

# Investigating systematic individual differences in sleep-deprived performance on a high-fidelity flight simulator

HANS P. A. VAN DONGEN

*Washington State University, Spokane, Washington*

and

JOHN A. CALDWELL, JR. and J. LYNN CALDWELL

*Air Force Research Laboratory, Brooks City-Base, San Antonio, Texas*

Laboratory research has revealed considerable systematic variability in the degree to which individuals' alertness and performance are affected by sleep deprivation. However, little is known about whether or not different populations exhibit similar levels of individual variability. In the present study, we examined individual variability in performance impairment due to sleep loss in a highly select population of military jet pilots. Ten active-duty F-117 pilots were deprived of sleep for 38 h and studied repeatedly in a high-fidelity flight simulator. Data were analyzed with a mixed-model ANOVA to quantify individual variability. Statistically significant, systematic individual differences in the effects of sleep deprivation were observed, even when baseline differences were accounted for. The findings suggest that highly select populations may exhibit individual differences in vulnerability to performance impairment from sleep loss just as the general population does. Thus, the scientific and operational communities' reliance on group data as opposed to individual data may entail substantial misestimation of the impact of job-related stressors on safety and performance.

It has been demonstrated that considerable individual differences exist in performance impairment during total sleep deprivation (Leproult et al., 2003; Van Dongen, Maislin, & Dinges, 2004) and that these individual differences constitute a *trait* (Van Dongen, Baynard, Maislin, & Dinges, 2004). Substantial individual differences in vulnerability to performance deficits from sleep loss have also been documented for conditions of chronic sleep restriction (Van Dongen, Maislin, Mullington, & Dinges, 2003). Such findings are important both conceptually (Kosslyn et al., 2002) and with regard to real-world operations in which industrial shift workers, medical per-

sonnel, disaster relief workers, military personnel, and others must regularly function effectively despite limited or disrupted sleep.

Performance impairment due to sleep loss is a significant issue within commercial and military aviation contexts (Caldwell, 2005), in which long-duration flights often involve truncated preflight sleep, lengthy periods of continuous wakefulness, and insufficient layover sleep. Pilots routinely face these contributors to on-the-job fatigue. Individuals who are comparatively resistant to the fatiguing effects of sleep loss may be better suited for such rigorous scheduling demands than those who are more vulnerable.

It is not yet known to what extent the trait individual differences in fatigue vulnerability found within the general population can also be observed in aviators and other highly select populations. Selection factors could moderate trait individual differences in populations that are frequently exposed to critical circumstances in which optimal performance must be maintained despite sleep deprivation. Pilots in general, and military pilots in particular, constitute a population that may be subject to such self-selection. It could be hypothesized that these groups consist largely of relatively fatigue-resistant individuals because those who are found to be vulnerable are somehow screened out during training, or because those who successfully complete training, but later find themselves unable to tolerate the highly fatiguing aviation schedules, opt out of the profession. It would be useful to determine

---

This work was supported by the U.S. Army Medical Research and Materiel Command, Fort Detrick, MD; the U.S. Air Force Office of Scientific Research (AFOSR); and the U.S. Air Force Research Laboratory. H.P.A.V.D. was supported by AFOSR Grant FA9550-05-1-0086 and NIH Grant R01-HL70154, and by the Institute for Experimental Psychiatry Research Foundation. The authors are grateful to the command and staff of the 49th Fighter Wing at Holloman Air Force Base, to Col. David L. Brown, to 1st Lt. Jennifer Smith, to the technicians from the Air Force Research Laboratory and the School of Aerospace Medicine, to every individual responsible for the operation of the Link F-117A Weapon System Trainer, to the F-117 instructor pilots who helped refine the simulator flight profile, to the F-117 pilots who selflessly dedicated their time to serve as research participants, and to Erik Olofsen and Poduri Rao for sharing their statistical expertise. Correspondence concerning this article should be addressed to H. P. A. Van Dongen, Sleep and Performance Research Center, Washington State University, P.O. Box 1495, Spokane, WA 99210-1495 (e-mail: hvd@wsu.edu).

whether or not this is the case for various reasons, such as exploring the existence of selection or adaptation factors responsible for fatigue resistance.

Thus, it is of interest to examine whether pilots are somehow considerably more homogeneous than other groups with regard to individual differences in vulnerability to impairment from sleep loss. In the present study, we examined individual differences in the impact of up to 38 h of continuous wakefulness on simulator flight performance and subjective states of fully qualified, active-duty U.S. Air Force fighter pilots. Group average results of this sleep deprivation study have been documented previously (Caldwell, Caldwell, Brown, & Smith, 2004). Here, we further analyzed the data from that study to determine whether or not this highly select population showed evidence of systematic individual differences in performance changes across time awake. Although there have been debates about how to properly analyze systematic individual differences across time (see Van Dongen, 2005), there is a substantial, consistent literature on the statistically valid approaches (see, e.g., Burton, Gurrin, & Sly, 1998; Cox & Solomon, 2004; Feldman, 1988; Goldstein, 1995; Khuri, Mathew, & Sinha, 1998; Rao, 1997; Singer, 1998; Van Dongen, Maislin, & Dinges, 2004; Van Dongen, Olofsen, Dinges, & Maislin, 2004). These include mixed-model ANOVA, which is the technique we applied in this study.

## METHOD

### Participants

Ten qualified F-117 pilots, ranging in age from 27 to 43 years ( $M = 35.7$  years,  $SD = 5.7$ ), who were members of the 49th Fighter Wing at Holloman Air Force Base, New Mexico, served as participants. Prior to the study, each participant's medical record was screened for current illnesses or disqualifying medications by the medical monitor or his designee at the Holloman Air Force Base medical clinic. Every participant was in possession of a current *up slip* (i.e., a U.S. Air Force medical permit to engage in military flying duties) at the time of admission, and each of them was current and qualified in the F-117A aircraft. All the participants were male (at the time, there were no female F-117A pilots).

Only pilots who had been working the day shift (those normally remaining awake approximately from 06:00–07:00 to 21:00–23:00) for at least 5 days prior to the beginning of the study were allowed to enroll. None of the participants was taking any kind of medication known to impact mental alertness (e.g., sedating antihistamines, sleep medications, prescription stimulants). One pilot was a tobacco user; he was allowed to use tobacco during breaks to avoid withdrawal effects, but not during testing sessions. None of the participants were habitual nappers or heavy caffeine consumers.

The experiment was approved by the Air Force Institutional Review Board for the safe conduct of research using human subjects, and all the participants gave informed consent.

### Experimental Design

The purpose of the study was to investigate systematic individual differences in performance impairment on a high-fidelity flight simulator during simulated sustained operations (i.e., missions in which sleep deprivation is common). The entire experiment was conducted inside the F-117 Weapon System Trainer (WST) facility at Holloman Air Force Base. Flight performance data were collected with the WST flight simulator and ancillary equipment. Other measures were

collected with various laboratory testing devices, which were set up in a sound-attenuated testing room within the simulator facility.

