Investigation into Motor Carrier Practices to Achieve Optimal Commercial Motor Vehicle Driver Performance: Phase I

FOREWORD

A laboratory study was conducted to examine the duration of the restart period for commercial motor vehicle drivers that would be needed to recycle to the work force with optimal performance. This report presents the design, methods, research findings, and conclusions of the study. This report may be of interest to anyone interested in fatigue and its management in commercial motor vehicle operations and other modes of transportation. This is the final report of the study.

NOTICE

Technical Report Documentation Page

1. Report No. **FMCSA-RRR-10-005**	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle **INVESTIGATION INTO MOTOR CARRIER PRACTICES TO ACHIEVE OPTIMAL COMMERCIAL MOTOR VEHICLE DRIVER PERFORMANCE: PHASE I**		5. Report Date **December 2010**
		6. Performing Organization Code
7. Author(s) **Hans P.A. Van Dongen, PhD, Gregory Belenky, MD**		8. Performing Organization Report No.
9. Performing Organization Name and Address **Sleep and Performance Research Center** **Washington State University** **Spokane, WA**		10. Work Unit No. (TRAIS)
		11. Contract or Grant No.
12. Sponsoring Agency Name and Address **U.S. Department of Transportation** **Federal Motor Carrier Safety Administration** **1200 New Jersey Ave SE** **Washington, DC 20590**		13. Type of Report **Final Report**
		14. Sponsoring Agency Code

16. Abstract
The objective of this research project was to determine the effectiveness of the 34-hour restart provision in the Hours-of-Service regulations governing property-carrying commercial motor vehicle drivers. A sample of 27 healthy subjects was studied in an in-residence laboratory study with frequent testing of cognitive performance and driving performance on a high-fidelity driving simulator. A "worst-case"–"best-case" between-groups comparison was made of two 5-day (14-hour/day) work periods separated by a 34-hour restart period. Half the sample was randomized to the "best-case" condition, which entailed daytime wakefulness and work (and nighttime sleep) throughout the study. The other half was randomized to the "worst-case" condition, which entailed nighttime wakefulness and work (and daytime sleep) during the two 5-day work periods, while transitioning back to a daytime schedule during the 34-hour restart period. In the "best-case" condition, performance was the same before and after the 34-hour restart. In the "worst-case" condition, however, performance after the 34-hour restart was impaired relative to performance before the restart, indicating that the restart period was not effective at maintaining performance in this condition. Driver fatigue could be reduced by establishing work schedules and adapting HOS regulations that allow greater flexibility (e.g., taking into account strategic napping and circadian timing).

17. Key Words **34-hour restart provision, circadian rhythm, commercial motor vehicle, CMV, fatigue, Hours-of-Service regulations, performance, sleep, worker recycling**		18. Distribution Statement **No restrictions**	
19. Security Classif. (of this report) **Unclassified**	20. Security Classif. (of this page) **Unclassified**	21. No. of Pages **75**	22. Price

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized.

SI* (MODERN METRIC) CONVERSION FACTORS

Table of APPROXIMATE CONVERSIONS TO SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
in	Inches	25.4	millimeters	mm
ft	Feet	0.305	meters	m
yd	Yards	0.914	meters	m
mi	Miles	1.61	kilometers	km
		AREA		
in²	square inches	645.2	square millimeters	mm²
ft²	square feet	0.093	square meters	m²
yd²	square yards	0.836	square meters	m²
ac	Acres	0.405	hectares	ha
mi²	square miles	2.59	square kilometers	km²
		VOLUME	Note: Volumes greater than 1000 L shall be shown in m³	
fl oz	fluid ounces	29.57	milliliters	mL
gal	Gallons	3.785	liters	L
ft³	cubic feet	0.028	cubic meters	m³
yd³	cubic yards	0.765	cubic meters	m³
		MASS		
oz	Ounces	28.35	grams	g
lb	Pounds	0.454	kilograms	kg
T	short tons (2000 lb)	0.907	megagrams (or "metric ton")	Mg (or "t")
		TEMPERATURE	Temperature is in exact degrees	
°F	Fahrenheit	5 × (F-32) ÷ 9 or (F-32) ÷ 1.8	Celsius	°C
		ILLUMINATION		
fc	foot-candles	10.76	lux	lx
fl	foot-Lamberts	3.426	candela/m²	cd/m²
		Force and Pressure or Stress		
lbf	Poundforce	4.45	newtons	N
lbf/in²	poundforce per square inch	6.89	kilopascals	kPa

Table of APPROXIMATE CONVERSIONS FROM SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
mm	Millimeters	0.039	inches	in
m	Meters	3.28	feet	ft
m	Meters	1.09	yards	yd
km	Kilometers	0.621	miles	mi
		AREA		
mm²	square millimeters	0.0016	square inches	in²
m²	square meters	10.764	square feet	ft²
m²	square meters	1.195	square yards	yd²
ha	Hectares	2.47	acres	ac
km²	square kilometers	0.386	square miles	mi²
		VOLUME		
mL	Milliliters	0.034	fluid ounces	fl oz
L	Liters	0.264	gallons	gal
m³	cubic meters	35.314	cubic feet	ft³
m³	cubic meters	1.307	cubic yards	yd³
		MASS		
g	Grams	0.035	ounces	oz
kg	Kilograms	2.202	pounds	lb
Mg (or "t")	megagrams (or "metric ton")	1.103	short tons (2000 lb)	T
		TEMPERATURE	Temperature is in exact degrees	
°C	Celsius	1.8C + 32	Fahrenheit	°F
		ILLUMINATION		
lx	Lux	0.0929	foot-candles	fc
cd/m²	candela/m²	0.2919	foot-Lamberts	fl
		Force & Pressure or Stress		
N	Newtons	0.225	poundforce	lbf
kPa	Kilopascals	0.145	poundforce per square inch	lbf/in²

* SI is the symbol for the International System of Units. Appropriate rounding should be made to comply with Section 4 of ASTM E380.
(Revised March 2003, Section 508-accessible version September 2009)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

Acronym	Definition
ANOVA	analysis of variance
CDDT	cardinal direction decision task
CMV	commercial motor vehicle
DSST	digit-symbol substitution task
EEG	electroencephalogram
EFFR	effort rating scale
FMCSA	Federal Motor Carrier Safety Administration
HOS	Hours of Service
KSS	Karolinska Sleepiness Scale
N1	non-REM sleep stage 1
N2	non-REM sleep stage 2
N3	slow-wave sleep (non-REM sleep stage 3)
PANAS	Positive Affect Negative Affect Schedule
PERF	performance rating scale
PSG	polysomnography
PVT	psychomotor vigilance test
REM	rapid eye movement (sleep)
REML	REM latency
SL	sleep latency
SWS	slow-wave sleep
SWSL	slow-wave sleep latency
VASM	visual analog scale of mood
WRAIR	Walter Reed Army Institute of Research

[This page intentionally left blank]

EXECUTIVE SUMMARY

The objective of this Phase I research project was to determine the recuperative effectiveness of the current 34-hour restart provision in the hours-of-service (HOS) regulations governing freight-carrying commercial motor vehicle (CMV) drivers. An in-residence laboratory study was conducted to address this issue.

A sample of 27 healthy subjects was studied in a "worst-case"–"best-case" between-groups comparison of two 5-day (14-hour/day) work periods separated by a 34-hour restart period. Half the sample was randomized to the "best-case" condition, which entailed daytime wakefulness and work (and nighttime sleep) throughout the study. The other half was randomized to the "worst-case" condition, which entailed nighttime wakefulness and work (and daytime sleep) during the two 5-day work periods, while transitioning back to a daytime schedule during the 34-hour restart period.

The main goal of the study was to evaluate whether the 34-hour restart period was effective at maintaining performance in both conditions. To this end, performance on a variety of neurobehavioral tasks and on a high-fidelity driving simulator was measured throughout the study. The primary performance outcome measure was the number of lapses (reaction times greater than 500 milliseconds) on a 10-minute psychomotor vigilance test (PVT), which is a validated metric of the performance consequences of fatigue.

In the "best-case" condition, average PVT performance in the 5-day work period after the 34-hour restart was the same as that in the 5-day work period before the 34-hour restart, indicating that the restart period was effective at maintaining performance in this condition. In the "worst-case" condition, however, average PVT performance in the 5-day work period after the 34-hour restart was impaired relative to performance in the 5-day work period before the 34-hour restart, indicating that the restart period was not effective at maintaining performance in the "worst-case" condition.

Subjects in the "worst-case" condition displayed a progressive increase in lane deviation over the hours of the night, which was accompanied by an increase of up to 1 percent in fuel use. These and other indices of driving impairment were predicted by performance on the PVT administered just before each of the simulator driving bouts. The objective observations of performance impairment during the study were not reflected in subjective measures of sleepiness and mood, which inaccurately suggested that some adaptation to the "worst-case" condition would occur.

Despite equal durations of sleep opportunity in the two study conditions, subjects in the "worst-case" condition did not manage to get enough sleep, due to the adverse circadian placement of the sleep periods. Perhaps due to the transitioning back to a normal nighttime sleep schedule during the 34-hour restart period, the restart intervention was not effective at mitigating the sleep loss and consequent performance impairment in the "worst-case" condition.

The study findings highlight the importance of considering circadian effects on sleep and performance in HOS regulations. This final report describes the methodology and the results of the research study, and provides further conclusions and recommendations.

[This page intentionally left blank]

1. INTRODUCTION

1.1 OBJECTIVE

The objective of the project was to determine the effectiveness of the current 34-hour restart provision in the hours-of-service (HOS) regulations governing freight-carrying commercial motor vehicle (CMV) drivers with regard to restoring performance. Specifically, the effectiveness of the 34-hour restart provision was evaluated using a "worst-case"–"best-case" in-laboratory experimental study design with testing of neurobehavioral task and driving simulator performance.

1.2 BACKGROUND

The current Federal Motor Carrier Safety Administration (FMCSA) HOS regulations for freight-carrying CMV drivers prescribe that drivers 1) may drive 11 hours within a 14-hour window after coming on duty following 10 consecutive hours off duty; 2) may not drive after 60/70 hours on duty in 7/8 consecutive days; 3) may restart a 60/70-hour period after taking 34 or more consecutive hours off duty (the 34-hour restart provision). The scientific evidence supporting the 34-hour restart provision is limited, and the provision may present problems in light of the current state of sleep science. Specifically, the provision does not take into account the well-described circadian rhythms in both performance and sleep propensity. Depending on circadian placement of both the sleep opportunity and the period of on-duty time, the current restart provision may be (in the case of adverse placement of sleep and work relative to circadian phase) inadequate to restore performance and may be (in the case of optimal placement of sleep and work relative to circadian phase) adequate or even unnecessary to sustain performance. The FMCSA has commissioned scientific studies to evaluate the efficacy of the current 34-hour restart provision and to provide information in support of possible revisions to this provision.

> Fatigue—as measured by both self-report and objective measures indicating degraded performance—is a function of sleep/wake history (time awake), circadian phase (time of day), and workload (time on task). These factors are clearly seen in laboratory studies of sleep loss.[1–3] The effects of all three factors during 38 hours of total sleep deprivation in 50 young healthy human volunteers can be seen in Source: Wesensten et al.[3]
>
> Figure 1. The circadian rhythm (the 24-hour rhythm of the endogenous biological clock) modulates alertness, performance, and sleep propensity.[4–7] The circadian rhythm in core body temperature is depicted in Source: Mistlberger and Rusak[9]

Figure 2. Sleep propensity displays the inverse of this rhythm in core body temperature. Sleep propensity is high when core body temperature is falling or low. Sleep propensity is low when core body temperature is rising or high. Similarly, but approximately 180 degrees out of phase with sleep propensity, performance is high when body temperature is rising or high and low when body temperature is falling or low. Light exposure is the primary factor setting circadian rhythm phase. Humans, being diurnal (daytime-waking) animals, are most productive and sleep best when they work during the day and sleep during the night. Conversely, human beings perform poorly and sleep poorly when working through the night and attempting to sleep during the day. Given the normal diurnal light exposure, even most permanent night-shift workers do not resynchronize their circadian rhythms in core body temperature and hence never adapt to

night-shift work. The average night-shift worker is able to sleep no more than ~5 hours/day,[8] because he or she is attempting to sleep at a time when sleep propensity is falling.
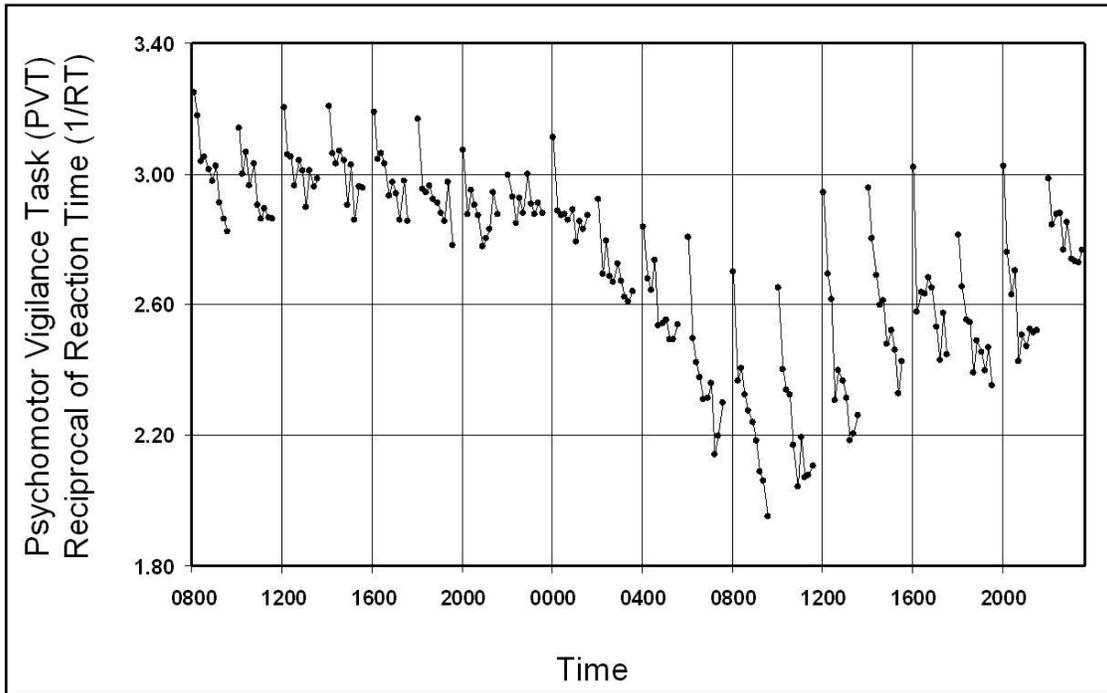
Source: Wesensten et al.[3]

Figure 1. Graph. Time on task, time of day, and time awake interact to degrade/ modulate cognitive performance over 38 hours of total sleep deprivation in 50 healthy volunteer subjects.
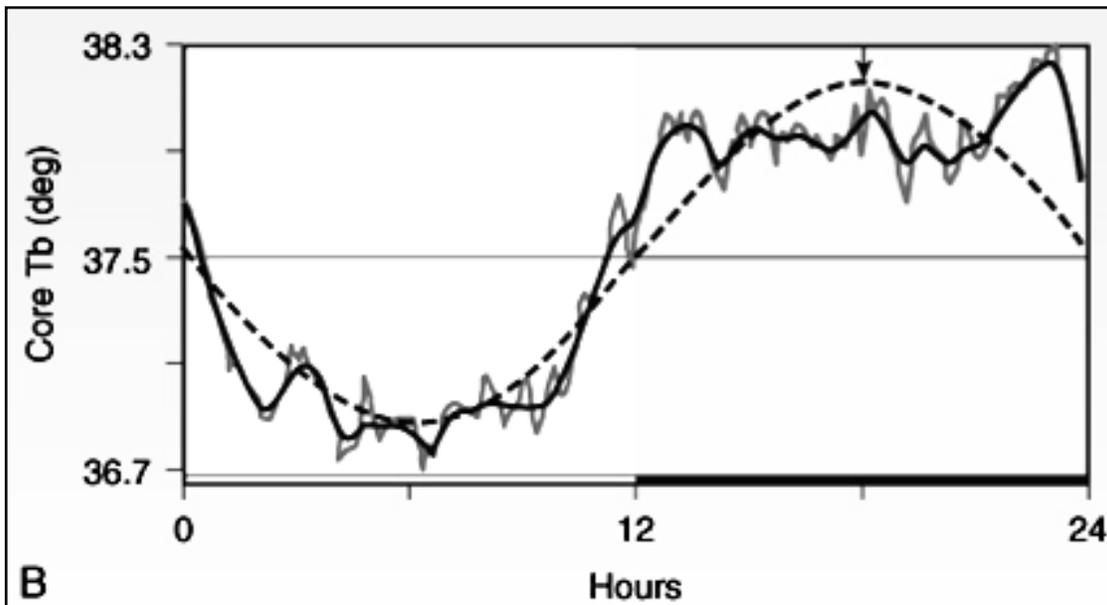
Source: Mistlberger and Rusak[9]

Figure 2. Graph. Core body temperature rises and falls in an approximately 24-hour (circadian) rhythm.

1.3 PHASE I PROJECT TASKS

The following tasks were completed for this project:

1.3.1 Task 1: Conduct kick-off meeting

The investigators conferred with the Government sponsors and provided an overview of the project and the planned strategy. A PowerPoint presentation was presented outlining the project tasks and the estimated timeline (October 22, 2008).

1.3.2 Task 2: Submit work plan

The investigators wrote a work plan for the project providing details for each of the project tasks, including the formation of a peer review committee of at least three members. The work plan specified the laboratory research in detail and provide a basis for Tasks 3 and 4 (October 22, 2008).

1.3.3 Task 3: Conduct peer review committee meeting

The investigators convened a meeting of the peer review committee via teleconference. The committee consisted of Dr. Thomas Balkin (Walter Reed Army Institute of Research), Dr. Siobhan Banks (University of Pennsylvania), and Dr. Melissa Mallis (Institutes for Behavioral Resources). The committee members were provided with copies of the work plan in advance of the meeting (November 21, 2008).

1.3.4 Task 4: Submit Institutional Review Board proposal

The investigators submitted a protocol to the Institutional Review Board of Washington State University, which subsequently approved the study. No amendment was needed following the peer review committee meeting (September 25, 2008).

1.3.5 Task 5: Perform laboratory research

A "worst-case"–"best-case" between-groups comparison of two 5-day work periods separated by a 34-hour restart period was conducted, as described in further detail in the Methods section below (completed August 11, 2009).

1.3.6 Task 6: Data reduction and analysis

Primary statistical testing involved a within-subjects comparison of performance in the first 5-day work period vs. the second 5-day work period in the "worst-case" condition as compared to the "best-case" condition, as described in further detail in the Methods section below (completed September 12, 2009).

1.3.7 Task 7: Submit draft final report

A draft final project report was written that detailed the methods used and the study results (September 21, 2009).

1.3.8 Task 8: Conduct second peer review committee meeting

The investigators convened another meeting of the peer review committee via teleconference (see Task 3). The peer review committee members were provided with copies of the draft final report in advance of the meeting (October 8, 2009).

1.3.9 Task 9: Submit final project report and technical brief

To complete the project, this report was submitted. In addition, a technical brief (PowerPoint) was produced and shared with the Government sponsors (October 26, 2009).

2. METHODS

2.1 EXPERIMENTAL DESIGN

Volunteer subjects were run in an in-residence laboratory research study of sleep and performance. A "worst-case"–"best-case" between-groups comparison of two 5-day work periods separated by a 34-hour restart period was conducted. Figure 3 illustrates the design of the study for the two conditions. These conditions had equal amounts of scheduled work (simulator driving and performance testing), scheduled wakefulness, and scheduled sleep. However, they differed in the timing of the sleep/wake/work periods, in that the "worst-case" condition entailed nighttime wakefulness and work (and daytime sleep) during the two 5-day work periods, whereas the "best-case" condition entailed daytime wakefulness and work (and nighttime sleep) throughout the study. Furthermore, the 34-hour restart period in the "worst-case" condition, in addition to constituting a temporary suspension of work, involved transitioning back to a daytime schedule, whereas the 34-hour restart period in the "best-case" condition involved just a one-day suspension of work.

The "worst-case" condition began with a day for practicing simulator driving and performance testing procedures, followed by 10 hours of time in bed for baseline sleep (day 1), and then a day which included a 5-hour scheduled napping period for transitioning to a nighttime work schedule (day 2). Subsequently, subjects were exposed to the first 5-day nighttime work period, in which there were five 14-hour duty periods with simulator driving and performance testing, each separated by 10 hours of diurnal time in bed for sleep (days 3–7). This first nighttime work period was followed, after the accumulation of 70 hours of duty time, by a 34-hour restart period, as permitted by the current HOS regulations. During the restart period, subjects transitioned back to a daytime schedule (days 7 and 8). The restart period began with a 5-hour scheduled napping period, followed by a 7-hour off-duty period of wakefulness and a 10-hour nocturnal sleep period. This was followed by a second 7-hour off-duty period of wakefulness, and another 5-hour scheduled napping period to transition back to a nighttime schedule for the second work period. Thus, the 34-hour restart period included a total of 20 hours of time in bed for sleep and 14 hours of wakefulness, the same as in the "best-case" condition (see below). After the 34-hour restart period, subjects in the "worst-case" condition were exposed to the second 5-day nighttime work period (days 9–13), which was identical to the first nighttime work period. After the accumulation of 70 hours of duty time, the schedule ended with a day that included a 5-hour scheduled napping period for transitioning back to a daytime schedule (day 13), and finally a recuperation day beginning with 10 hours of time in bed for recovery sleep (day 14).

The "best-case" condition also began with a day for practicing simulator driving and performance testing procedures, followed by 10 hours of' time in bed for baseline sleep (day 1). The next day was a transitioning day as in the "worst-case" condition, but since there was no need to transition to a nighttime schedule, this second day was similar to the other days in the "best-case" condition—but without scheduled work, and it included a 10-hour scheduled nocturnal sleep period (day 2). Subsequently, subjects were exposed to the first 5-day daytime work period, in which there were five 14-hour duty periods with simulator driving and performance testing, each separated by 10 hours of nocturnal time in bed for sleep (days 3–7). This first daytime work period was followed, after the accumulation of 70 hours of duty time, by

a 34-hour restart period, as permitted by the current HOS regulations. The restart period included a 10-hour scheduled nocturnal sleep period, a 14-hour period of off-duty wakefulness, and another 10-hour scheduled nocturnal sleep period (day 8). The 34-hour restart period included a total of 20 hours of time in bed for sleep and 14 hours of wakefulness, the same as in the "worst-case" condition. After the 34-hour restart period, subjects in the "best-case" condition were exposed to the second 5-day daytime work period (days 9–13), which was identical to the first daytime work period. After the accumulation of 70 hours of duty time, the schedule ended with a "recuperation" day beginning with 10 hours of time in bed for sleep (day 14).

Figure 3. Chart. Study design for the "worst-case" (top schematic) and "best-case" (bottom schematic) conditions of the study, with hours of the day (midnight to midnight) progressing from left to right (see top numbers), days of the study progressing from top to bottom (see right side), gray indicating scheduled wakefulness, and black indicating scheduled sleep periods.

The experiment was conducted in the Sleep and Performance Research Center at Washington State University Spokane. This 10,000-square-foot facility includes a state-of-the-art sleep research laboratory and a high-fidelity job task simulation laboratory connected to form a single contiguous, isolated space with a high degree of laboratory control for the precise and efficient investigation of fatigue and performance responses to prescribed sleep/wake/work schedules. The facility contains a sleep suite (four bedrooms, lounge area, medical exam room, monitoring room, and two full baths), a simulation laboratory with two high-fidelity driving simulators and several other critical job task simulators, and office and work space for investigators, post-doctoral fellows, graduate students, undergraduate research assistants, and technical personnel. The laboratory is fully equipped and staffed for polysomnographic recording of sleep and for computer- and simulator-based measurement of performance during long-term (days to weeks) in-residence studies.

Subjects in both conditions were in the laboratory continuously for a total of 317 hours (14 days, 13 nights). They were monitored continuously by direct observation or, when they were in their room for performance testing or for sleep, through a camera system (with infrared light for observation in darkness during scheduled sleep). No visitors, phone calls, text messaging, e-mail, live radio and television, or other external, potentially confounding influences were allowed inside the laboratory; only our trained research assistants interacted with the subjects. During scheduled sleep periods, lights in the bedrooms were off and subjects were not allowed to engage in any activities other than sleeping (or resting if they could not sleep). Ambient temperature was kept at 72 ± 2 degrees F, and light levels during scheduled wakefulness were fixed at below 50 lux. Meals were served at regular intervals, and breakfast, lunch, and dinner were shifted by 12 hours during nighttime wake periods in the "worst-case" condition.

2.2   MEASUREMENTS

During the 5-day work periods, four times per duty day, subjects drove a 40-minute route on a PatrolSim IV driving simulator (MPRI, Salt Lake City, UT). This is a high-fidelity driving simulator widely used to train professional drivers. For this project, we developed and installed additional hardware and software to be able to capture driving performance data at high resolution, so as to turn this training device into a research tool.[10] Furthermore, we developed a standardized driving scenario, involving rural highway driving with five to seven randomly located encounters with pedestrians or dogs crossing the road. Braking responses to these unexpected crossing events were recorded to capture any lapses of attention. In addition, 10 straight, uneventful road segments in the scenario ("straightaways") were used to extract unconfounded data on lane deviation and other performance measures potentially indicative of fatigued driving. The speed limit throughout the scenario was 55 mi/h.

The following outcome variables were extracted for each simulator drive: average and variability (standard deviation) of speed in the straightaways; lane deviation (standard deviation of lane position) in the straightaways; reaction time of braking for the pedestrian/dog crossing events; total number of braking errors (i.e., braking unnecessarily in the straightaways, or braking more than once or not at all around the pedestrian/dog crossing events); and fuel use (as calculated by the simulator's internal fuel use model) across the straightaways.

Each 40-minute driving simulator session was preceded and followed by a 10-minute psychomotor vigilance test (PVT). This is a simple reaction-time task with high stimulus density. It is a well-validated and widely used standard assay of fatigue.[11] As in previous studies,[2,12,13] the number of performance lapses was extracted, with performance lapses being defined as reaction times greater than 500 ms. Because of high sensitivity to fatigue and favorable statistical properties,[14] these PVT lapse counts were used as the primary outcome measure of the study.

Other assessments of neurobehavioral performance were performed during the study. Each driving simulator session (with preceding and subsequent PVT bout) was paired with administration of a brief (~12-minute) neurobehavioral test battery. The battery included computerized versions of the Karolinska Sleepiness Scale (KSS),[15] a visual analog scale of mood (VASM),[12] the Positive Affect Negative Affect Schedule (PANAS),[16] a digit symbol substitution task (DSST),[17] performance and effort rating scales (PERF and EFFR, respectively),[18] and a cardinal direction decision task (CDDT).[19] The KSS, VASM, PANAS, PERF, and EFFR yielded subjective assessments of sleepiness, mood, and effort. For each, an overall score was extracted, except for the PANAS, for which both positive and negative affect scores were determined. The DSST involved matching numbers to symbols. The computer screen showed a key with a set of nine symbols and their corresponding digits (1–9). When given a symbol in another, fixed location on the screen, subjects were required to type its corresponding number. After the response, a new symbol was presented immediately. The number of correct responses in the 3-minute task duration was extracted, yielding a measure of cognitive throughput. The CDDT required participants to make judgments about where a target appeared. The stimulus consisted of a first-person view where a single target was highlighted in a set of eight objects arranged in a circular target field. This view was presented adjacent to an allocentric perspective (i.e., a map with the target field at the center), which indicated the viewing perspective. The task was to identify the portion of the target field containing the target. A total of 25 trials was presented per test bout, and responses were self-paced. From the number of attempts needed to complete them the number of error responses was calculated (i.e., the number of attempts minus 25), which served as the outcome measure.

