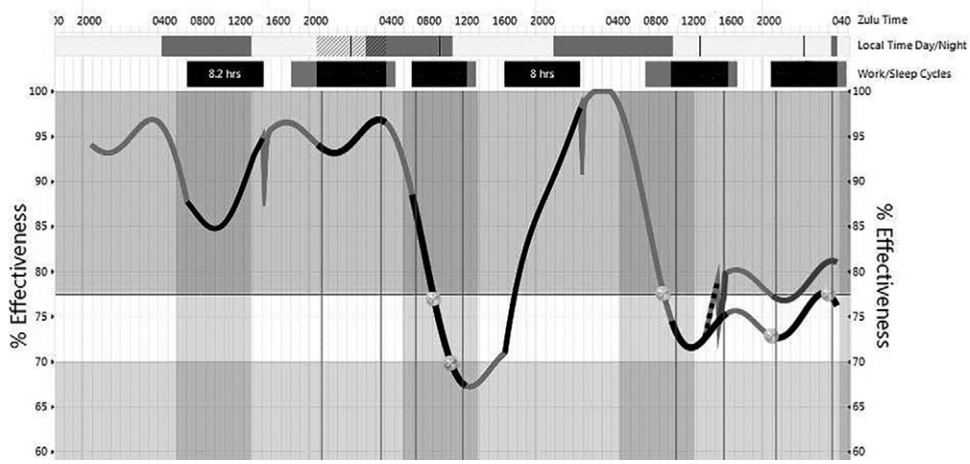

Fig. 1. AvORM mission effectiveness graph. The graph is separated into three effectiveness bands (green: values >77.5; yellow: >70 and ≤77.5; red: ≤70). This graph includes both an effectiveness line without naps (the predominant line) and an augmented crew effectiveness line (the divergent line on the right side of the graph) where a nap is taken (dashed line) during a flight.

computer and Internet connection throughout the mission, they can view updated ME graphs that show mission schedule changes and new prescriptive sleep schedules. During operations, however, crew do not always have access to authorized computers, so they may not be able to access these updated schedules.

C-17 aircrew are especially susceptible to fatigue given operational missions that last multiple days and cross several time zones, and duty periods that can be 16 (basic crew) to 24 h (augmented crew) in duration. In the current study, we examined whether AMC C-17 mobility aircrew members followed prescriptive sleep schedules from the fatigue mitigation tool and suggest implications in terms of fatigue and potential improvements in the fatigue mitigation tool. Specifically, we compared aircrew sleep during missions measured from actigraph watches to prescriptive sleep schedules generated by the mission scheduling tool in their fatigue risk management program.

METHODS

Subjects

Participating in the study were 44 volunteer C-17 aircrew (M_{age} = 28.23; SD_{age} = 4.23; $Range_{age}$ 21–39; $Proportion_{male}$ = 77.27%). Participants included aircraft commanders, instructor pilots, evaluator pilots, pilots, copilots, instructor load masters, and load masters with a mean of 1381.26 flight hours (SD = 1003) in the C-17. This study was approved by the Air Force Research Laboratory (AFRL) Institutional Review Board under the common rule (32 CFR 219), DoDD 3216.2, and AFI 40-402, protocol number: FWR20160111H.

Materials and Equipment

Participants completed three different questionnaires throughout the study that were administered during initial on-boarding, premission, and postmission. We will only report pertinent postmission questionnaire items relating to self-reported sleep, fatigue, and AvORM ME graph perceptions and referencing in the current paper. These items can be found in **Table I**.

Participants wore a Micro Motionlogger® actigraph watch [Ambulatory Monitoring Inc. (AMI), Ardsley, NY, USA] to record actigraph data for sleep estimation. The watch was initialized and data was downloaded using the Motionlogger WatchWare 1.99.17.4 software (AMI). Data was collected at 1-min epochs with the Zero Crossing Mode (ZCM), a measure of frequency of movement. AMI suggests that ZCM is the best mode for sleep estimation.[2] We also collected a Life Measures data channel, a proprietary technique that helps detect when a watch is off-wrist, and an Event Marker data channel (denoted by press of a button on the watch). Participants did not have to change the watch battery during data collection.

Actigraph data was scored with Action-W 2.7.2 (AMI). The software automatically calculates sleep vs. wake minutes once the actigraph data is opened. The software was set to use the Cole-Kripke sleep algorithm[6] and latency to persistent sleep criterion as "No more than 1 minute wake in a 10 minute period" based on a previous actigraph study[1] with mobility aircrew. Bad data periods were trimmed and marked, and down intervals (in-bed behavior) were manually set in the software to provide sleep period associated statistics calculations such as total sleep minutes, total wake minutes, sleep efficiency, and so on. Given that we did not collect sleep log information from participants to decrease participant burden, we used event markers and activity levels to develop down intervals within the software. A ZCM value of 200 typically denotes out of bed behavior.[2] As a result, the beginning of a down interval was based on a ZCM value of 200 or above that then decreased into a sleep period, and the ending of a down interval was based on a ZCM value of 200 or above where the preceding values increased from a sleep period. Nap periods

Table I. Relevant Post-Mission Questions.