The participants completed three adaptation/training sessions on the day before the experiment. After sleeping at home and engaging in normal waking activities the next day, they came to the flight simulator at 18:00 and were kept awake there until approximately 21:00 the next day. There were five testing sessions, which covered the final 24 h of this period of sustained wakefulness. Only 1 participant was studied per experimental session. See Caldwell et al. (2004) for a full description of all tests and procedures used in this study.

On the training day preceding the experiment, the participants arrived at the simulator facility at approximately 12:30. They were informed about the flight profile to be flown in the flight simulator and the manner in which each simulated flight would be conducted. It was explained that a staff member would sequence the volunteer through each of the flight maneuvers in the flight profile. In order to reduce practice-related improvements during the actual experiment, three training flights were conducted, at 14:00, 17:00, and 20:00. Interspersed among these flights were laboratory tests; the ones reported here are the Fitness Impairment Tester (FIT; PMI, Rockville, MD), the Profile of Mood States (POMS; McNair, Lorr, & Droppleman, 1971), and a visual analog scale (VAS) of sleepiness adapted from Penetar et al. (1993). There were three iterations of the POMS and VAS and two FIT evaluations on the training day. Data from the third training flight, from the third iterations of the POMS and VAS, and from the second evaluation on the FIT were considered baseline measurements.

The participants were released from training at approximately 21:00 to spend the night at home. They were asked to wear an Actiwatch 64 wrist actigraph (Mini Mitter, Bend, OR) while outside the simulator facility. The actigraph data were later subjected to automated analysis using the Actiware-Sleep software Version 3.3 (Mini Mitter) to estimate sleep duration during the night after training (i.e., before the experiment) and to assure that the participants complied with requested bedtimes. The participants were instructed to spend 8 h in bed and to wake up no later than 07:00 the next morning. The participants were also asked to avoid napping following awakening, and to refrain from caffeine intake after 10:00.

The participants returned to the simulator facility at 18:00. They stayed awake through the night and into the next day. Every 5 h, starting at 21:00 (i.e., 3 h after arrival in the facility), the participants underwent a series of tests. In each session, they first performed a series of laboratory tests, including the FIT, the POMS, and the VAS. Upon completing the nonflight tests (approximately 2 h after beginning these tests), the participants were escorted into the flight simulator. They were asked to fly a sequence of 13 maneuvers, together taking approximately 1 h to complete. Table 1 shows the flight profile, which was the same for each test session. The participants were instructed to fly all maneuvers at an airspeed of 300 knots, to perform all climbs and descents at a rate of 1,000 ft/min, and to fly all turns at a 30° angle of bank. No performance feedback was provided during the flight maneuvers.

At the conclusion of the flight, the participants were allowed a 2-h break, after which the next test session began. There was a total of five test sessions, conducted at 5-h intervals. Following the last test session, the pilots were debriefed and released after 38 h of sustained wakefulness. They were driven home by a staff member or a family member at the end of the experimental day.

### Flight Performance Assessments

The F-117A WST (L-3 Communications/Link Simulation and Training, Binghamton, NY), which is a high-fidelity flight simulator employed by the U.S. Air Force, was used to conduct the flight performance assessments. The WST is a stationary digital device that simulates the characteristics and operations of the F-117A stealth fighter aircraft currently in the U.S. Air Force inventory. The F-117A aircraft is a twin-turbofan, low-radar, ground-attack fighter with a single-seat cockpit. The WST provides a fully functioning replica of the interior

**Table 1**  
**Simulator Flight Profile**

Maneuver Abbreviation	Description	Flight Parameters Evaluated
R360	right 360° turn at an altitude of 11,000 ft MSL	alt, IAS, roll
S&L	straight and level on a heading of 345° at 11,000 ft MSL	alt, IAS, hdg
L360	left 360° turn at an altitude of 11,000 ft MSL	alt, IAS, roll
CLMB	straight climb from 11,000 to 13,000 ft MSL	IAS, IVS, hdg
S&L	straight and level on a heading of 345° at 13,000 ft MSL	alt, IAS, hdg
DRT	descending right 360° turn to an altitude of 10,000 ft MSL	IAS, IVS, roll
S&L	straight and level on a heading of 345° at an altitude of 10,000 ft MSL	alt, IAS, hdg
CLT	climbing left 540° turn to an altitude of 15,000 ft MSL	IAS, IVS, roll
S&L	straight and level on a heading of 165° at 15,000 ft MSL	alt, IAS, hdg
R360	right 360° turn at an altitude of 15,000 ft MSL	alt, IAS, roll
S&L	straight and level on a heading of 165° at 15,000 ft MSL	alt, IAS, hdg
L720	left 720° turn at an altitude of 15,000 ft MSL	alt, IAS, roll
DESC	straight descent from 15,000 to 13,000 ft MSL	IAS, IVS, hdg

Note—Abbreviations for flight parameters are explained in Table 2. MSL, mean sea level.

cockpit of the actual aircraft, including all primary and secondary flight controls, aural cues (engine sounds), and cockpit lighting. It faithfully simulates the aircraft to the extent that training in the WST is directly transferable in terms of instrument flights, takeoffs and landings, instrument navigation, system operations, and air-to-ground attack procedures. It is typically used as a training tool and as a method for sustaining pilot proficiency in the F-117.

In the present study, only the instrument-flight simulation capability of the WST was utilized. All flights were set up for night illumination conditions with zero visibility and no visible lighting on the horizon. In addition, the WST was set up to generate zero air turbulence with no wind gusts in order to prevent flight-path deviations unrelated to pilot performance. The autothrottle and autopilot modes were disengaged to force all participants to “hand fly” the simulator.

Flight performance skill was assessed using the Coherent Automated Simulation Test Environment (CoASTE) software, which sampled flight parameters at a rate of 2 Hz throughout each of the flights. The specific parameters analyzed for each of the maneuvers are indicated in Table 1 and explained in Table 2. For every maneuver, root mean square (RMS) errors were calculated as a measure of control variability for each of the considered flight parameters. The RMS errors were computed over the deviations from the assigned flight paths (evaluated at a rate of 2 Hz), thus expressing how much on average each parameter deviated from the target value for the maneuver. To minimize inflation of the RMS errors due to improper maneuver setup rather than increased control variability, all the participants were required to be within 2° of the target heading, 5 knots of the target airspeed, and 20 ft of the target altitude before each maneuver was started.

There were two right 360° turns and five straight and level segments. Previous analyses (Caldwell et al., 2003) had shown that there were no statistically significant differences among these repeated maneuvers. Therefore, the two right turns were combined with each other by averaging the RMS errors within individuals. Likewise, the five straight and level segments were combined by averaging the RMS errors within individuals. In the end, for these two maneuver types as well as the others, there were three RMS error scores (one for each of the three different parameters considered) for every test session (i.e., at 23:00, 04:00, 09:00, 14:00, and 19:00).

