Sleep at Simulated 2438 m: Effects on Oxygenation, Sleep Quality, and Postsleep Performance

J. Michael Muhm, T. Leigh Signal, Paul B. Rock, Stephen P. Jones, Karyn M. O’Keeffe, Matthew R. Weaver, Shuying Zhu, Philippa H. Gander, and Greg Belenky

Two flight crews and two cabin crews were used on ultra long-range commercial airplane flights. While one is on duty, the other has the opportunity to sleep in onboard areas in which the barometric pressure can be as low as 75.3 kPa (565 mmHg), equivalent to a terrestrial altitude of 2438 m (8000 ft). In-flight studies of crewmembers conducted during ultra-long-range flights have documented sleep disturbances which have been attributed to various causes such as noise, turbulence, temperature, humidity, inadequate bedding, or the presence of other individuals in the sleep area (15, 16).

Unacclimatized travelers to high terrestrial altitudes experience sleep-related respiratory disturbances, hypoxemia, impaired quantity and quality of sleep, and impaired postsleep performance (19, 20). The altitude and SpO2 thresholds for the onset of altitude-related sleep disturbances, if thresholds exist, are not readily ascertained from the current literature. It is possible that the hypoxemia occurring during sleep at 2438 m also contributes to the sleep disturbances observed in flight-crew members.

We conducted a study in a hypobaric chamber to evaluate the effect of sleep in barometric pressures equivalent to altitudes of 2438 m and 305 m (1000 ft) on blood oxygenation and sleep quantity and quality. Because sleep disturbances, regardless of cause, adversely affect neurobehavioral performance and mood (18), we also examined those outcomes. If adverse effects were found at the barometric pressure equivalent to 2438 m, our intention was to determine if sleep in the barometric equivalent of 305 m would ameliorate those effects. We therefore included a condition in which pre- and postsleep altitudes were 2438 m and the sleep altitude was 305 m.

METHODS

The protocol was approved in advance by the Institutional Review Boards of Oklahoma State University Center for Health Sciences and The Boeing Company. Each subject provided written informed consent after being informed of the study’s purpose, procedures, time commitment, and risks and benefits of participation.

Male volunteers 30 - 60 yr of age were recruited from the general population of Tulsa, OK [altitude 198 m (650 ft)]. Potential subjects were required to meet FAA Class III physical standards (except for vision and hearing requirements) that exclude significant past or present cardiovascular, pulmonary, and sinus disease. Their body mass index was required to be 35 or less and their height 1.88 m (74 inches) or less. Subjects’ race and ethnicity were not considered.

From the The Boeing Company, Seattle, WA; Sleep/Wake Research Center, Research School of Public Health, Massey University, Wellington, New Zealand; Oklahoma State University, Tulsa, OK; and Washington State University, Spokane, WA.

Address reprint requests to: J. Michael Muhm, M.D., M.P.H., The Boeing Company, P.O. Box 3707, MC 7H-90, Seattle, WA 98124-2207; Mike.muhm@boeing.com.

Reprint & Copyright © by the Aerospace Medical Association, Alexandria, VA.

DOI: 10.3357/ASEM.2327.2009
All subjects completed a screening questionnaire investigating issues of general health, mood, and sleep, which included the Epworth Sleepiness Scale (9), the Pittsburgh Sleep Quality Index (3), the Horne and Ostberg Morning-Eveningness Scale (8), and the 21-question Depression, Anxiety, and Stress Scales (10). Those whose responses indicated the presence of a sleep disorder, excessive daytime sleepiness, poor quality sleep, or extreme circadian phase were excluded. In addition, individuals whose mood was assessed as possibly affecting sleep and those who reported using stimulants, including any nicotine or more than 3 caffeinated beverages per day (approximately 300 mg caffeine) were excluded.

Subjects were instructed to maintain stable sleep patterns during the week preceding each study session and to wear actiwatches (Mini Mitter, Bend, OR) and maintain sleep diaries to assess compliance. On the day of each study session, subjects were screened for acute disqualifying conditions: infectious disease, recent major surgical procedure, dental abscess, scuba diving within 24 h of the study, and drug or alcohol intoxication.