QUESTION	1	2	3	4	5	
Rate your sleep quality during the mission	Extremely poor	Poor	Average	Good	Excellent	--
*What were your sleeping accommodations?	Home	Hotel	Open Bay	Other	--	--
Rate your fatigue during the mission (including ground time).	None	Very little	Some	Quite a bit	Substantial	--
Rate your fatigue during the mission flight(s) (not including ground time).	None	Very little	Some	Quite a bit	Substantial	--
Was fatigue an issue for you during the mission flight(s) (not including ground time)?	Not at all	Slightly	Somewhat	Moderately	Extremely	--
Rate how valid AvORM mission effectiveness graphs are in predicting fatigue.	Not at all	Slightly	Somewhat	Moderately	Extremely	N/A
†,‡Did you reference the AvORM mission effectiveness graph for this mission?	Yes	No	N/A	--	--	--
†Rate how well the AvORM mission effectiveness graph matched your fatigue experiences during the mission.	Extremely poor	Poor	Average	Good	Excellent	--
†Did you use the AvORM effectiveness graph predictions to help plan fatigue mitigation strategies during the mission?	Yes	No	--	--	--	--
‡,††Did you have access to an updated AvORM effectiveness graph if mission changes occurred?	Yes	No	No mission changes occurred	--	--	--
††,‡‡Did you reference the updated AvORM effectiveness graph?	Yes	No	--	--	--	--
‡‡Did you use the updated AvORM effectiveness graph predictions to help plan fatigue mitigation strategies?	Yes	No	--	--	--	--

*Select all that apply. †Participants received conditional questions if they answered with "Yes" to the first question denoted with †. ‡Participants received conditional questions if they did not answer with "N/A" to the first question denoted with ‡. ††Participants received conditional questions if they answered with "Yes" to the first question denoted with ††. ‡‡Participants received conditional questions if they answered with "Yes" to the first question denoted with ‡‡.

had to be at least 15 min in duration to create a down interval for that period.

Procedure

Aircrew interested in learning more about the study were briefed on the study purpose and protocol as a group. Those who wanted to participate then completed an informed consent document and subsequently completed an initial questionnaire assessing demographic information, fatigue-related perceptions, and fatigue mitigation behaviors. Participants received an actigraph watch and associated instructions along with mission-related materials and information they would commonly receive from the Flight Safety Officer 1 to 2 d before a mission. Participants were instructed to wear the watch 2 d prior to, throughout, and for 2 d following their mission. They were instructed to use an event marker button on the watch to denote sleep and nap periods. Participants could take off the watch periodically for specific activities (e.g., showers, swimming). Participants completed a premission and a postmission questionnaire with items assessing self-reported sleep and fatigue, ME graph referencing and perceptions, and fatigue mitigation behaviors. Participants returned the actigraph watch to the Flight Safety Officer 2 d after the mission.

Statistical Analyses

To calculate performance effectiveness values from participants' actual sleep schedules, we input sleep estimates from the actigraph data into an *R* (version 3.4.1; R Core Team, Vienna, Austria) implementation of the SAFTE[9,11] model. This implementation was coded by our research team to derive performance effectiveness estimates and was verified against another dataset of sleep periods and performance effectiveness estimates from a FAST implementation, suggesting similar outputs.

To calculate performance effectiveness values from the mission prescriptive sleep schedules, our team collected ME graphs after the conclusion of each participant's mission. Missions consisted of either a basic or augmented crew. Augmented crew missions allow for nap opportunities for aircrew on appropriate flights (see Fig. 1) and are built into the ME graph to show performance effectiveness values for aircrew members who follow this schedule. The performance effectiveness lines without the nap periods are also included on the graph (referred to as Augmented Crew/No Nap in the current study), resulting in two performance effectiveness curves present in the graph. Unfortunately, some ME graphs were not available due to technical issues with the AvORM program. Useable actigraph data with a corresponding prescriptive sleep schedule was provided by 20 participants. This resulted in 25 instances (some participants completed multiple missions) of actigraph data with 80 flights total. We input the prescriptive sleep schedules from the ME graphs into the R version of the SAFTE model to derive performance effectiveness values. The AvORM scheduling tool implementation is proprietary and does not provide effectiveness data, but only a graph (i.e., image) of the effectiveness line. As a result, using our single implementation of SAFTE in R was preferred to make sure the effectiveness values were calculated consistently across the actual (i.e., from the actigraph watch) and prescriptive sleep schedules.