#### Other Measurements

Various nonflight measurements were taken (Caldwell et al., 2004), including electroencephalography, electrooculography, electrocardiography, an oculomotor test, and a number of cognitive performance tests and subjective evaluations. Only the FIT, the POMS, and a VAS of sleepiness are considered here.

For the FIT—a computerized fitness-for-duty test—the participants were required to peer into a device presenting visual stimuli (both moving and stationary). The device detected changes in pupil size of 0.05 mm or more and movements of the eye of 1° or more, in response to controlled flashes of light and moving light targets. Measures of saccadic velocity (SV), pupil diameter (PD), pupil contraction latency (CL), and pupil constriction amplitude (CA) were calculated. For 1 participant, no baseline data were available for the FIT.

Subjective evaluations of mood and sleepiness were made with the POMS and VAS, respectively. Only the vigor-activity (VA) and fatigue-inertia (FI) subscales of the POMS are considered here. For 1 participant, no POMS data were available for the second test session during the experimental protocol. The two data points that were therefore missing (i.e., one for the VA subscale and one for the FI subscale) did not need to be replaced for any of the data analyses.

The VAS measured subjective experiences as marked by the participants on several 100-mm lines, each of which was anchored at one end with the words *not at all* and at the other end with the word *extremely*. Centered under each line was an adjective; here, we report only on the VAS labeled *sleepy* (slpy).

#### Statistical Analyses

Statistical analyses were performed using the mixed-model ANOVA (Burton et al., 1998; Feldman, 1988; Singer, 1998; Van Dongen, 2005; Van Dongen, Olofsen, et al., 2004), which is a technique specifically suited for the analysis of systematic individual differences across time. Every outcome variable was subjected to a mixed-model ANOVA with measurement time as a fixed effect (five

**Table 2**  
**Measured Simulator Flight Parameters**

Parameter	Abbreviation	Range
Indicated altitude	alt	0–30,000 ft
Indicated airspeed	IAS	30–600 knots
Indicated vertical speed	IVS	0 ± 5,000 ft/min
Magnetic heading	hdg	0°–360°
Roll angle	roll	0° ± 90°
Pitch angle		0° ± 90°
Slip		0 ± 2 balls
Localizer/course deviation		0 ± 2 dots
Glideslope/course deviation		0 ± 2 dots

Note—Abbreviations are given only for flight parameters analyzed in the present study.

levels), participants as a random effect ( $n = 10$ ), and a compound symmetry correlation structure (Burton et al., 1998; Van Dongen, Olofsen, et al., 2004) as necessary for variance components analysis (see below). This mixed-model ANOVA was performed using SAS Release 8.02 (SAS Institute, Cary, NC) to assess the overall effect of time using an  $F(4,36)$  test. In order to be certain that outcome variables showed a dynamic response to sleep deprivation, only variables for which the effect of time was statistically significant were analyzed further. (Recall that the primary reason for the present analyses was not to document the impact of sleep loss on performance, but rather to explore the extent of individual differences in the vulnerability to performance impairment from sleep loss in a group of highly trained professionals.)

A variance components analysis was done by repeating the mixed-model ANOVA and estimating the between-subjects variance  $\omega^2$  (systematic interindividual variability) and the within-subjects variance  $\sigma^2$  (residual intraindividual variability), assuming independence and normality for both. For this variance components analysis, performed using NONMEM Version V Level 1.1 (GloboMax LLC, Hanover, MD), the Laplace method for maximum likelihood estimation was employed. Statistical significance of  $\omega^2$  was assessed by means of a likelihood ratio  $\chi^2(1)$  test. Empirical best linear unbiased predictors (EBLUPs) for the random effect in the mixed-model ANOVA were plotted to assess the individual participants' overall standings relative to each other.

From the estimated variance components  $\omega^2$  and  $\sigma^2$ , the intraclass correlation coefficient (ICC; Cox & Solomon, 2004; Rao, 1997; Van Dongen, Maislin, & Dinges, 2004) was computed as a measure of systematic individual variability, expressed as a percentage [ $ICC = 100\% \times \omega^2 / (\omega^2 + \sigma^2)$ ]. The 95% confidence interval (CI) for the ICC was assessed on the basis of a likelihood ratio  $\chi^2(1)$  test. Note that the likelihood-based test of statistical significance for the ICC is equivalent to that for the between-subjects variance  $\omega^2$ .

The variance components analysis was repeated with the baseline measurement as a covariate, to make it possible to distinguish systematic individual differences in the response to sleep deprivation from individual differences already present at baseline (e.g., due to aptitude differences). The between-subjects variance  $\omega^2$  (systematic interindividual variability) and its statistical significance, the within-subjects variance  $\sigma^2$ , the ICC and its 95% CI, and the EBLUPs were reassessed. In addition, the statistical significance of the baseline covariate was investigated with an  $F(1,36)$  test.

In order to compare the magnitude of systematic individual differences in the present sample with that observed in an earlier sleep deprivation study (Van Dongen, Baynard, et al., 2004) of a sample of 21 healthy young adults from the general population (around Philadelphia, PA), we considered the VAS of sleepiness. This measure had also been used in the earlier study, but in the form of a nine-point integer scale rather than the 100-mm line used in the present study. To allow for an exploratory comparison of the two samples, the VAS data from the present study were converted to the same nine-point scale. The analyses described above were repeated, both with and without baseline covariate, to estimate the between-subjects variance and the ICC for the converted data. The results were compared with those of the earlier study (Van Dongen, Baynard, et al., 2004), for which the VAS results are reported here for the first time.

## RESULTS

### Temporal Profiles of Responses to Sleep Loss

For each of the variables considered in this study, Tables 3 (simulator flight parameters) and 4 (other measurements) show the overall effects of time ( $F$  statistics and  $p$  values). A number of variables did not show a statistically significant effect of time, and therefore it was not

**Table 3**  
Effect of Time in Mixed-Model ANOVA  
for Simulator Flight Parameters

Maneuver	Parameter	$F$	$p^*$
S&L	alt	1.96	.122
	IAS	0.71	.590
	hdg	8.70	<.001
R360	alt	1.90	.132
	IAS	1.31	.285
	roll	3.43	.018
L360	alt	3.41	.018
	IAS	3.81	.011
	roll	2.69	.046
L720	alt	7.58	<.001
	IAS	4.06	.008
	roll	8.94	<.001
DESC	IAS	2.26	.082
	IVS	1.26	.304
	hdg	0.93	.457
CLMB	IAS	3.13	.026
	IVS	1.31	.285
	hdg	2.39	.069
DRT	IAS	3.08	.028
	IVS	6.08	.001
	roll	2.11	.100
CLT	IAS	3.39	.019
	IVS	3.42	.018
	roll	3.70	.013

