MILITARY MEDICINE, 172, 9:982, 2007

Effect of Caffeine on Simulator Flight Performance in Sleep-Deprived Military Pilot Students

Guarantor: Tuomo K. Leino, MD PhD
Contributors: Jouni J. Lohi, MD PhD*; Kerttu H. Huttunen, PhD†; Taija M. M. Lahtinen, MB‡; Airi A. Kilpeläinen, MSocSc§; Arto A. Muhli, BSc¶; Tuomo K. Leino, MD PhD‖

Caffeine has been suggested to act as a countermeasure against fatigue in military operations. In this randomized, double-blind, placebo-controlled study, the effect of caffeine on simulator flight performance was examined in 13 military pilots during 37 hours of sleep deprivation. Each subject performed a flight mission in simulator four times. The subjects received either a placebo (six subjects) or 200 mg of caffeine (seven subjects) 1 hour before the simulated flights. A moderate 200 mg intake of caffeine was associated with higher axillary temperatures, but it did not affect subjectively assessed sleepiness. Flight performance was similar in both groups during the four rounds flown under sleep deprivation. However, subjective evaluation of overall flight performance in the caffeine group tended to be too optimistic, indicating a potential flight safety problem. Based on our results, we do not recommend using caffeine pills in military flight operations.

Introduction

Military preparedness may require human functioning 24 hours a day and, hence, timing of sleep often deviates from its biologically natural nocturnal placement. Military pilots especially may be exposed to a combination of extended work hours, reduced sleep hours, night work, and circadian dysrhythmia caused by shift work during flight operations. However, military operations are characterized by the need to maintain a constant high level of performance. A pilot's performance is negatively affected by sleep deprivation (SD).

Symptoms of SD include extreme sleepiness, lapses in attention, decreased working memory and visual perception, irritability, lack of initiative, susceptibility to accidents, decreased decision-making ability, and decreased self care.[1-3] Increased sleepiness may also be related to impaired psychomotor functions, increased risk-taking behavior, and a higher accident risk during shift work.[4,5]

Caffeine has been used as a countermeasure against fatigue. Caffeine is rapidly absorbed in 30 to 120 minutes and it has an almost immediate impact on vigor and fatigue scores. This impact usually lasts from 5 to 7 hours, since the half-life of caffeine is ~250 minutes.[6] Caffeine is known to increase alertness, anxiety, and blood pressure, reduce simple reaction times, and improve sustained attention, dual task performance, and encoding of new information.[7] However, high (usually over 250 mg) doses of caffeine can cause unwanted side effects (trembling, jitteriness, tension, anxiety, etc.).

In various studies, caffeine has been found to improve physical performance,[8] attenuate fatigue,[9] diminish deterioration of reaction times and cognitive performance during extended wakefulness,[10–12] and maintain a pilot's cognitive performance during the overnight period.[13] The effect of caffeine on the flight performance of sleep-deprived pilots has been poorly studied. In this study, we examined the effects of caffeine on the performance of sleep-deprived military pilots in a Hawk simulator.

Methods

Participants

The participants of this study were 13 Finnish male military pilots. Originally, 15 pilots were recruited into the study, but due to little recent flight experience with a BAe Hawk Mk 51, two subjects did not participate in the simulator flight experiments. The age of the 13 pilots ranged from 23 to 24 years. All of the subjects were healthy and passed a medical prescreen (to rule out, e.g., significant illnesses of any type, sleep difficulties, and allergic reactions to medications) before the SD period. Their average flight experience was 190 flight hours (range, 170–210).

The subjects were asked whether they were regular coffee drinkers or daily used any other products, like beverages, tea, or chocolate, containing caffeine. The number of habitual caffeine users was 11 of the 13 test subjects. Smoking and snuffing were monitored and marked down during the SD period. Of the 13 subjects, 6 used nicotine products—1 used tobacco, 1 snuff, and 4 used both of them. According to a sleep diary, the subjects slept for 4 to 8 hours, on average 7 hours, on the night preceding the tests.