TABLE II. Descriptive Statistics of Differences Between Schedules.

	ALL CREW TYPES (80 FLIGHTS)	BASIC CREW (25 FLIGHTS)	AUGMENTED CREW (55 FLIGHTS)	AUGMENTED CREW/NO NAPS
Average absolute difference	5.89% (SD = 4.53)	4.04% (SD = 2.54)	6.61% (SD = 5.30)	4.25% (SD = 3.81)
Average absolute max difference	14.66%	8.16%	17.20%	12.24%
Average absolute RMSE	7.51%	4.93%	8.52%	5.78%
Minutes ME higher (total)	74.27%	64.97%	77.88%	37.08%
Minutes ME higher (average difference)	6.75%	4.54%	7.49%	2.47%
Minutes ME higher (max difference)	47.56%	16.15%	47.56%	34.20%
1 band difference	27.35%	16.50%	31.57%	31.17%
2 band difference	7.94%	5.01%	9.08%	0.67%

Descriptive statistics of performance effectiveness value differences between the ME schedule and actual (Watch) sleep schedules are based on minute-by-minute analysis. Minutes ME higher (total) refers to the total count of minutes (% of all minutes) during which ME is higher. The following metrics examine the magnitude of the difference when ME is higher: Minutes ME higher (average difference) refers to the difference in values between ME and Watch when ME is higher (on average), and Minutes ME higher (max difference) refers to the average maximum difference between ME and Watch when ME is higher (takes the max difference for each watch comparison and then averages those max values). Augmented Crew/No Naps refers to the effectiveness line if no suggested naps were taken during the mission.

To examine differences in performance effectiveness estimates between aircrew members' actual sleep and prescriptive sleep schedules from the ME graphs, we performed several analyses. First, performance effectiveness values from the two sources were compared on a minute-by-minute basis in terms of percentage difference. We then examined interactions among various factors (e.g., type of schedule, flight leg, and effectiveness band) with linear mixed effects models using the lme4 package[3] in R.

RESULTS

Descriptive statistics and three sets of inferential analyses were conducted to compare ME prescriptive sleep schedules and the actual aircrew member sleep schedules on a minute-by-minute basis. When describing these analyses below (including figures and tables), we refer to these two schedules as ME and Watch, respectively. Descriptive statistics for the schedules, as well as basic vs. augmented crew schedules, can be found in **Table II**. We also included statistics for the Augmented Crew/No Nap effectiveness lines.

ME graphs depict performance effectiveness curves overlaid on colored bands that highlight increased levels of risk of error as effectiveness decreases. At the top is a green band for performance effectiveness values between 77.5% and 100%, followed by a yellow middle band for values between 70% and 77.5%, and finally a bottom red band for values between 0 and 70%. The first set of analyses comparing ME to Watch examined the proportion of time (minutes) that ME and Watch performance effectiveness values fell into green, yellow, or red bands across legs of flight (see **Fig. 2** for best fit lines of the data). There was a significant three-way interaction among Type × Leg × Band [$F(14, 813.42) = 2.51$, $P < 0.01$, see **Table III**]. In the green band, by the second leg, there were significant differences between ME and Watch proportions. In the red band, there were differences by the fifth flight leg (see Fig. 2 and Contrasts in Table III). These results suggest that the distribution of effectiveness values falling within the three bands differed between the ME graph and actual sleep schedule predictions, such that the actual sleep

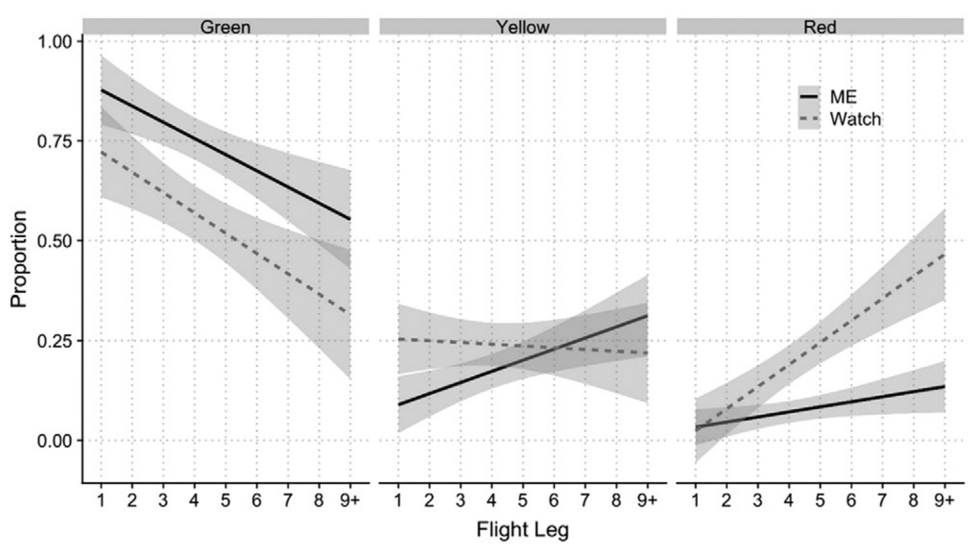

Fig. 2. Distribution of ME and Watch schedule predictions by flight leg and ME Band (green, values >77.5; yellow, >70 and ≤77.5; red, ≤70) as best fit lines.