Figure 4 shows the timing of the 1-hour blocks consisting of a 10-minute PVT, 40-minute simulator driving, and 10-minute PVT, as marked by the asterisk triplets. The figure also shows the timing of the brief neurobehavioral performance test bouts, as marked by lower-case x symbols. On off-duty days, no driving occurred, but the neurobehavioral performance test battery was administered a few times, augmented with a 10-minute PVT (as marked with capital X symbols). This off-duty performance monitoring served to gauge fatigue levels during the baseline, restart, and recovery periods bracketing the two 5-day work periods; they were not used for the present analyses. Driving simulator and performance testing practice occurred on the first day; these practice sessions were also not used for analysis.

Two driving simulators were available in the laboratory; up to four subjects could be participating in the study at a given time. Therefore, subjects were randomly assigned consistently either to do the driving (preceded and followed by the PVT) first and undergo the neurobehavioral performance testing second, or the other way around. There was always a 45-minute break between the driving/PVT blocks and the neurobehavioral performance test bouts. Figure 4 illustrates the simulator driving and performance testing schedule for the subjects who underwent the driving/PVT block first and the neurobehavioral performance testing second.

During scheduled sleep periods, sleep stages were recorded using digital equipment (Nihon Kohden, Foothill Ranch, CA) for polysomnography (PSG). Scalp and skin electrodes were used to record brain waves (bipolar electroencephalogram, EEG), eye movement (electrooculogram), muscle activity (submental myogram), and heart beats (electrocardiogram). The EEG electrodes were placed at frontal (F3, F4), central (C3, C4), and occipital (O1, O2) locations, referenced against the mastoids (M1, M2). Sleep stages were scored using standard criteria promulgated by the American Academy of Sleep Medicine.[20] Every third day, electrodes were removed to give subjects an opportunity to take a shower and to heal any skin irritation that might have occurred due to electrode attachment. Figure 4 shows which sleep periods were recorded.
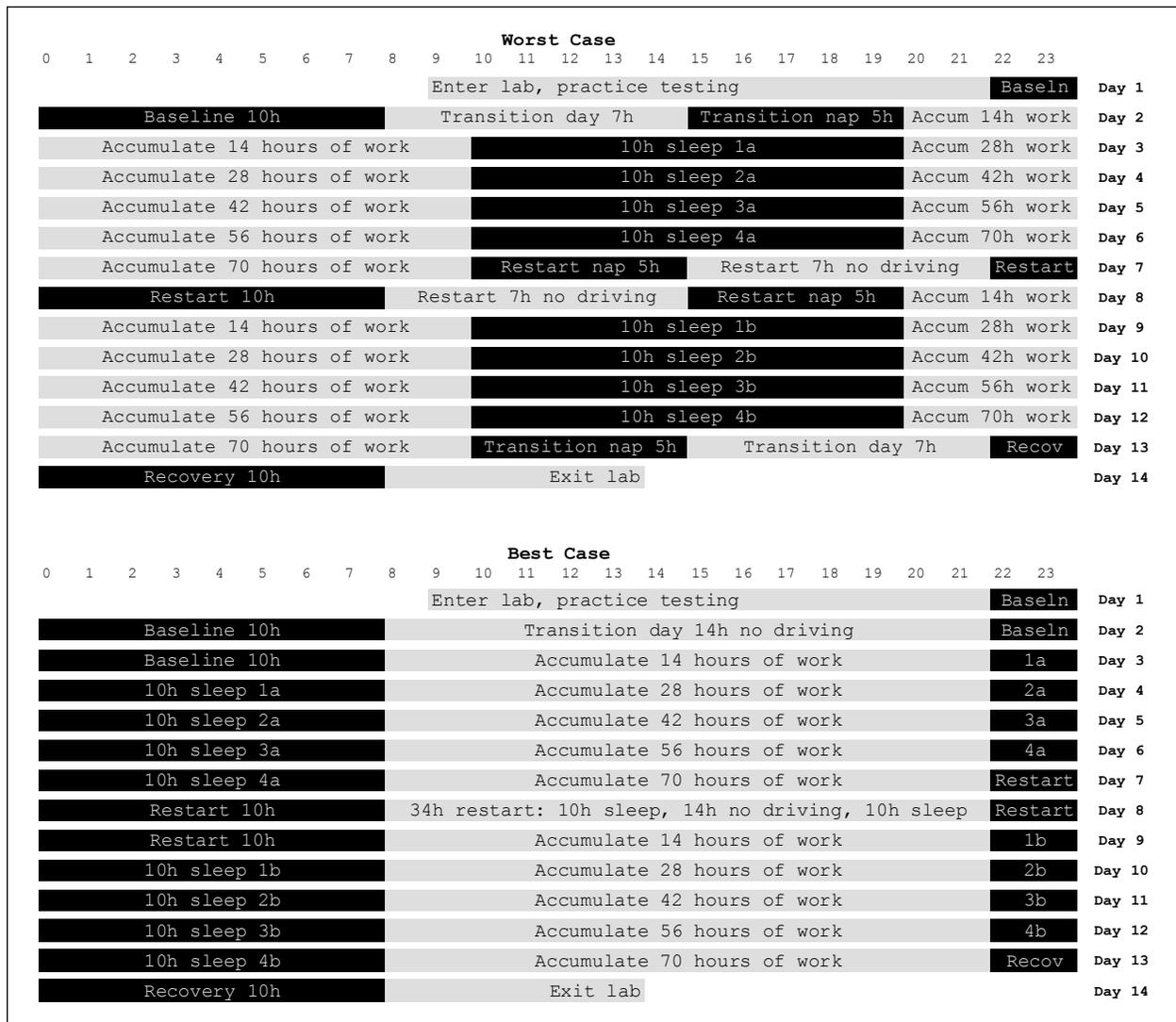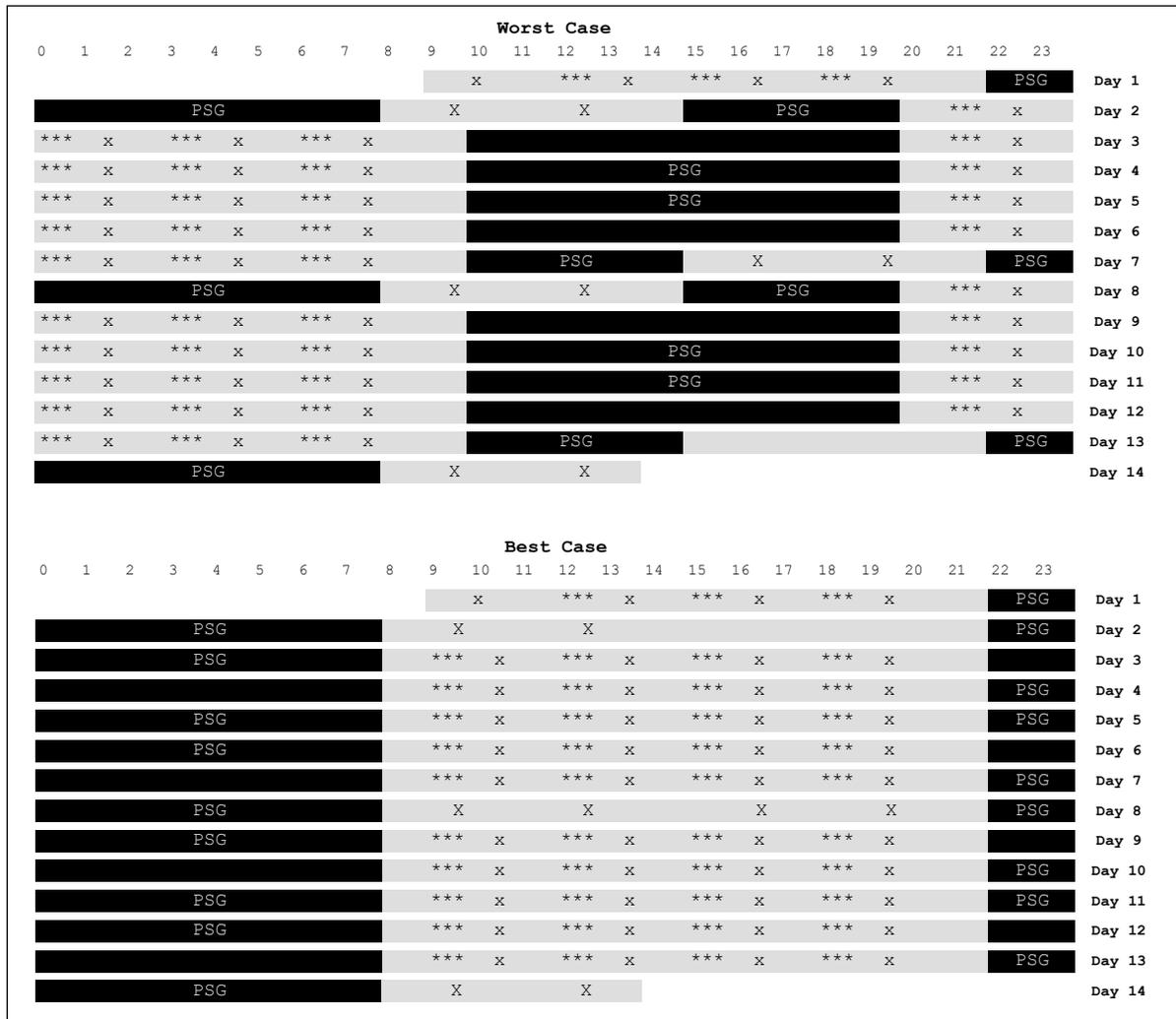

Figure 4. Chart. Measurement scheme for the "worst-case" (top schematic) and "best- case" (bottom schematic) conditions of the study, with hours of the day (midnight to midnight) progressing from left to right (see top numbers); days of the study progressing from top to bottom (see right side); gray indicating scheduled wakefulness; black indicating scheduled sleep periods; triple asterisks marking 40-minute driving simulator sessions each preceded and followed by a 10-minute PVT; lower-case x marking brief neurobehavioral performance test bouts; capital X marking brief neurobehavioral performance test bouts augmented with a PVT (during off-duty days); and PSG marking polysomnography (i.e., sleep recordings).

9

2.3   STATISTICAL METHODS AND POWER CALCULATION

The primary statistical design involved within-subject comparison of performance during the first 5-day work period with performance during the second 5-day work period (i.e., repeated measures), and between-groups comparison of the "worst-case" condition and the "best-case" condition. Therefore, we employed two-way mixed-effects analysis of variance (ANOVA) and focused on the interaction of group (condition) by work period (session). This interaction tests the null hypothesis that the 34-hour restart period is effective at maintaining performance in both conditions (regardless of the presence of any baseline differences). Additional analyses useful for interpretation of the main results involved three-way mixed-effects ANOVA of condition by session by day (to look at changes over days within sessions), and three-way mixed-effects ANOVA of condition by session by time of day (to examine the effects of circadian timing). The primary outcome measure we considered was PVT lapses (see above). Additional neurobehavioral outcomes and driving simulator performance outcomes (accounting for subjects' assignment to either simulator #1 or #2) were used for secondary analyses.

For a power calculation performed in advance of the study, it was noted that the literature suggests that daily sleep duration for work days in the "worst-case" condition would be expected to be ~5 hours.[8] It was therefore useful to consider performance impairment in the 5-hour sleep restriction condition of a study[1] conducted by the Walter Reed Army Institute of Research (WRAIR) as a basis for effect size estimation. In that study, performance was measured diurnally, and should thus be less impaired than would be expected in the nocturnally oriented "worst-case" condition of the present project. In other words, the WRAIR study would yield a conservative effect size estimate.

Assuming that the "best-case" condition would show no change in performance across days (i.e., would serve as a control condition) and could therefore be ignored for the purpose of power calculation, the primary comparison for the power calculation would involve a comparison of performance in the first 5-day work period with performance in the second 5-day work period in the "worst-case" condition. It was not known a priori what level of impairment the second 5-day work period would show, but we would consider it substantively deteriorated if it resembled the 3-hour sleep restriction condition in the WRAIR study. In order to obtain an expected effect size estimate, we compared performance across the first 5 sleep restrictions days of the 5-hour and 3-hour conditions of the WRAIR study ($n = 16$ for the 5-hour group; $n = 18$ for the 3-hour group), using PVT lapses as the primary outcome measure. Performing linear mixed-effects regression across days and examining the interaction with condition in this subset of the WRAIR study data yielded as the coefficient for the difference in slope between the two conditions 2.06 ± 0.29 (estimate ± standard error). The effect size was then calculated from the corresponding t statistic: $t_{168} = 7.12$ (p < 0.001). This yielded Cohen's $d = 2 \times 7.12 / \sqrt{168} = 1.10$, which is considered to be a *large* effect size.

In the present study, the comparison between the first and second 5-day work periods is within subjects. Based on the above effect size estimate, the power calculation showed that with a type I error rate of 0.05 (one-sided), a substantive deterioration of performance from the first to the second 5-day work period should be detectable with at least 95 percent statistical power (i.e., a type II error rate of 0.05) if there were 12 subjects in the "worst-case" condition. An equal number of subjects would be needed in the "best-case" (control) condition. We therefore

determined that a total sample size of at least $N = 24$ was needed to test whether or not the 34-hour restart period provided enough recovery in the "worst-case" condition to prevent cumulative performance degradation across work periods, with at least 95 percent statistical power.

Statistical analyses of polysomnographic measures of sleep were driven by which combinations of sleep periods corresponded best with each other between the two study conditions (see Figure 3 and Figure 4). A series of one-way ANOVAs was performed for comparison of the first (10-hour) baseline night between the two conditions; comparison of the first (10-hour) baseline night plus the second (10-hour) baseline night in the "best-case" condition with the first (10-hour) baseline night plus the first (5-hour) transition nap in the "worst-case" condition; comparison of the sum of the two polysomnographically recorded (10-hour) sleep periods in the first 5-day work period between the two conditions; comparison of the combination of the two (10-hour) nights in the 34-hour restart period of the "best-case" condition with the combination of the two (5-hour) transition naps plus the (10-hour) nocturnal sleep period in the 34-hour restart period of the "worst-case" condition; comparison of the sum of the two polysomnographically recorded (10-hour) sleep periods in the second 5-day work period between the two conditions; and comparison of the (10-hour) recovery night at the end of the "best-case" condition with the (5-hour) transition nap plus the (10-hour) recovery night at the end of the "worst-case" condition. Furthermore, a two-way ANOVA was performed focusing on the interaction of condition (group) by work period (session).

2.4 SUBJECT RECRUITMENT AND SCREENING

Subjects were healthy young individuals in the age range 22 to 40 years. This population was selected because of their relative homogeneity in sleep/wake and circadian physiology (e.g., minimal aging effects and no sleep disorders). This is beneficial (and necessary) for obtaining sufficient statistical power to demonstrate any effects of the "best-case" and "worst-case" conditions on performance. If any performance impairments are found in this population, then even greater deficits would be expected in the more heterogeneous population of CMV drivers.

The inclusion/exclusion criteria for the study were as follows:

- Physically and psychologically healthy (i.e., no clinical disorders and/or illnesses), as determined by physical exam, history, and questionnaires.

- No current medical or drug treatment (excluding oral contraceptives), as determined by history and questionnaire.

- No clinically significant abnormalities in blood and urine, and free of traces of drugs, as determined by blood chemistry and urinalysis, as well as a urine drug test upon entering the study.

- Free of traces of alcohol, as verified with a breathalyzer during screening and upon entering the study.

- No history of psychiatric illness, as determined by history and questionnaire.

- No history of drug or alcohol abuse in the past year, and no history of methamphetamine abuse, as determined by history and questionnaire.

- Not a current smoker, as determined by questionnaire.

- No history of moderate to severe brain injury, as determined by history and questionnaire.

- No history of a learning disability, as determined by questionnaire.

- Not susceptible to simulator adaptation syndrome, as determined by supervised test driving of the simulator followed by questionnaire and interview.

- No previous adverse reaction to sleep deprivation, as determined by history and questionnaire.

- Not vision-impaired (unless corrected to normal), as determined by questionnaire.

- Not pregnant, as determined by history and questionnaire.

- No sleep or circadian disorder, as determined by history, suite of questionnaires, and baseline polysomnography.

- Good habitual sleep, between 6 and 10 hours in duration, as determined by questionnaire and verified with wrist actigraphy and diary in the week before the study.

- Regular bedtimes, habitually getting up between 06:00 am and 09:00 am, as determined by questionnaire and verified with actigraphy and diary in the week before the study.

- Neither an extreme morning-type nor an extreme evening-type, as determined by questionnaire.

- No travel across time zones within one month of entering the study, as determined by questionnaire.

- No shift work within one month of entering the study, as determined by questionnaire.

- Native English speaker, as determined by questionnaire.

- Proficient driver, as determined by valid driver's license and supervised test driving of the simulator.

- Age from 22 to 40 years, as verified by date of birth on driver's license.

Prospective subjects were recruited with advertisements in local newspapers and on the Internet. A total of 760 individuals responded to the advertisements and were interviewed by telephone. Those who met key selection criteria such as age were screened during two laboratory-based screening sessions, beginning with an informed consent procedure. Screening procedures included a physical exam, blood and urine samples, supervised test driving of the simulator, and a variety of questionnaires (see inclusion/exclusion criteria above). Subject recruitment continued until the sample size required for the study was reached.

3. RESULTS

3.1 SUBJECTS

Data were available for a total of $N = 27$ healthy subjects who met the inclusion/exclusion criteria (see above). An additional seven subjects began the study but their data could not be used because of development of flu-like symptoms (one subject); polysomnographic evidence of a sleep disorder (two subjects); suspected non-compliance (two subjects); and subject-initiated withdrawal from the study (two subjects).

Of the sample of 27 subjects, 14 subjects were randomized to the "best-case" condition, and 13 subjects were randomized to the "worst-case" condition. The subjects in the "best-case" condition were seven men and seven women, ages 23–38 (mean ± standard deviation: 27.5 ± 5.6). The subjects in the "worst-case" condition were six men and seven women, ages 22–39 (mean ± standard deviation: 27.0 ± 5.4). There was no significant difference in age distribution between the two groups ($F_{1,25} = 0.055$, $p = 0.82$).

3.2 PERFORMANCE OUTCOMES

3.2.1 Primary Performance Outcome

The primary outcome measure for the study was the number of lapses on the PVT. The primary statistical analysis focused on the interaction of group ("worst-case" vs. "best-case" condition) by 5-day work period (pre-restart vs. post-restart session), collapsed over days and times of day within sessions. This interaction tests the null hypothesis that the 34-hour restart period is effective at maintaining performance in both conditions. This interaction was statistically significant for PVT lapses ($F_{1,2129} = 20.06$, $p < 0.001$). There was also a significant main effect of session ($F_{1,2129} = 21.79$, $p < 0.001$), but no significant main effect of group ($F_{1,2129} = 1.31$, $p = 0.25$). Figure 5 displays the interaction, showing that whereas performance did not vary significantly across the study in the "best-case" condition (simple effect of session: $F_{1,2129} = 0.02$, $p = 0.89$), there was significant deterioration of performance from before the 34-hour restart to after it in the "worst-case" condition (simple effect of session: $F_{1,2129} = 40.38$, $p < 0.001$).
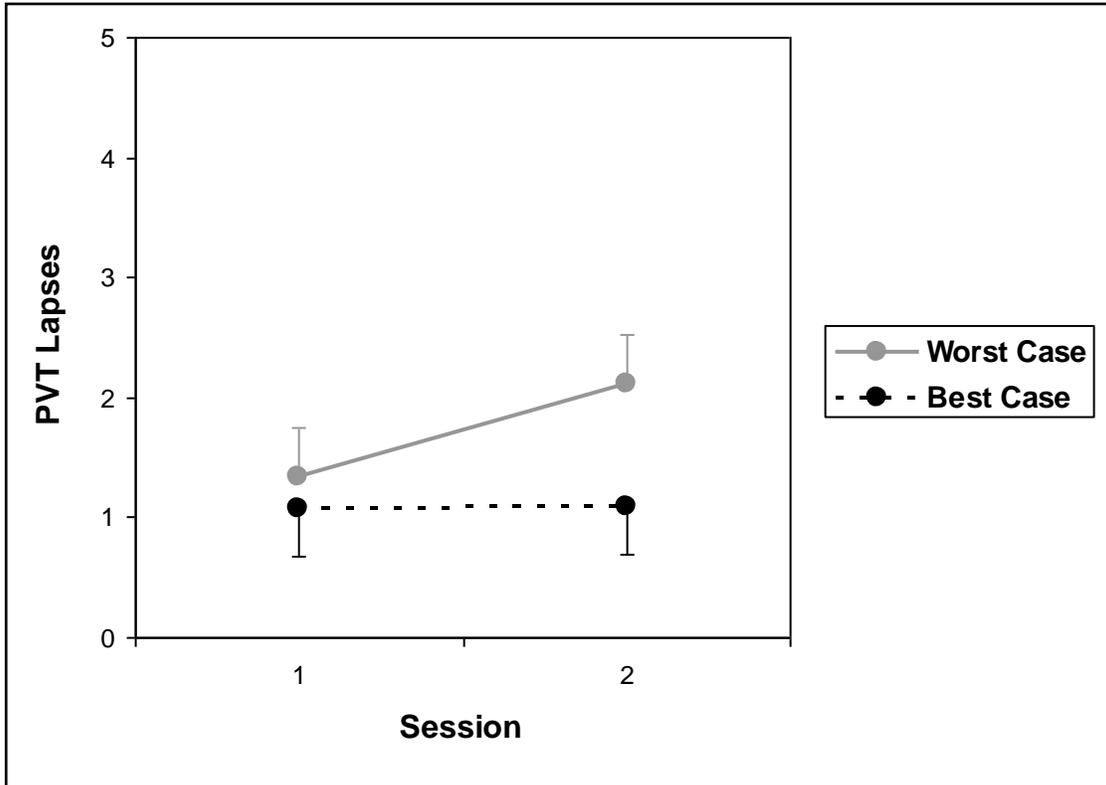
Figure 5. Graph. Lapses (number of reaction times greater than 500 ms) on the 10-minute PVT in the pre-restart 5-day work period (session 1) as compared to the post-restart 5-day work period (session 2) for the "worst-case" and "best-case" conditions (error bars indicate standard error derived from mixed-effects ANOVA).

In order to look at changes in performance over days within sessions, a further analysis of PVT lapses investigated the interaction of group by work period by day, collapsed over time of day. This three-way interaction was not statistically significant ($F_{4,2113} = 0.79$, $p = 0.53$). The main effect of day was also not significant ($F_{4,2113} = 0.70$, $p = 0.59$), nor were the two-way interactions of group by day ($F_{4,2113} = 1.50$, $p = 0.20$) and work period by day ($F_{4,2113} = 0.69$, $p = 0.60$). Figure 6 below displays the data by day. Although it might seem that performance deteriorated progressively from the first to the third day of the second 5-day work period in the "worst-case" condition, this was not borne out by the statistical results. Thus, it may be concluded that the effects seen in Figure 6 above were relatively stable over days within the 5-day work periods.
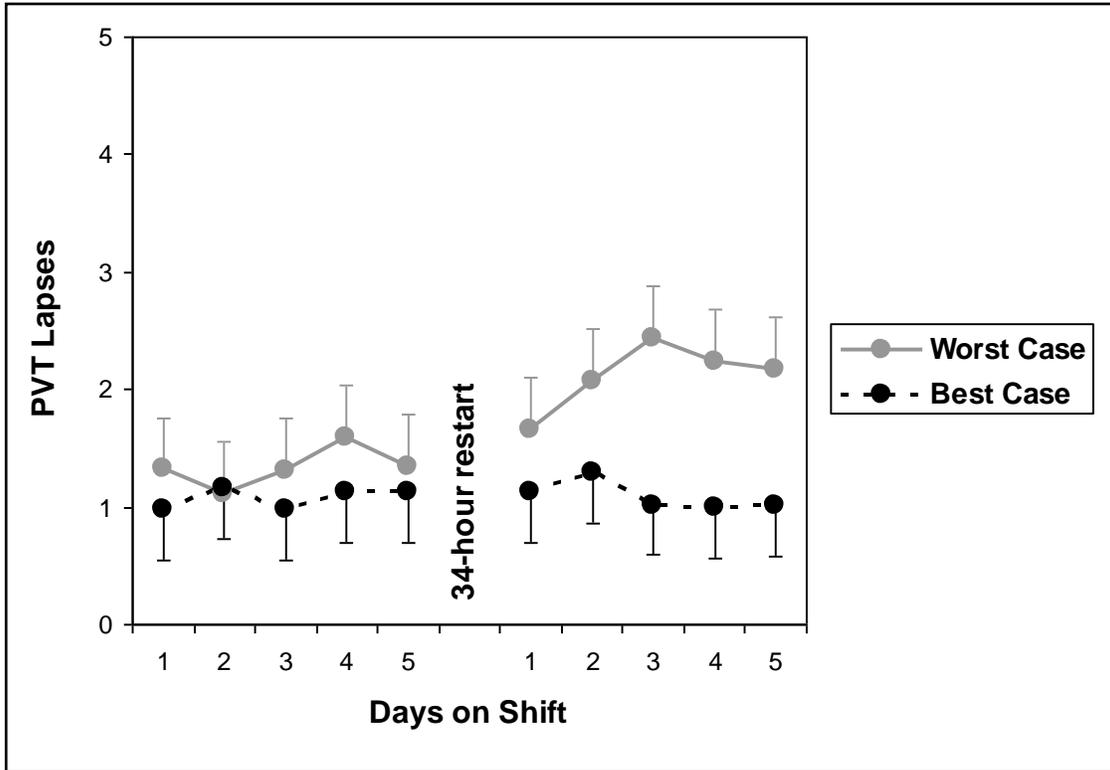
Figure 6. Graph. Lapses on the 10-minute PVT as a function of days in the 5-day work periods before and after the 34-hour restart period, for the "worst-case" and "best-case" conditions (vertical error bars indicate standard error derived from mixed-effects ANOVA).

In order to study performance in the two conditions as a function of time of day, another secondary analysis of PVT lapses investigated the interaction of group by work period by time of day, collapsed over days within each 5-day work period. The three-way interaction was not statistically significant ($F_{7,2101} = 0.44$, $p = 0.88$). Yet, there was a significant main effect of time of day ($F_{7,2101} = 16.32$, $p < 0.001$) and a significant interaction of group by time of day ($F_{7,2101} = 10.59$, $p < 0.001$), although the interaction of work period by time of day was not significant ($F_{7,2101} = 0.22$, $p = 0.98$). Figure 7 displays the data as a function of time of day (collapsed over days within each work period). The figure shows that performance was relatively stable across time of day in the "best-case" condition. However, there was a clear time-of-day effect in the "worst-case" condition, with performance degrading progressively across time of night.

Figure 7. Graph. Lapses on the 10-minute PVT as a function of time of day, collapsed over the 5-day work periods before and after the 34-hour restart period (error bars indicate standard error derived from mixed-effects ANOVA). Times of day are through the night (9:05 p.m.–6:55 a.m.) for the "worst-case" condition and through the day (9:05 a.m.–6:55 p.m.) for the "best-case" condition; 1 hour should be added for subjects who were assigned to performance testing first and driving second.

3.2.2   Secondary Performance Outcomes

Secondary performance outcomes were derived from a computerized neurobehavioral test battery which included, in order of presentation, the KSS, VASM, PANAS (both subscales were analyzed), DSST, PERF, EFFR, and CDDT. Results for these outcome measures are presented here in that order.

For the KSS, the primary analysis focusing on the interaction of group by session (collapsed over days and over times of day within sessions) yielded a significant interaction effect ($F_{1,1050} = 8.34$, $p = 0.004$) and a significant effect of group ($F_{1,1050} = 8.11$, $p = 0.005$), but no significant effect of session ($F_{1,1050} < 0.01$, $p = 0.99$). Figure 8 displays these findings, showing that subjects in the "worst-case" condition were overall sleepier than those in the "best-case" condition. In addition, subjects in the "worst-case" condition were slightly less sleepy after the 34-hour restart than before, whereas subjects in the "best-case" condition were slightly sleepier after the 34-hour restart than before.