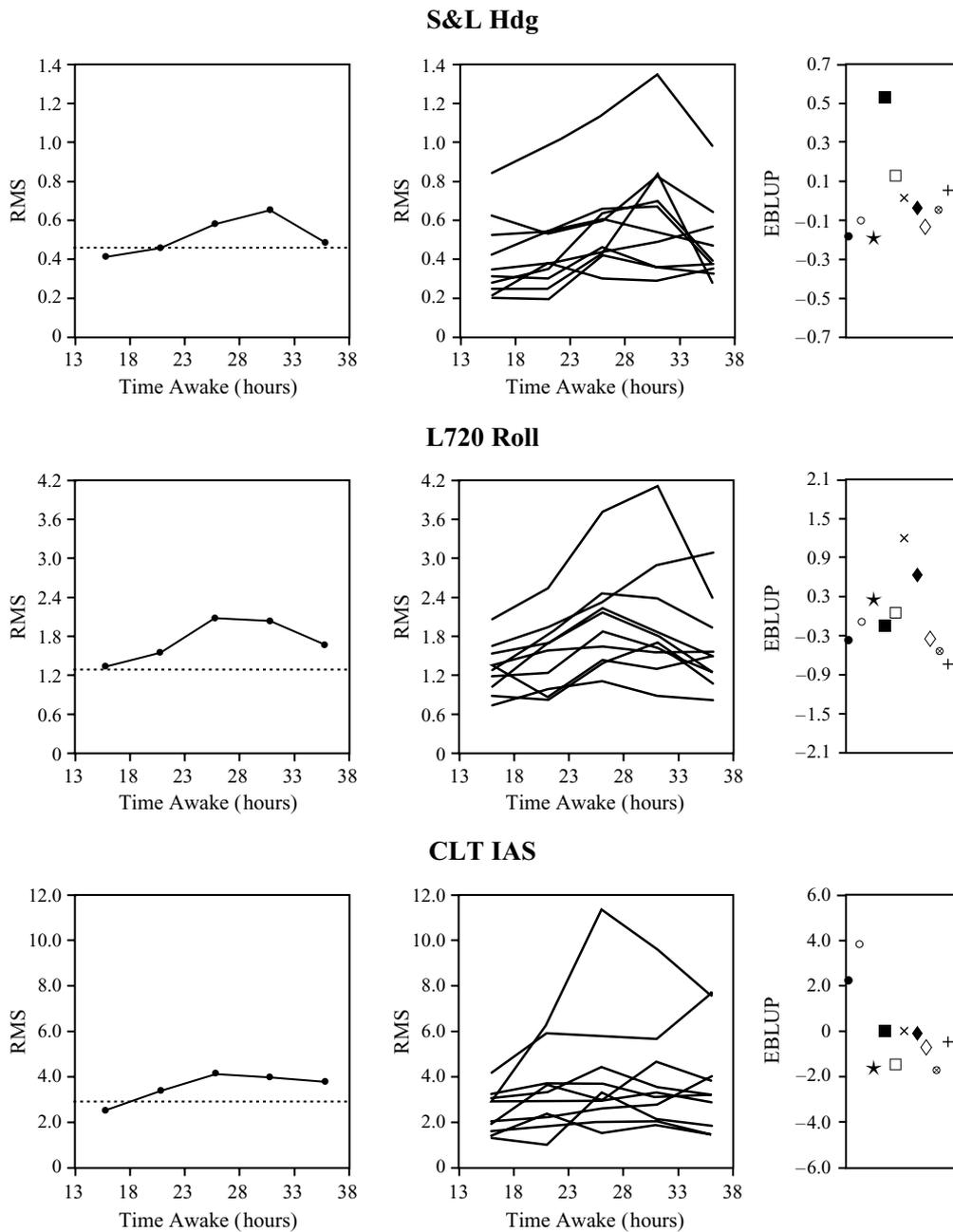
Note—See Tables 1 and 2 for abbreviations. \*Only variables with  $p < .05$  (boldface) were subjected to further analyses.

clear if these variables were dynamically responsive to sleep deprivation. Since the purpose of this study was to investigate individual differences in responses to sleep deprivation, only the 19 variables that did exhibit a significant effect of time were subjected to further analyses. For 6 of these variables, Figures 1 and 2 illustrate the temporal profiles of responses to sleep loss for the overall sample (left panels) as well as for the individual participants (center panels).

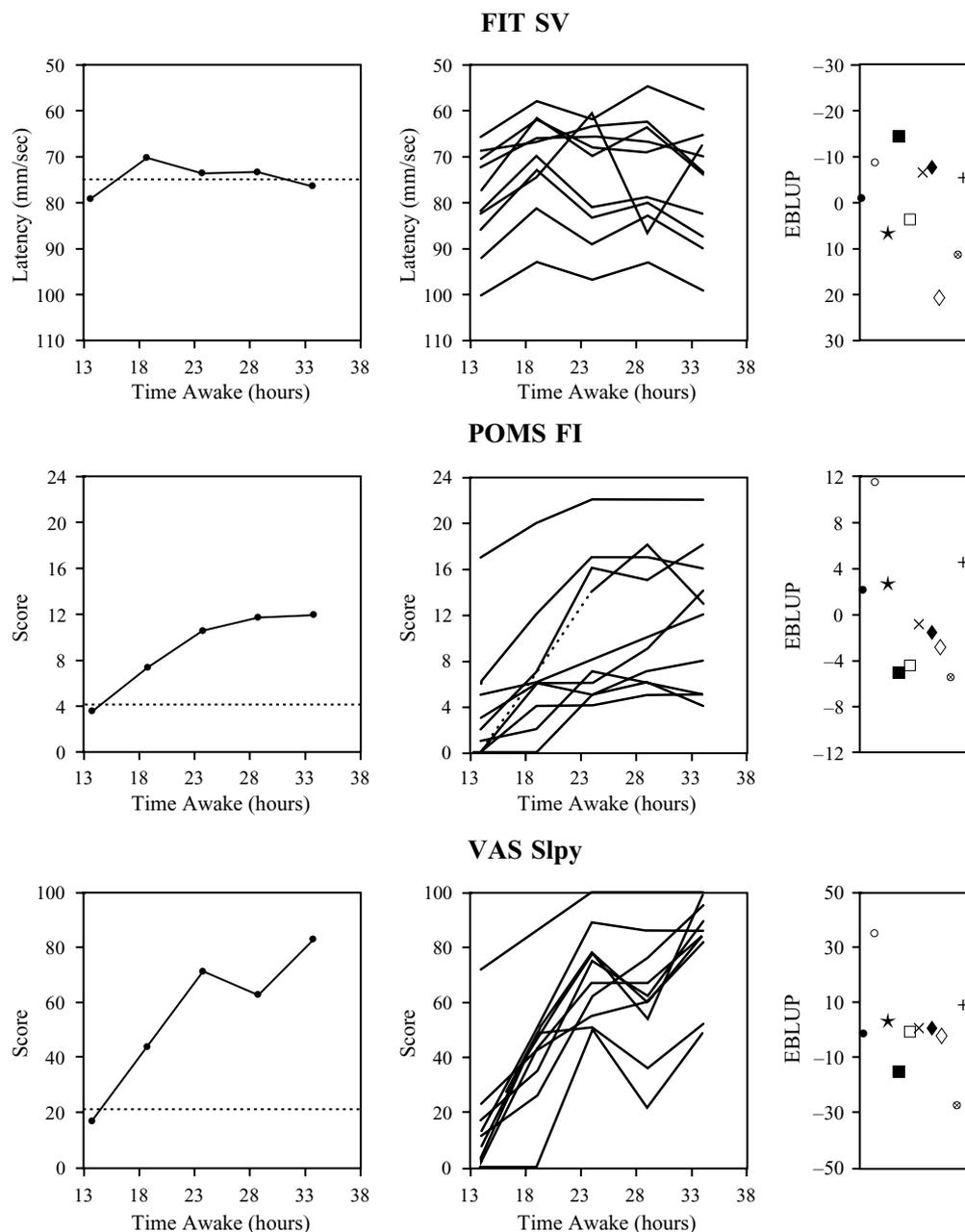
### Individual Variability in Responses to Sleep Loss