TABLE III. Linear Mixed Effects Model for Type × Leg × Band and Interaction Contrasts.

3a FACTOR	*df*	*F*	3b BAND	LEG	ESTIMATE	SE	*t* RATIO
Type	1, 813.42	0.00	Green	1	0.00	0.09	0.00
Leg	7, 819.29	0.00		2	0.20	0.09	2.29*
Band	2, 813.42	192.50****		3	0.36	0.09	4.05****
Type × Leg	7, 813.42	0.00		4	0.19	0.10	1.92
Type × Band	2, 813.42	21.79****		5	0.21	0.11	1.88
Leg × Band	14, 813.42	12.05****		6	0.29	0.12	2.49*
Type × Leg × Band	14, 813.42	2.51***		7	0.10	0.12	0.85
				8	0.16	0.12	1.29
			Yellow	1	0.00	0.09	0.00
				2	−0.22	0.09	−2.50*
				3	−0.27	0.09	−3.02**
				4	−0.07	0.10	−0.77
				5	0.02	0.11	0.18
				6	−0.04	0.12	−0.36
				7	0.14	0.12	1.23
				8	0.13	0.12	1.08
			Red	1	0.00	0.09	0.00
				2	0.02	0.09	0.22
				3	−0.09	0.09	−1.03
				4	−0.11	0.10	−1.15
				5	−0.23	0.11	−2.06*
				6	−0.25	0.12	−2.13*
				7	−0.24	0.12	−2.09*
				8	−0.30	0.12	−2.37*

This table shows 3a linear mixed effects models main effects and interactions of proportion of effectiveness values falling within each band (green, yellow, red) by flight leg and type of schedule (ME vs. Watch). It also shows 3b linear mixed effects models contrasts. SE = standard error. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, ****$P < 0001$.

schedule estimates more flight time in the red performance band for later legs compared to the ME graph schedule predictions.

The second set of analyses comparing ME graph and actual sleep schedules examined the magnitude of the difference between the estimates by Band and Leg. Deviations between ME and Watch predictions were computed so that if ME was in one band and Watch in another, the difference was relative to ME (i.e., ME minus Watch). The deviation between ME and Watch increased across flight legs both when ME predictions were in the green and yellow bands (see **Fig. 3**). In particular, while the difference between the two schedules was relatively small in the first flight leg (about 1% difference), by the second flight leg, the difference grew to about 5% and further to 10% by the eighth flight leg when ME predictions fell within the green band. The statistical analysis found no significant interaction between Band and Leg [$F(12, 172.65) = 0.49$, $P = 0.91$], but there were significant main effects of Band [$F(2, 175.95) = 9.57$, $P < 0.001$] and Leg [$F(6, 178.23) = 2.45$, $P < 0.05$]. The differences were higher when ME predictions fell within the green band as compared to the yellow ($P < 0.05$) or red ($P < 0.001$) bands, and the differences gradually increased with flight leg in the green band ($P < 0.05$).

Whereas the second set of analyses comparing ME and Watch predictions focused on the absolute deviation between the two, the third set of analyses compared 'grade' band (green, yellow, or red) deviations between the two schedule types. **Fig. 4** and **Table IV** depict the proportion of all flight minutes during each flight leg that the executed ME graph predicted an effectiveness value that was equal to 1 grade higher, or 2 grades higher, than the values from aircrew's actual sleep schedules (and vice versa). In general, during the first flight leg, both ME and Watch schedules have an effectiveness value falling within the same grade. However, this quickly falls by the second flight leg and further declines by the third and fourth flight legs, during which the ME graph schedule has a considerably higher predicted effectiveness value than the values based on the Watch schedules. Furthermore, as time passes, the rift between the two schedules grows to encompass two grades, whereby ME predicts an effectiveness value falling within the green band while Watch falls within the red band some of the time (depicted by the ME 2 Grades Higher line in Fig. 4). Analysis of these data reveals a significant interaction between Band Difference Type and Leg [$F(24, 547.63) = 4.09$, $P < 0.001$] and a significant main effect of Band Difference Type [$F(3, 547.63) = 126.36$, $P < 0.001$]. In particular, there is a significant increase in the proportion of flight leg minutes during which the ME graph predictions are 1 grade band higher than the actual sleep predictions after the first flight leg, suggesting the schedules begin to diverge enough to potentially impact risk assessments. Furthermore, while there is a significantly higher proportion of flight leg minutes during which the ME graph and actual sleep values fell within the same grade band as compared to different bands, it is important to note that it is significantly more likely in these data that the ME graph predicts higher effectiveness than predicted by the watch data than vice versa. Two band discrepancies between the predictions are also less likely.