The study was conducted in a hypobaric chamber (C.G.S. Scientific Corporation) located in the Center for Aerospace and Hyperbaric Medicine at Oklahoma State University Center for Health Sciences in Tulsa, OK. The chamber, ventilated at no less than 10 cfm/occupant, was equipped with an airlock that allowed ingress and egress of personnel and supplies without affecting the barometric pressure in the main chamber. The barometric pressure used to simulate 305 m was 97.7 kPa (733 mmHg), and 75.3 kPa (565 mmHg) to simulate 2438 m. Temperature and relative humidity were measured once per minute during the 18 h of each study session. The main compartment of the chamber was fitted with four beds in which the subjects slept. The beds were arranged against the walls of the chamber, end-to-end and two on a side with an aisle between. Opaque curtains were pulled between the beds during the sleep period for privacy. Each subject was given soft ear plugs and sleep masks to wear during the sleep period. Dinner was provided 4 h before the beginning of the sleep period, breakfast at the beginning of the postsleep period. Non-caffeinated beverages were available throughout the study session. A sink and toilet were located in the airlock.

The study used a prospective, double-blind, crossover design in which each subject was scheduled to participate in three 18-h study sessions starting at 1700, each consisting of a 6-h presleep period, a 7-h sleep period, and a 5-h postsleep period. The study sessions differed by the combinations of altitudes simulated in the three periods. In condition GND, all three periods were conducted at a simulated altitude of 305 m; in condition ALT, all periods at a simulated altitude of 2438 m; and in condition MIX the presleep and postsleep pressures were conducted at a simulated altitude of 2438 m and the intervening sleep period at 305 m. The study design called for the order of exposure to the study conditions to be balanced across study subjects to minimize possible order effects. Study sessions for individual subjects were separated by no less than 3 wk to allow for resolution of any possible altitude acclimatization occurring during the preceding study session. To maintain subject blinding to different barometric pressures, the barometric pressure was transiently changed to 87.5 kPa [656 mmHg (1219 m, 4000 ft)] at the beginning and end of the study session and at the beginning and end of the sleep period in condition GND, and at the beginning and end of the sleep period in condition ALT. To maintain blinding of the research staff while they entered or exited the chamber, the airlock was decompressed to 87.5 kPa (1219 m) before being recompressed in conditions GND and MIX. No formal assessment of the success of the blinding was made.

Measurements of blood oxygenation included oxygen saturation measured by pulse oximetry (SpO₂), desaturation index (the number of times per hour a sleeping individual’s SpO₂ decreased ≥ 4% below the preceding baseline for ≥ 10 s), and percent of time in which SpO₂ was less than 90% during sleep. SpO₂ and heart rate (HR) were measured using fingertip probes and oximeters (Nonin PalmSat Model 2500). During both the pre- and postsleep periods SpO₂ and HR were recorded at a frequency of 0.25 Hz for approximately 2 min on three separate occasions at 1 h 45 min intervals. SpO₂ and HR were recorded continuously at a frequency of 0.25 Hz throughout the 7-h sleep period.

Electroencephalographic (EEG), extraocular (EOG), and electromyographic (EMG) recordings were obtained from each subject during the 7-h sleep period using an ambulatory recorder (A10, Embla) and the electrode montage: C4-A1, C3-A2, EOG R-A1, EOG L-A2, EMG. Signals were filtered (0.5 – 90 Hz), stored on a data card, and downloaded at the end of the protocol. Sleep, awakenings, and arousals from sleep were scored according to the standard criteria (1,14). All sleep recordings were viewed by two independent, experienced sleep scorers (intrarrater reliability > 90% for all recordings). A post-sleep questionnaire developed by Massey University that contained nine questions assessing perception of sleep quantity and quality was administered at the beginning of the postsleep period.

Neurobehavioral performance was assessed using the Psychomotor Vigilance Test (PVT) (6) (PVT-192 Ambulatory Monitoring, Ardsley, NY), and the Simple Reaction, Two Choice Reaction Time, Tower of Hanoi, Switching, and Continuous Performance tests of the Automated Neuropsychological Assessment Metrics (ANAM) test battery that assess vigilance, memory, and executive functioning (2). The Simple Reaction test was administered twice—one at the beginning and once at the end—during each ANAM test battery. Mood was assessed with the Profile of Mood States (POMS) (11). The PVT, ANAM battery, and POMS were administered at hours 1, 2, and 4 after the beginning of the pre- and post-sleep periods.