For the variables showing a significant dynamic response to sleep loss, Table 5 shows the estimated between-subjects variance  $\omega^2$  (interindividual variability) and within-subjects variance  $\sigma^2$  (residual intraindividual variability) with standard errors ( $SEs$ ), the statistical significance of the between-subjects variance  $\omega^2$  (likelihood ratio  $\chi^2$  statistic and  $p$  value), and the estimated ICC (as a measure of systematic individual variability) with 95% CI. The between-subjects variance was statistically significant for every variable considered, indicating that the participants differed systematically over time for all of these responses to sleep loss. This result held true even after we corrected for multiple comparisons by means of the false discovery rate procedure (Curran-Everett, 2000).

The ICC expresses the proportion of variance in the data that is explained by systematic individual variability. Theoretically, it can range from 0% (no systematic individual differences) to 100% (perfectly stable individual differences). The ICC values resulting from the experi-



**Figure 1.** Profiles of simulator flight performance during sleep deprivation. The left panels show group averages for each of five measurement times at 5-h intervals (connected dots) plotted across time awake (in hours). The left panels also show the group average at baseline (dotted line). The center panels show raw data (curves) plotted across time awake for each of the 10 participants. The right panels show the overall responses to sleep deprivation across the five measurement times for each of the participants (marked by different symbols) relative to the corresponding group averages. These are the participants' empirical best linear unbiased predictors (EBLUPs) associated with the between-subjects variance. Each of the symbols (offset horizontally for clarity) reflects the same individual in all three right panels. The top panels show root mean square (RMS) error data for the magnetic heading (hdg) parameter of the straight and level (S&L) maneuver. The middle panels show RMS error data for the roll angle (roll) parameter of the 720° left turn (L720). The bottom panels show RMS error data for the indicated airspeed (IAS) parameter of the climbing left turn (CLT). Upward corresponds to greater impairment in every panel.



**Figure 2.** Profiles of nonflight measurements during sleep deprivation. The top panels show saccadic velocity (SV) data (in mm/sec) for the Fitness Impairment Tester (FIT). The ordinates of these panels are reversed, so that upward corresponds to greater impairment in every panel. The middle panels show fatigue-inertia (FI) subscale scores for the Profile of Mood States (POMS). The dashed line in the center panel interpolates over one missing data point for 1 individual. The bottom panels show sleepiness (sply) scores (on a 100-mm line) for the visual analog scale (VAS). Other details are the same as in Figure 1. The nonflight measurements were taken over the 2-h period preceding each simulator flight.

ment ranged from 35.6% to 84.9% (35.6%–81.3% for the simulator flight parameters and 60.8%–84.9% for the other measurements). According to published benchmarks (Landis & Koch, 1977), the ICC values were “fair” (20%–40%) for three variables, “moderate” (40%–60%) for four variables, “substantial” (60%–80%) for nine variables, and “almost perfect” (80%–100%) for the remain-

ing three variables. Even though the 95% CIs for the ICC were relatively wide (owing to the small sample size), these results indicate that, overall, individual differences in the responses to sleep deprivation were considerable.

Figures 1 and 2 illustrate the systematic individual differences among pilots by showing the EBLUPs associated with the between-subjects variance (right panels). These

**Table 4**  
**Effect of Time in Mixed-Model ANOVA**  
**for Other Measurements**

Measurement	Parameter	<i>F</i>	<i>p</i> <sup>†</sup>
FIT	SV	5.64	<b>.001</b>
	PD	0.92	.463
	CL	2.95	<b>.033</b>
	CA	0.44	.779
POMS*	VA	14.16	<b>&lt;.001</b>
	FI	20.65	<b>&lt;.001</b>
VAS	slpy	52.49	<b>&lt;.001</b>

Note—See text for explanation of abbreviations. \*These variables had one missing data point, so that an *F*(4,35) test was performed instead of an *F*(4,36) test. †Only variables with *p* < .05 (boldface) were subjected to further analyses.

represent the pilots' overall impairment levels relative to each other, as derived by the separation of the between-subjects variance (i.e., systematic interindividual variability) from the within-subjects variance (i.e., residual intraindividual variability) in the temporal profiles of responses to sleep loss (corresponding center panels).

**Individual Variability Controlling for Baseline**

The analyses were repeated with the baseline measurement as a covariate. The baseline covariate was statistically significant (*p* < .05) for 12 of the 19 variables under consideration. Baseline differences affected the overall response to sleep deprivation for at least one parameter of every simulator flight maneuver and for all other measurements except the VAS. Nonetheless, systematic individual differences in the magnitude of impairment from sleep deprivation prevailed. Table 6 shows the reestimated values of  $\omega^2$  and  $\sigma^2$  (with *SEs*), statistical significance of

$\omega^2$  (likelihood ratio  $\chi^2$  statistic and *p* value), and value of the ICC (with 95% CI) for each of the variables that showed a significant response to sleep deprivation over time (note that the effect of time was not altered by the baseline covariate).

When baseline differences were taken into account, the between-subjects variance  $\omega^2$  was systematically reduced for all variables—in some cases by an order of magnitude. The between-subjects variance (and thus the ICC) was no longer statistically significant for the roll angle (roll) of the right 360° turn (R360), the indicated airspeed (IAS) of the left 360° turn (L360), and the IAS of the left 720° turn (L720). However, the between-subjects variance remained statistically significant for 16 of the 19 variables considered, even after we corrected for multiple comparisons by means of the false discovery rate procedure (Curran-Everett, 2000). The statistically significant baseline-corrected ICC values ranged from 25.4% to 64.3% (25.4%–63.9% for the simulator flight parameters and 36.5%–64.3% for the other measurements). These ICC values were “fair” (20%–40%) for 6 variables, “moderate” (40%–60%) for 7 variables, and “substantial” (60%–80%) for 3 variables.