In addition to analyses of performance effectiveness scores, we also examined participants' perceptions of sleep, fatigue, and AvORM ME graphs based on data from their self-reported

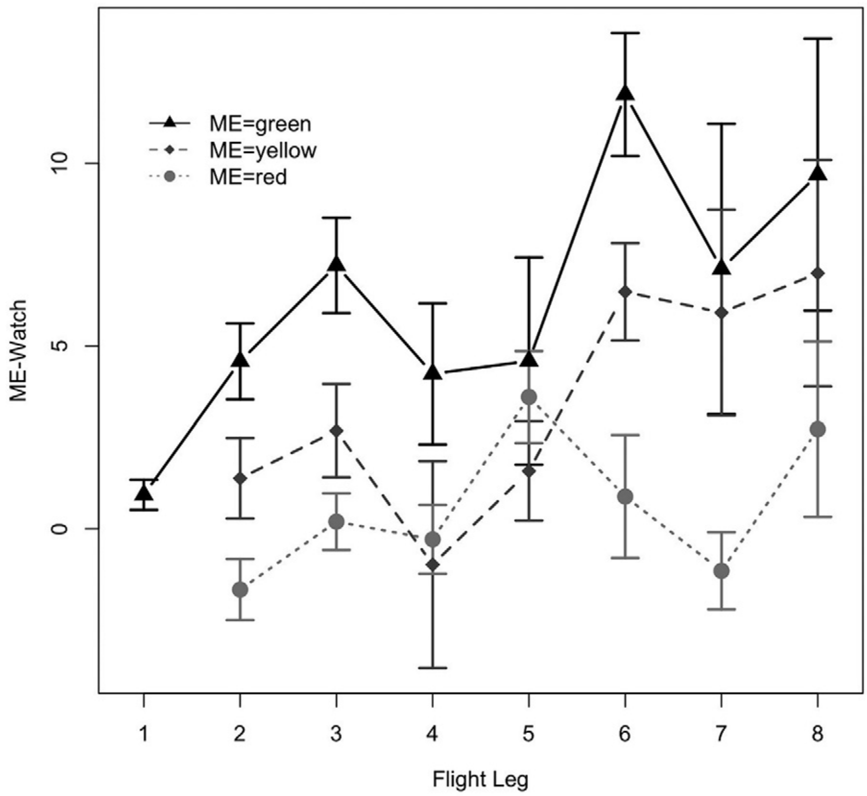

Fig. 3. Differences in effectiveness value between ME and Watch schedule predictions by flight leg and ME Band (green: values >77.5; yellow: >70 and ≤77.5; red: ≤70).

postmission questionnaires. There were 13 participants who completed this portion of the study. If a participant completed multiple postmission questionnaires, we only include their first questionnaire completion as to not bias the data. In general, these aircrew members tended to rate their sleep as "Average" during the mission (M = 2.77; SD = 0.73). All participants reported sleeping in hotels during the mission, with some participants sleeping in other locations as well (one – pilot lounge; one – temporary lodging; three – aircraft).

Participants tended to experience "Quite a bit" of fatigue during the mission (including ground time; M = 3.85; SD = 0.69) or just during flights (M = 3.92; SD = 0.76) and felt that fatigue was "Moderately" an issue during mission flights (M = 3.15; SD = 0.90). They also tended to view the AvORM ME graph as being "Somewhat" valid in predicting fatigue (M = 3.43; SD = 1.13); however, five participants answered with "N/A" for this item. Only four participants referenced the ME graph for the mission (two individuals answered with "N/A"). Of the four participants who referenced the graph, they had varied perceptions on how well the graph matched their fatigue experiences (M = 3.25, SD = 0.96), and only two used the graph to help plan fatigue mitigation strategies. For those individuals who did not answer with "N/A" in terms of referencing the ME graph, only two had access to an updated ME graph with mission changes (one individual reported there were no mission changes). However, neither of these individuals referenced the updated ME graph.

DISCUSSION

The purpose of this study was to compare prescriptive sleep schedules to actual aircrew sleep schedules during missions in a sample of C-17 aircrew. We collected actigraph data to derive performance effectiveness values based on estimations of actual aircrew sleep and compared these to prescriptive sleep schedule performance effectiveness predictions.