Statistical Analysis: We estimated that 19 male subjects were sufficient to provide 80% power to detect a 25 ms difference in PVT response times (corresponding to 0.71
standard deviations) or a 30-min difference in total sleep time (0.77 standard deviations) at the 0.05 confidence level. Our target of 24 subjects allowed for a dropout rate of up to 20%. Mixed linear models were used to determine if altitude affected SpO2, heart rate, sleep quantity and quality, the post-sleep minus presleep difference in neurobehavioral performance, and subjective assessments of mood and sleep (SAS for Windows v9.1). The independent variables in each model were: study condition (GND, ALT, MIX), order in which the condition was completed (1st, 2nd, 3rd), total sleep in the 24 h prior to the protocol (number of hours as measured by the actigraph), age (years), and the interaction of study condition and order. Because multiple trials of SpO2, heart rate, neurobehavioral performance, and mood were administered each session, the models of these outcomes also included trial number and the interaction of trial number and study condition as independent variables. Analyses of PVT data were conducted using both reaction time (RT) and speed (1/RT). For both measurements, the mean, median, standard deviation, 10th percentile, 90th percentile, and rate of change over the 10-min test session were analyzed. In addition, a Poisson regression model was used to analyze the number of lapses, defined as a reaction time greater than 500 ms. In this paper we primarily report the analysis of mean reaction time. The Wilcoxon signed-rank test was used to compare medians of nonparametrically distributed variables.

RESULTS

Eighty-five males were screened for the study after giving informed consent, 58 of whom were excluded due to one or more of the exclusion criteria. Twenty-seven were selected to participate, 3 of whom did not due to scheduling conflicts, resulting in 24 approved subjects. One removed himself for personal reasons and did not participate in any sessions. Data from 3 who participated were excluded: 1 who did not comply with the study protocol and participated in only 1 session, 2 who did not participate in any sessions. Data from 3 who participated were excluded: 1 who did not comply with the study protocol and participated in only 1 session, and 2 who had desaturation indices greater than 15 during sleep at 305 m, suggesting previously undetected respiratory disturbance, resulting in 20 subjects who completed a total of 57 subject-sessions (1 subject participating in 1 session is one subject-session). The polysomnographic recording during one subject-session could not be analyzed because of technical problems. Subjects’ ages were (mean ± std dev) 44.1 ± 8.8 yr; height, 70.1 ± 1.9 inches; weight, 181.8 ± 29.2 pounds; BMI, 26.1 ± 4.2 kg · m⁻²; ground level SpO2, 97.5 ± 1.3%; resting ground level heart rate, 71.1 ± 12.4 bpm.

Data was collected between 10/17/06 and 2/17/07. The range of average temperatures in the chamber over the 20 study sessions was 18.1°C – 23.7°C (64.6°F – 74.6°F), and the range of average relative humidity was 22.7–64.7%. The allocation of subjects to study condition was not completely balanced across order of exposure: 9 of 19 experienced condition ALT on the first night; 10 of 19 experienced MIX on the second night; and 9 of 20 experienced GND on the third night.

Mixed model analysis showed average SpO2 to be significantly related to condition, trial number, and the interaction between condition and trial number (Table I). The desaturation index at GND [1.9 (0.4 – 8.9)] [median (range)] differed significantly from that at ALT [14.4 (7.9 – 51.2)] (P < 0.001), but not from that at MIX [2.3 (0.7 – 10.7)] (P = 0.07) (Fig. 1). The percent of time during sleep in which SpO2 was below 90% at GND [0.0 (0.0 – 0.1)] differed significantly from that at ALT [44.4% (3.6–86.9%)] (P < 0.001) and at MIX [0.1% (0.0–22.9%)] (P < 0.001) (Fig. 2). Visual inspection of SpO2 data indicated that baseline SpO2 remained relatively stable across the sleep period in the 2438-m condition after an initial drop following sleep onset.

Heart rate was not significantly related to condition but was significantly related to age, trial number, and the interaction of condition and trial number (Table I). Heart rate decreased as age increased.