16

Figure 8. Graph. Subjective sleepiness score on the KSS in the pre-restart 5-day work period (session 1) as compared to the post-restart 5-day work period (session 2) for the "worst-case" and "best-case" conditions (error bars indicate standard error derived from mixed-effects ANOVA).

In order to investigate changes in subjective sleepiness over days within sessions, a further analysis of the KSS examined the interaction of group by work period by day, collapsed over time of day. The three-way interaction was not statistically significant ($F_{4,1034} = 0.92$, $p = 0.45$). However, the main effect of day was significant ($F_{4,1034} = 5.72$, $p < 0.001$), as was the two-way interaction of group by day ($F_{4,1034} = 4.01$, $p = 0.003$), though not the two-way interaction of work period (session) by day ($F_{4,1034} = 0.87$, $p = 0.48$). Figure 9 displays the data by day, showing that subjective sleepiness in the "worst-case" condition diminished across days in the first 5-day work period, increased immediately after the 34-hour restart period, and then diminished again across days in the second 5-day work period.
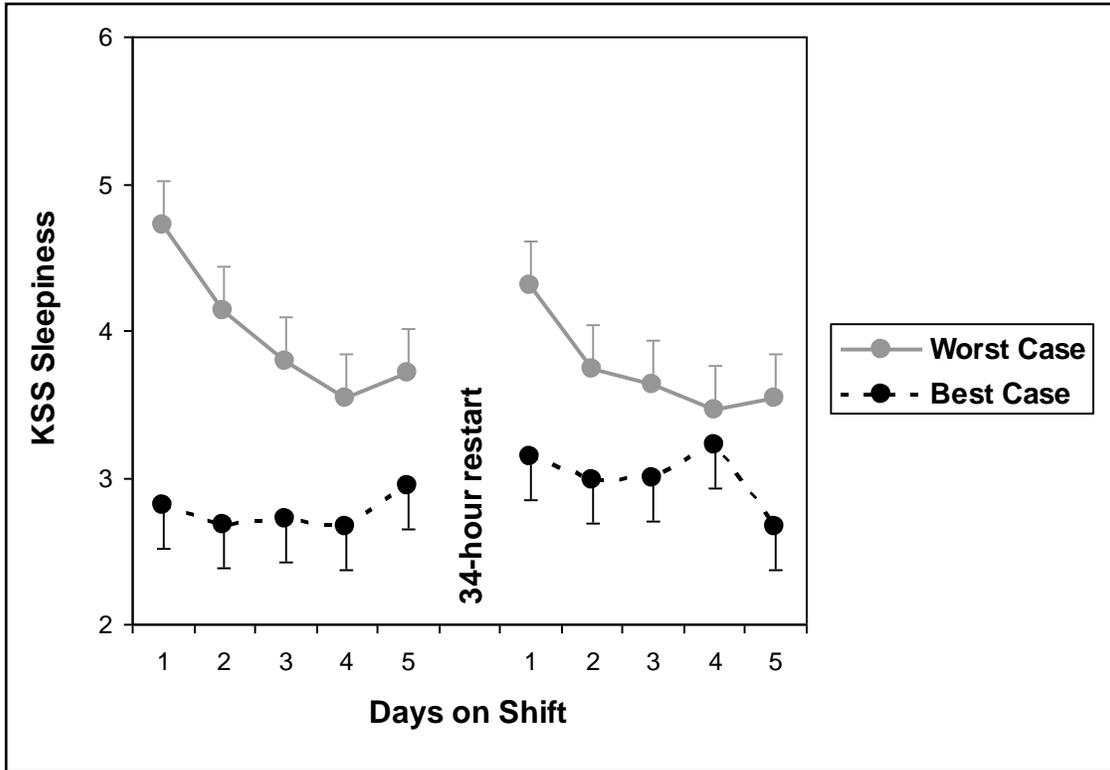
Figure 9. Graph. Subjective sleepiness on the KSS as a function of days in the 5-day work periods before and after the 34-hour restart period, for the "worst-case" and "best-case" conditions (error bars indicate standard error derived from mixed-effects ANOVA).

In order to investigate subjective sleepiness in the two conditions as a function of time of day, a final analysis of the KSS examined the interaction of group by work period by time of day, collapsed over days within each 5-day work period. The three-way interaction was not statistically significant ($F_{3,1038} = 1.86$, $p = 0.13$). Yet, there was a significant main effect of time of day ($F_{3,1038} = 75.87$, $p < 0.001$), a significant interaction of group by time of day ($F_{3,1038} = 49.28$, $p < 0.001$), and a significant interaction of work period by time of day ($F_{3,1038} = 5.91$, $p = 0.005$). Figure 10 displays the data as a function of time of day (collapsed over days within each work period). The figure shows that subjective sleepiness was relatively stable across time of day in the "best-case" condition. However, there was a clear time-of-day effect in the "worst- case" condition, with subjective sleepiness worsening progressively across time of night, particularly in the 5-day work period preceding the 34-hour restart.
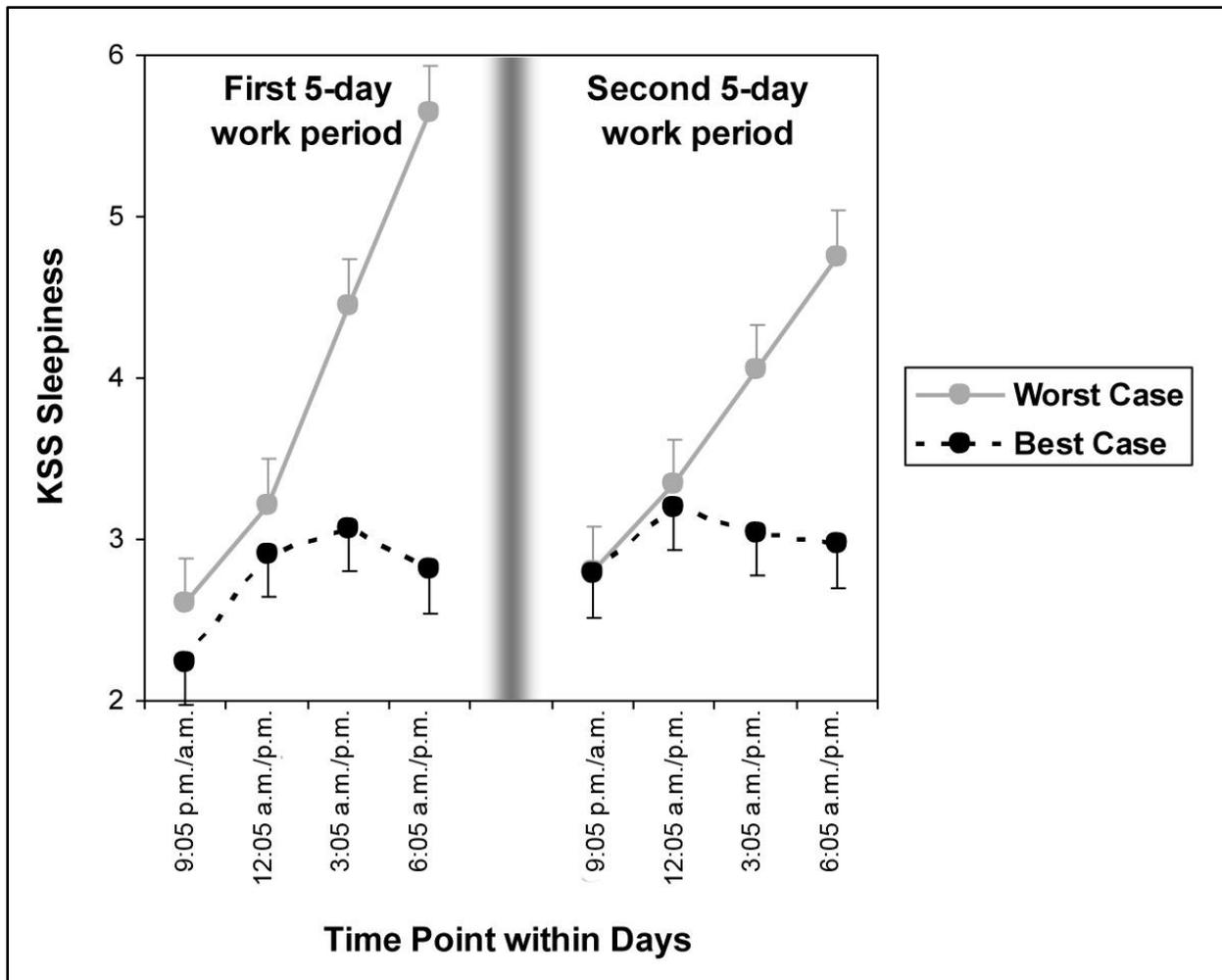
Figure 10. Graph. Subjective sleepiness on the KSS as a function of time of day, collapsed over the 5-day work periods before and after the 34-hour restart period (error bars indicate standard error derived from mixed-effects ANOVA). Times of day are through the night (9:05 p.m.–6:05 a.m.) for the "worst-case" condition and through the day (9:05 a.m.–6:05 p.m.) for the "best-case" condition; 1 hour and 50 minutes should be added to each time point for subjects who were assigned to driving first and performance testing second.

For the VASM, the primary analysis focusing on the interaction of group by session (collapsed over days and over times of day within sessions) yielded only a statistically significant effect of session ($F_{1,1051} = 13.01$, $p < 0.001$), with no significant interaction effect ($F_{1,1051} = 0.01$, $p = 0.92$) and no significant effect of group ($F_{1,1051} = 0.14$, $p = 0.71$). Figure 11 displays these findings, showing that there was a minor degradation of subjective mood across the study, but no differences between the "worst-case" and "best-case" conditions.

Figure 11. Graph. Subjective mood score on the VASM in the pre-restart 5-day work period (session 1) as compared to the post-restart 5-day work period (session 2) for the "worst-case" and "best-case" conditions.

Further analyses of the VASM to investigate changes in subjective mood over days within sessions, and as a function of time of day, partially corroborated the absence of statistically significant differences between the two groups. Mixed-effects ANOVA focusing on the interaction of group by work period by day, collapsed over time of day, yielded no significant three-way interaction ($F_{4,1035} = 0.98$, $p = 0.42$) and no significant interaction of group by day ($F_{4,1035} = 0.90$, $p = 0.46$), although there was a trend for a main effect of day ($F_{4,1035} = 2.34$, $p = 0.053$) and a significant interaction of work period by day ($F_{4,1035} = 4.64$, $p = 0.001$). However, mixed-effects ANOVA focusing on the interaction of group by work period by time of day, collapsed over days within each 5-day work period, yielded a significant interaction of group by time of day ($F_{3,1039} = 6.02$, $p < 0.001$), as well as a significant interaction of work period by time of day ($F_{3,1039} = 3.83$, $p = 0.010$) and a main effect of time of day ($F_{3,1039} = 14.09$, $p < 0.001$), although there was no significant three-way interaction ($F_{3,1039} = 1.15$, $p = 0.33$). Figure 12 displays the subjective mood data as a function of time of day (collapsed over days within each work period). The figure shows that subjective mood was relatively stable across time of day in

the "best-case" condition, but gradually declined across time of night in the "worst-case" condition. Furthermore, both groups displayed slightly degraded mood at the beginning of the wake period in the 5-day work period after the 34-hour restart period as compared to before the 34-hour restart period.
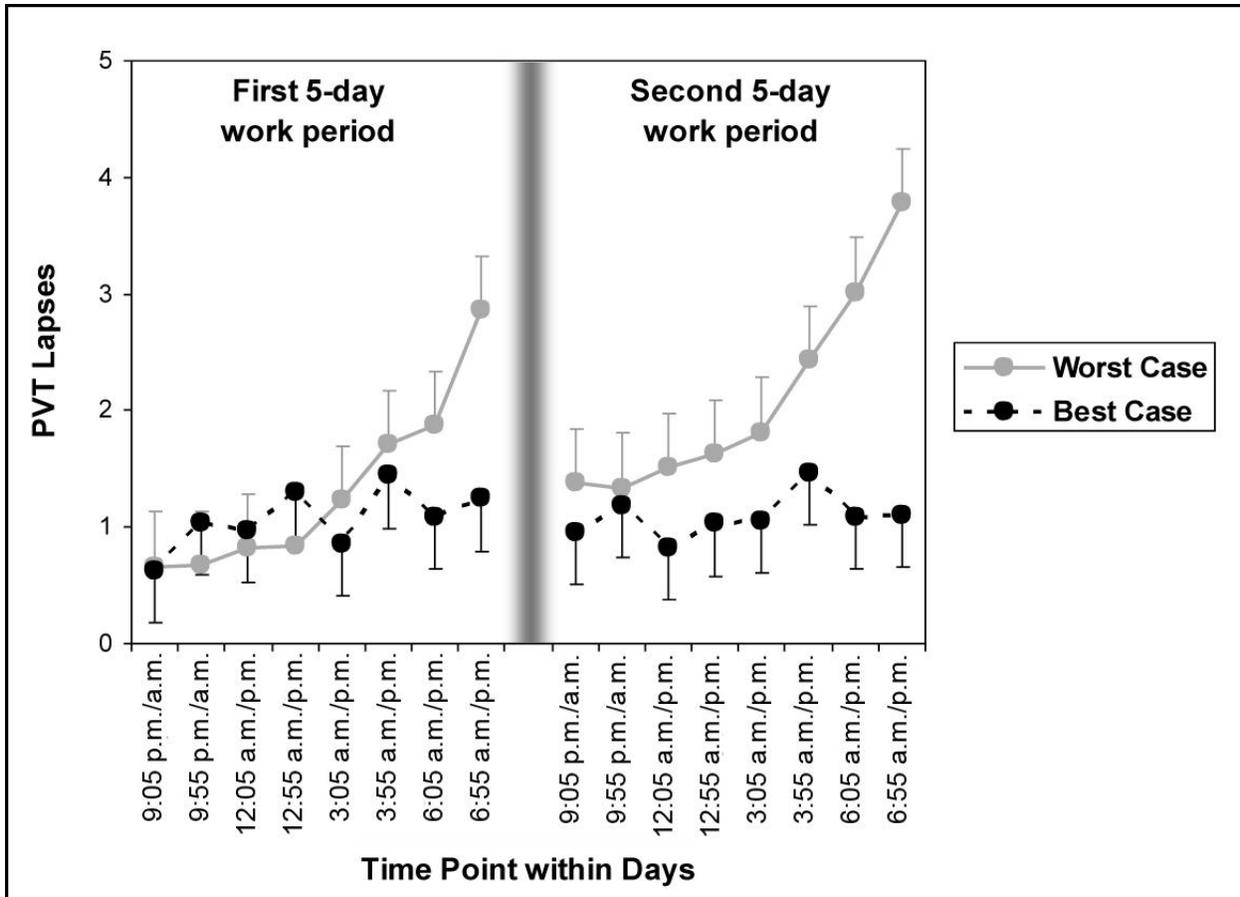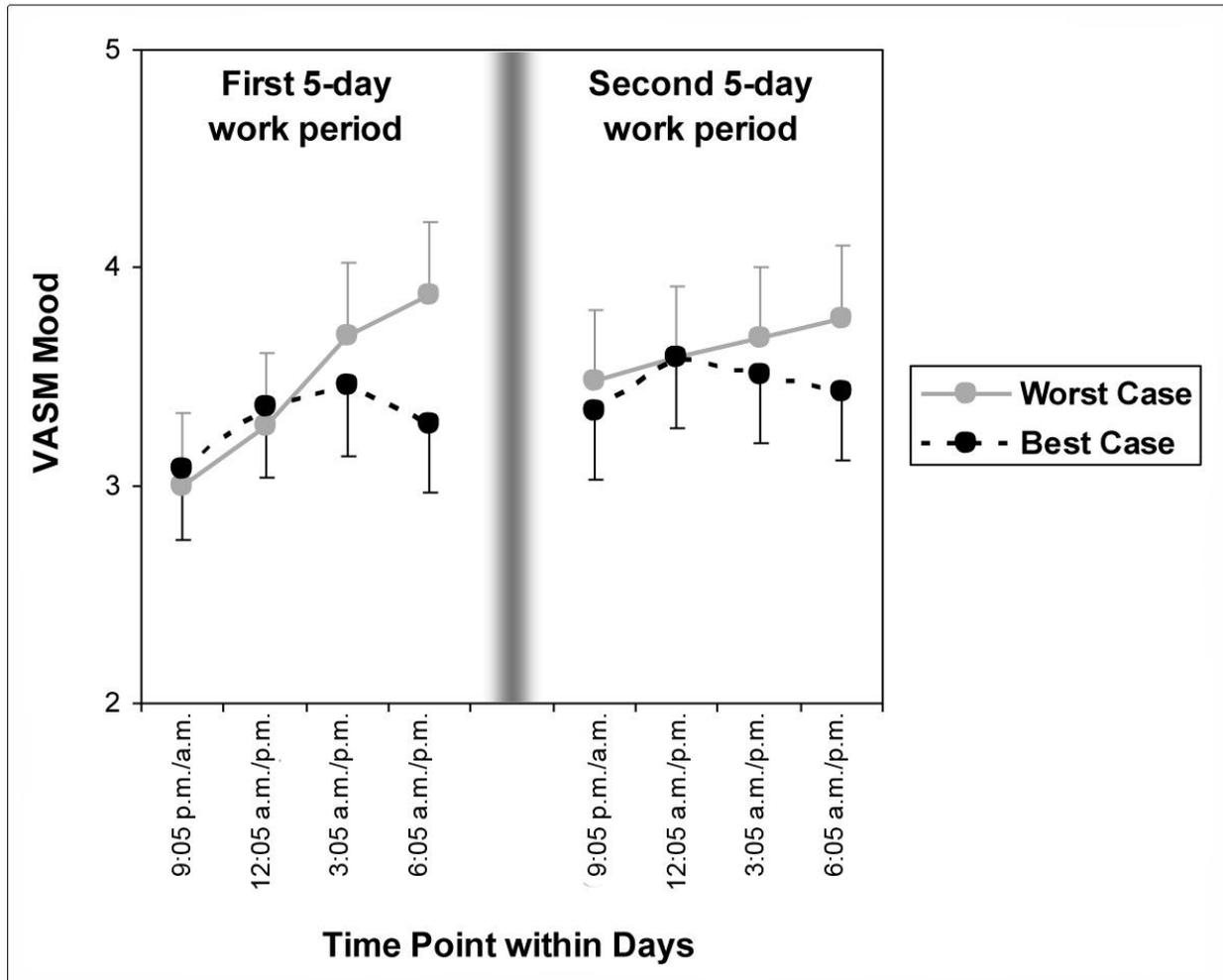

Figure 12. Graph. Subjective mood on the VASM as a function of time of day, collapsed over the 5-day work periods before and after the 34-hour restart period. Times of day are through the night (9:05 p.m.–6:05 a.m.) for the "worst-case" condition and through the day (9:05 a.m.–6:05 p.m.) for the "best-case" condition; 1 hour and 50 minutes should be added to each time point for subjects who were assigned to driving first and performance testing second.

The results for the positive affect scale of the PANAS mimicked those of the VASM. The primary analysis focusing on the interaction of group by session (collapsed over days and over times of day within sessions) yielded only a statistically significant effect of session ($F_{1,1051} = 20.48$, $p < 0.001$), with no significant interaction effect ($F_{1,1051} = 1.26$, $p = 0.26$) and no significant effect of group ($F_{1,1051} = 0.82$, $p = 0.37$). Figure 13 displays these findings, showing that there was a minor degradation of positive mood across the study. The difference between the "worst-case" and "best-case" conditions was not statistically significant.
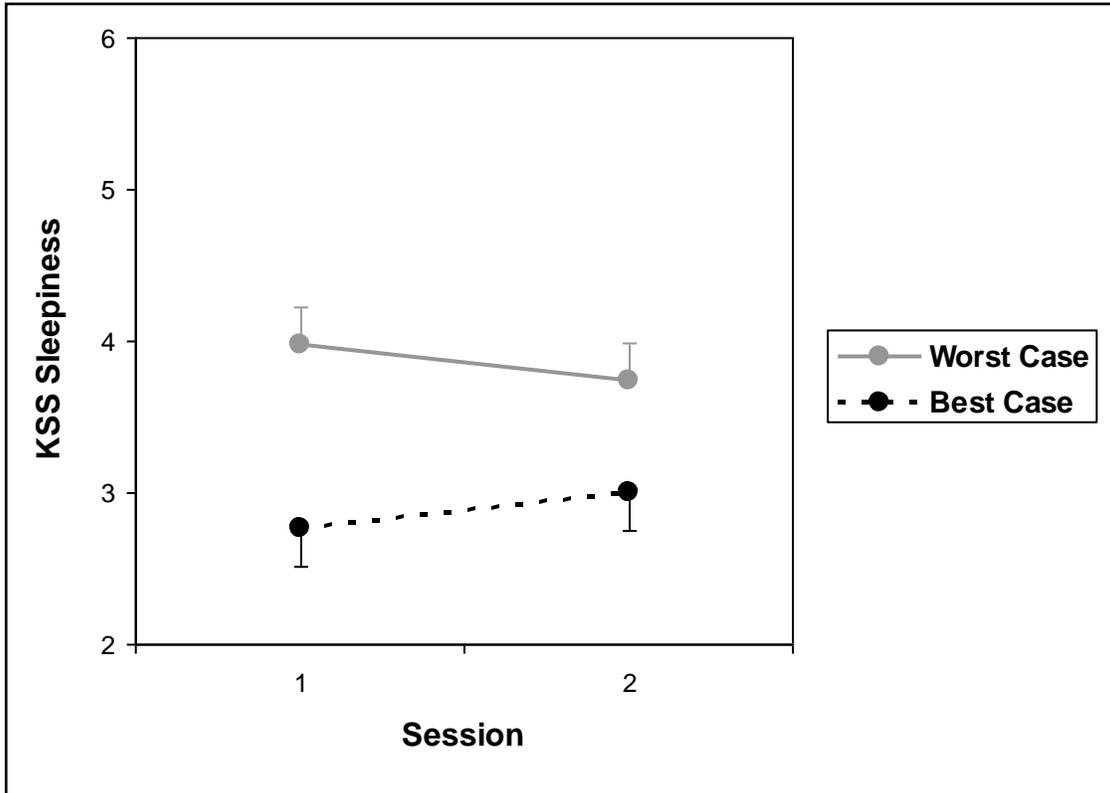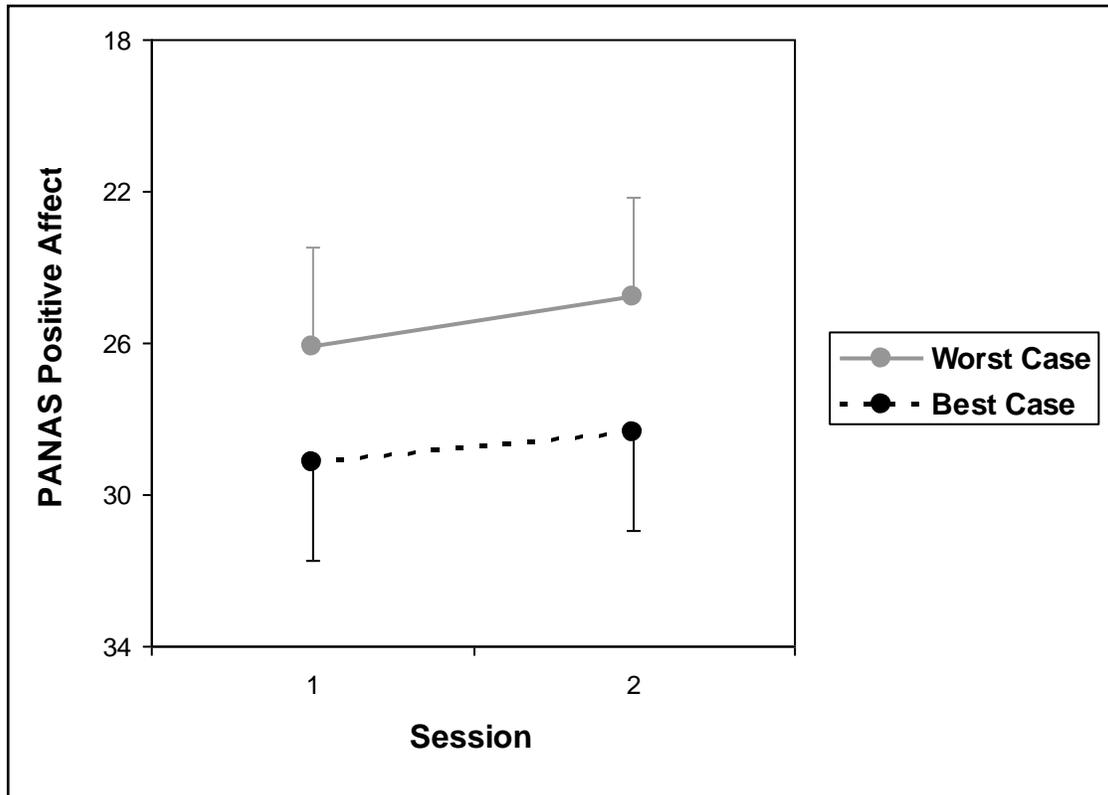
Figure 13. Graph. Positive affect score on the PANAS in the pre-restart 5-day work period (session 1) as compared to the post-restart 5-day work period (session 2) for the "worst-case" and "best-case" conditions (the vertical scale is inverted; upward direction corresponds to less positive affect).

Further analyses of the positive affect scale of the PANAS partially corroborated the absence of statistically significant differences between the two groups, as was the case also for the VASM. Mixed-effects ANOVA focusing on the interaction of group by work period by day, collapsed over time of day, yielded no significant three-way interaction ($F_{4,1035} = 0.91$, $p = 0.46$), no significant interaction of group by day ($F_{4,1035} = 1.48$, $p = 0.21$), and no main effect of day ($F_{4,1035} = 1.15$, $p = 0.33$), although there was a trend for an interaction of work period by day ($F_{4,1035} = 3.57$, $p = 0.007$). Mixed-effects ANOVA focusing on the interaction of group by work period by time of day, collapsed over days within each 5-day work period, yielded a significant interaction of group by time of day ($F_{3,1039} = 20.37$, $p < 0.001$), as well as a significant interaction of work period by time of day ($F_{3,1039} = 6.12$, $p < 0.001$) and a main effect of time of day ($F_{3,1039} = 24.00$, $p < 0.001$), although there was no significant three-way interaction ($F_{3,1039} = 0.93$, $p = 0.42$). As a function of time of day, the positive affect data considered (collapsed over days within each work period) paralleled the subjective mood data (see Figure 12).

For the negative affect scale of the PANAS, the primary analysis focusing on the interaction of group by session (collapsed over days and over times of day within sessions) yielded no significant interaction effect ($F_{1,1051} = 0.05$, $p = 0.83$), no significant effect of group ($F_{1,1051} = 0.20$, $p = 0.65$), and no significant effect of session ($F_{1,1051} = 0.53$, $p = 0.46$). Overall, the subjects in both conditions showed low negative affect throughout the study.

Yet, analysis of the interaction of group by work period by day, collapsed over time of day, showed a significant three-way interaction ($F_{4,1035} = 3.12$, $p = 0.014$) and a significant interaction of group by day ($F_{4,1035} = 2.76$, $p = 0.027$). There was no significant interaction of work period by day ($F_{4,1035} = 1.59$, $p = 0.018$), and no significant main effect of day ($F_{4,1035} = 0.57$, $p = 0.69$). Figure 14 displays the data by day, showing that in the first days of the study, negative affect was slightly elevated in the "worst-case" condition and slightly diminished in the "best-case" condition. However, starting on the third day of the first 5-day work period, the two groups were no longer statistically significantly different.

Figure 14. Graph. Negative affect score on the PANAS as a function of days in the 5-day work periods before and after the 34-hour restart period, for the "worst-case" and "best-case" conditions.

Analysis of the interaction of group by work period by time of day, collapsed over days within each 5-day work period, yielded no significant results—no three-way interaction ($F_{3,1039} = 0.17$, $p = 0.92$), no interaction of group by time of day ($F_{3,1039} = 0.21$, $p = 0.89$), no interaction of work period by time of day ($F_{3,1039} = 0.24$, $p = 0.87$), and no main effect of time of day ($F_{3,1039} = 0.50$, $p = 0.68$). Thus, negative affect was relatively stable across time of days in both conditions.