Aircrew members did not, or were unable to, always follow the prescriptive sleep schedule from the ME graph. As a result, a large percentage (on average 74%) of performance effectiveness values from the ME graph during flight minutes overestimated effectiveness compared to effectiveness estimates based upon the actual sleep schedules of operators. In other words, aircrew were more fatigued during missions than suggested by the mission plans. Also, in many cases the effectiveness values fell into a lower band of performance effectiveness, sometimes in the red band, while the ME graphs estimated performance to be in the green band. This is most likely due to missed or shortened sleep periods. There was a slight difference between basic vs. augmented crews in how much the prescriptive schedule overestimated effectiveness values (see Table II). This can be partially attributed to augmented crews having in-flight nap opportunities scheduled during their mission. These extra nap opportunities were intended to mitigate fatigue, but the larger discrepancy between the prescriptive and actual schedules suggests that those nap opportunities were not taken as often as hoped.

The overestimation in performance effectiveness values for ME relative to Watch predictions tended to increase further into the mission (in later flight legs) as sleep opportunities were missed, shortened, or taken at different times, and fatigue was compounded. The ME graph predicted effectiveness values were higher in more than 65% of all flight leg minutes in each flight leg and potentially as high as 90% of the flight minutes in later flight legs as aircrew actual sleep schedules tended to deviate more from the prescriptive schedule.

Differences between the prescriptive sleep schedules and aircrew members' actual sleep aligns with the issue that most participants did not reference the ME graph and that some participants did not have access to an updated graph even though mission changes had occurred. Additionally, the augmented performance effectiveness curve is based on all possible nap opportunities across each flight leg. These nap opportunities are managed at the discretion of the aircraft commander. As a result, it is possible that some aircrew

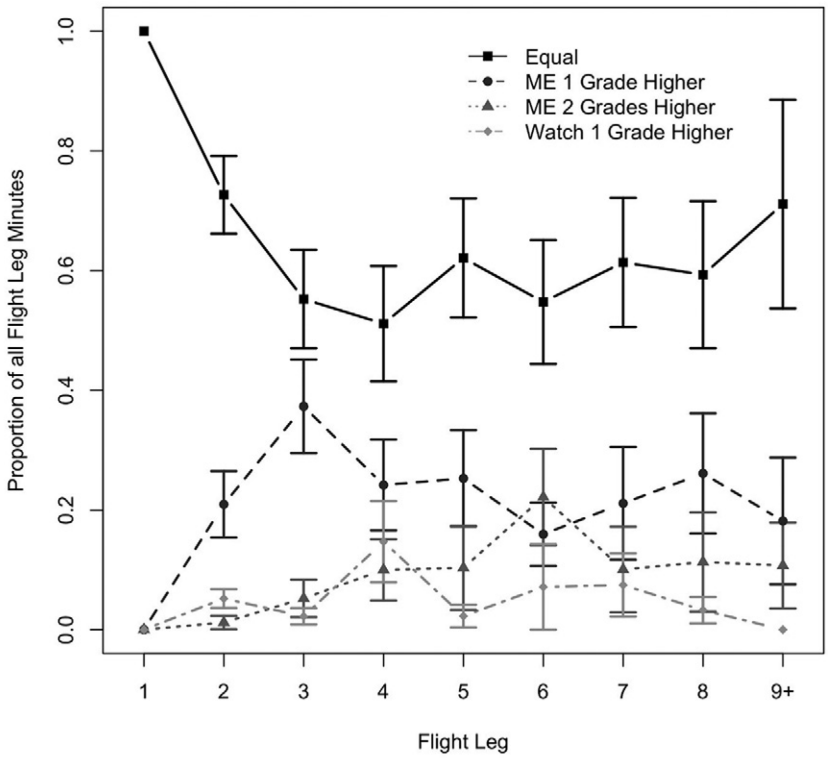

Fig. 4. Distribution of grade differences between ME and Watch effectiveness values by flight leg. "ME 1 Grade Higher" indicates the ME schedule's effectiveness values were either in the green and Watch values were in the yellow or ME values fell into the yellow and the Watch values were in the red. "ME 2 Grades Higher" indicates the ME schedule's effectiveness values were in the green while the Watch values fell in the red. "Watch 1 Grade Higher" corresponds to the reverse relationship as per the above.

might partake in some nap opportunities, but not others, for a host of mission-related and other reasons, which causes misalignment with the prescriptive schedule. Furthermore, while collecting ME graph information from AvORM, we realized that several of the missions that technically had two or more smaller missions linked together were not actually linked in AvORM. As a result, the AvORM graphs only showed performance effectiveness for the specific component of the mission and did not take prior mission information into account when creating the ME graph for the subsequent components. This can create performance effectiveness curves that are of limited use to the aircrew. Given these issues, aircrew might believe that the ME graphs will not be very useful and, thus, do not commonly reference them before or during missions. Supporting this, we found that participants tended to find the ME graph predictions only somewhat valid and had varying perceptions of how well the graph matched their actual fatigue experiences. This suggests a need for individualized ME graphs based on actual aircrew sleep and other individual difference factors that can be updated in real time to produce accurate fatigue estimates during missions and help aircrew assess fatigue and apply effective mitigation strategies based on current fatigue levels, prescribed sleep schedules, and mission requirements.