In order to put the magnitude of the systematic individual variability in each of the responses to sleep loss into perspective, the between-subjects *SD* (i.e., the square-root of the between-subjects variance  $\omega^2$ ) was expressed as a percentage of the range (maximum minus minimum) of the group average response to sleep deprivation over time. Table 7 shows the results for the 16 variables that exhibited statistically significant between-subjects variance after controlling for baseline. The percentages varied around 50% (range, 21%–91%), indicating that the systematic individual variability of impairment during sleep

**Table 5**  
**Results of Variance Components Analysis**

Variable	Between-Subjects Variance				Within-Subjects Variance		ICC (%)	
	$\omega^2$	<i>SE</i>	$\chi^2$	<i>p</i>	$\sigma^2$	<i>SE</i>	Estimate	95% CI
S&L hdg	0.0427	0.0280	52.69	<.001	0.0098	0.0021	81.3	62.2–93.1
R360 roll	0.120	0.045	26.81	<.001	0.072	0.018	62.6	35.9–84.5
L360 alt	278	110	34.17	<.001	122	65	69.5	44.5–87.9
L360 IAS	0.544	0.230	8.76	.003	0.982	0.255	35.6	9.5–67.4
L360 roll	0.256	0.087	23.58	<.001	0.178	0.039	59.0	31.7–82.6
L720 alt	309	81	37.24	<.001	121	40	71.9	47.9–89.0
L720 IAS	0.77	0.22	12.76	<.001	1.00	0.39	43.6	16.2–73.1
L720 roll	0.314	0.143	42.75	<.001	0.100	0.033	75.8	53.4–90.7
CLMB IAS	0.79	0.32	10.05	.002	1.26	0.46	38.4	11.7–69.4
DRT IAS	2.24	1.53	34.03	<.001	0.99	0.25	69.3	44.4–87.8
DRT IVS	2.52	0.86	16.69	<.001	2.53	0.98	49.9	22.2–77.3
CLT IAS	2.92	1.40	38.25	<.001	1.10	0.61	72.6	48.9–89.3
CLT IVS	2.71	0.88	24.11	<.001	1.83	0.85	59.7	32.4–82.9
CLT roll	0.086	0.059	8.79	.003	0.154	0.049	35.7	9.6–67.4
FIT SV	106	43	60.84	<.001	19	10	84.9	68.4–94.5
FIT CL	149	50	25.19	<.001	96	26	60.8	33.8–83.5
POMS VA	15.8	6.8	35.29	<.001	6.4	1.3	71.0	46.5–88.6
POMS FI	25.8	5.4	54.69	<.001	5.4	1.4	82.7	64.4–93.7
VAS slpy	247	142	32.78	<.001	115	20	68.2	43.0–87.3

Note—The maneuver/measurement and parameter columns in Tables 3 and 4 are combined here to yield a single abbreviation for each variable. CI, confidence interval.

**Table 6**  
**Results of Variance Components Analysis With Baseline as a Covariate**

Variable	Between-Subjects Variance				Within-Subjects Variance		ICC (%)	
	$\omega^2$	<i>SE</i>	$\chi^2$	<i>p</i>	$\sigma^2$	<i>SE</i>	Estimate	95% CI
S&L hdg	0.00862	0.00431	14.71	<.001	0.00979	0.00208	46.8	19.2–75.3
R360 roll	0.0139	0.0093	2.10	.148	0.0716	0.0175	16.3	0.0–49.6
L360 alt	180	77	24.06	<.001	122	65	59.6	32.4–82.9
L360 IAS	0.178	0.197	1.87	.171	0.982	0.255	15.3	0.0–48.6
L360 roll	0.061	0.064	4.71	.030	0.178	0.039	25.4	1.9–58.7
L720 alt	85	54	11.63	.001	121	40	41.4	14.4–71.7
L720 IAS	0.23	0.14	2.59	.108	1.00	0.39	18.4	0.0–51.8
L720 roll	0.137	0.068	22.49	<.001	0.100	0.033	57.8	30.3–81.8
CLMB IAS	0.77	0.31	9.76	.002	1.26	0.46	37.8	11.3–69.0
DRT IAS	1.75	1.28	28.02	<.001	0.99	0.25	63.9	37.4–85.1
DRT IVS	1.63	0.95	10.48	.001	2.53	0.98	39.2	12.5–70.1
CLT IAS	1.53	0.59	22.92	<.001	1.10	0.61	58.2	30.9–82.1
CLT IVS	1.08	0.66	9.44	.002	1.83	0.85	37.1	10.7–68.5
CLT roll	0.085	0.059	8.77	.003	0.154	0.049	35.7	9.5–67.4
FIT SV	11.4	5.5	8.23	.004	19.8	10.7	36.5	9.04–69.6
FIT CL	57.2	21.6	12.23	<.001	70.1	17.2	44.9	16.2–75.4
POMS VA	11.6	4.4	27.72	<.001	6.4	1.3	64.3	37.7–85.3
POMS FI	5.27	2.00	15.97	<.001	5.39	1.38	49.4	21.5–77.0
VAS slpy	183	70	25.79	<.001	115	20	61.4	34.6–83.9

Note—CI, confidence interval.

deprivation, controlling for baseline differences, was almost of the same order of magnitude as the group average level of impairment during sleep deprivation.

### Comparison With Healthy Nonpilots

On the basis of the VAS of sleepiness, systematic individual differences in the present sample of pilots were compared with trait individual differences in a sleep-deprived sample of nonpilots studied by Van Dongen, Baynard, et al. (2004). In the present study, after converting the VAS to the nine-point integer scale used in the earlier study, the between-subjects variance  $\omega^2$  was  $1.68 \pm 0.95$  (estimate  $\pm SE$ ) and the ICC was 60%. After controlling for baseline,  $\omega^2$  was  $1.34 \pm 0.53$  and the ICC was 54%. In the earlier study,  $\omega^2$  was  $2.15 \pm 0.87$  and the ICC was 69%. After controlling for baseline,  $\omega^2$  was  $1.03 \pm 0.47$  and the ICC was 60%. The estimated between-subjects variances (and their *SEs*) are illustrated in Figure 3. The similarity of the results of the two studies suggests that systematic individual differences in subjective sleepiness due to sleep deprivation were not very different for the pilots and for healthy young adults from the general population.