For the number of correct responses on the DSST, the primary analysis focusing on the interaction of group by session (collapsed over days and over times of day within sessions) yielded only a statistically significant effect of session ($F_{1,1050} = 292.92$, $p < 0.001$), with no significant interaction effect ($F_{1,1050} = 0.28$, $p = 0.60$) and no significant effect of group ($F_{1,1050} = 0.78$, $p = 0.38$). Figure 15 displays these findings, showing that performance on the DSST improved significantly across the study in both conditions.
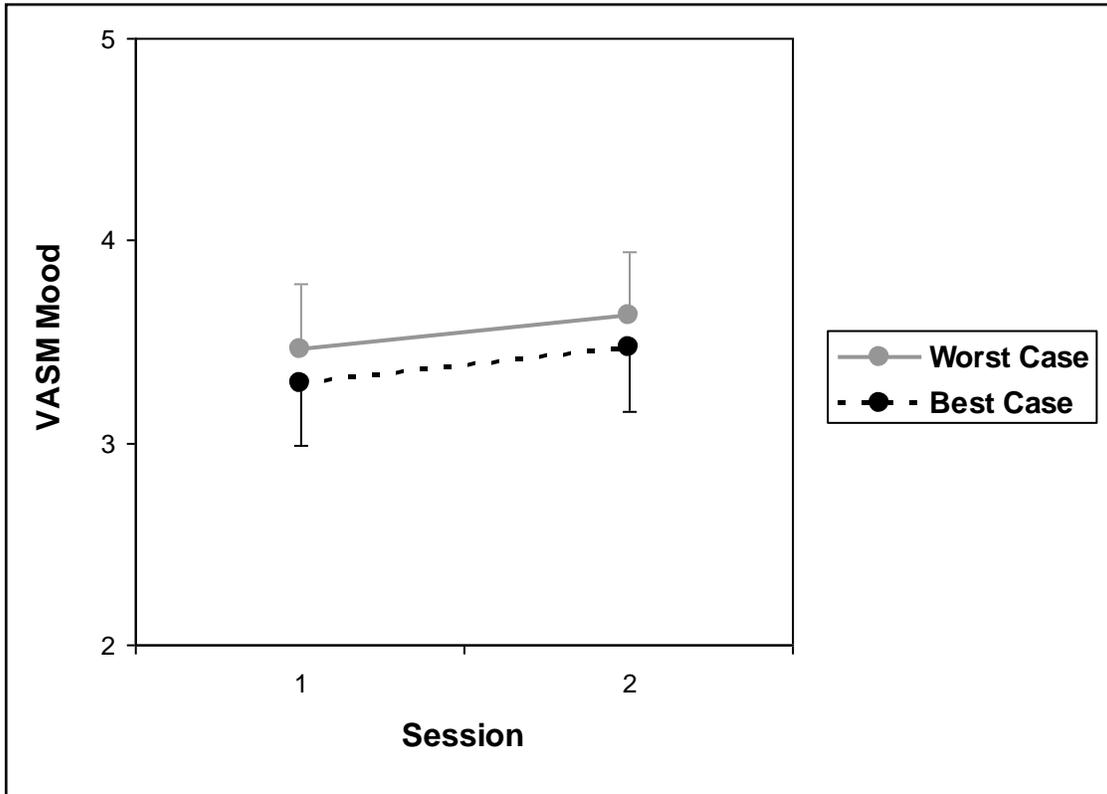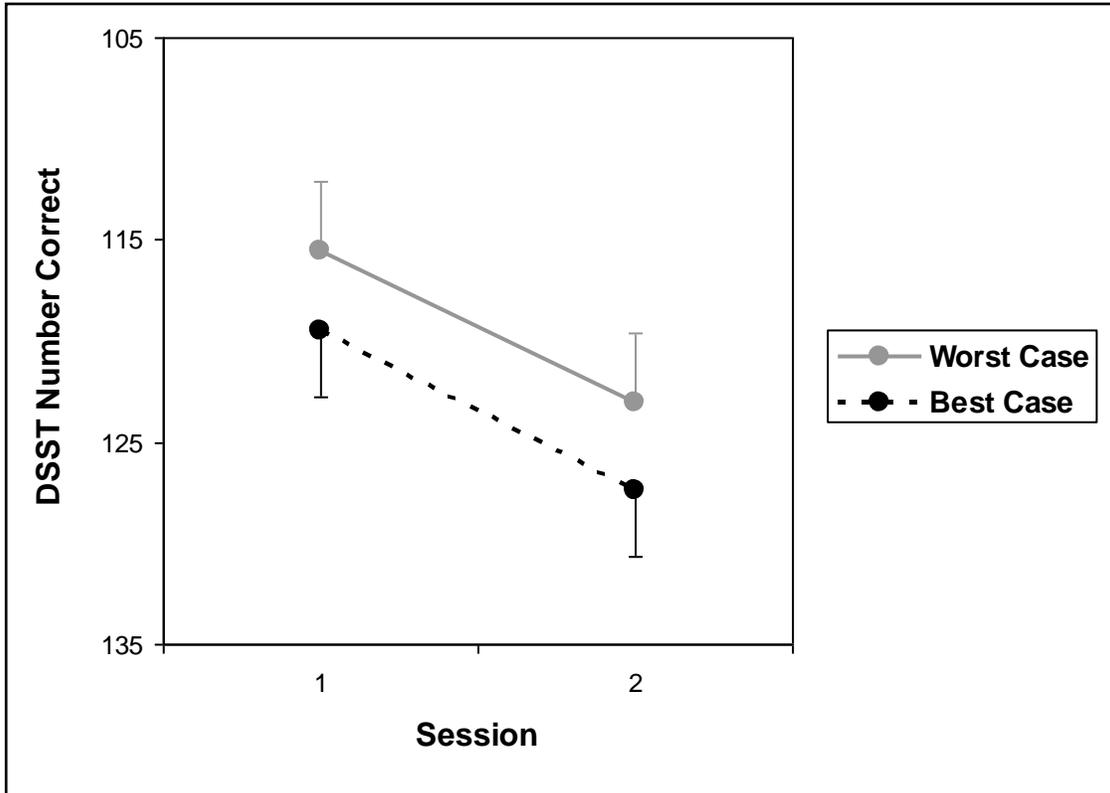
Figure 15. Graph. Number of correct responses on DSST in the pre-restart 5-day work period (session 1) as compared to the post-restart 5-day work period (session 2) for the "worst-case" and "best-case" conditions (the vertical scale is inverted; downward direction corresponds to performance improvement).

In order to investigate further the changes in DSST performance across days in the study, an additional analysis examined the interaction of group by work period by day, collapsed over time of day. The three-way interaction was not statistically significant ($F_{4,1034} = 0.82$, $p = 0.52$), nor was the two-way interaction of group by day ($F_{4,1034} = 0.51$, $p = 0.73$). However, the main effect of day was significant ($F_{4,1034} = 29.84$, $p < 0.001$), as was the two-way interaction of work period (session) by day ($F_{4,1034} = 6.13$, $p < 0.001$). Figure 16 displays the data by day, showing that performance on the DSST improved steadily across days in both conditions, in accordance with the well-known practice effect for this task.
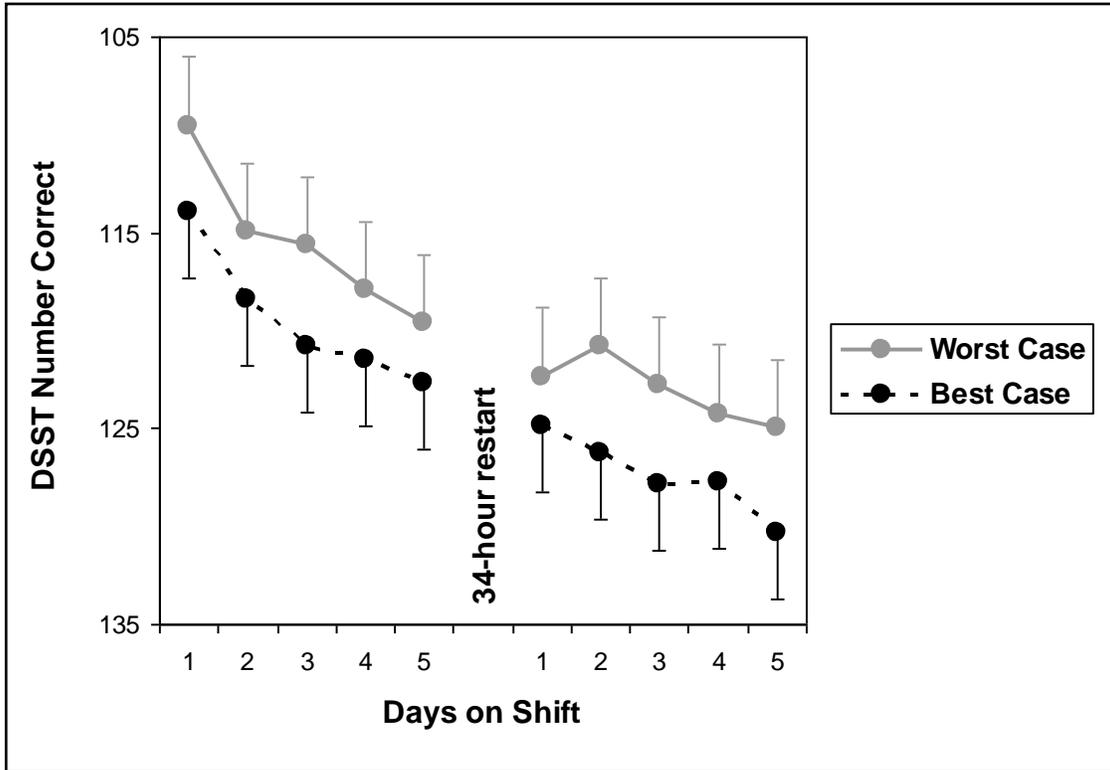
Figure 16. Graph. Number of correct responses on DSST as a function of days in the 5-day work periods before and after the 34-hour restart period, for the "worst-case" and "best-case" conditions (the vertical scale is inverted; downward direction corresponds to performance improvement).

Analysis of the interaction of group by work period by time of day, collapsed over days within each 5-day work period, yielded a significant two-way interaction of group by time of day ($F_{3,1038} = 2.94$, $p = 0.032$), but no significant three-way interaction ($F_{3,1038} = 0.74$, $p = 0.53$), no significant interaction of work period by time of day ($F_{3,1038} = 1.47$, $p = 0.22$), and no significant main effect of time of day ($F_{3,1038} = 0.12$, $p = 0.95$). Figure 17 shows the data as a function of time of day (collapsed over days within each work period). The figure shows the combined influences of the practice effect, promoting performance improvement across test bouts, and of time of day, promoting performance degradation across time of night in the "worst-case" condition (see figure 7).
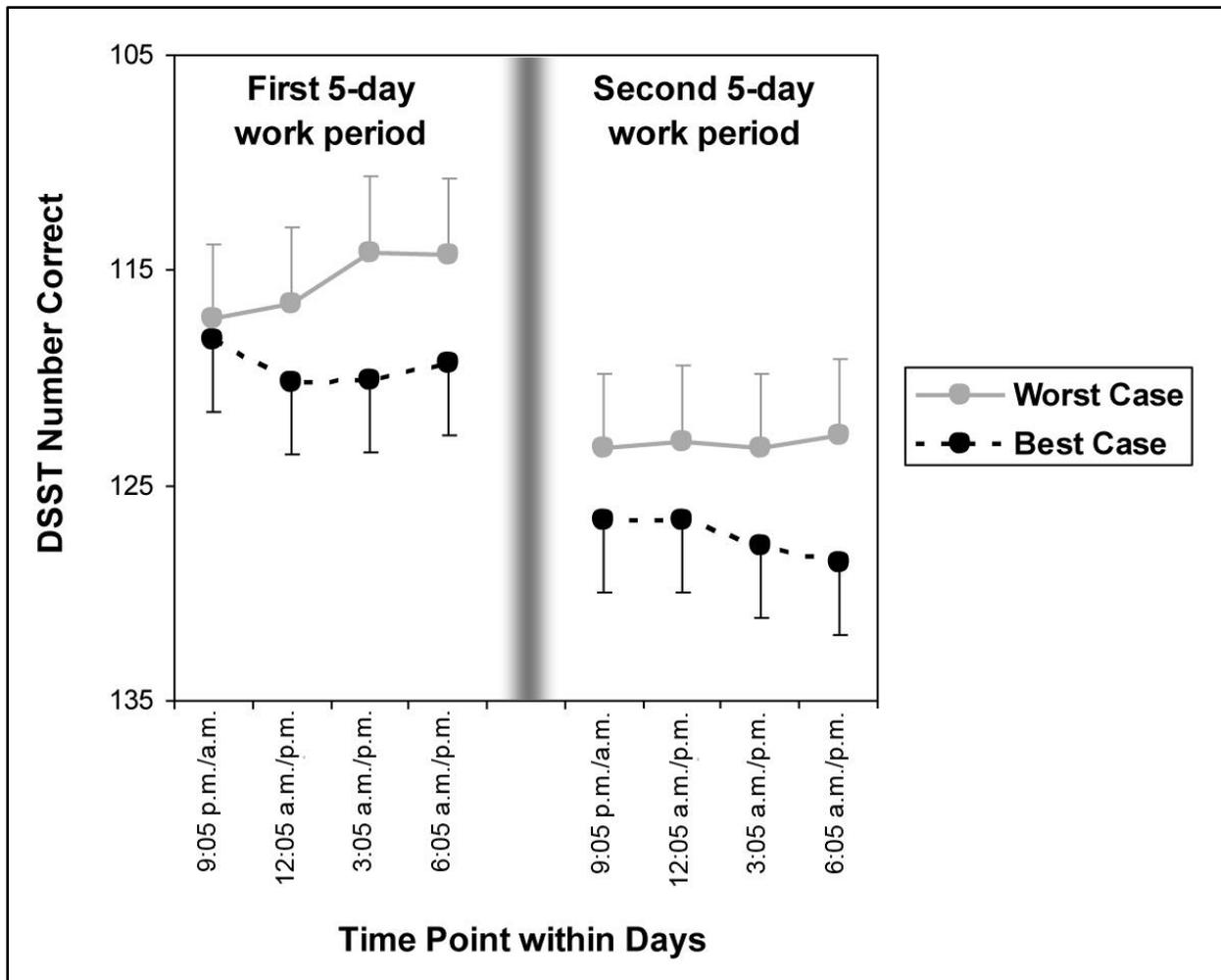
Figure 17. Graph. Number of correct responses on the DSST as a function of time of day, collapsed over the 5-day work periods before and after the 34-hour restart period (the vertical scale is inverted; downward direction corresponds to performance improvement). Times of day are through the night (9:05 p.m.–6:05 a.m.) for the "worst-case" condition and through the day (9:05 a.m.–6:05 p.m.) for the "best-case" condition; 1 hour and 50 minutes should be added to each time point for subjects who were assigned to driving first and performance testing second.

For the PERF administered near the end of each neurobehavioral test bout, the primary analysis focusing on the interaction of group by session (collapsed over days and over times of day within sessions) yielded a significant effect of group ($F_{1,1051} = 7.29$, $p = 0.007$), but no significant interaction effect ($F_{1,1051} = 0.14$, $p = 0.71$) and no significant effect of session ($F_{1,1051} = 0.28$, $p = 0.60$). Figure 18 displays these findings, showing that subjects in the "worst-case" condition consistently rated their performance as poorer than did subjects in the "best-case" condition.
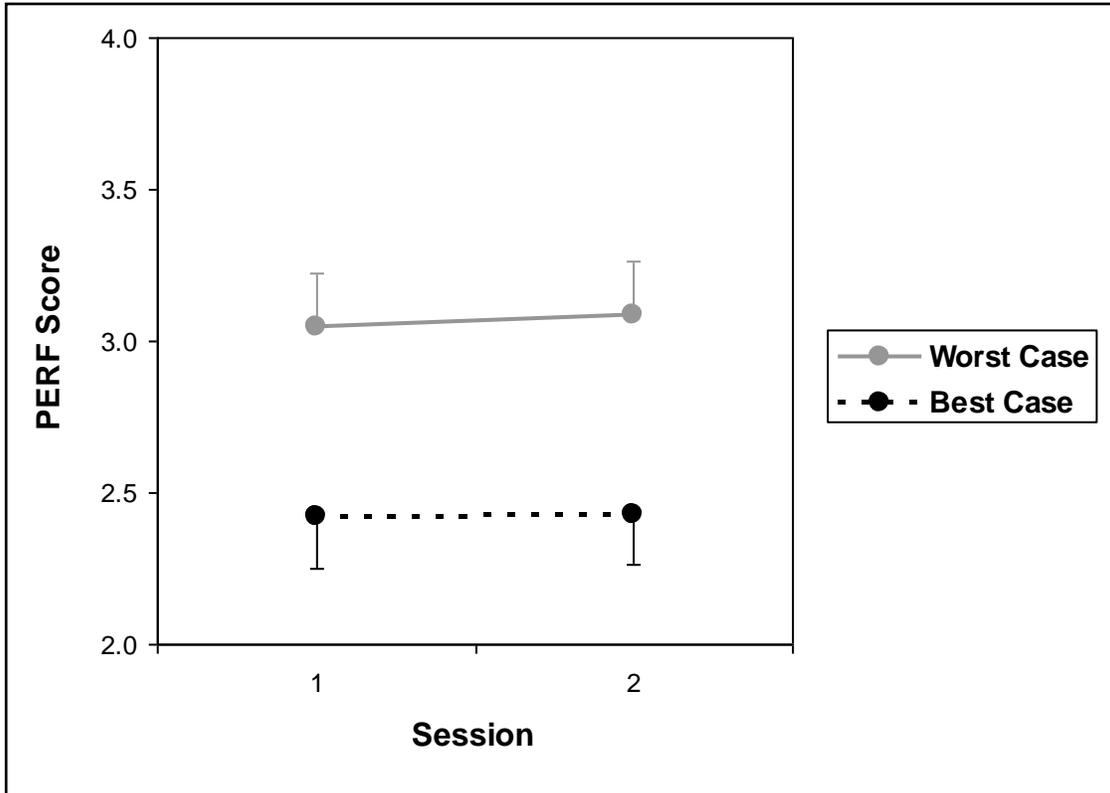
Figure 18. Graph. Retrospective self-evaluation score on the PERF in the pre-restart 5-day work period (session 1) as compared to the post-restart 5-day work period (session 2) for the "worst-case" and "best-case" conditions (upward direction corresponds to poorer performance).

In order to investigate changes in performance ratings over days within sessions, a further analysis of the PERF examined the interaction of group by work period by day, collapsed over time of day. The three-way interaction was not statistically significant ($F_{4,1035} = 1.60$, $p = 0.17$), and neither was the two-way interaction of group by day ($F_{4,1035} = 0.73$, $p = 0.57$). However, the main effect of day was significant ($F_{4,1035} = 2.66$, $p = 0.032$), and there was a trend for a two-way interaction of work period (session) by day ($F_{4,1035} = 2.18$, $p = 0.069$). Figure 19 displays the data by day, revealing primarily that in the 5-day work period after the 34-hour restart, both groups rated their performance as slightly improving across the days.

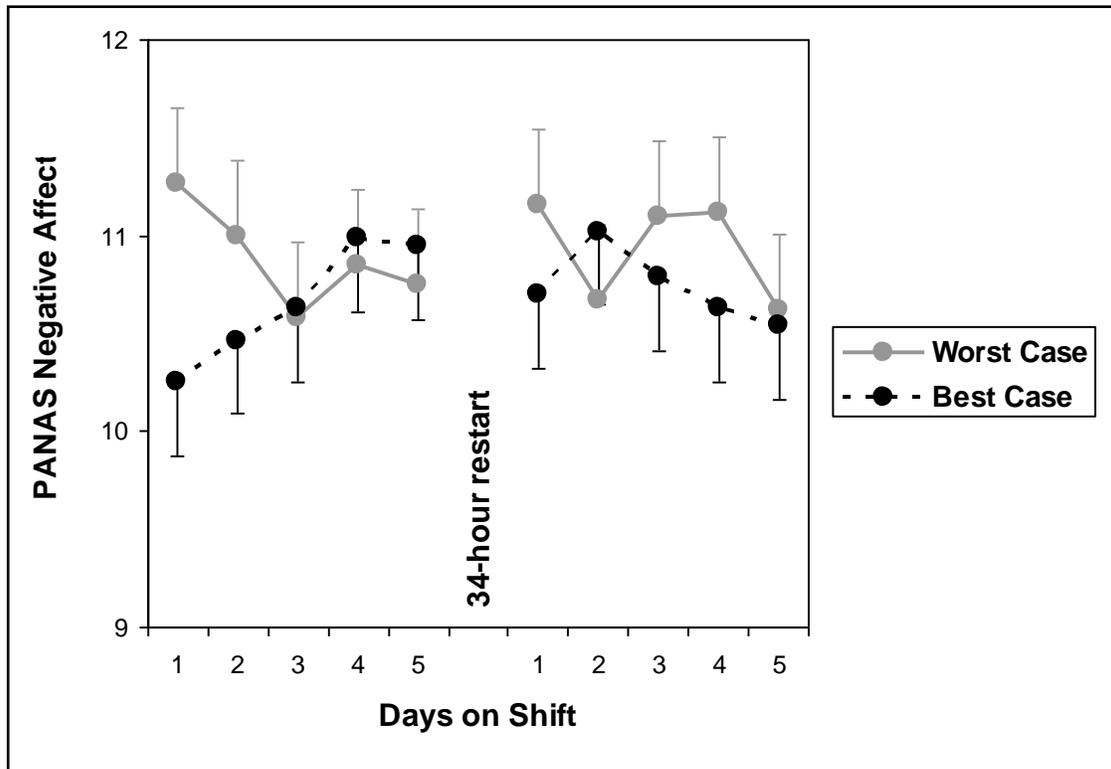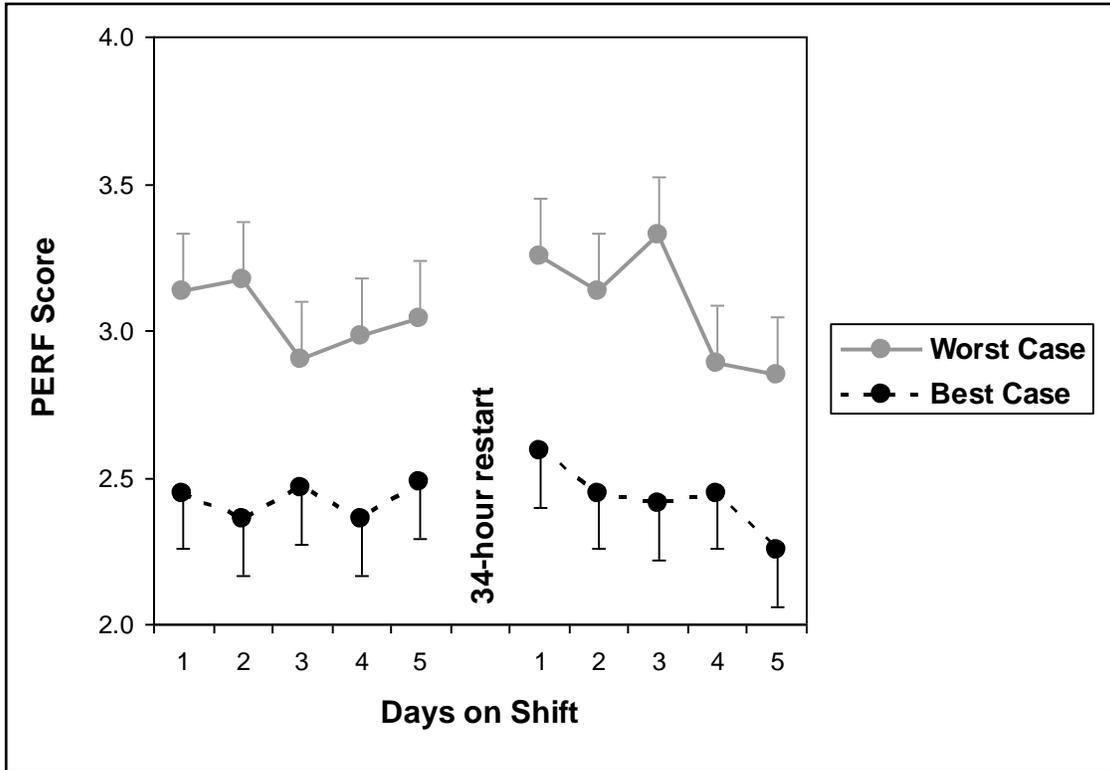**Figure 19. Graph. Retrospective self-evaluation score on the PERF as a function of days in the 5-day work periods before and after the 34-hour restart period, for the "worst-case" and "best-case" conditions (upward direction corresponds to poorer performance).**

In order to investigate changes in performance ratings in the two conditions as a function of time of day, an additional analysis of the PERF examined the interaction of group by work period by time of day, collapsed over days within each 5-day work period. The three-way interaction was not statistically significant ($F_{3,1039} = 0.20$, $p = 0.89$), nor were the two-way interaction of group by time of day ($F_{3,1039} = 0.68$, $p = 0.56$) and the main effect of time of day ($F_{3,1039} = 1.53$, $p = 0.21$). However, there was a significant two-way interaction of work period by time of day ($F_{3,1039} = 3.27$, $p = 0.021$), which is shown in Figure 20. During the 5-day work period preceding the 34-hour restart, subjects rated their performance as slightly deteriorating during the later hours of the waking period, regardless of condition. This effect was no longer apparent in the 5-day work period following the 34-hour restart.
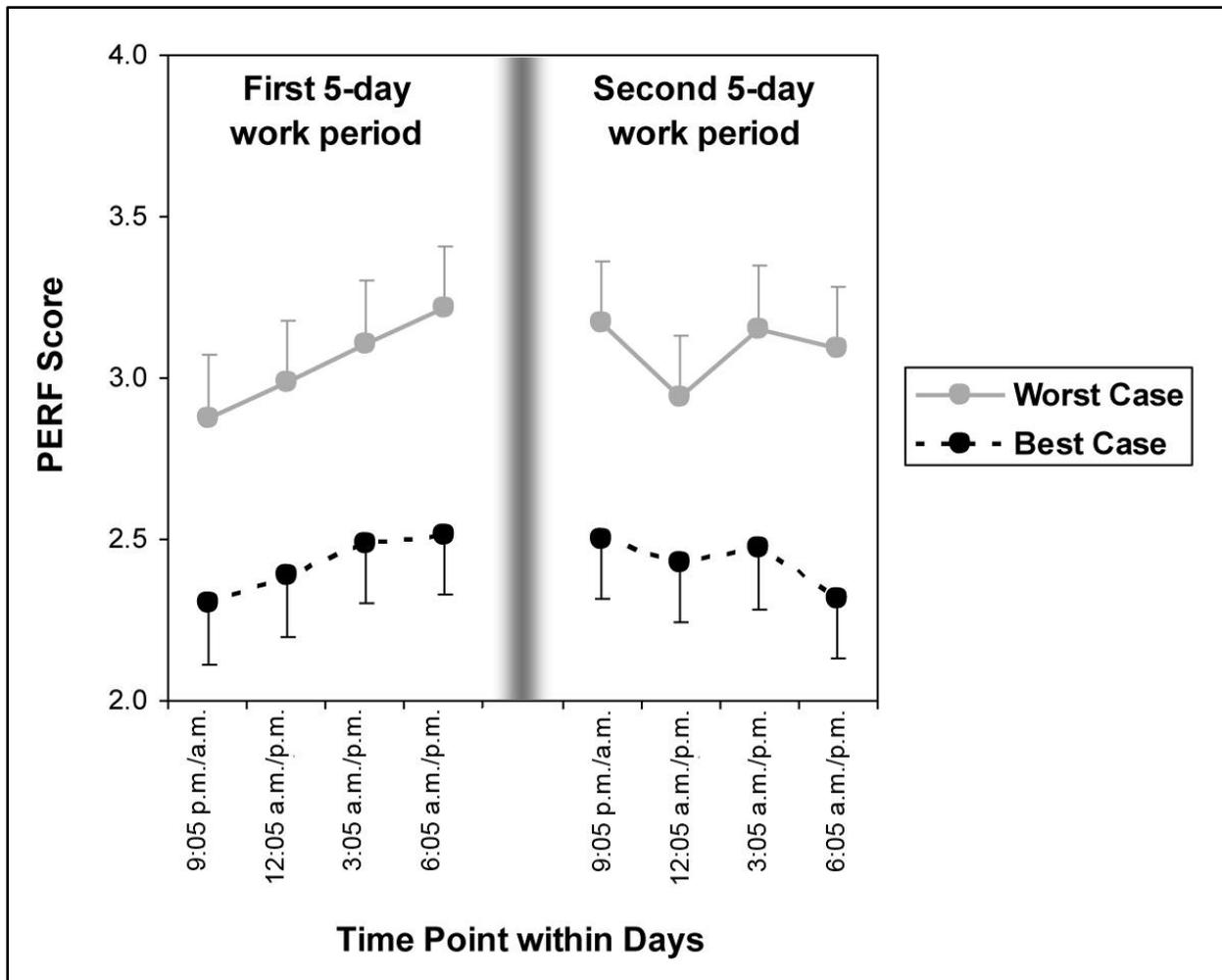
Figure 20. Graph. Retrospective self-evaluation score on the PERF as a function of time of day, collapsed over the 5-day work periods before and after the 34-hour restart period (upward direction corresponds to poorer performance). Times of day are through the night (9:05 p.m.–6:05 a.m.) for the "worst-case" condition and through the day (9:05 a.m.–6:05 p.m.) for the "best-case" condition; 1 hour and 50 minutes should be added to each time point for subjects who were assigned to driving first and performance testing second.