Procedure

The subjects were asked to refrain from caffeine or products containing caffeine during the 12 hours preceding the start of the trial. On the morning of the first test day, wake-up at 6:00 a.m. was followed by breakfast. Four flight rounds were performed. In each round, the subjects were scheduled for the simulator flight one after the other, with a 30-minute break between the subjects. On the first test day, the first flight round was performed between 9:00 a.m. and 12:00 p.m. and the second flight round between 6:30 p.m. and 9:30 p.m. The third flight round took place between 2:00 a.m. and 5:00 a.m. and the last flight round between 9:00 a.m. and 12:00 p.m. on the second test day. Each simulator flight lasted 27 minutes,

*Lapland Air Command, Finnish Air Force, Rovaniemi, Finland.
†Department of Finnish, Information Studies and Logopedics, University of Oulu, P.O. Box 1000, FI-90014 Oulu, Finland.
‡Department of Otorhinolaryngology, University of Oulu, P.O. Box 5000, FI-90014 Oulu, Finland.
§Department of Psychology, University of Jyväskylä, P.O. Box 35, FI-40014 Jyväskylä, Finland.
¶IT Administration Services, University of Oulu, P.O. Box 8300, FI-90014 Oulu, Finland.
‖Air Force Academy, Finnish Air Force, P.O. Box 7, FI-41161 Tikkakoski, Finland.
This manuscript was received for review in January 2007. The revised manuscript was accepted for publication in May 2007.

on average (range, 24–33 minutes). The subjects also took part in repeated vigilance, hearing and speech production tests that were a part of the test protocol and scheduled individually for each subject. After staying awake for 37 hours, the subjects were free to go to bed at 7:00 p.m. on the second test day.

In this double-blind study, all of the subjects received a placebo before the first simulator flight to obtain the baseline for each subject's individual flight performance. Before the subsequent flight rounds, the intake of caffeine and a placebo was randomized, so that six subjects received a placebo and seven received caffeine. The subjects were given two capsules with water at each dosage time exactly 1 hour before the simulator flight. Each active capsule contained 100 mg of caffeine, hence one peroral dose comprised a placebo or 200 mg of caffeine. The dosage levels were not adjusted according to body weights. The body weights of the participants in the caffeine group ranged from 72 kg to 87 kg. Hence, each dosage of caffeine was only moderate, ranging from 2.3 mg/kg to 2.8 mg/kg.

Flight Procedure

The subjects performed an instrument flight rule (IFR) flight mission with the BAe Hawk Mk 51 (HWS-2) simulator four times within 30 hours, in all 52 sorties. English was used as the communication language during the simulated flights. The simulator flight mission included the following flight phases: (1) start at 1,500 feet instrument meteorological condition and climb to flight level (FL) 100; (2) IFR vectors at FL 100 monitoring one enemy aircraft using VHF omnidirectional radiorange (VOR)/distance measuring equipment (DME) information; (3) IFR vectors at FL 100 monitoring two enemy aircraft using VOR/DME information and an emergency procedure; (4) IFR vectors at FL 100 monitoring two enemy aircraft using VOR/DME information; (5) IFR vectors and descent to 2,300 feet for initial approach; (6) IFR base leg before instrument landing system (ILS) approach; (7) ILS localizer established; and (8) ILS outer marker inbound in minimum weather conditions and landing. The flight mission can be described as a moderately demanding IFR flight for recruited pilot students. The above-mentioned flight phases were chosen to represent varying levels of cognitive workload from the viewpoint of situation awareness (SA), information processing, and decision-making pressure. To evaluate situation awareness, during flight phases 2 to 4, the subjects were asked to report the heading to one or two enemy aircraft once a minute. Two flight routes were used and they were counterbalanced across the flight rounds for all of the subjects. The cognitive workload was similar in all four flight rounds, but the position and heading of the enemy aircraft was changed to minimize the learning effect.

Simultaneously with the pilots performing each flight, the pilot students were evaluated by a flight instructor and a flight surgeon (T. K. Leino). Flight data information was used in subjective evaluations that were performed using both a visual analog scale (VAS) and a scale from 1 to 5 with 0.25 intervals. On the interval scale, 1 means the worst possible and 5 the best possible performance. Scores were given for situation awareness, heading error, altitude error, IFR flight performance, ILS flight performance, emergency procedures, and overall flight performance.[14] The subjects also gave a self-evaluation of overall flight performance immediately after each flight using a VAS. Essential flight data were also videorecorded during each flight for later analyses.

Subjective Sleepiness and Body Temperature

Subjective sleepiness was evaluated approximately 30 minutes after each simulated flight using the Stanford Sleepiness Scale (SSS),[15] which is known to be sensitive to the effects of caffeine.[16] The scale ranged from 1 ("Feeling active and vital; alert; wide awake" to 7 "Almost in reverie; sleep onset soon; losing struggle to remain awake"). The body temperatures of all of the participants were recorded. Axillary temperature was used in the analyses, because it is confirmed to represent core temperature in the study of Shann and Mackenzie.[17]

Data Analysis

SAS 9.1.3 proprietary software (SAS Institute, Cary, North Carolina) was used to import and synchronize the original data sources. Data analysis for the smoothed splines in Figures 1 to 7 and 9 was conducted using SAS/GRAPH software (SAS Institute). Smoothed splines were chosen to best demonstrate the underlying nature of circadian variation. A first-order autoregressive covariance structure was chosen by comparing Schwarz's Bayesian information criteria fit statistics from each considered covariance structure. A SAS/STAT mixed procedure was used in this analysis.