There are several limitations regarding the current study that readers should be aware of. First, we did not have aircrew complete a daily activity and sleep log, but only had them press the event marker button on the actigraph watch when they attempted to sleep and upon waking up. Several participants did not press the event markers consistently. This resulted in difficulty analyzing the actigraph data. As a result, the sleep estimates from the actigraph data are likely to have some noise given the limitations of actigraph watches, especially in regard to assessing wakefulness. We opted to not have aircrew keep a daily log because past studies implementing the daily log had very low participant compliance in terms of filling out the log and this appeared to be

Table IV. Proportion of Grade Differences by Grade Differences.

FLIGHT LEG	EQUAL GRADE	WATCH 1 GRADE HIGHER	WATCH 2 GRADES HIGHER	ME 1 GRADE HIGHER	ME 2 GRADES HIGHER	ME HIGHER THAN WATCH	COUNT
1	1.00	0.00	0.00	0.00	0.00	0.71	25
2	0.73	0.05	0.00	0.21	0.01	0.82	24
3	0.55	0.02	0.00	0.37	0.05	0.71	23
4	0.51	0.15	0.00	0.24	0.10	0.70	20
5	0.62	0.02	0.00	0.25	0.10	0.72	15
6	0.55	0.07	0.00	0.16	0.22	0.83	14
7	0.61	0.07	0.00	0.21	0.10	0.65	14
8	0.59	0.03	0.00	0.26	0.11	0.90	12
9+	0.71	0.00	0.00	0.18	0.11	0.69	4

Proportion of grade differences based on whether Watch or ME predictions were higher broken out by flight leg. ME higher than Watch indicates proportion of the flight leg minutes during which ME predicted a higher effectiveness value even if it was not across grades.

associated with compliance in wearing the watch. We thought limiting the demands of the current effort might result in better participant compliance in wearing the watch and, generally, this occurred.

Another limitation is that we used a different implementation of SAFTE to calculate performance effectiveness values for the ME graphs and aircrew's actual sleep schedules, as opposed to the AvORM FAST version. This could especially affect the third set of analyses we report comparing ME to Watch predictions in the green, yellow, and red bands. These results should be interpreted with some caution as the actual effectiveness values from the AvORM version of the scheduling tool might be slightly different, resulting in a small proportion of the values falling into different bands. However, visual comparison of the actual ME graphs from AvORM compared to our calculations indicated they were very similar. Additionally, we were not able to capture updated ME graphs after each leg of flight to see how mission changes affected aircrew sleep schedules; instead, we focused on comparing the performance effectiveness values based on the executed mission times to aircrew members' actual sleep schedules.

The data and analyses presented in the current study do not speak to important details regarding the causes or implications of the results. Additionally, we did not assess the likelihood that differences in effectiveness over the course of a mission impact the overall risk assessment for that mission. Even if fatigue is higher, the overall risk category can remain the same based on the risk assessment criteria. Our results suggest that it might be useful for risk assessments to include consideration of reduced and misaligned sleep that could negatively impact the safety and effectiveness of missions, especially for longer duration missions where it becomes increasingly likely that the actual sleep obtained by aircrew will fall short of the mission schedule sleep prescriptions.

Fatigue is an important factor mobility aircrew face during missions that can have serious safety and health implications. The current study suggests that fatigue is still an issue for mobility aircrew, even though they have fatigue mitigation resources such as performance effectiveness estimates from prescriptive sleep schedules based on mission flight times. Performance effectiveness predictions from prescriptive sleep schedules were generally higher than what aircrew were actually experiencing based on watch-based estimates of their sleep, and the overestimation in effectiveness values tended to increase further into the mission. This is most likely due to incorrect assumptions in the prescriptive sleep schedules as well as issues with aircrew having access to updated effectiveness graphs. AMC has recently developed a new AvORM application for the electronic flight bag that will provide more access for aircrews to see updated ME graphs; however, this system offers no ability to adjust sleep times based on an individual aircrew member's actual sleep experience. Consequently, there is a need to develop tools that allow for real-time individualization of performance effectiveness estimates. Additionally, since mission planners use these performance effectiveness predictions when planning missions, these schedules may underestimate risks to safety and performance, impacting the success of missions and the overall health of aircrew. Future research should continue to collect data from these mobility samples, as well as other aircrew samples across the aviation community, to replicate these findings and inform needed improvements in current scheduling tools and biomathematical fatigue models.