For the EFFR administered near the conclusion of each neurobehavioral test bout, the primary analysis focusing on the interaction of group by session (collapsed over days and over times of day within sessions) yielded no statistically significant effects: no interaction ($F_{1,1051} = 2.19$, $p = 0.14$), and no main effects of group ($F_{1,1051} = 0.96$, $p = 0.33$) and session ($F_{1,1051} = 0.83$, $p = 0.36$). Figure 21 displays these results, showing no differences between conditions, or before and after the 34-hour restart, in the amount of effort subjects felt was necessary to complete the neurobehavioral test bouts.
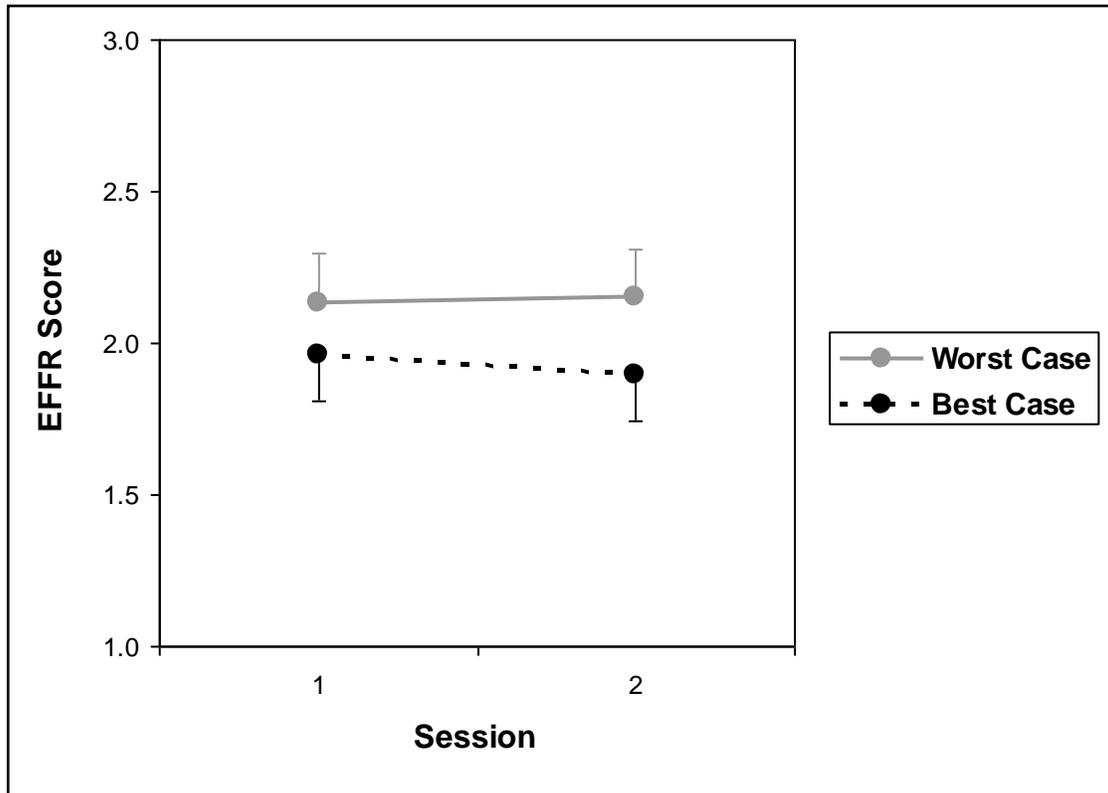
Figure 21. Graph. Retrospective self-evaluation score on the EFFR in the pre-restart 5-day work period (session 1) as compared to the post-restart 5-day work period (session 2) for the "worst-case" and "best-case" conditions (upward direction corresponds to greater subjective effort needed to complete the neurobehavioral test bouts).

Further analyses of the EFFR partially corroborated the absence of statistically significant differences between the two groups. Mixed-effects ANOVA focusing on the interaction of group by work period by day, collapsed over time of day, yielded no significant three-way interaction ($F_{4,1035} = 0.86$, $p = 0.49$), no significant interaction of group by day ($F_{4,1035} = 0.91$, $p = 0.45$), no significant interaction of work period by day ($F_{4,1035} = 1.32$, $p = 0.26$), and no main effect of day ($F_{4,1035} = 0.37$, $p = 0.83$). Mixed-effects ANOVA focusing on the interaction of group by work period by time of day, collapsed over days within each 5-day work period, yielded a significant interaction of group by time of day ($F_{3,1039} = 4.63$, $p = 0.003$) and a trend for a main effect of time of day ($F_{3,1039} = 2.43$, $p = 0.064$), although there was no significant three-way interaction ($F_{3,1039} = 0.78$, $p = 0.50$) and no significant interaction of work period by time of day ($F_{3,1039} = 0.48$, $p = 0.70$). As a function of time of day, the subjective effort data paralleled the subjective mood data (see Figure 12) and the positive affect data.

For the number of error responses on the CDDT, the primary analysis focusing on the interaction of group by session (collapsed over days and over times of day within sessions) yielded a statistically significant effect of session ($F_{1,1051} = 17.78$, $p < 0.001$) and a trend for the interaction effect ($F_{1,1051} = 3.35$, $p = 0.067$), but no significant effect of group ($F_{1,1051} = 1.48$, $p = 0.22$). Figure 22 displays these findings, showing that performance on the CDDT improved across the study in both conditions, but tended to do so more in the "best-case" condition.
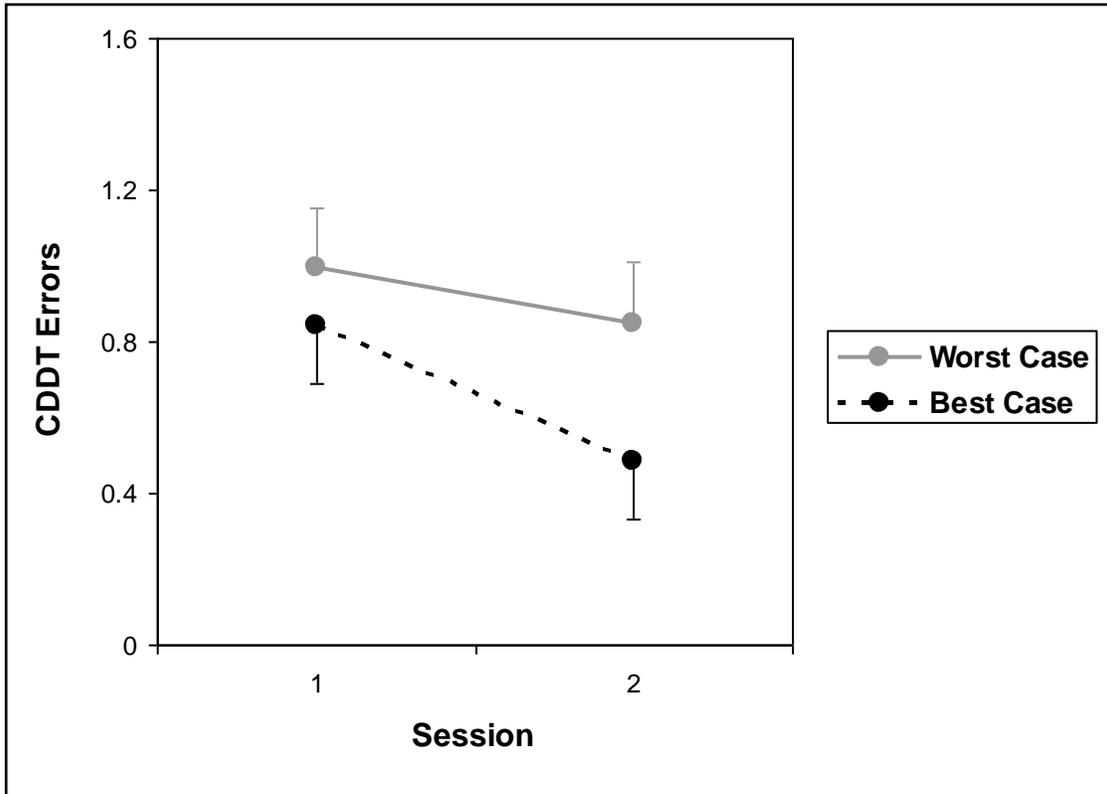
Figure 22. Graph. Number of error responses on the CDDT in the pre-restart 5-day work period (session 1) as compared to the post-restart 5-day work period (session 2) for the "worst-case" and "best-case" conditions.

In order to investigate further CDDT performance across days in the study, a secondary analysis examined the interaction of group by work period by day, collapsed over time of day. The three-way interaction was not statistically significant ($F_{4,1035} = 1.09$, $p = 0.36$), nor was the two-way interaction of group by day ($F_{4,1035} = 1.24$, $p = 0.29$). However, the main effect of day was significant ($F_{4,1035} = 4.00$, $p = 0.003$), as was the two-way interaction of work period (session) by day ($F_{4,1035} = 2.63$, $p = 0.033$). Figure 23 displays the data by day, showing that in both conditions, performance on the CDDT improved steadily across days in the 5-day work period preceding the 34-hour restart, and then leveled off in the 5-day work period following the 34-hour restart.
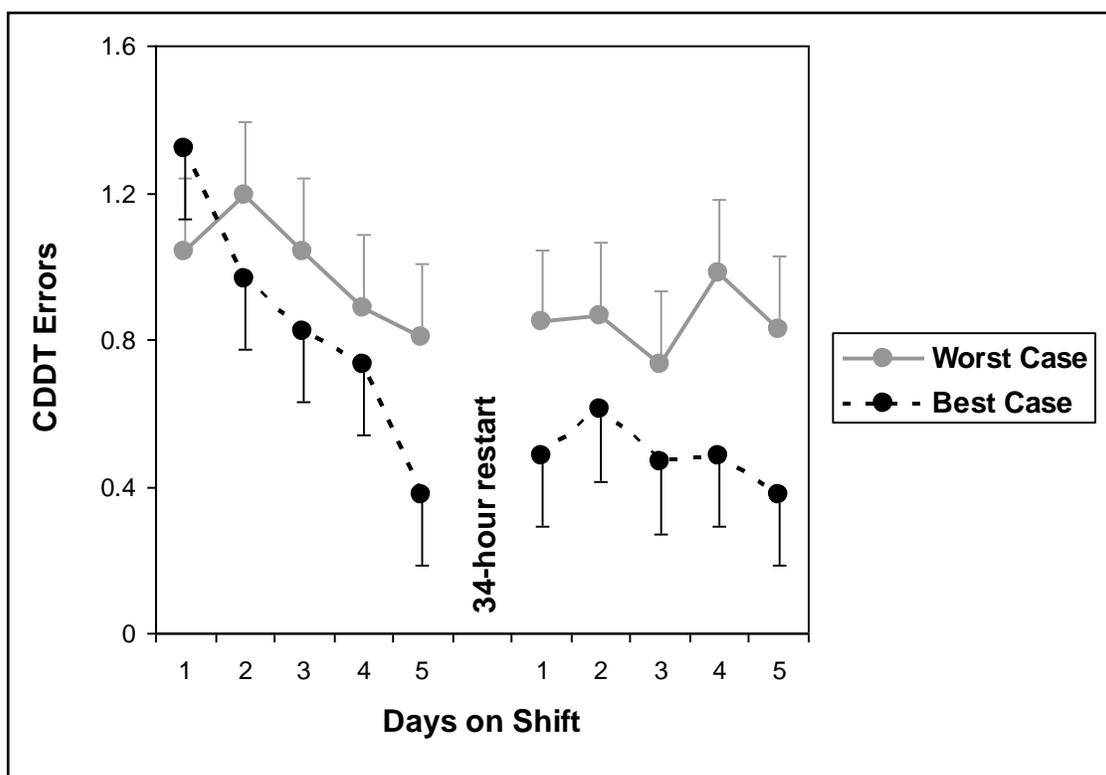
Figure 23. Graph. Number of error responses on the CDDT as a function of days in the 5-day work periods before and after the 34-hour restart period, for the "worst-case" and "best-case" conditions.

Analysis of the interaction of group by work period by time of day, collapsed over days within each 5-day work period, yielded only a trend for a two-way interaction of work period by time of day ($F_{3,1039} = 2.30$, $p = 0.076$). There was no significant three-way interaction ($F_{3,1039} = 0.65$, $p = 0.59$), no significant two-way interaction of group by time of day ($F_{3,1039} = 0.47$, $p = 0.70$), and no significant main effect of time of day ($F_{3,1039} = 1.59$, $p = 0.19$). In other words, CDDT performance was relatively stable within days over time of day.

3.2.3 Simulator Driving Performance

Analyses of driving simulator performance were based on data from 12 subjects randomized to the "best-case" condition and 13 subjects randomized to the "worst-case" condition—driving performance data from two subjects were unavailable as a result of data-capturing problems with the simulators. Driving simulator outcome variables were subjected to the same primary and secondary analyses as were the neurobehavioral performance outcomes described above, but subjects' assignment to simulator #1 or #2 was added as a covariate to account for any possible simulator hardware differences. Further analyses were run with pre-driving PVT performance (lapses) as an added covariate, to examine the potential of the PVT for predicting simulator driving performance.

For average driving speed in the straightaways, the primary analysis focusing on the interaction of group by session (collapsed over days and over times of day within sessions) yielded a significant effect of session ($F_{1,970} = 47.97$, $p < 0.001$) and a trend for an interaction effect

$(F_{1,970} = 3.51, p = 0.061)$. There was no significant effect of group $(F_{1,970} = 0.63, p = 0.43)$. Figure 24 displays these findings, showing that subjects in both groups stayed close to the speed limit of 55 mi/h, with a very minor increase in average speed from before to after the 34-hour restart. Average simulator driving speed was not significantly predicted by pre-driving PVT lapses $(F_{1,968} = 0.07, p = 0.80)$.

Figure 24. Graph. Average simulator driving speed in the pre-restart 5-day work period (session 1) as compared to the post-restart 5-day work period (session 2) for the "worst-case" and "best-case" conditions.

Investigation of average driving speed across days in the study revealed a significant two-way interaction of session by day $(F_{4,954} = 2.65, p = 0.032)$ and a main effect of day $(F_{4,954} = 8.90, p < 0.001)$, but there was no statistical significance for the three-way interaction of group by session by day $(F_{4,954} = 0.77, p = 0.54)$ and the two-way interaction of group by day $(F_{4,954} = 0.66, p = 0.62)$. Figure 25 displays the data by day, showing that in both conditions, average speed increased marginally across the pre-restart session and leveled off in the post-restart session.
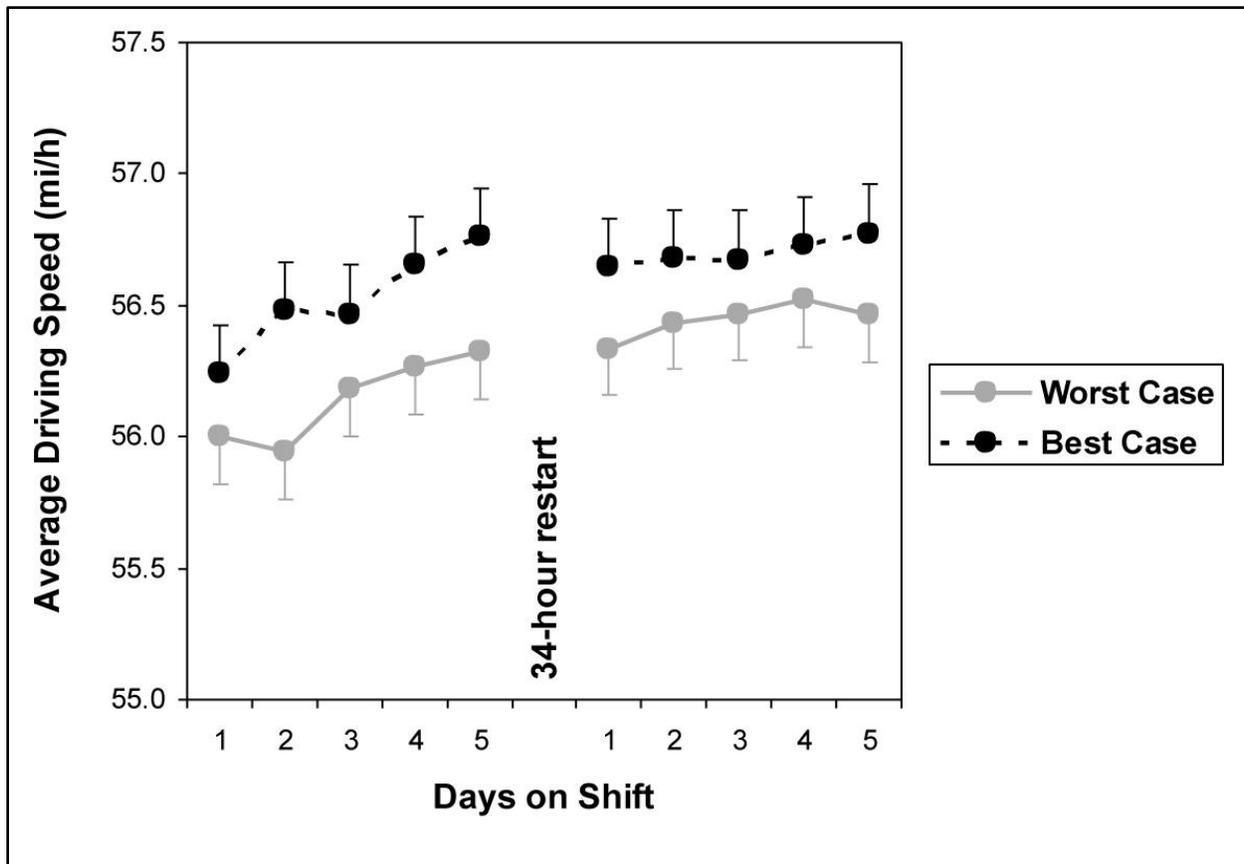
Figure 25. Graph. Average simulator driving speed as a function of days in the 5-day work periods before and after the 34-hour restart period, for the "worst-case" and "best-case" conditions.

Analysis of the interaction of group by work period by time of day, collapsed over days within each 5-day work period, revealed that there were no statistically significant effects of time of day for average driving speed. There was no significant three-way interaction ($F_{3,958} = 0.51$, $p = 0.68$), no significant two-way interaction of group by time of day ($F_{3,958} = 0.18$, $p = 0.91$), no two-way interaction of work period by time of day ($F_{3,958} = 1.53$, $p = 0.21$), and no significant main effect of time of day ($F_{3,958} = 0.29$, $p = 0.83$).

For variability (standard deviation) of driving speed across the straightaways, the primary analysis of the interaction of group by session (collapsed over days and over times of day within sessions) yielded a significant effect of session ($F_{1,970} = 43.84$, $p < 0.001$), but no significant interaction effect ($F_{1,970} = 0.16$, $p = 0.69$) and no significant effect of group ($F_{1,970} = 0.29$, $p = 0.59$). Variability of simulator driving speed was significantly predicted by pre-driving PVT lapses ($F_{1,968} = 19.85$, $p < 0.001$)—on average, an increase of 1 in PVT lapses corresponded to an increase of 0.022 mi/h (approximately 3.5 percent) in the standard deviation of driving speed.

Investigation of variability in driving speed across days in the study revealed a significant main effect of day ($F_{4,954} = 4.05$, $p = 0.003$). There was no significant three-way interaction of group by session by day ($F_{4,954} = 0.80$, $p = 0.52$), no significant two-way interaction of session by day ($F_{4,954} = 1.42$, $p = 0.22$), and no significant two-way interaction of group by day ($F_{4,954} = 1.55$,

$p = 0.19$). Figure 26 shows that in both conditions, speed variability decreases marginally but steadily across the days of the study, suggesting a minor practice effect in driving the simulator.
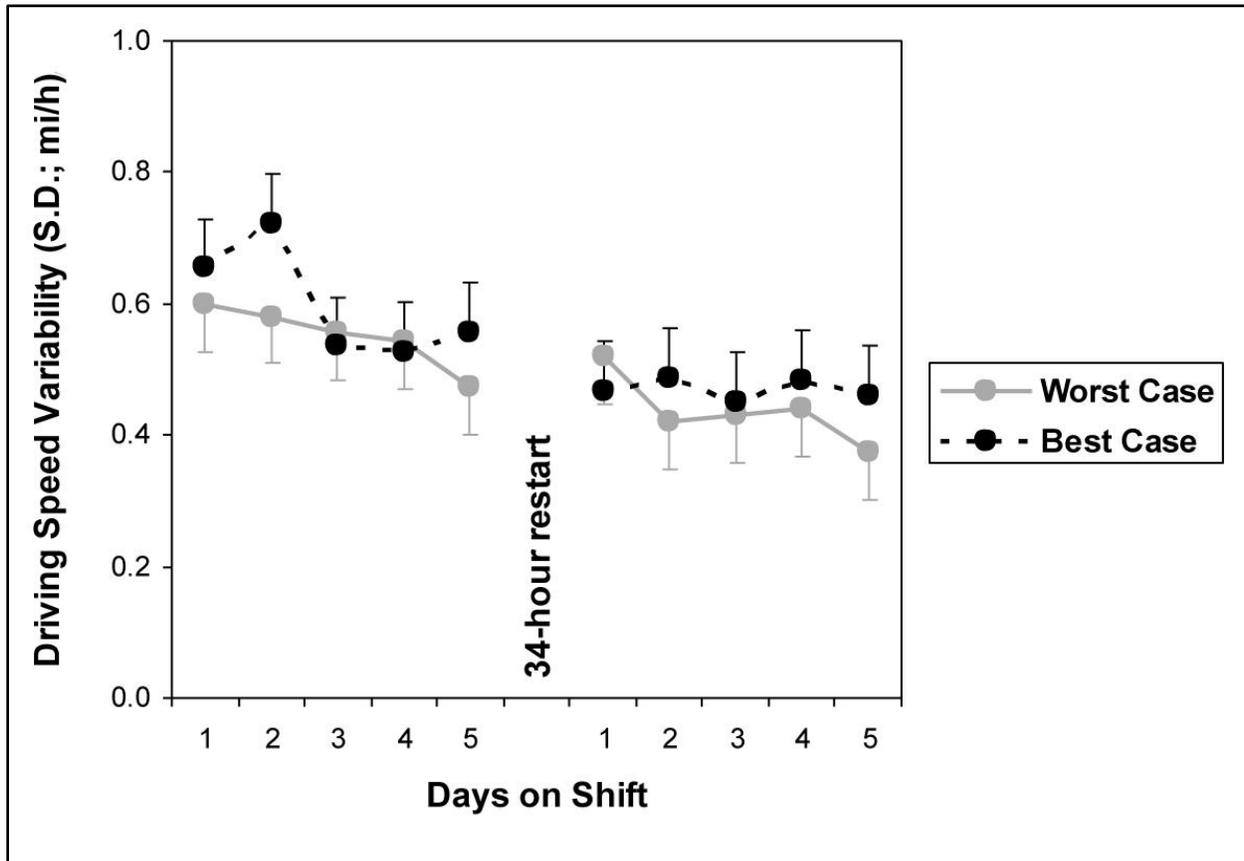
Figure 26. Graph. Variability (standard deviation) of simulator driving speed as a function of days in the 5-day work periods before and after the 34-hour restart period, for the "worst-case" and "best-case" conditions (upward direction corresponds to driving less steadily).

Analysis of the interaction of group by work period by time of day, collapsed over days within each 5-day work period, demonstrated a significant two-way interaction of group by time of day ($F_{3,958} = 4.95$, $p = 0.002$). There were no other significant effects related to time of day: no significant three-way interaction ($F_{3,958} = 1.21$, $p = 0.31$), no significant two-way interaction of work period by time of day ($F_{3,958} = 0.15$, $p = 0.93$), and no significant main effect of time of day ($F_{3,958} = 1.03$, $p = 0.38$). Figure 27 shows that variability in driving speed decreased over the hours of the day in the "best-case" condition, whereas it increased over the hours of the night in the "worst-case" condition.
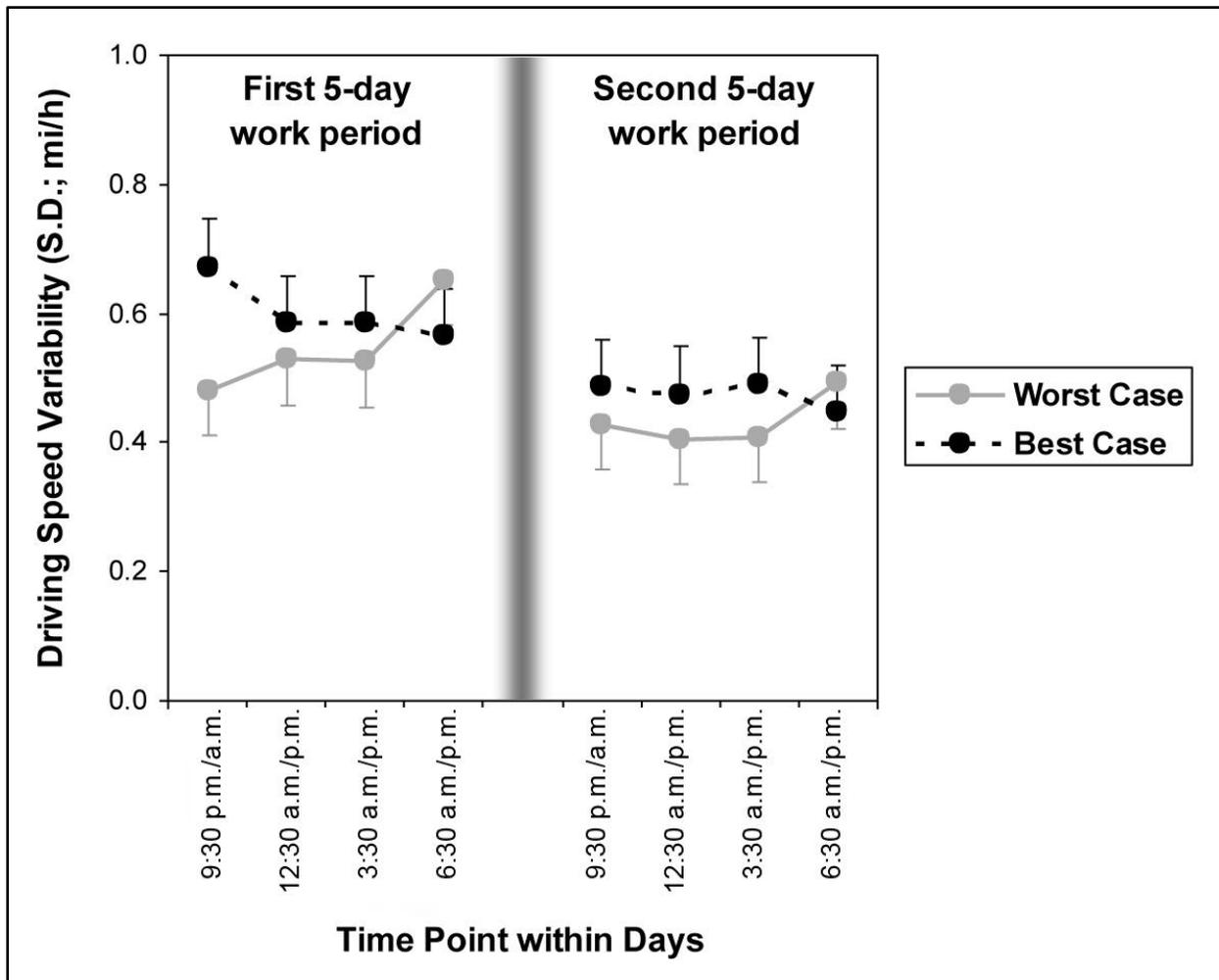
Figure 27. Graph. Variability (standard deviation) of simulator driving speed as a function of time of day, collapsed over the 5-day work periods before and after the 34-hour restart period (upwards corresponds to driving less steadily). Times of day are through the night (9:30 p.m.–6:30 a.m.) for the "worst-case" condition and through the day (9:30 a.m.–6:30 p.m.) for the "best-case" condition; 1 hour should be added to each time point for subjects who were assigned to performance testing first and driving second.