## DISCUSSION

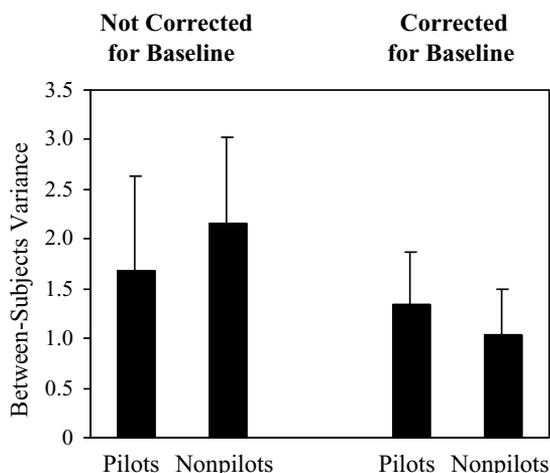
This study demonstrated systematic individual differences in the impact of 38 h of continuous wakefulness on simulator flight performance and subjective states for a sample of fully qualified, active-duty U.S. Air Force fighter pilots. Our goal was not to establish the effects of sleep deprivation on simulator flight performance per se, since this has already been done (Caldwell et al., 2004). Rather, for outcome variables showing dynamic responsiveness to sleep loss across five measurement times during

the last 24 h of a 38-h sleep deprivation period, we focused on quantifying persistent differences among the pilots in terms of their responses to sleep deprivation. This was done using the ICC, which is a valid measure of the degree of systematic individual variability (Cox & Solomon, 2004; Rao, 1997; Van Dongen, Maislin, & Dinges, 2004). The ICC varied over outcome variables, reaching values as high as 64.3% after correcting for baseline measurements. Thus, depending on the variable at hand, up to approximately two

**Table 7**  
**Individual Variability Versus Group Average Response to Sleep Deprivation**

Variable	Between-Subjects <i>SD</i> *	Range of Group Average Response†	Ratio (%)
S&L hdg	0.093	0.240	39
L360 alt	13	17	80
L360 roll	0.25	0.59	42
L720 alt	9.2	25.0	37
L720 roll	0.37	0.73	51
CLMB IAS	0.88	1.58	55
DRT IAS	1.3	1.4	91
DRT IVS	1.3	3.2	40
CLT IAS	1.2	1.6	76
CLT IVS	1.0	2.2	47
CLT roll	0.29	0.56	52
FIT SV	3.4	9.1	37
FIT CL	7.6	11.2	68
POMS VA	3.4	7.4	46
POMS FI	2.3	8.3	28
VAS slpy	14	66	21

Note—See the text for explanation of abbreviations. \*Square root of the systematic between-subjects variance  $\omega^2$  during sleep deprivation, controlling for baseline (see Table 6). †Maximum minus minimum of the group average response as observed across the period of sleep deprivation.



**Figure 3.** Systematic individual differences in subjective sleepiness during sleep deprivation in two different samples. The graph shows the estimated between-subjects variance (and standard error) for a visual analog scale of sleepiness (employing a nine-point scale), both before and after controlling for baseline individual differences. The present sample of pilots and a sample from the general population studied earlier (Van Dongen, Baynard, et al., 2004) are compared.

thirds of the variance was explained by systematic individual differences in impairment levels. The magnitude of the systematic individual differences after baseline correction was considerable, hovering around 50% of the full range of the group average changes from the beginning of the night through the end of the next day of sleep deprivation.

In theory, these individual differences could have resulted from enduring state differences among the study participants. However, important state factors were experimentally controlled, including caffeine intake in the hours before and during the sleep deprivation experiment, light exposure during measurements, physical activity and posture during measurements (the simulator environment allowed little variation in these factors), practice effects (the study participants were fully qualified pilots, and a training day preceded the experiment), and demand characteristics (the participants were instructed and trained before the experiment, sequenced through the flight maneuvers, and supervised by an experimenter). It is therefore unlikely that the observed individual differences could be explained entirely by enduring state differences.

The participants were instructed to stay in bed for a period of 8 h the night before the sleep deprivation experiment, and they slept on average 7.4 h as estimated with actigraphy. This amount of sleep would restore the bulk of performance capability after a period of chronic sleep restriction (Belenky et al., 2003). Yet there were small differences in the duration of sleep during the night before the experiment (the *SD* for the actigraphy data over individuals was 0.6 h). Even differences as large as 4.1 h daily over 7 days, however, have been shown to exert a relatively small effect on the magnitude of neurobehavioral impairment during subsequent sleep deprivation (Van Dongen,

Baynard, et al., 2004). Thus, any prior chronic sleep loss should not have affected the results of the present study significantly.

Visual inspection of the data (as illustrated in Figures 1 and 2) brought to light the observation that in various cases 1 pilot exhibited particularly severe performance impairment during the night of sleep deprivation. This might be seen as an indication that the individual gave up performing to his maximum ability. However, the participant with the greatest impairment was a different person, depending on the outcome variable (cf. Figure 1, right panels). In each instance, this person's performance profile showed the typical circadian rebound (Johnson et al., 1992; Van Dongen & Dinges, 2000) when wakefulness was maintained into the subsequent day. Considering also that all simulator flights were monitored by an experimenter, it would appear that the cases of severe impairment did not inevitably reflect a lack of effort by the respective study participants, but rather involved true vulnerability to sleep deprivation in specific aspects of performance.

There were substantial baseline differences among the pilots in several simulator flight parameters and in self-reported fatigue. These baseline differences could have resulted from differences in prior state, since baseline measurements were taken on the training day, which was not preceded by experimental control of sleep history, caffeine intake, or other factors potentially affecting task performance and subjective evaluations. In addition, the baseline differences in simulator flight performance may have resulted from differences in flight experience or may have reflected differences in aptitude and baseline cognitive ability (see, e.g., Kane & Engle, 2002). For most of the simulator flight parameters and for self-reported sleepiness and mood, controlling for baseline individual differences did not eliminate the systematic individual differences during sleep deprivation. Since the baseline period always preceded the sleep deprivation period, however, it is possible that an order effect confounded the baseline-corrected results. Thus, the results in Table 6 should be considered preliminary.

Whether baseline individual differences were controlled for or not, at least part of the systematic individual variability in simulator flight performance and subjective states during sleep deprivation appear to be due to idiosyncratic, trait-like characteristics of the individuals who participated in our study. This is in agreement with preliminary results from a recent Boeing 747-400 simulator study of 6 commercial airline pilots flying at night (Mallis et al., 2002), and in particular with results from recent laboratory studies (Leproult et al., 2003; Van Dongen, Baynard, et al., 2004) showing that individuals vary systematically in the magnitude of neurobehavioral impairment from sleep loss. The study by Van Dongen, Baynard, et al., which involved healthy young adults sampled from the general population, demonstrated that this variability constitutes a trait, which was referred to as *differential vulnerability*.

Comparison of the results from the study by Van Dongen, Baynard, et al. (2004) with the findings from the present study with regard to a common outcome variable—a VAS

of sleepiness—suggested that the magnitudes of differential vulnerability to subjective sleepiness caused by sleep deprivation were similar for the highly trained U.S. Air Force fighter pilots and the healthy young adults from the general population (see Figure 3). However, in order to compare the magnitudes of individual variability between different populations quantitatively (i.e., statistically), further research involving direct comparison in a single experiment with objective measures of performance impairment is needed.

A practical implication of the present findings is that significant variations in fatigue vulnerability remained among the fighter pilots despite the demands of their profession. Thus, managers and schedulers need to understand that optimal mission assignments and work/rest schedules cannot be based on traditional, grouped research data, because such averages will paint an overly optimistic picture for some personnel (indicating that they can fly longer missions or sleep less than they actually can) and an overly pessimistic picture for others (resulting in shorter-than-necessary missions for pilots who are capable of flying longer). Similar difficulties would be present in other occupations in which long duty periods and/or shortened sleep episodes threaten to impair operator performance.