ACKNOWLEDGMENTS

The views presented in this paper are those of the authors and do not represent the official position of the U.S. Air Force or Department of Defense. Dr. Megan Morris completed a portion of this work as a contractor with Ball Aerospace. Capt. Alex Gaines completed a majority of this work while at the Air Force Research Laboratory. Capt. Helen Jantscher completed a majority of this work while at the Assessments and Lessons Learned Division at Air Mobility Command (AMC/A9L). Distribution A: Approved for public release; distribution unlimited. 88ABW Cleared 11/13/2020; 88ABW-2020-3564.

Financial Disclosure Statement: This research was funded in part with support from the Defense Health Agency, Joint Program Committee on Military Operation Medicine, Fatigue Mechanisms Working Group. The authors have no competing interests to declare.

Authors and Affiliations: Megan B. Morris, Ph.D., and Glenn Gunzelmann, Ph.D., Air Force Research Laboratory, Wright-Patterson AFB, OH, USA; Bella Z. Veksler, Ph.D., TiER1 Performance Solutions, Covington, KY, USA; Michael A. Krusmark, M.A., L3Harris Technologies, Dayton, OH, USA; Alex R. Gaines, M.S., U.S. Air Force Academy, CO, USA; and Helen L. Jantscher, M.S., Air Force Joint Test Program Office, Nellis AFB, NV, USA.

REFERENCES

1. Amaddio KM, Howland JP. Fatiguing the force: using operational data to improve the United States Air Force's mission effectiveness model. Proceedings of the 19th International Symposium on Aviation Psychology; 2017 May 8–11; Dayton, OH, USA. [Accessed 1 Sept. 2020]. Available from https://corescholar.libraries.wright.edu/isap_2017/95/.
2. Ambulatory Monitoring Inc. User's guide: Micro Motionlogger® Watch and Watchware Software (version 1.94.0.0 and higher). Ardsley (NY): Ambulatory Monitoring Inc.; 2013.
3. Bates D, Mullen KM, Nash JC, Varadhan R. lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-13. 2014. [Accessed Aug. 9, 2021]. Available from https://CRAN.R-project.org/package=lme4.
4. Caldwell JA. Fatigue in aviation. Travel Med Infect Dis. 2005; 3(2):85–96.
5. Caldwell JA, Mallis MM, Caldwell JL, Paul MA, Miller JC, Neri DF. Fatigue countermeasures in aviation. Aviat Space Environ Med. 2009; 80(1):29–59.
6. Cole RJ, Kripke DF, Gruen W, Mullaney DJ, Gillin JC. Automatic sleep/wake identification from wrist actigraphy. Sleep. 1992; 15(5):461–469.
7. Gaines AR, Morris MB, Gunzelmann G. Fatigue-related aviation mishaps. Aerosp Med Hum Perform. 2020; 91(5):440–447.
8. Gander P, Hartley L, Powell D, Cabon P, Hitchcock E, et al. Fatigue risk management: organizational factors at the regulatory and industry/company level. Accid Anal Prev. 2011; 43(2):573–590.
9. Hursh S. Evaluating task effectiveness based on sleep pattern. United States patent US7207938B2. 2007April 24. [Accessed Aug. 9, 2021]. Available from https://patents.google.com/patent/US7207938B2/en.
10. Hursh SR, Balkin TJ, Miller JC, Eddy DR. The fatigue avoidance scheduling tool: modeling to minimize the effects of fatigue on cognitive performance. SAE Trans. 2004; 113(1):111–119.

11. Hursh SR, Redmond DP, Johnson ML, Thorne DR, Belenky G, et al. Fatigue models for applied research in warfighting. Aviat Space Environ Med. 2004; 75(3):A44–A53.
12. Krueger GP. Sustained military performance in continuous operations: combatant fatigue, rest and sleep needs. Fort Rucker (AL): U.S. Army Aeromedical Research Laboratory; 1991. [Accessed 1 Sept. 2020]. Available from https://www.usaarl.army.mil/assets/docs/techReports/91-19.PDF.
13. Miller JC, Melfi ML. Causes and effects of fatigue in experienced military aircrew. Brooks City-Base (TX): Air Force Research Laboratory; 2006. Report No: AFRL-HE-BR-TR-2006-0071. [Accessed 1 Sept. 2016]. Available from https://apps.dtic.mil/dtic/tr/fulltext/u2/a462989.pdf.
14. Morris MB, Wiedbusch MD, Gunzelmann G. Fatigue incident antecedents, consequences, and aviation operational risk management resources. Aerosp Med Hum Perform. 2018; 89(8):708–716.