SPSS version 11.5.0 software (SPSS, Chicago, Illinois) was used in analyzing the relationship between the two overall flight performance scores given by a flight instructor—one on a VAS and one on an interval scale from 1 to 5. Spearman's correlation coefficient was used in the analysis. The same method was used when comparing the overall flight performance score on a VAS given by a flight instructor with the self-evaluation score of the pilot students. SPSS software was also used when examining the within group (Wilcoxon-signed ranks test) and between-group differences (Kruskal-Wallis test) of the SSS.

Results

Flight Performance

Situation awareness during the flight was covered by three parameters in our analyses: reported heading error to enemy aircraft (difference in degrees) and two situation awareness scores given by the flight instructor (one on a scale of 1–5 and one on a VAS). Figure 1 shows that the SA score on the VAS varies between −38 and + 27 compared with the individual VAS scores in the first flight round. The group mean values of the caffeine and placebo groups did not differ from each other in their ability to sustain SA in any of the flight rounds.

As an objective measurement calculated by the simulator system, the reported heading error to enemy aircraft was, on average, 25 degrees, ranging from 7 to 107 degrees. Figure 2 illustrates the change in the average heading error to enemy aircraft compared with the first flight. As can be seen from the figure, despite some individual variation, no significant difference between the groups was found in any of the flight rounds.

IFR flight performance was also assessed in our study. Figure 3 shows that the altitude maintaining score (on the VAS) tended

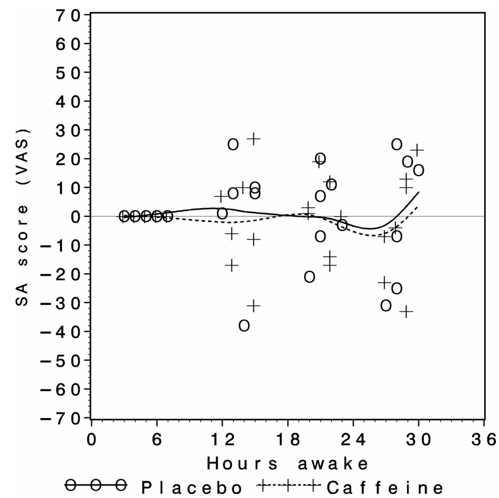

Fig. 1. SA score (on a VAS) compared with the score of each subject in the first flight round.

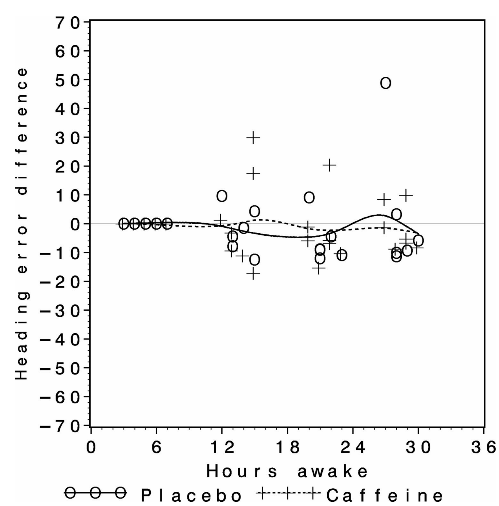

Fig. 2. Reported heading error to enemy aircraft (in degrees, difference compared to heading error made by each subject in the first flight round).

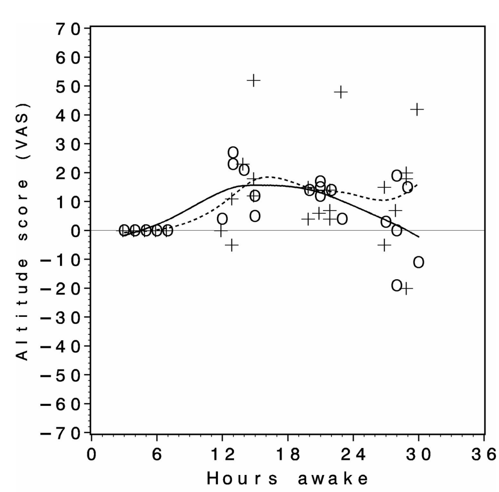

Fig. 3. Altitude maintaining score (on a VAS) compared with the score of each subject in the first flight round.