This issue could be quite complex if it turns out that individuals' impairment levels relative to each other may vary from one performance measure to another, as was suggested by the laboratory observations of Van Dongen, Baynard, et al. (2004). The ranking of individuals in terms of the magnitude of their performance impairment varied across outcome variables in the present study as well (cf. Figures 1 and 2, right panels), and a structured analysis of this matter is currently being undertaken.

In conclusion, the sleep deprivation study on which we have reported here involved simulated real-world operational demands in the form of flight performance on a high-fidelity F-117 flight simulator. Even though the sample size of the study was relatively small ( $n = 10$ ), there were multiple statistically significant (i.e., reliable) findings of systematic individual variability in flight performance impairment and other indicators of fatigue among fighter pilots. No evidence was found that self-selection factors had eliminated the differential vulnerability to sleep deprivation for these active-duty pilots. Thus, military commanders should be advised that individual differences may play a significant role in flight performance during sustained operations. Similar considerations apply in other operational environments, such as shift work, in which sleep loss is common and differential vulnerability may be considerable despite possible self-selection mechanisms. The results of this study made clear that the scientific and operational communities' reliance on group data as opposed to individual data can undermine the accurate characterization of factors dramatically affecting job-related safety and performance.

## REFERENCES

- BELENKY, G., WESENSTEN, N. J., THORNE, D. R., THOMAS, M. L., SING, H. C., REDMOND, D. P., ET AL. (2003). Patterns of performance degradation and restoration during sleep restriction and subsequent recovery: A sleep dose-response study. *Journal of Sleep Research*, **12**, 1-12.
- BURTON, P., GURRIN, L., & SLY, P. (1998). Extending the simple linear regression model to account for correlated responses: An introduction to generalized estimating equations and multi-level mixed modeling. *Statistics in Medicine*, **17**, 1261-1291.
- CALDWELL, J. A. (2005). Fatigue in aviation. *Journal of Travel Medicine & Infectious Disease*, **3**, 85-96.
- CALDWELL, J. A., CALDWELL, J. L., BROWN, D. L., & SMITH, J. K. (2004). The effects of 37 hours of continuous wakefulness on the physiological arousal, cognitive performance, self-reported mood, and simulator flight performance of F-117A pilots. *Military Psychology*, **16**, 163-181.
- CALDWELL, J. A., CALDWELL, J. L., BROWN, D. L., SMYTHE, N. K., SMITH, J. K., MYLAR, J. T., ET AL. (2003). *The effects of 37 hours of continuous wakefulness on the physiological arousal, cognitive performance, self-reported mood, and simulator flight performance of F-117A pilots* (Rep. No. AFRL-HE-BR-TR-2003-0086). Brooks City-Base, TX: Air Force Research Laboratory.
- COX, D. R., & SOLOMON, P. J. (2004). *Components of variance*. Boca Raton: Chapman & Hall/CRC.
- CURRAN-EVERETT, D. (2000). Multiple comparisons: Philosophies and illustrations. *American Journal of Physiology: Regulatory, Integrative, & Comparative Physiology*, **279**, R1-R8.
- FELDMAN, H. A. (1988). Families of lines: Random effects in linear regression analysis. *Journal of Applied Physiology*, **64**, 1721-1732.
- GOLDSTEIN, H. (1995). *Multilevel statistical models*. London: Arnold.
- JOHNSON, M. P., DUFFY, J. F., DIJK, D.-J., RONDA, J. M., DYAL, C. M., & CZEISLER, C. A. (1992). Short-term memory, alertness and performance: A reappraisal of their relationship to body temperature. *Journal of Sleep Research*, **1**, 24-29.
- KANE, M. J., & ENGLE, R. W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychonomic Bulletin & Review*, **9**, 637-671.
- KHURI, A. I., MATHEW, T., & SINHA, B. K. (1998). *Statistical tests for mixed linear models*. New York: Wiley.
- KOSSLYN, S. M., CACIOPPO, J. T., DAVIDSON, R. J., HUGDAHL, K., LOVALLO, W. R., SPIEGEL, D., & ROSE, R. (2002). Bridging psychology and biology: The analysis of individuals in groups. *American Psychologist*, **57**, 341-351.
- LANDIS, J. R., & KOCH, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, **33**, 159-174.
- LEPROULT, R., COLECCHIA, E. F., BERARDI, A. M., STICKGOLD, R., KOSSLYN, S. M., & VAN CAUTER, E. (2003). Individual differences in subjective and objective alertness during sleep deprivation are stable and unrelated. *American Journal of Physiology: Regulatory, Integrative, & Comparative Physiology*, **284**, R280-R290.
- MALLIS, M. M., NERI, D. F., OYUNG, R., COLLETTI, L., NGUYEN, T., & DINGES, D. F. (2002). Stability of behavioral alertness in pilots repeating simulated night flights. *Sleep*, **25**(Suppl.), A443.
- McNAIR, D. M., LORR, M., & DROPPLEMAN, L. F. (1971). *Manual for the Profile of Mood States*. San Diego: Educational and Industrial Testing Service.
- PENETAR, D., McCANN, U., THORNE, D., KAMIMORI, G., GALINSKI, C., SING, H., ET AL. (1993). Caffeine reversal of sleep deprivation effects on alertness and mood. *Psychopharmacology*, **112**, 359-365.
- RAO, P. S. R. S. (1997). *Variance components estimation*. London: Chapman & Hall.
- SINGER, J. D. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational & Behavioral Statistics*, **24**, 323-355.
- VAN DONGEN, H. P. A. (2005). Analysis of inter- and intra-individual variability. *Journal of Sleep Research*, **14**, 201-203.

- VAN DONGEN, H. P. A., BAYNARD, M. D., MAISLIN, G., & DINGES, D. F. (2004). Systematic interindividual differences in neurobehavioral impairment from sleep loss: Evidence of trait-like differential vulnerability. *Sleep*, **27**, 423-433.
- VAN DONGEN, H. P. A., & DINGES, D. F. (2000). Circadian rhythms in fatigue, alertness, and performance. In M. H. Kryger, T. Roth, & W. C. Dement (Eds.), *Principles and practice of sleep medicine* (3rd ed., pp. 391-399). Philadelphia: Saunders.
- VAN DONGEN, H. P. A., MAISLIN, G., & DINGES, D. F. (2004). Dealing with inter-individual differences in the temporal dynamics of fatigue and performance: Importance and techniques. *Aviation, Space, & Environmental Medicine*, **75**, A147-A154.
- VAN DONGEN, H. P. A., MAISLIN, G., MULLINGTON, J. M., & DINGES, D. F. (2003). 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.
- VAN DONGEN, H. P. A., OLOFSEN, E., DINGES, D. F., & MAISLIN, G. (2004). Mixed-model regression analysis and dealing with interindividual differences. *Methods of Enzymology*, **384**, 139-171.

(Manuscript received March 18, 2005;  
accepted for publication April 27, 2005.)