Measurements of sleep quantity (total sleep time and wake time), and sleep quality (sleep efficiency, sleep latency, awakenings/hour, arousals/hour, and sleep architecture) were not significantly related to condition. Some were significantly related to order and age—

<table>
<thead>
<tr>
<th>Outcome</th>
<th>GND (305/305/305) (N = 20)</th>
<th>ALT (2438/2438/2438) (N = 19)</th>
<th>MIX (2438/305/2438) (N = 19)</th>
<th>Condition¹</th>
<th>Order²</th>
<th>Prior Sleep³</th>
<th>Age⁴</th>
<th>Trial Number⁵</th>
<th>Condition × Order⁶</th>
<th>Condition × Trial Number⁷</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Presleep*</td>
<td>96.2 ± 1.9</td>
<td>90.7 ± 2.0</td>
<td>90.5 ± 2.0</td>
<td>&lt; 0.001</td>
<td>0.148</td>
<td>0.668</td>
<td>0.988</td>
<td>&lt; 0.001</td>
<td>0.692</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>SpO2 (%) Sleep**</td>
<td>92.3 ± 2.0</td>
<td>86.1 ± 2.0</td>
<td>90.7 ± 2.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post-sleep*</td>
<td>95.9 ± 1.9</td>
<td>92.0 ± 2.0</td>
<td>91.5 ± 2.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heart Rate Presleep* (bpm)</td>
<td>76.1 ± 12.1</td>
<td>77.9 ± 12.2</td>
<td>77.7 ± 12.5</td>
<td>0.099</td>
<td>0.938</td>
<td>0.094</td>
<td>0.031</td>
<td>&lt; 0.001</td>
<td>0.446</td>
<td>0.030</td>
</tr>
<tr>
<td>Sleep**</td>
<td>67.0 ± 12.1</td>
<td>70.5 ± 12.2</td>
<td>67.5 ± 12.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post-sleep*</td>
<td>75.8 ± 12.1</td>
<td>80.4 ± 12.2</td>
<td>79.0 ± 12.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Mean and standard errors calculated from fitted model.
¹ Categorical 3-level variable (ALT, MIX, GND for Condition, 1, 2, 3 for Order).
² Continuous variable (hours for Prior Sleep, years for Age).
³ Categorical 8-level variable (1,2, ..., 8).
4 Measured at 0.25 Hz for approximately 2 min on three separate occasions 1 h 45 min apart.
5 Measured at 0.25 Hz continuously during sleep period.
general sleep was more disturbed during the first study session than in subsequent sessions, and older subjects had more disturbed sleep (Table II). The responses to one of the nine questions on the postsleep questionnaire were found to be related to condition with sleep reported as being of poorer quality at ALT than at GND. A "conservative" subset of the results was created for reanalysis of measurements of sleep quantity and quality by excluding all data from sessions in which the total time for analysis (time in bed) was outside 420 ± 1 min. The analyses of sleep quantity and quality based on this subset support the reported analyses.

Tests of neurobehavioral performance and mood were not significantly related to condition (Table III). Trial number had a significant effect on PVT reaction time, with longer reaction times tending to occur in trials toward the end of the simulated flight compared with those at the beginning of the flight. Speed of PVT response was not significantly related to condition. Run order had a significant effect on some of the ANAM tests, and in all cases poorer performance tended to occur in the first simulated flight compared with flights 2 and 3. Age had a significant effect on performance on the Two Choice test, with older subjects tending to have longer reaction times. Trial number had a significant effect on Tower of Hanoi, with shorter reaction times tending to occur in later trials. Trial number had a significant effect on some POMS axes, with scores for Vigor lower and for Fatigue higher for the trials immediately prior to sleep, and those for tension and depression highest during the preflight trial. Age had a significant effect on the score for Vigor, with the score increasing as age increased.

**DISCUSSION**

We found that sleep in barometric pressures equivalent to 2438 m was accompanied by moderately severe hypoxemia but not by adverse effects on sleep quantity, sleep quality, postsleep neurobehavioral performance, or mood in 30 – 56 yr old healthy men. From studies at lower altitudes, we expected this level of hypoxemia to be consistent with sleep fragmentation and neurobehavioral performance impairment (12,18).

The study had sufficient statistical power to detect functionally important differences in total sleep time and neurobehavioral performance. The lower limit of the 95% confidence interval for the difference in total sleep time between 2438 m and 305 m was -28 min, greater than the -30 min difference considered a priori to be functionally important. Similarly, the upper limit of the 95% confidence interval for the difference in the neurobehavioral performance measure of primary interest, mean PVT response time between 2438 m and 305 m, was 2.2 ms, implying that the effect of sleep at an altitude of 2438 m was unlikely to be as high as 25 ms, the amount considered functionally important. Therefore we conclude that it is unlikely that an important effect...
on mean PVT response time or sleep quantity was missed due to small sample size (7).