For lane deviation (standard deviation of lane position) in the straightaways, the primary analysis of the interaction of group by session (collapsed over days and times of day within sessions) yielded a significant interaction effect ($F_{1,970} = 9.15$, $p = 0.003$) and a significant main effect of session ($F_{1,970} = 75.23$, $p < 0.001$), but no significant main effect of group ($F_{1,970} = 0.88$, $p = 0.35$). Figure 28 displays these findings; subjects in the "best-case" condition showed improvement in holding a steady lane position from the pre-restart session to the post-restart session (from a standard deviation of 13.6 cm to a standard deviation of 12.0 cm), and subjects in the "worst-case" condition showed only half as much improvement (from a standard deviation of 14.5 cm to a standard deviation of 13.8 cm). Simulator lane deviation was significantly predicted by pre-driving PVT lapses ($F_{1,968} = 69.19$, $p < 0.001$)—on average, an increase of 1 in PVT lapses corresponded to an increase of 0.003 m (approximately 2.0 percent) in the standard deviation of lane position.

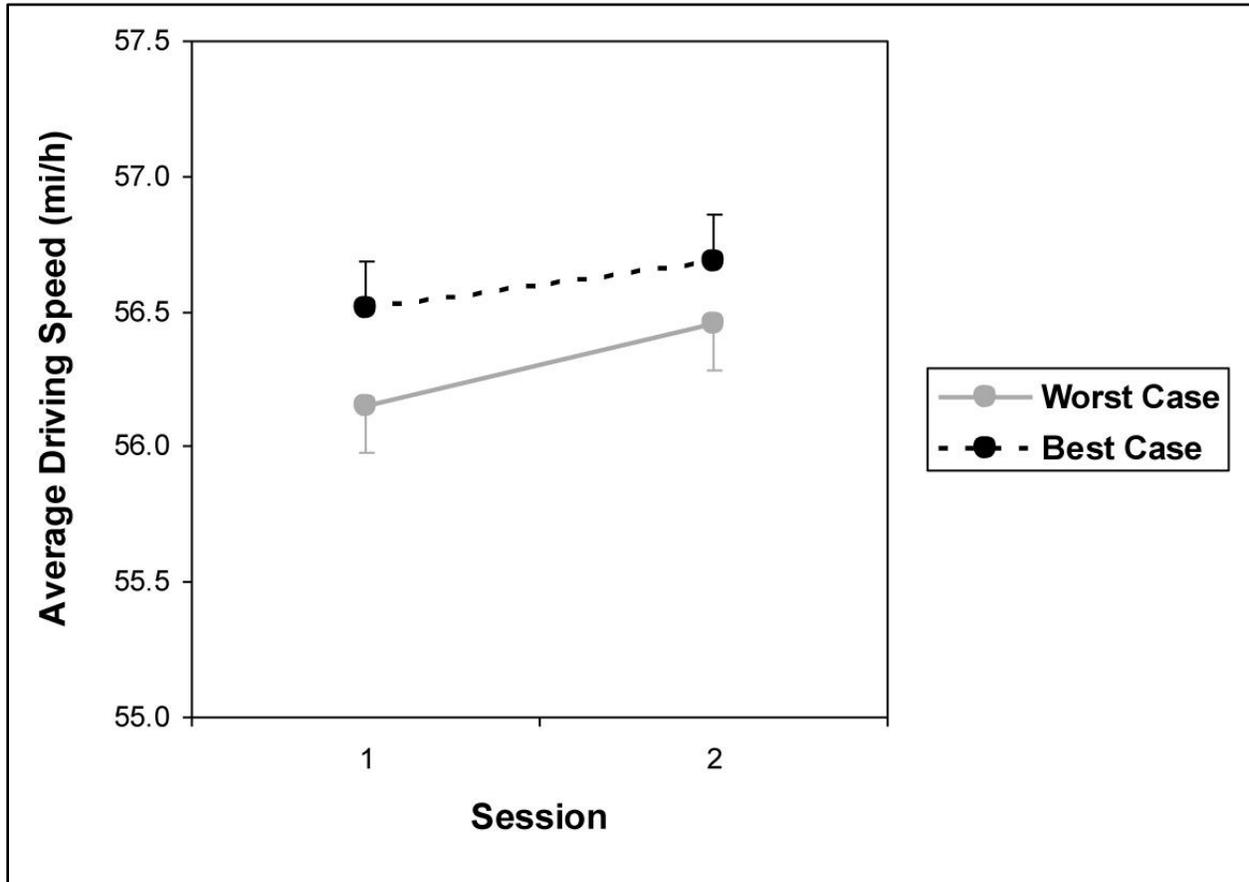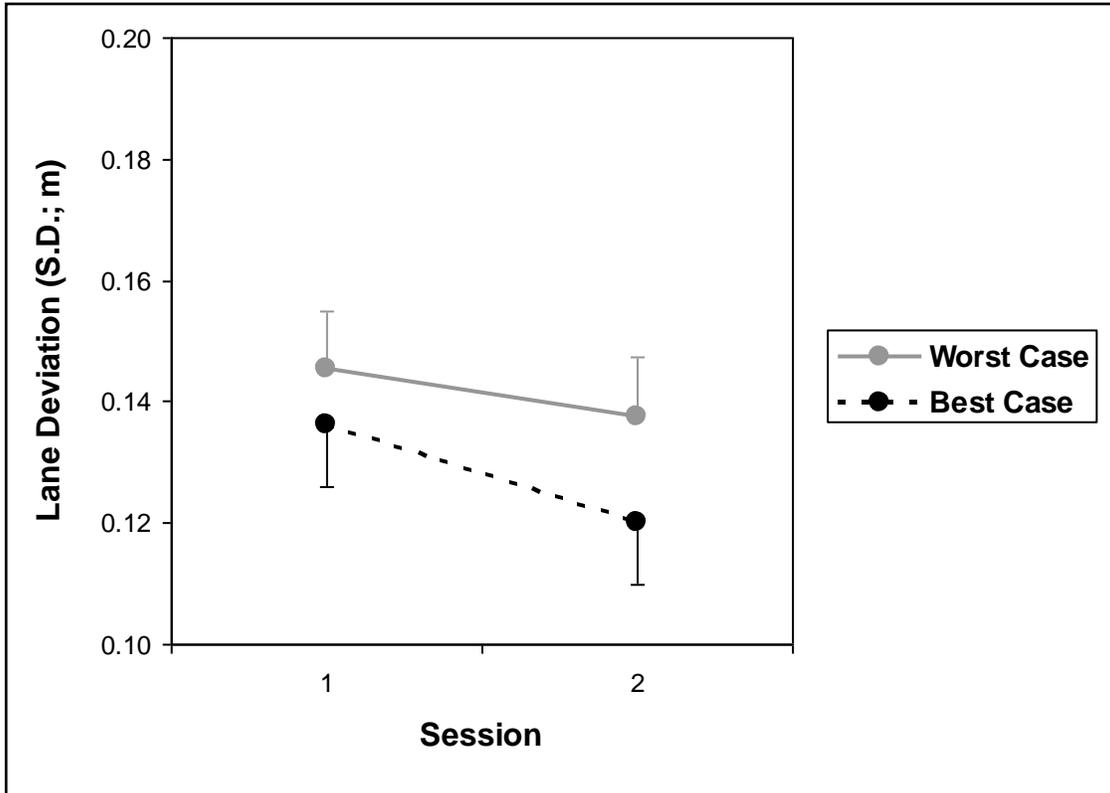**Figure 28. Graph. Lane deviation (standard deviation of lane position) on the driving simulator in the pre-restart 5-day work period (session 1) as compared to the post-restart 5-day work period (session 2) for the "worst-case" and "best-case" conditions.**

Investigation of lane deviation across days in the study yielded a significant two-way interaction of session by day ($F_{4,954} = 2.67$, $p = 0.031$) and a significant main effect of day ($F_{4,954} = 10.47$, $p < 0.001$), but no significant three-way interaction of group by session by day ($F_{4,954} = 0.34$, $p = 0.85$) and no significant two-way interaction of group by day ($F_{4,954} = 0.40$, $p = 0.81$). Examination of lane deviation by time of day yielded a significant two-way interaction of group by time of day ($F_{3,958} = 10.25$, $p < 0.001$) and a significant main effect of time of day ($F_{3,958} = 5.98$, $p < 0.001$), although there was no significant three-way interaction ($F_{3,958} = 0.63$, $p = 0.59$) and no significant two-way interaction of session by time of day ($F_{3,958} = 1.20$, $p = 0.31$). Figure 29 shows that lane deviation gradually increased across time of night in the "worst-case" condition.
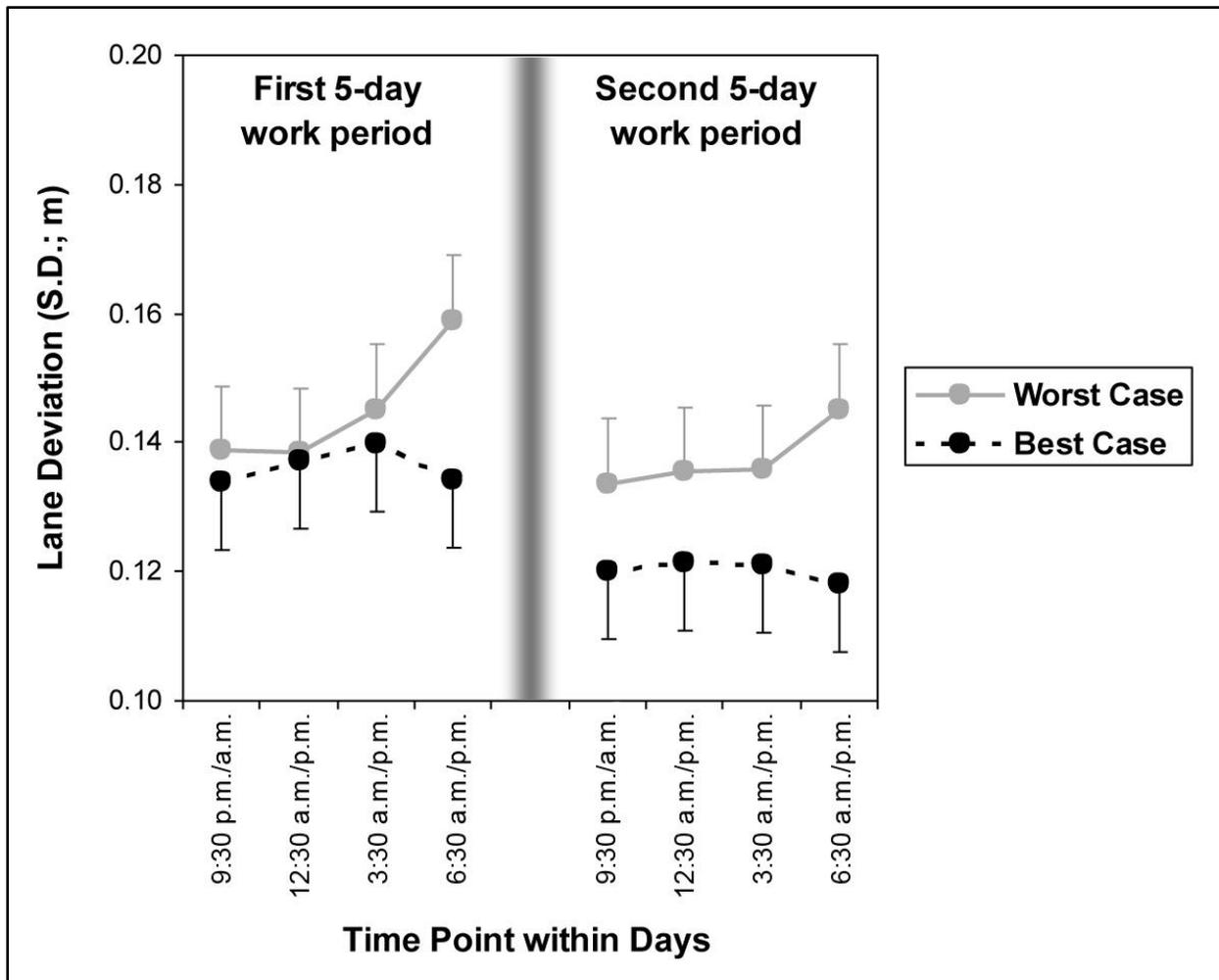
Figure 29. Graph. Lane deviation (standard deviation of lane position) on the driving simulator as a function of time of day, collapsed over the 5-day work periods before and after the 34-hour restart period. Times of day are through the night (9:30 p.m.–6:30 a.m.) for the "worst-case" condition and through the day (9:30 a.m.–6:30 p.m.) for the "best-case" condition; 1 hour should be added to each time point for subjects who were assigned to performance testing first and driving second.

For the reaction time of emergency braking for the pedestrian/dog crossing events during simulator driving, the primary analysis of the interaction of group by session (collapsed over days and times of day within sessions) yielded no significant effects: no interaction ($F_{1,970} < 0.01$, $p = 0.98$), no effect of session ($F_{1,970} = 1.43$, $p = 0.23$), and no effect of group ($F_{1,970} = 1.06$, $p = 0.30$). Secondary analyses investigating the interactions with day and with time of day did not yield any statistically significant effects, either. However, simulator braking reaction time was significantly predicted by pre-driving PVT lapses ($F_{1,968} = 15.24$, $p < 0.001$)—on average, an increase of 1 in PVT lapses corresponded to an increase of 0.011 s (approximately 0.9 percent) in the reaction time for braking.

Results were not much different for braking errors (i.e., braking unnecessarily in the straightaways, or failing to brake or braking more than once around the pedestrian/dog crossing events). The primary analysis of the interaction of group by session yielded a significant (albeit

negligible) main effect of session ($F_{1,970} = 6.01$, $p = 0.014$), but no significant interaction of group by session ($F_{1,970} = 0.56$, $p = 0.45$) and no significant main effect of group ($F_{1,970} = 2.43$, $p = 0.12$). There was a trend for simulator braking errors to be predicted by pre-driving PVT lapses ($F_{1,968} = 2.75$, $p = 0.098$). Examination of braking errors across days in the study revealed no further significant effects: no significant interaction of group by session by day ($F_{4,954} = 1.19$, $p = 0.32$), no interaction of session by day ($F_{4,954} = 1.36$, $p = 0.025$), no significant interaction of group by day ($F_{4,954} = 0.84$, $p = 0.50$), and no significant main effect of day ($F_{4,954} = 0.57$, $p = 0.69$). Investigation of braking errors by time of day yielded a trend for a three-way interaction of group by session by time of day ($F_{3,958} = 2.42$, $p = 0.065$). There was no significant interaction of group by time of day ($F_{3,958} = 0.13$, $p = 0.94$), no significant interaction of session by time of day ($F_{3,958} = 0.97$, $p = 0.41$), and no main effect of time of day ($F_{3,958} = 0.11$, $p = 0.95$).

For computed (simulated) fuel use, the primary analysis of the interaction of group by session (collapsed over days and times of day within sessions) yielded a significant main effect of session ($F_{1,970} = 35.97$, $p < 0.001$), but no significant interaction effect ($F_{1,970} = 0.09$, $p = 0.76$) and no significant main effect of group ($F_{1,970} = 0.76$, $p = 0.38$). On average, subjects in the "best-case" condition used 0.248 gallons of fuel to traverse the 10 straightaways before the 34-hour restart, and 0.247 gallons after the 34-hour restart. Similarly, subjects in the "worst-case" condition used 0.249 gallons of fuel before the 34-hour restart, and 0.248 gallons after the 34-hour restart. Simulator fuel use was significantly predicted by pre-driving PVT lapses ($F_{1,968} = 7.16$, $p = 0.008$), but on average, an increase of 1 in PVT lapses corresponded to an increase of only 0.0001 gallon of fuel use (approximately 0.04 percent).

Investigation of fuel use across days in the study revealed a significant interaction of session by day ($F_{4,954} = 4.51$, $p = 0.001$) and a significant main effect of day ($F_{4,954} = 9.94$, $p < 0.001$), but no significant interaction of group by session by day ($F_{4,954} = 0.49$, $p = 0.75$) and no significant interaction of group by day ($F_{4,954} = 0.65$, $p = 0.63$). Investigation of fuel use by time of day, on the other hand, yielded a significant interaction of group by time of day ($F_{3,958} = 3.89$, $p = 0.009$). There was no significant three-way interaction ($F_{3,958} = 1.05$, $p = 0.37$), no significant interaction of session by time of day ($F_{3,958} = 0.99$, $p = 0.39$), and no main effect of time of day ($F_{3,958} = 0.73$, $p = 0.54$). Figure 30 shows that fuel use increased slightly over time of night in the "worst-case condition.
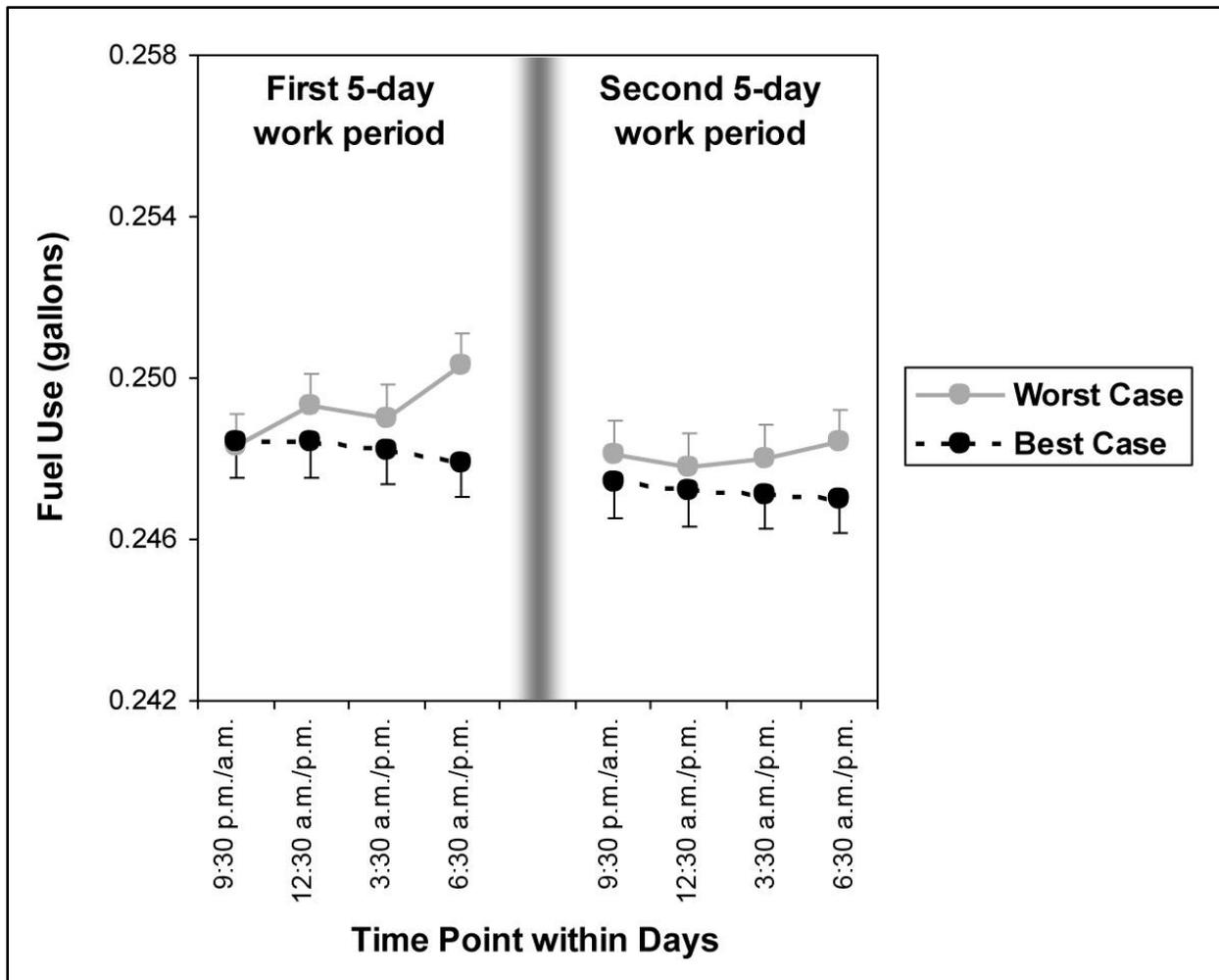
Figure 30. Graph. Computed fuel use on the driving simulator as a function of time of day, collapsed over the 5-day work periods before and after the 34-hour restart period. Times of day are through the night (9:30 p.m.–6:30 a.m.) for the "worst-case" condition and through the day (9:30 a.m.–6:30 p.m.) for the "best-case" condition; 1 hour should be added to each time point for subjects who were assigned to performance testing first and driving second.

3.3   SLEEP AND POLYSOMNOGRAPHY

Figure 31 illustrates the primary comparisons being made between the "best-case" and "worst-case" conditions for polysomnographic measures of sleep (see figure 4). Besides comparison of the first 10-hour night that the two conditions had in common (performed solely for control purposes), these focused on comparison of the combined first and second 10-hour nights in the "best-case" condition with the combined first 10-hour night and first 5-hour transition nap in the "worst-case" condition ("Baseline"); comparison of the combined two polysomnographically recorded 10-hour sleep periods in the first 5-day work period between the two conditions ("Session 1"); comparison of the combined two 10-hour nights in the 34-hour restart period of the "best-case" condition with the combined two 5-hour transition naps and intervening 10-hour nocturnal sleep period in the 34-hour restart period of the "worst-case" condition ("Restart"); comparison of the combined two polysomnographically recorded 10-hour sleep periods in the

second 5-day work period between the two conditions ("Session 2"); and comparison of the 10-hour recovery night at the end of the "best-case" condition with the combined 5-hour transition nap and 10-hour recovery night at the end of the "worst-case" condition ("Recovery").

Figure 31. Chart. Comparison scheme for the polysomnographically recorded sleep periods in the "worst-case" (top schematic) and "best-case" (bottom schematic) conditions of the study, with hours of the day progressing from left to right (see top numbers), days of the study progressing from top to bottom (see right side), gray indicating scheduled wakefulness, black indicating scheduled sleep periods, and ××× marking sleep periods that were not recorded polysomnographically. Days are shown from 10 p.m. to 10 p.m. (rather than from midnight to midnight) for easier visual comparison of the scheduled sleep periods between the two conditions. Text labels indicate which sleep periods were combined to form the Baseline, Session 1, Restart, Session 2, and Recovery periods that were being compared between the two conditions.

Total time in bed was 20 hours for each comparison set except during the Baseline period, where the "worst-case" condition was shorted by 5 hours as necessary for the transition to a nocturnal work schedule—and during the Recovery period, which was only one night of 10 hours of time

in bed in the "best-case" condition, but involved a total of 15 hours of time in bed in the "worst-case" condition to make up for the difference during the Baseline period (see Figure 32).

Figure 32. Graph. Combined time in bed during the Baseline, Session 1, Restart, Session 2, and Recovery sets of sleep periods, comparing them for "worst-case" and "best-case" conditions.

During the first (10-hour) baseline night, which the study conditions had in common (see Figure 31), there were no statistically significant differences between the two conditions in total sleep time ($F_{1,25} = 2.72$, $p = 0.11$), non-rapid eye movement (non-REM) stage 1 (N1; $F_{1,25} = 1.55$, $p = 0.22$), non-REM stage 2 (N2; $F_{1,25} = 1.72$, $p = 0.20$), slow-wave sleep (SWS or N3; $F_{1,25} = 0.12$, $p = 0.73$), REM ($F_{1,25} = 0.13$, $p = 0.72$), sleep latency (SL; $F_{1,25} = 0.06$, $p = 0.81$), slow-wave sleep latency (SWSL; $F_{1,25} = 0.33$, $p = 0.57$), and REM latency (REML; $F_{1,25} = 1.33$, $p = 0.26$). Thus, the two subject groups were not significantly different in these indices of baseline sleep architecture.

Figure 33 shows the comparisons between the "best-case" and "worst-case" conditions across the study for total sleep time. As expected, the "worst-case" condition had significantly less combined total sleep time than the "best-case" condition during the Baseline period ($F_{1,25} = 119.57$, $p < 0.001$). The "worst-case" condition also had significantly less total sleep time during the Session 1 period ($F_{1,24} = 11.99$, $p = 0.002$). There was a trend for less combined total sleep time in the "worst-case" condition as compared to the "best-case" condition during the Restart period ($F_{1,24} = 3.61$, $p = 0.070$). The "worst-case" condition again had significantly less total sleep time during the Session 2 period ($F_{1,24} = 8.76$, $p = 0.007$). However, the "worst-case" condition had significantly more combined total sleep time than the "best-case" condition during the Recovery period ($F_{1,24} = 20.50$, $p < 0.001$), as was also expected.
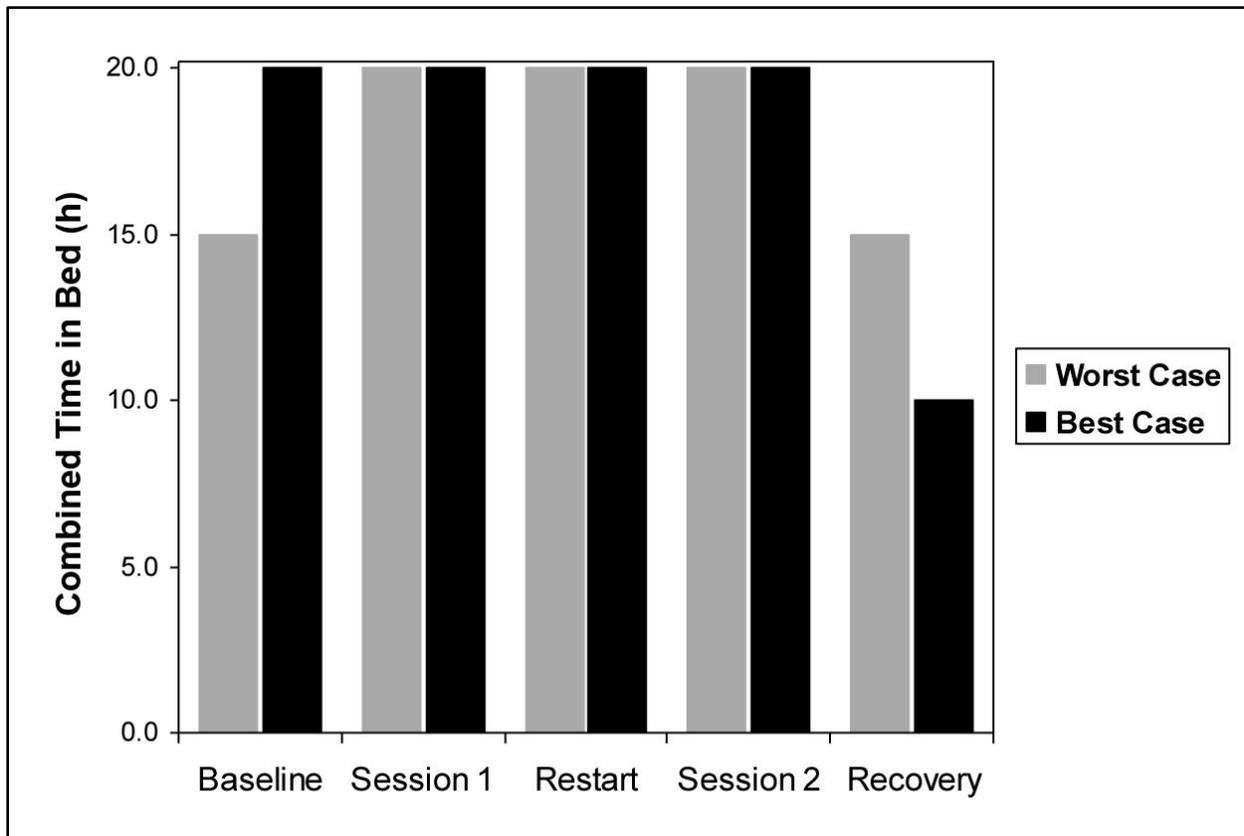
Figure 33. Graph. Combined total sleep time during the Baseline, Session 1, Restart, Session 2, and Recovery sets of sleep periods, comparing them for "worst-case" and "best-case" conditions.