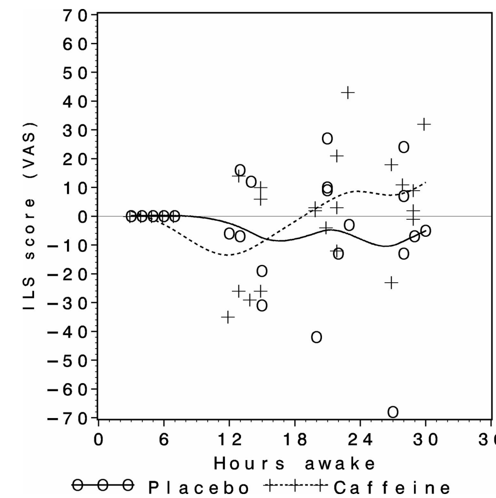

Fig. 4. ILS flight performance score (on a VAS) compared with the score of each subject in the first flight round.

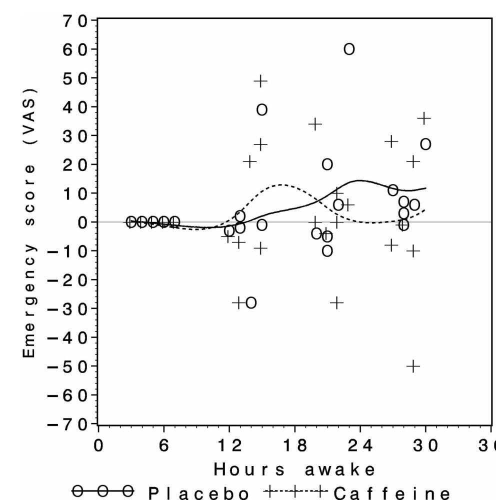

Fig. 5. Emergency procedure score (on a VAS) compared with the score of each subject in the first flight round.

to be slightly better during the second and third flight rounds, and that no clear difference was seen between the two groups. ILS flight performance (instructor score on the VAS, Fig. 4) and emergency procedure performance (Fig. 5) were also rather similar in the two groups studied.

Overall flight performance analysis was measured by three parameters: instructor score for overall flight performance (on a scale from 1–5 and on a VAS) and self-evaluation score for overall flight performance (on a VAS). Like other parameters measured in our study both on an interval scale and on a VAS, the flight instructor score on the VAS correlated statistically significantly with the score on the interval scale from 1 to 5 (Spearman's correlation coefficient, 0.88; p = 0.000) and with the self-evaluation score of the pilot students (Spearman's correlation coefficient, 0.57; p = 0.000). Figure 6 shows that caffeine made no difference in overall flight performance compared with flight performance in the placebo group. However, after ~20 hours of SD, the subjects who received caffeine evaluated their overall flight performance to be slightly better than in the first and the second flight rounds (Fig. 7).

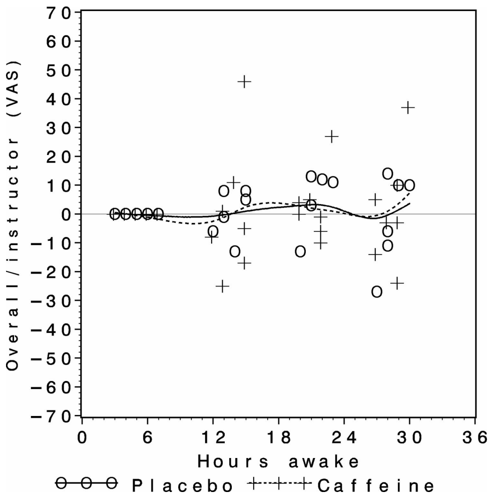

Fig. 6. Overall flight performance score (given by the flight instructor on a VAS) compared with the score of each subject in the first flight round.

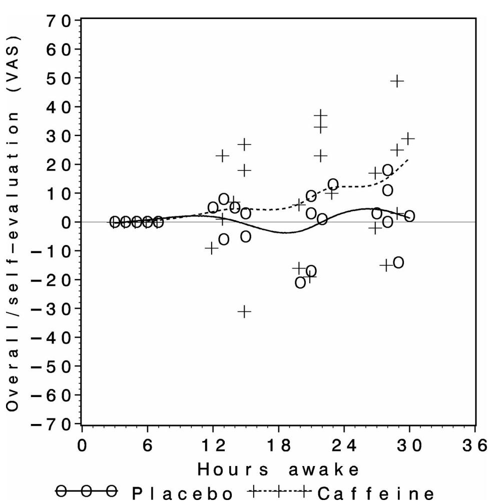

Fig. 7. Overall self-evaluation score (on a VAS) compared with the score of each subject in the first flight round.