Our failure to achieve a balanced participation by order and condition introduced two biases that tended to increase the apparent differences in measurements of sleep quantity and quality between 305 m and 2438 m. The purpose of balancing was to equalize the contribution of the deleterious "first night" effects on sleep in unaccustomed surroundings across the three conditions. In our study, 47% of subjects experienced the 2438-m sleep period during their first study session, when sleep disruption from adaptation was minimal. A second bias resulting from adaptation was minimal. A second bias due to these biases.

It is unlikely that we overestimated the duration of desaturation events, and therefore underestimate the number of desaturation events and percent of sleep time \( \leq 90\% \) (5). Variations in the timing and duration of the transient pressure changes introduced at the beginning and end of the sleep periods may explain the small apparent differences in the desaturation index and percent of time below 90% observed between conditions GND and MIX—conditions in which the assigned altitude during sleep was 305 m (Figs. 1 and 2). No measurements of respiratory physiology were made so the mechanisms underlying the hypoxic episodes cannot be known from this study. Thomas et al. found that altitude-induced respiratory abnormalities at 3962 m (13,000 ft) include both central and obstructive events (17).

Sleep at 305 m in our study more closely resembled sleep in members of an aircrew at a layover hotel than normal sleep in 40 – 49 yr old men. More time was spent in the lighter stages of sleep and there were more awakenings per hour than in normal sleep. On the other hand, sleep during rest periods in flight was shorter, lighter, and more disturbed than observed at 2438 m in our study, raising the possibilities that the effect of altitude was underestimated in our study, or that factors other than altitude affect sleep during flight (16).

Thomas et al. found that sleep at 3962 m was not associated with decrements in working memory or simple reaction time in healthy nonsmoking men and women with a mean age of 27 yr and a mean BMI of 23 (17).
Their findings together with ours suggest that the magnitude of hypoxemia observed during sleep in healthy people at 2438 – 3962 m is insufficient by itself to produce neurobehavioral performance decrements. It is unknown if this can be generalized to those with illnesses that affect oxygen delivery.

The changes in oxygen saturation found in our study suggest that respiration during sleep is disturbed at 2438 m, but the lack of concomitant sleep disturbances makes the functional significance of this finding for flight crew unclear. Our results indicate that when altitude exposure is acute (1 night only), neurobehavioral performance and sleep are not significantly affected. Nevertheless, respiratory disturbances during sleep at 2438 m and associated hypoxemia may adversely affect cardiovascular health if sufficiently severe and prolonged (4,13). Given that ultra-long-range flight crew members may sleep at altitude several times a week during a 30–40 yr career, these findings merit further investigation. In particular, more information is needed concerning the physiologic mechanisms underlying the respiratory disturbances, the consequences of the resulting changes in pulmonary ventilation, and the altitude threshold at which respiratory disturbances and sleep changes occur.

In summary, we found a pronounced degree of oxygen desaturation during sleep at 2438 m, but no significant differences in objective measurements of sleep quantity, sleep quality, postsleep neurobehavioral performance, nor subjective measurements of mood and sleep quality in 30 – 56 yr-old healthy men. Further study is needed to determine if the same effects occur in women and to characterize the changes in respiratory physiology that occur during sleep at 2438 m altitude in both men and women.

ACKNOWLEDGMENTS

The authors thank Margo J. van den Berg for the technical assistance she provided in the initial stages of this study.

Authors and affiliations: J. Michael Muhm, M.D., M.P.H., Stephen P. Jones, Ph.D., and Shuying Zhu, Ph.D., The Boeing Company, Seattle, WA; T. Leigh Signal, Ph.D., Karyn M. O’Keeffe, B.S., and Philippa H. Gander, Ph.D., Sleep/Wake Research Center, Research School of Public Health, Massey University, Wellington, New Zealand; Paul B. Rock, D.O., Ph.D., and Matthew R. Weaver, B.S., Oklahoma State University, Tulsa, OK; and Greg Belenky, M.D., Washington State University, Spokane, WA.

REFERENCES