In a two-way ANOVA of combined total sleep time focusing on the interaction of condition ("best-case" compared to "worst-case") by session (Session 1 compared to Session 2), there was a significant main effect of condition ($F_{1,48} = 20.52$, $p < 0.001$), but no significant interaction effect ($F_{1,48} = 0.05$, $p = 0.82$) and no significant main effect of session ($F_{1,48} = 0.78$, $p = 0.38$). Thus, the difference in total sleep time between the two conditions (with the "best-case" condition exhibiting more total sleep time than the "worst-case" condition) appeared to be stable from Session 1 to Session 2. In the aggregate over all the polysomnographically recorded sleep periods (90 h time in bed in total), the "worst-case" condition obtained 66.2 h total sleep time, and the "best-case" condition obtained 74.3 h total sleep time (a statistically significant difference of 8.1 h: $F_{1,21} = 16.47$, $p < 0.001$).

For slow-wave sleep (non-REM sleep stage N3), the comparisons between the "best-case" and "worst-case" conditions across the study are shown in Figure 34. The "worst-case" condition had significantly less combined N3 than did the "best-case" condition during the Baseline period ($F_{1,25} = 6.75$, $p = 0.016$). There was no statistically significant difference between the conditions during the Session 1 period ($F_{1,24} = 0.03$, $p = 0.86$), Restart period ($F_{1,24} = 0.49$, $p = 0.49$), or Session 2 period ($F_{1,24} = 0.01$, $p = 0.91$). However, there was a trend for the "worst-case" condition to have more combined N3 than the "best-case" condition during the Recovery period ($F_{1,24} = 2.93$, $p = 0.10$). In two-way ANOVA of combined N3 focusing on the interaction of condition by session, there was no significant interaction effect ($F_{1,48} = 0.05$, $p = 0.83$), no significant main effect of condition ($F_{1,48} < 0.01$, $p = 0.96$), and no significant main effect of

session ($F_{1,48} = 0.05$, $p = 0.83$). In the aggregate over all the polysomnographically recorded sleep periods, the "worst-case" condition accumulated 10.1 hours of stage N3 sleep, and the "best-case" condition obtained 10.5 hours of stage N3 sleep (a non-significant difference: $F_{1,21} = 0.03$, $p = 0.86$).

Figure 34. Graph. Combined slow-wave sleep (non-REM sleep stage N3) during the Baseline, Session 1, Restart, Session 2, and Recovery sets of sleep periods, comparing them for "worst-case" and "best-case" conditions.

For REM sleep, the comparisons between the "best-case" and "worst-case" conditions across the study are shown in Figure 35. The "worst-case" condition had significantly less combined REM than the "best-case" condition during the Baseline period ($F_{1,25} = 60.03$, $p < 0.001$) and during the Session 1 period ($F_{1,24} = 5.69$, $p = 0.025$). There was no statistically significant difference between the conditions during the Restart period ($F_{1,24} = 2.48$, $p = 0.13$) or during the Session 2 period ($F_{1,24} = 0.80$, $p = 0.38$). However, the "worst-case" condition had significantly more combined REM than the "best-case" condition during the Recovery period ($F_{1,24} = 7.19$, $p = 0.013$). In two-way ANOVA of combined REM focusing on the interaction of condition by session, there was no significant interaction effect ($F_{1,48} = 1.04$, $p = 0.31$), but there were significant main effects of condition ($F_{1,48} = 5.31$, $p = 0.026$) and session ($F_{1,48} = 8.92$, $p = 0.004$). In the aggregate over all the polysomnographically recorded sleep periods, the "worst-case" condition accumulated 17.1 h of REM sleep, and the "best-case" condition obtained 19.7 h of REM sleep (a significant difference of 2.6 h: $F_{1,21} = 8.31$, $p = 0.009$).
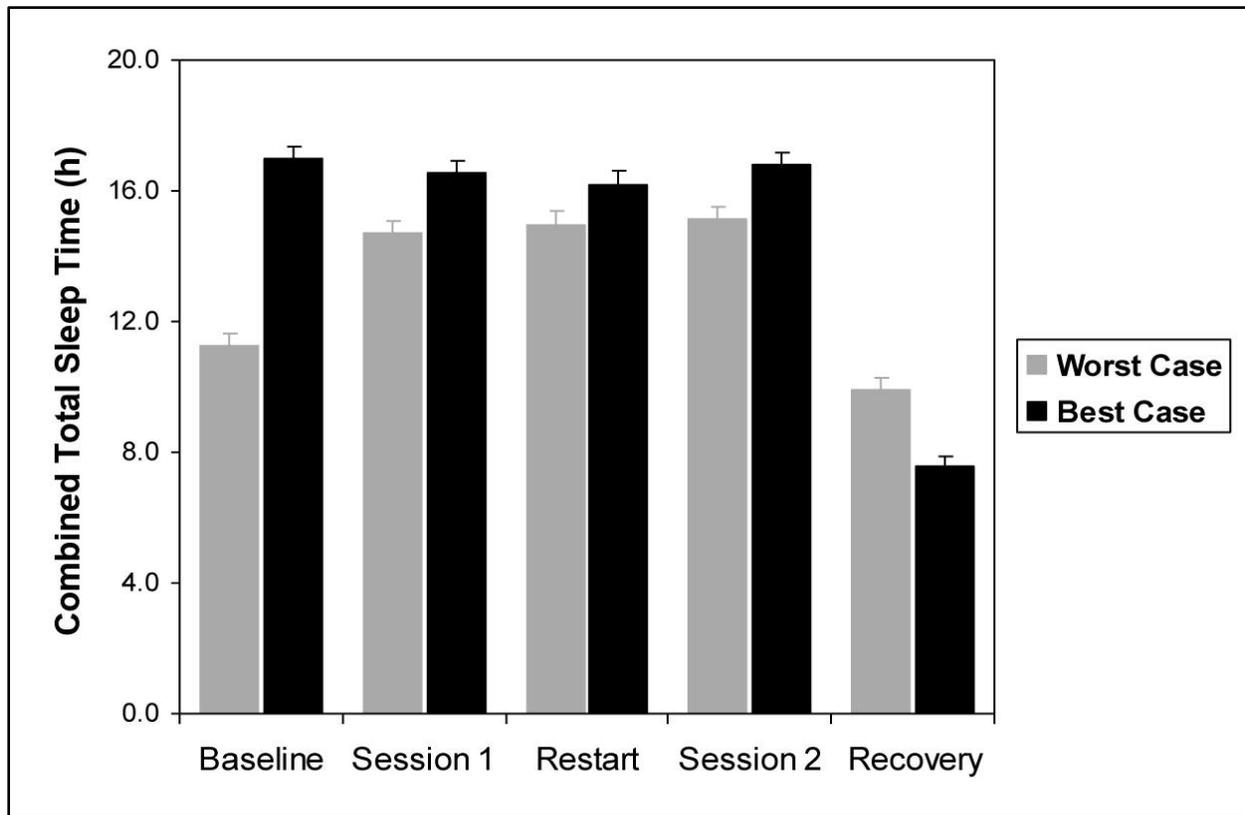
Figure 35. Graph. Combined REM sleep during the Baseline, Session 1, Restart, Session 2, and Recovery sets of sleep periods, comparing them for "worst-case" and "best-case" conditions.

For non-REM sleep stage N2, the comparisons between the "best-case" and "worst-case" conditions across the study are shown in Figure 36. The "worst-case" condition had significantly less combined N2 than the "best-case" condition during the Baseline period ($F_{1,25} = 46.57$, $p < 0.001$), during the Session 1 period ($F_{1,24} = 5.64$, $p = 0.026$), and during the Session 2 period ($F_{1,24} = 7.39$, $p = 0.012$). There was no statistically significant difference between the conditions during the Restart period ($F_{1,24} = 1.23$, $p = 0.28$). However, the "worst-case" condition had significantly more combined N2 sleep than the "best-case" condition during the Recovery period ($F_{1,24} = 5.34$, $p = 0.030$). In two-way ANOVA of combined N2 focusing on the interaction of condition by session, there was no significant interaction effect ($F_{1,48} < 0.01$, $p > 0.99$) and no significant main effect of session ($F_{1,48} < 0.01$, $p = 0.95$), but there was a significant main effect of condition ($F_{1,48} = 12.80$, $p < 0.001$). In the aggregate over all the polysomnographically recorded sleep periods, the "worst-case" condition accumulated 35.3 h of stage N2 sleep, and the "best-case" condition obtained 40.8 h of stage N2 sleep (a significant difference of 5.5 h: $F_{1,21} = 5.29$, $p = 0.032$).
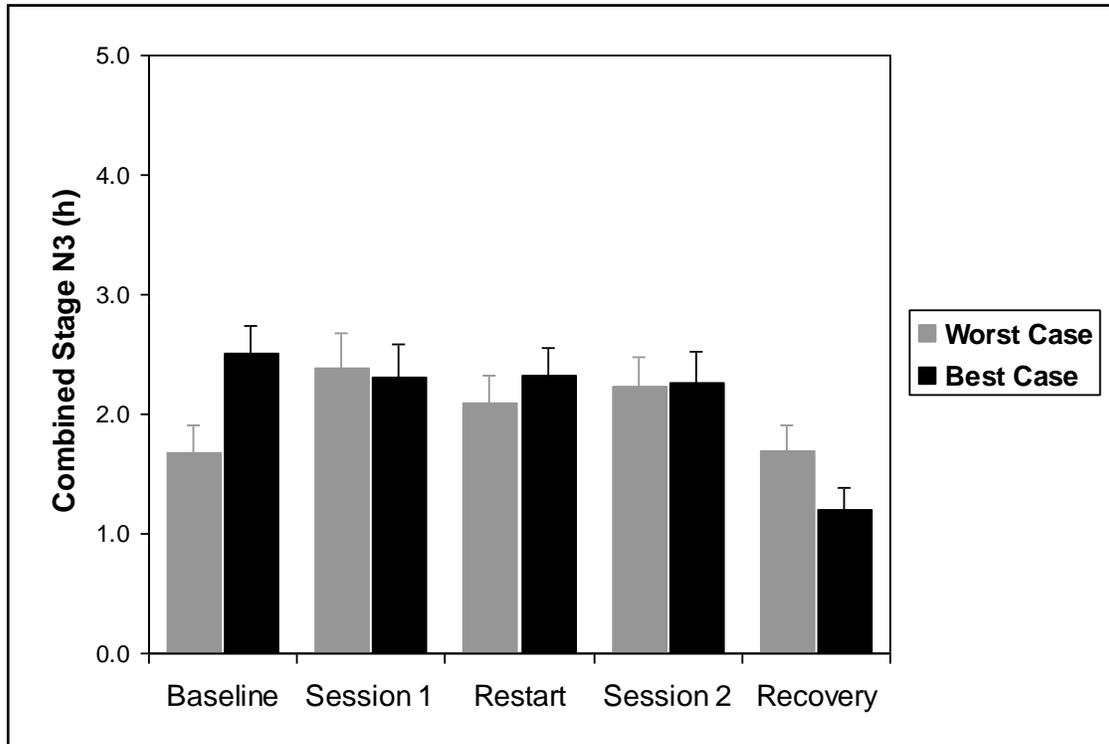
Figure 36. Graph. Combined non-REM sleep stage N2 during the Baseline, Session 1, Restart, Session 2, and Recovery sets of sleep periods, comparing them for "worst-case" and "best-case" conditions.

For non-REM sleep stage N1, the comparisons between the "best-case" and "worst-case" conditions across the study are shown in Figure 37. There were no statistically significant differences between the conditions during the Baseline period ($F_{1,25} = 0.06$, $p = 0.80$), Session 1 period ($F_{1,24} = 0.73$, $p = 0.40$), Restart period ($F_{1,24} = 0.13$, $p = 0.72$), and Session 2 period ($F_{1,24} = 0.35$, $p = 0.56$). However, the "worst-case" condition had significantly more combined N1 sleep than the "best-case" condition during the Recovery period ($F_{1,24} = 4.79$, $p = 0.039$). In two-way ANOVA of combined N1 focusing on the interaction of condition by session, there were no significant effects: no interaction ($F_{1,48} = 0.07$, $p = 0.79$), no main effect of session ($F_{1,48} = 0.11$, $p = 0.74$), and no main effect of condition ($F_{1,48} = 1.07$, $p = 0.31$). In the aggregate over all the polysomnographically recorded sleep periods, the "worst-case" condition accumulated 3.6 h of stage N1 sleep, and the "best-case" condition obtained 3.3 h of stage N1 sleep (a non-significant difference: $F_{1,21} = 0.60$, $p = 0.45$).
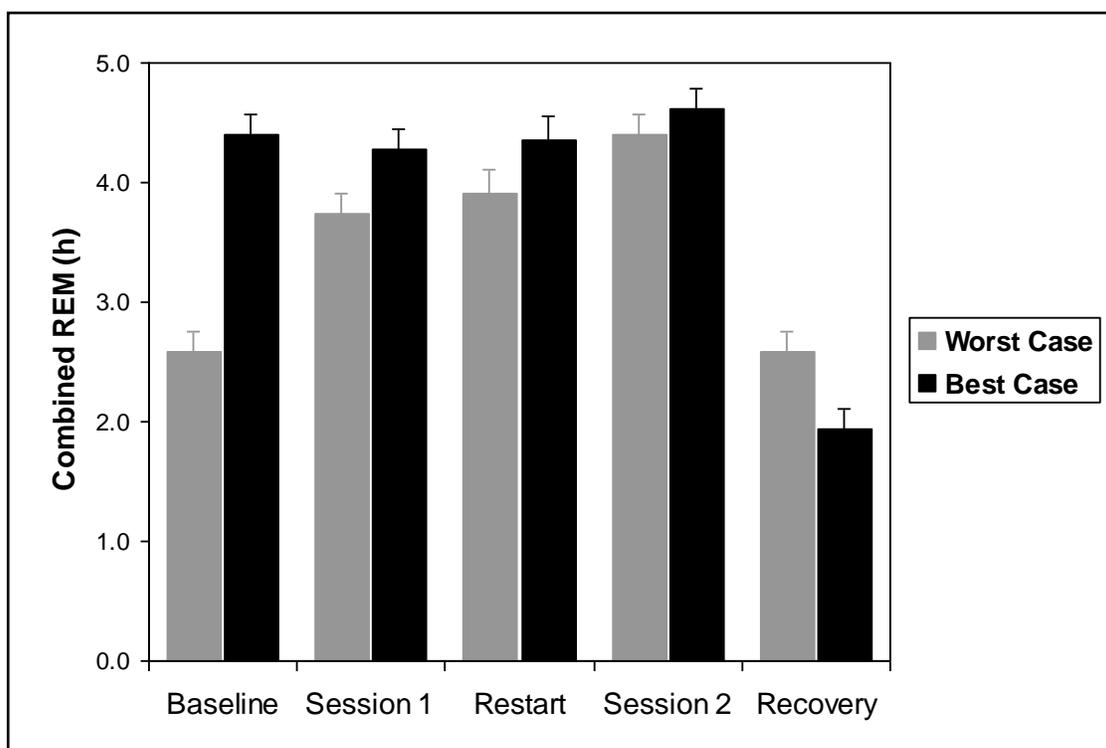
Figure 37. Graph. Combined non-REM sleep stage N1 during the Baseline, Session 1, Restart, Session 2, and Recovery sets of sleep periods, comparing them for "worst-case" and "best-case" conditions.

For sleep latency (SL), slow-wave sleep latency (SWSL), and REM latency (REML), analyses focused on averages within the Baseline, Session 1, Restart, Session 2, and Recovery sets of sleep periods. Averages are more relevant for latency variables than combined totals as used above for the sleep stages. For average SL, the comparisons between the "best-case" and "worst-case" conditions are shown in Figure 38. There was no statistically significant difference between the conditions during the Baseline period ($F_{1,25} = 0.04$, $p = 0.85$). However, there were significant differences during the Session 1 period ($F_{1,24} = 23.61$, $p < 0.001$), Session 2 period ($F_{1,24} = 42.74$, $p < 0.001$) and Recovery period ($F_{1,24} = 4.60$, $p = 0.042$), where average SL was consistently shorter in the "worst-case" condition than in the "best-case" condition. However, there was no significant difference between conditions in the Restart period ($F_{1,24} = 0.20$, $p = 0.66$). In two-way ANOVA of average SL, there was no significant interaction effect of condition by session ($F_{1,48} = 0.39$, $p = 0.53$) and no main effect of session ($F_{1,48} = 0.13$, $p = 0.72$), but there was a significant main effect of condition ($F_{1,48} = 57.77$, $p < 0.001$).

Figure 38. Graph. Average SL during the Baseline, Session 1, Restart, Session 2, and Recovery sets of sleep periods, comparing them for "worst-case" and "best-case" conditions.

For average SWSL, the comparisons between the "best-case" and "worst-case" conditions are shown in Figure 39. There were no statistically significant differences between the conditions during the Baseline period ($F_{1,25} = 2.07$, $p = 0.16$), Session 1 period ($F_{1,24} = 1.20$, $p = 0.28$), Session 2 period ($F_{1,24} = 1.60$, $p = 0.22$), and Recovery period ($F_{1,24} = 0.05$, $p = 0.82$). However, average SWSL was significantly longer in the "worst-case" condition than in the "best-case" condition during the Restart period ($F_{1,23} = 10.06$, $p = 0.004$). In two-way ANOVA of average SWSL, there was no significant interaction effect ($F_{1,48} = 0.50$, $p = 0.48$), no main effect of session ($F_{1,48} = 0.50$, $p = 0.48$), and no main effect of condition ($F_{1,48} = 2.56$, $p = 0.12$).
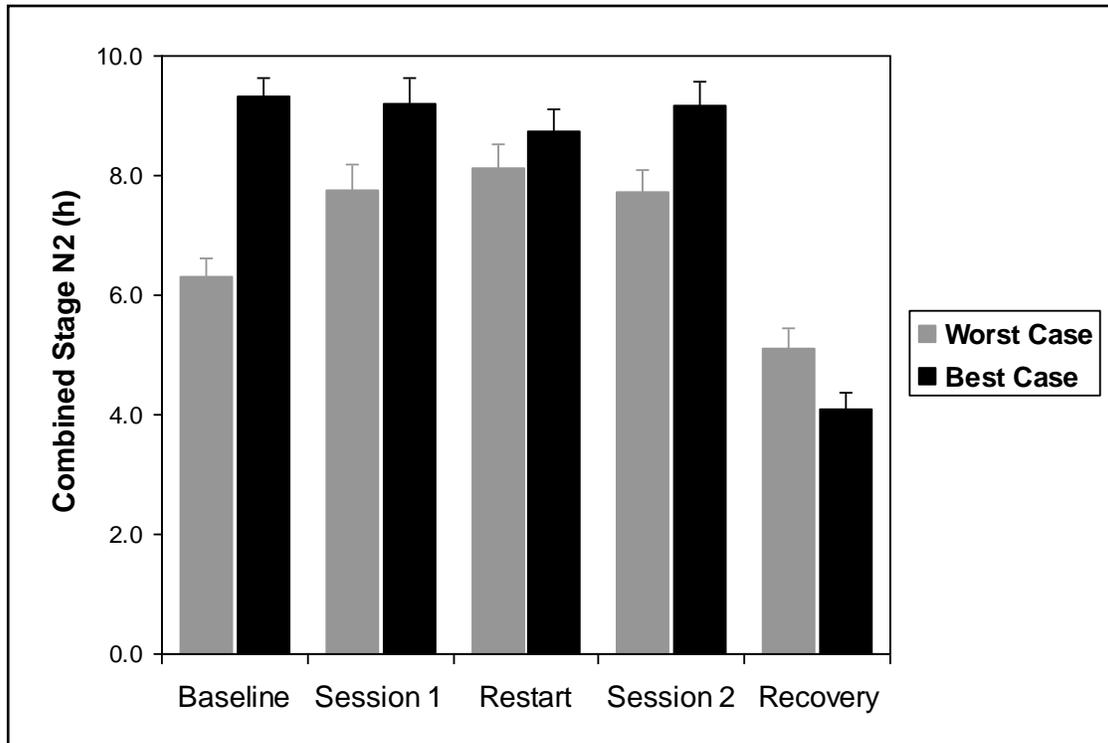
Figure 39. Graph. Average SWSL during the Baseline, Session 1, Restart, Session 2, and Recovery sets of sleep periods, comparing them for "worst-case" and "best-case" conditions.

For average REML, the comparisons between the "best-case" and "worst-case" conditions are shown in Figure 40. There was a trend for a difference between the conditions during the Baseline period ($F_{1,25} = 3.58$, $p = 0.070$), a significant difference during the Session 1 period ($F_{1,24} = 9.64$, $p = 0.005$) and during the Session 2 period ($F_{1,24} = 10.52$, $p = 0.004$), and a significant difference during the Recovery period ($F_{1,24} = 4.46$, $p = 0.045$), with REML being shorter in the "worst-case" condition than in the "best-case" condition during these periods. There was no statistically significant difference between the two conditions during the Restart period ($F_{1,23} = 1.99$, $p = 0.17$). In two-way ANOVA of average REML, there was no significant interaction of condition by session ($F_{1,48} = 0.29$, $p = 0.59$) and no main effect of session ($F_{1,48} = 0.02$, $p = 0.89$), but there was a significant main effect of condition ($F_{1,48} = 19.64$, $p < 0.001$).
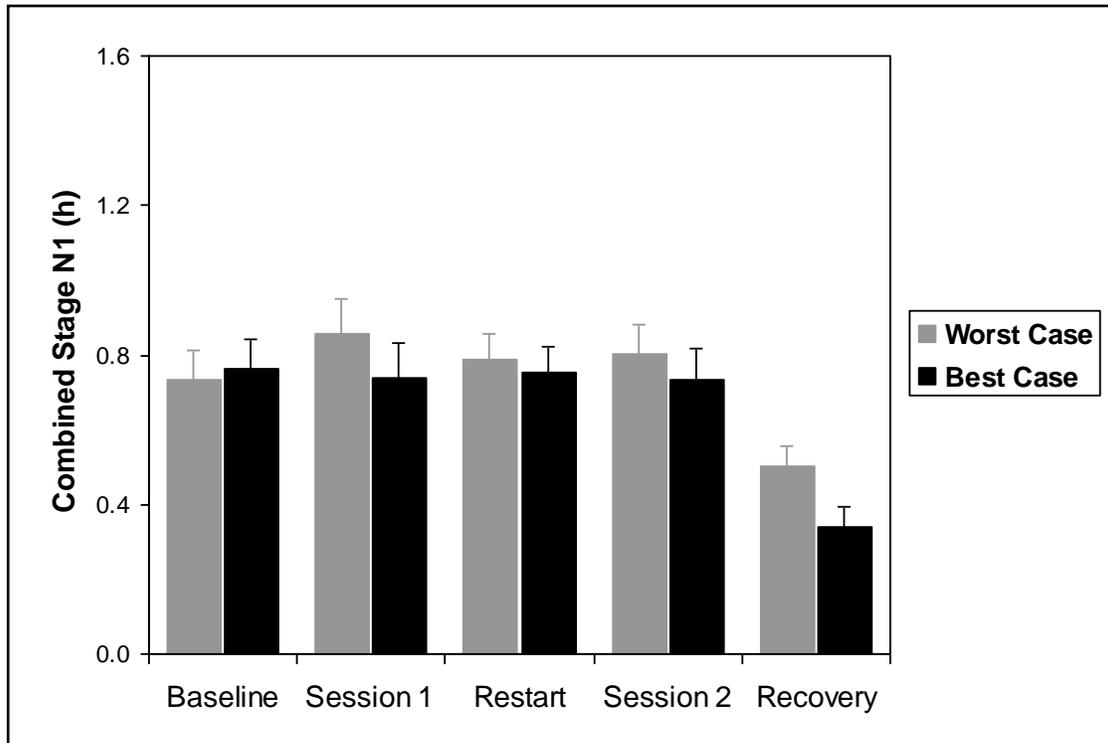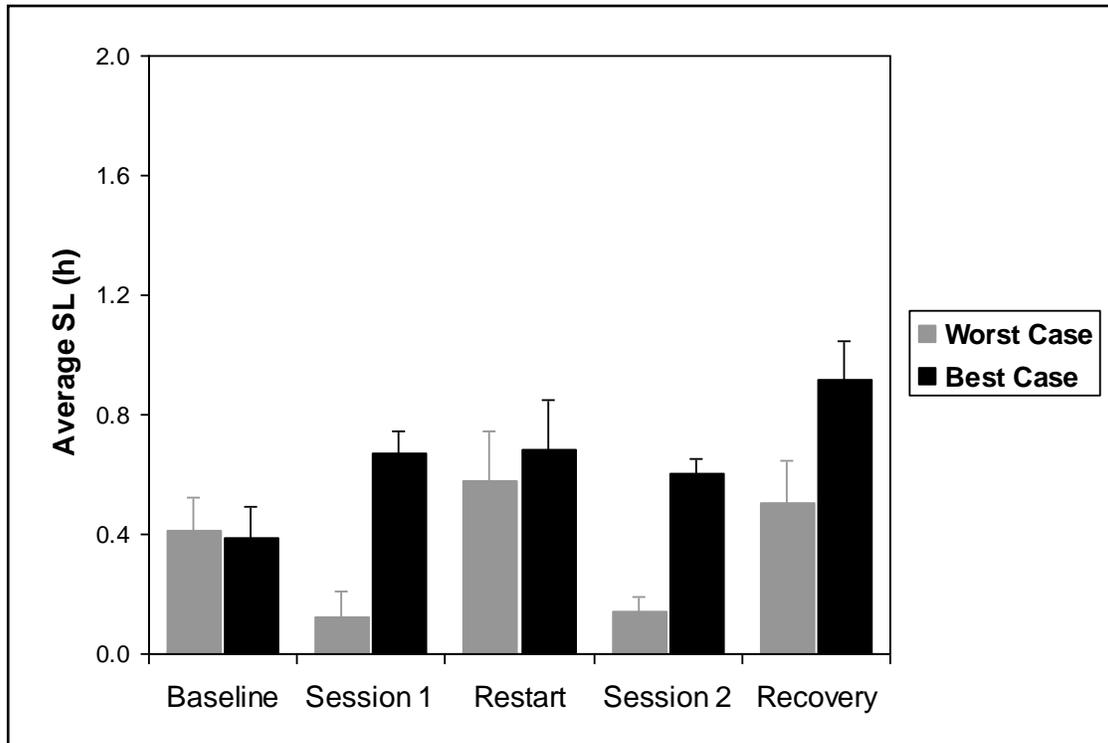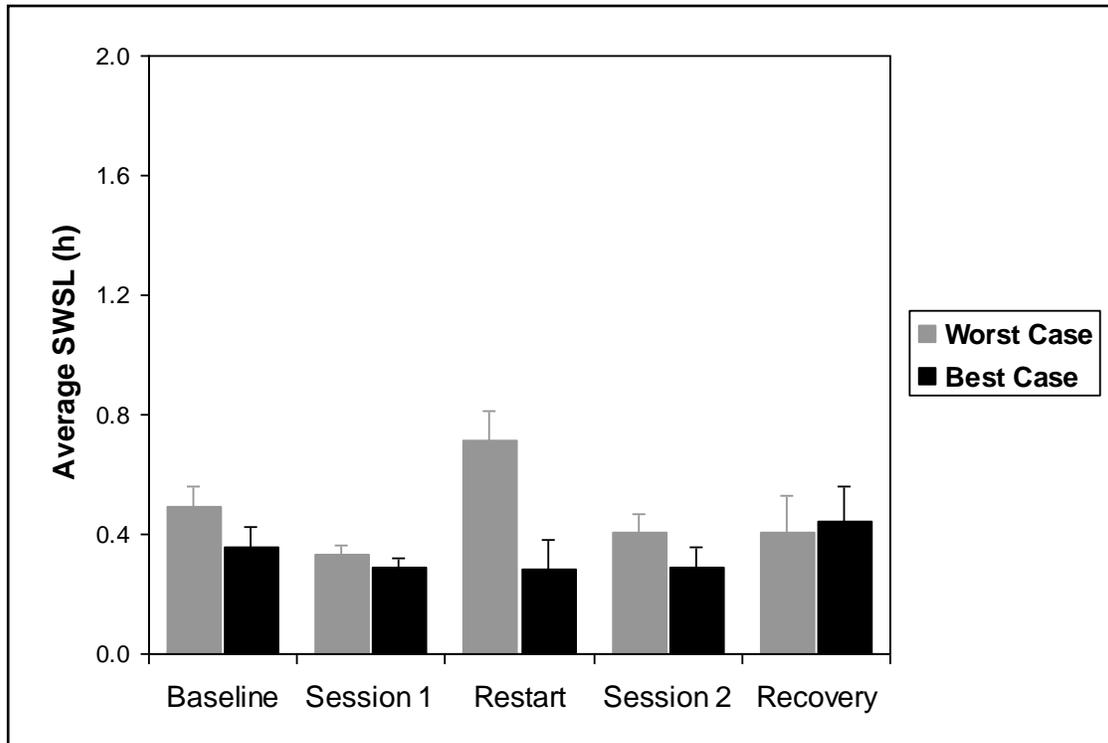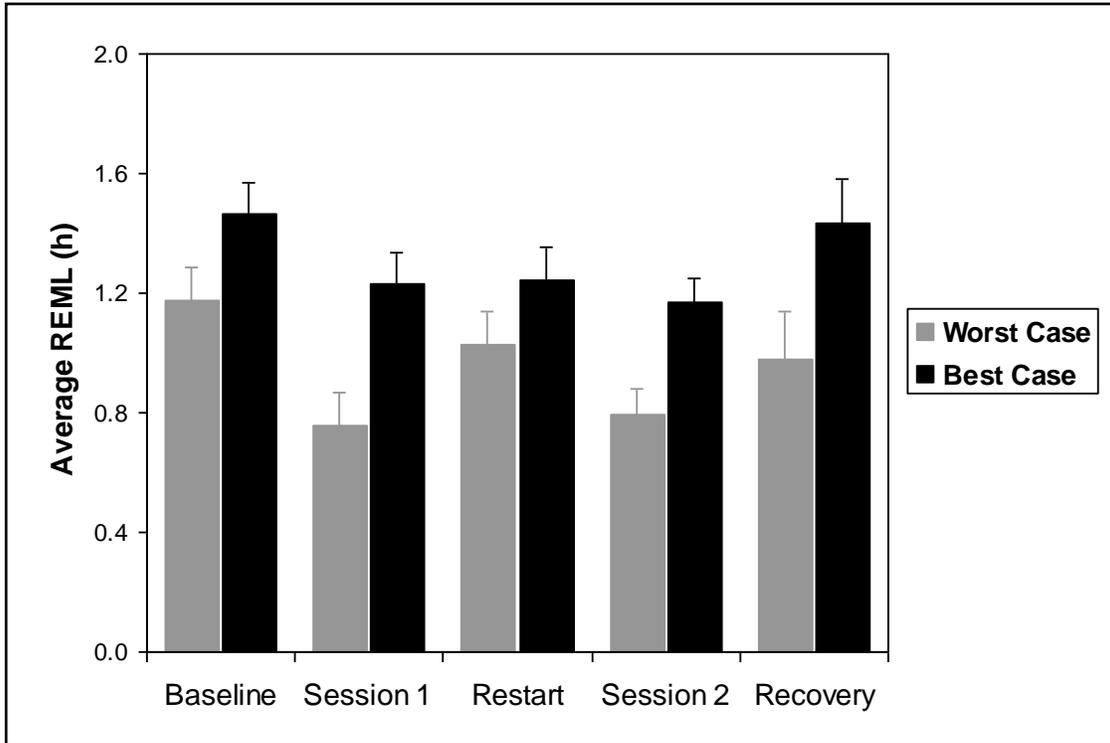
Figure 40. Graph. REML during the Baseline, Session 1, Restart, Session 2, and Recovery sets of sleep periods, comparing them for "worst-case" and "best-case" conditions.

4. CONCLUSIONS

4.1 KEY FINDINGS

The objective of the project was to determine the effectiveness of the current 34-hour restart provision in the HOS regulations governing freight-carrying commercial motor vehicle (CMV) drivers with regard to restoring performance. An in-residence laboratory research study was conducted to address this issue. Running the study in the laboratory (as opposed to in the field) helped to eliminate environmental confounds, allowed for the use of sensitive laboratory performance measures, simplified the logistics, and moderated the sample size requirement as corroborated by a power calculation performance in advance of the study.

A sample of $N = 27$ healthy subjects was studied in a "worst-case"–"best-case" between-groups comparison of two 5-day (14-hour/day) work periods separated by a 34-hour restart period. Half the sample ($n = 14$) was randomized to the "best-case" condition, which entailed daytime wakefulness and work (and nighttime sleep) throughout the study. The other half ($n = 13$) was randomized to the "worst-case" condition, which entailed nighttime wakefulness and work (and daytime sleep) during the two 5-day work periods, while transitioning back to a daytime schedule during the 34-hour restart period (see Figure 3). Performance on a variety of neurobehavioral tasks and on a high-fidelity driving simulator was measured throughout the study (see figure 4). The main goal was to evaluate whether the 34-hour restart period was effective at maintaining performance in both conditions.

The primary performance measure for the study was the number of lapses (reaction times greater than 500 ms) on a 10-minute PVT, which was administered eight times per day in the working periods. In the "best-case" condition, average PVT performance in the 5-day work period after the 34-hour restart was the same as that in the 5-day work period before the 34-hour restart, indicating that the restart period was effective at maintaining performance in this condition. In the "worst-case" condition, however, average PVT performance in the 5-day work period after the 34-hour restart was impaired relative to average PVT performance in the 5-day work period before the 34-hour restart, indicating that the restart period was not effective at maintaining performance in the "worst-case" condition (see Figure 4). That being the case, the null hypothesis that the 34-hour restart period would be effective at maintaining performance in both conditions was rejected.

PVT performance deficits in the "worst-case" condition were most prominent in the later hours of the night, as is consistent with principles of sleep/wake physiology.[13] Two major underlying processes are involved in sleep/wake physiology:[5] a homeostatic process that builds up a pressure for sleep during wakefulness and dissipates that pressure during sleep, and a circadian process that exerts a pressure for wakefulness during the daytime hours and withdraws that pressure during the night. During normal daytime wakefulness and nighttime sleep, the homeostatic pressure for sleep and the circadian pressure for wakefulness act in opposition, which results in stable wakefulness during the day and consolidated sleep during the night.[21] However, when wakefulness is scheduled during the night, as was the case in the "worst-case" condition, then the homeostatic and circadian processes become desynchronized. This results in a

progressive deterioration of performance during the nighttime waking period,[22] as was seen in the present study (Figure 7).

In addition, when wakefulness is scheduled during the night, sleep must be obtained during the day, when the circadian pressure for wakefulness is high. This results in sleep curtailment, as observed in the present study (see Figure 33). Indeed, total sleep time was consistently reduced in the "worst-case" condition relative to the "best-case" condition during the two 5-day work periods. Moreover, during the 34-hour restart period, when subjects in the "worst-case" condition transitioned back to a normal nighttime sleep schedule and had transition nap opportunities, they still accumulated less total sleep time than the subjects in the "best-case" condition (who maintained a normal nighttime sleep routine throughout the study). Thus, despite equal amounts of sleep opportunity (time in bed), subjects in the "worst-case" condition did not manage to get enough sleep, due to the inopportune circadian placement of the sleep periods, and did not manage to maintain optimal performance during the nighttime waking period, due to the temporal desynchronization of the homeostatic pressure for sleep and the circadian pressure for wakefulness. Perhaps due to the temporary transitioning back to a normal nighttime sleep schedule during the 34-hour restart period, this intervention was not effective at mitigating the sleep loss and consequent performance impairment—and, in fact, might even have contributed to the increased impairment in the post-restart 5-day work period. If a nighttime wake schedule were to be maintained during the 34-hour restart period, it is possible that gradual circadian adjustment would have occurred, potentially making the 34-hour restart period more effective. The real-world utility of this possibility is questionable, though, as it is improbable that many individuals would elect to maintain a permanent night-shift schedule if given the choice.

The level of PVT performance impairment reached in the "worst-case" condition (with group-average lapse counts not exceeding 4) was modest when compared to the documented effects[1,2] of one night of acute total sleep deprivation or a week of sustained sleep restriction to 6 hours per day (which are in the range of six to eight lapses). One reason this may be the case was the restriction of duty time to 14 hours per day (per the current HOS regulations), leaving up to 10 hours per day for recuperation. Another reason may be the presence of nap opportunities preceding the transitions to nighttime wakefulness, which effectively served as prophylactic naps known to be useful as countermeasures for neurobehavioral performance impairment.[23,24]

Without strategic napping, performance in the "worst-case" condition would probably have shown more considerable deficits. Furthermore, the effects of chronic sleep restriction are to some extent cumulative,[1,2] with stabilization taking weeks to months,[25] which implies that additional cycles of a 34-hour restart period followed by a nocturnal 5-day work period could result in further increases in neurobehavioral impairment. It should also be noted that our research subjects were healthy young adults with no sleep disorders. Had we studied a sample of patients with sleep apnea—widely reported to be a common sleep disorder among CMV drivers[26]—or other medical conditions, the expected performance deficits in the "worst-case" condition would have been greater still, strengthening the conclusion that the 34-hour restart period was not effective at maintaining performance across multiple 5-day nighttime work periods.

4.2 ANCILLARY FINDINGS

Subjects in the "worst-case" condition were overall subjectively sleepier than subjects in the "best-case" condition, but did not show an increase in sleepiness after the 34-hour restart period that would correspond with the PVT findings. In fact, subjects in the "worst-case" condition showed reduction of sleepiness across days in each of the two 5-day work periods, suggesting that they subjectively adapted to the nighttime wakefulness schedule but that the 34-hour restart period undid some of this subjective adaptation (see figure 9). A dissociation between subjective sleepiness and objective performance impairment under conditions of sustained sleep curtailment has been reported previously,[2] although to our knowledge this is the first time it is shown for a night-shift schedule. Across the hours of each night, however, subjective sleepiness increased in a way similar to the within-day profile seen for PVT lapses (see Figure 7 and Figure 10), suggesting that subjective adaptation took place gradually over days but not acutely within days. Self-report measures of affect, mood, and effort did not systematically vary much between the two conditions or across the study days, with the exception of subjective mood, which showed the same profile of degradation over time of night in the "worst-case" condition as seen for subjective sleepiness and PVT lapses.

On the DSST and the CDDT, subjects in both the "best-case" and the "worst-case" conditions showed expected practice effects (Figure 16 and Figure 23). For the DSST, the practice effect was somewhat counteracted by increasing impairment across time of night in the "worst-case" condition (see Figure 17), but otherwise there were no substantive differences between the two conditions for either the DSST or the CDDT. This suggests that cognitive throughput (measured by the DSST) and spatial orientation (measured by the CDDT) were not adversely affected much by subjects being on a night shift schedule, nor did the 34-hour restart period seem to have much of an impact, positive or negative, on these cognitive functions. There is increasing evidence that the effects of sleep loss and circadian misalignment vary as a function of the components of cognition (e.g., attention, working memory) involved in the neurobehavioral performance task at hand.[12,27–30] It is therefore not particularly surprising that the results for the DSST and the CDDT did not line up with those for the PVT. However, the results for the PVT, which measures reaction time and sustained attention, may ultimately be more significant for CMV drivers. This possibility was in part corroborated by the simulator driving data, which are discussed next.

There was a modest practice effect for driving performance on the high-fidelity driving simulator in terms of speed variability and lane deviation. However, subjects in the "worst-case" condition displayed only half the reduction in lane deviation across the study that subjects in the "best-case" condition showed (see Figure 28). Furthermore, subjects in the "worst-case" condition displayed an increase in lane deviation over the hours of the night (Figure 29) similar to the increase seen in PVT lapses (Figure 7). This was accompanied by an increase of up to 1 percent in fuel use across the night in the "worst-case" condition (see Figure 30). This suggests that the effects of sleep loss and circadian misalignment associated with a night work schedule may impair driving not only in terms of road safety, but also in terms of economic cost.

Indices of driving impairment—speed variability, lane deviation, emergency braking reaction time, and fuel use—were significantly predicted by lapses on the PVT administered just before the simulator driving bouts. Fuel use in the simulator was shown to be a function of driving impairment; the poorer one drives, the more one tends to use more fuel. Although the

predictiveness of PVT performance for driving performance was not perfect, this finding suggests that the effects of sleep loss and circadian misalignment on driving performance reflect, at least in part, deficits in sustained attention (as captured by the PVT). This is consistent with an earlier finding that performance on a driving simulator was predicted by sleep latency on a maintenance of wakefulness test.[31] The present finding pertaining to the PVT is a potentially important new contribution, because it indicates that a large volume of literature using PVT performance as an index of fatigue-related impairment[11,13,14,32] is relevant for driving performance—including a recently published biomathematical model for the prediction of fatigue and performance that was calibrated on PVT lapses.[25]

Because of the between-groups design of the research study, the driving simulator scenarios (roads, routes, events, conditions, etc.) were standardized across the "worst-case" and "best-case" conditions. The scenarios did not, therefore, control for the increased traffic density typically encountered during daytime driving (for the "best-case" condition") or the reduced visibility associated with nighttime driving (for the "worst-case" condition). That being the case, the driving simulator findings should be interpreted as indicating basal capability for driving rather than actual driving performance in the real world. Further research is needed to compare the "worst-case" and "best-case" schedules in terms of real-world driving performance, safety, and cost.

Some observations on the structure of sleep are noteworthy. First, despite curtailment of total sleep time in the "worst-case" condition, slow-wave sleep (non-REM sleep stage N3) was conserved relative to the "best-case" condition (Figure 34). This phenomenon has been reported in sustained sleep restriction experiments,[1,2,33] but to our knowledge this is the first time that it is being reported for inverted sleep/wake schedules. Second, both REM sleep (Figure 35) and non-REM sleep stage N2 (Figure 36) varied hand in hand with total sleep time, whereas non-REM sleep stage N1 (Figure 37) did not show any substantive differences between conditions or across the days of the study. No single sleep stage could be identified as being responsible for the performance impairments in the "worst-case" condition, making it more parsimonious to assume that overall curtailment of total sleep time was the main effector of the observed performance deficits. This is consistent with conclusions drawn from other experimental findings.[2,34] Third, the "worst-case" condition resulted in reduced REM latencies during the daytime sleep periods (see Figure 40), which is consistent with the known circadian regulation of REM sleep expression.[35] Finally, once sleep loss was incurred in the "worst-case" condition, sleep latency was consistently shorter in this condition than it was in the "best-case" condition (see Figure 38). Thus, sleep curtailment in the "worst-case" condition did not result from difficulty initiating sleep, but rather from difficulty *maintaining* sleep (i.e., increased intermittent wakefulness and early termination of sleep) due to the growing circadian pressure for wakefulness across daytime sleep periods.

5. RECOMMENDATIONS

On the basis of the results of the study as described in this report, a number of recommendations can be made.

- The key finding of this laboratory research study was that the 34-hour restart provision in the HOS regulations governing property-carrying CMV drivers was sufficiently effective (even, perhaps, unnecessary) for maintaining optimal performance in subjects randomized to a daytime wake/work schedule, but was insufficiently effective—*and perhaps even detrimental*—for maintaining performance in subjects randomized to a nighttime wake/work schedule. The main reason for this is that the 34-hour restart provision does not take into account the circadian aspect of sleep/wake/work scheduling. HOS regulations could be improved by taking circadian timing into account.

- PVT performance was shown to be predictive of simulator driving performance, suggesting that mathematical models of fatigue and performance predictive of PVT performance may be predictive of driving performance as well. Validation of fatigue and performance models for accurately predicting real-world driving performance would allow for the development of model-based fatigue risk management approaches, which would help to make HOS regulations sensitive to circadian timing, and could ultimately be mandated as a safe alternative for prescriptive HOS regulations.

- Objective performance deficits in subjects assigned to a nighttime wake/work schedule were not accurately reflected in subjective sleepiness and mood states. This discrepancy between actual impairment and introspective awareness is a common finding in sleep and performance research, and presents a safety problem in real-world settings because individuals cannot be relied upon to accurately self-identify fatigue-induced impairment. Ongoing work on the development of individualized drowsy-driver detection technologies may contribute to addressing this issue. The present findings provide support for using lane deviation as an objective performance measure in such technologies. CMV drivers should also be educated with regard to the dangers of drowsy driving and the difficulties of self-evaluating the ability to continue driving safely.

- The research subjects in this study were healthy young adults with no sleep disorders. However, sleep apnea is widely reported to be a common sleep disorder among CMV drivers and other medical conditions are common among CMV drivers. Although it may be inferred that the results of the present study set a lower limit for levels of impairment to be expected in the CMV driver population, validation of the study findings in a sample of drivers in a real-world field study (such as that currently being sponsored by the FMCSA in conjunction with Transport Canada) is important.

- Performance deficits in the nighttime work periods increased as a function of time of night, and in this study, the sleep curtailment underlying these deficits involved difficulty maintaining sleep rather than difficulty initiating sleep. Such findings could provide a basis for research on judicious application of fatigue countermeasures, whether pharmacological (including caffeine) or behavioral. Strategic napping is a particularly powerful fatigue countermeasure, and the authors recommend consideration of adapting HOS regulations to allow for greater flexibility in split sleep schedules.

- The finding that the modest increase in indices of driving impairment on the high-fidelity driving simulators across time of night was associated with an increase of up to 1 percent in computed fuel use suggests that drowsy driving may be associated with direct economic costs. Demonstrating the validity of fuel use on the driving simulators as an accurate indicator of fuel use in real trucks on real roads (which could be pursued in closed track settings) would provide carriers with a bottom-line incentive for actively managing fatigue in CMV drivers.

Several of these recommendations could be explored in follow-up research in Phase II of this project. The peer review committee for the project underwrote the following specific suggestions for Phase II:

- Document the effectiveness of the 34-hour restart if optimal combinations of anchor sleep and naps (i.e., split sleep) are placed at circadian-appropriate times during both duty days and during the 34-hour restart period. There is considerable interest in CMV operations to explore the possibilities of strategic use of split sleep schedules. It would be a natural extension of Phase I to examine split sleep schedules in relation to the restart provision of the HOS regulations in Phase II.

- Investigate whether caffeine would improve performance in the "worse-case" condition. There is an outstanding need to investigate the effectiveness of caffeine for counteracting fatigue in CMV operations. A placebo-controlled study of caffeine effects in the context of the restart provision would help to address this issue in Phase II.

ACKNOWLEDGEMENTS

[This page intentionally left blank]

REFERENCES

1. Belenky G, Wesensten NJ, Thorne DR, Thomas ML, Sing HC, Redmond DP, et al. Patterns of performance degradation and restoration during sleep restriction and subsequent recovery: a sleep dose-response study. *Journal of Sleep Research 12*: 1–13, 2003.

2. Van Dongen HPA, Maislin G, Mullington JM, & Dinges DF. The cumulative cost of additional wakefulness: Dose-response effects on neurobehavioral functions and sleep physiology from chronic sleep restriction and total sleep deprivation. *Sleep 26*: 117–126, 2003.

3. Wesensten NJ, Belenky G, Thorne DR, Kautz MA, & Balkin TJ. Modafinil versus caffeine: Effects on fatigue during sleep deprivation. *Aviation, Space, and Environmental Medicine 75*: 520–525, 2004.

4. Achermann P, & Borbély AA. Simulation of daytime vigilance by the additive interaction of a homeostatic and a circadian process. *Biological Cybernetics 71*: 115–121, 1994.

5. Borbély AA, & Achermann P. Sleep homeostasis and models of sleep regulation. *Journal of Biological Rhythms 14*: 557–568, 1999.

6. Dijk DJ, & Czeisler CA. Contribution of the circadian pacemaker and the sleep homeostat to sleep propensity, sleep structure, electroencephalographic slow waves, and sleep spindle activity in humans. *Journal of Neuroscience 15*: 3526–3538, 1995.

7. Van Dongen HPA, & Dinges DF. Circadian rhythms in sleepiness, alertness, and performance. In M Kryger, T Roth, and WC Dement (Eds.) *Principles and Practice of Sleep Medicine*, 4th Edition, pp. 435–443, Philadelphia, PA: Elsevier, 2005.

8. Åkerstedt T. Shift work and disturbed sleep/wakefulness. *Occupational Medicine 53*: 89–94, 2003.

9. Mistlberger RE, Rusak B. Circadian rhythms in mammals: Formal properties and environmental influences. In M Kryger, T Roth, and WC Dement (Eds.) *Principles and Practice of Sleep Medicine*, 4th Edition, pp. 321–334, Philadelphia, PA: Elsevier, 2005.

10. Moore JM, Van Dongen HPA, Belenky G, Mott CG, Huang L, & Vila B. Use of a driving simulator to assess fuel inefficiency as a downstream effect of driver sleepiness in controlled laboratory experiments. *Sleep 32*: A387–A388, 2009.

11. Dorrian J, Rogers NL, & Dinges DF. Psychomotor vigilance performance: Neurocognitive assay sensitive to sleep loss. In CA Kushida (Ed.) *Sleep Deprivation. Clinical Issues, Pharmacology, and Sleep Loss Effects*, pp. 39–70, New York, NY: Marcel Dekker, 2005.

12. Van Dongen HPA, Baynard MD, Maislin G, & Dinges DF. Systematic interindividual differences in neurobehavioral impairment from sleep loss: Evidence of trait-like differential vulnerability. *Sleep 27*: 423–433, 2004.

13. Van Dongen HPA, & Dinges DF. Sleep, circadian rhythms, and psychomotor vigilance. *Clinics in Sports Medicine 24*: 237–249, 2005.

14. Lim J, & Dinges DF. Sleep deprivation and vigilant attention. *Annals of the New York Academy of Sciences 1129*: 305–322, 2008.

15. Åkerstedt T, & Gillberg M. Subjective and objective sleepiness in the active individual. *International Journal of Neuroscience 52*: 29–37, 1990.

16. Watson D, Clark LA, & Tellegen A. Development and validation of brief measures of positive and negative affect: The PANAS scale. *Journal of Personality and Social Psychology 54*: 1063–1070, 1988.

17. Wechsler D. Wechsler Adult Intelligence Scale–Revised. New York, NY: Psychological Corp., 1981.

18. Dinges DF, Kribbs NB, Steinberg KN, & Powell JW. Do we lose the willingness to perform during sleep deprivation? *Sleep Research 21*: 318, 1992.

19. Gunzelmann G, Anderson JR, & Douglass S. Orientation tasks with multiple views of space: Strategies and performance. *Spatial Cognition and Computation 4*: 207–253, 2004.

20. Iber C, Ancoli-Israel S, Chesson Jr AL, & Quan DF (Eds.). *The AASM manual for the scoring of sleep and associated events. Rules, terminology and technical specifications.* Westchester, IL: American Academy of Sleep Medicine, 2007.

21. Dijk DJ, & Czeisler CA. Paradoxical timing of the circadian rhythm of sleep propensity serves to consolidate sleep and wakefulness in humans. *Neuroscience Letters 166*: 63–68, 1994.

22. Van Dongen HPA. Shift work and inter-individual differences in sleep and sleepiness. *Chronobiology International 23*: 1139–1147, 2006.

23. Dinges DF, Orne MT, Whitehouse WG, & Orne EC. Temporal placement of a nap for alertness: Contributions of circadian phase and prior wakefulness. *Sleep 10*: 313–329, 1987.

24. Bonnet MH, & Arand DL. Consolidated and distributed nap schedules and performance. *Journal of Sleep Research 4*: 71–77, 1995.

25. McCauley P, Kalachev LV, Smith AD, Belenky G, Dinges DF, & Van Dongen HPA. A new mathematical model for the homeostatic effects of sleep loss on neurobehavioral performance. *Journal of Theoretical Biology 256*: 227–239, 2009.

26. Pack AI, Maislin G, Staley B, Pack FM, Rogers WC, George CFP, & Dinges DF. Impaired performance in commercial drivers. Role of sleep apnea and short sleep duration. *American Journal of Respiratory and Critical Care Medicine 174*: 446–454, 2006.

27. Harrison Y, & Horne JA. The impact of sleep deprivation on decision making: A review. *Journal of Experimental Psychology: Applied 6*: 236–249, 2000.

28. Drummond SPA, Meloy MJ, Yanagi MA, Orff HJ, & Brown GG. Compensatory recruitment after sleep deprivation and the relationship with performance. *Psychiatry Research: Neuroimaging 140*: 211–223, 2005.

29. Ratcliff R, & Van Dongen HPA. Sleep deprivation affects multiple distinct cognitive processes. *Psychonomic Bulletin & Review 16*: 742–751, 2009.

30. Tucker AM, Whitney P, Belenky G, Hinson JM, & Van Dongen HPA. Effects of sleep deprivation on dissociated components of executive functioning. *Sleep 33*: 47–57, 2010.

31. Banks S, Catcheside P, Lack LC, Grunstein RR, & McEvoy RD. The Maintenance of Wakefulness Test and driving simulator performance. *Sleep 28*: 1381–1385, 2005.

32. Balkin TJ, Bliese PD, Belenky G, Sing H, Thorne DR, Thomas M, et al. Comparative utility of instruments for monitoring sleepiness-related performance decrements in the operational environment. *Journal of Sleep Research 13*: 219–227, 2004.

33. Brunner DP, Dijk DJ, & Borbély AA. Repeated partial sleep deprivation progressively changes the EEG during sleep and wakefulness. *Sleep 16*: 100–113, 1993.

34. Mollicone DJ, Van Dongen HPA, Rogers NL, & Dinges DF. Response surface mapping of neurobehavioral performance: Testing the feasibility of split sleep schedules for space operations. *Acta Astronautica 63*: 833–840, 2008.

35. Czeisler CA, Zimmerman JC, Ronda JM, Moore-Ede MC, & Weitzman ED. Timing of REM sleep is coupled to the circadian rhythm of body temperature in man. *Sleep 2*: 329–346, 1980.