Sleepiness Scale

As measured by the SSS 30 minutes after each simulated flight, the prolonged SD time statistically or almost statistically significantly increased subjective sleepiness (Wilcoxon-signed ranks test; Z from −2.71 to −3.0; $p < 0.01$ or $p < 0.05$) except between the SSS after the second and the third simulated flights. There were no significant differences in subjective sleepiness between the caffeine and placebo groups (Kruskall Wallis test; $\chi^2 = 0.006–0169$; $df = 1$; p from 0.68–0.94; see Fig. 8).

Temperature Results

Axillary temperatures were successfully obtained from altogether 12 subjects: from all of the subjects in the caffeine group and from five of the six subjects assigned to the placebo group. During the first 12 hours of SD, when all of the subjects received a placebo, the groups did not differ statistically significantly from each other, although the caffeine group had, on average, 0.4°C ($p = 0.08$), higher axillary temperatures compared with the placebo group. Figure 9 presents the individual axillary

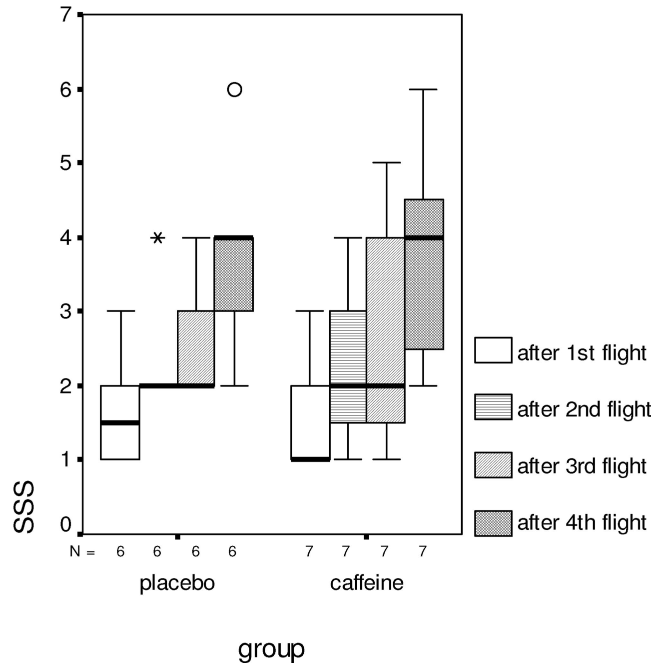

Fig. 8. Subjective sleepiness on the SSS scale, with group medians and quartiles, in the placebo and caffeine groups 30 minutes after each simulated flight.

temperatures measured during the study. Expectedly, the highest mean axillary temperatures were recorded during the afternoon of the first test day and the lowest in the small hours of the second test day. Axillary temperatures decreased in both groups congruently with SD time, with the main effect of SD being −0.3°C ($p = 0.197$), but the drop was greater in the placebo group. The main effect of caffeine was found to be +0.48°C ($p = 0.004$). A statistically significant negative correlation was found between the SSS score and axillary temperature (Spearman's correlation coefficient −0.38, $p = 0.007$). Figure 10 shows mean axillary temperature in different phases of the SD period. Simulator flights caused a notable rise in axillary temperature in part because of the flight vest the subjects wore during the simulated flights.

Discussion

The pilot students we studied reported themselves to be from "not fully alert" to "sleepy" at the end of our study. Despite this, their simulator flight performance was not impaired during the procedure. It is possible that the duration of SD used in our study (27–30 hours during the last simulator flight) was not long enough for the detrimental effects of sleep loss to emerge. The selected group of young adults in demanding military training may also be more resistant than others to the effects of SD. Furthermore, performing a simulated flight was a very stimulating activity for the young pilots. Thus, the test situation itself may explain why the subjects flew well even though they were quite tired. A more demanding flight procedure and a higher information load may result in different performance in a flight simulator during SD. During this study, the subjects performed the flight mission four times, so some learning effect was expected to emerge, although the flight route and enemy aircraft positions differed between the flight rounds. A learning effect was observed in two individuals especially in the reported heading error (Fig. 2).

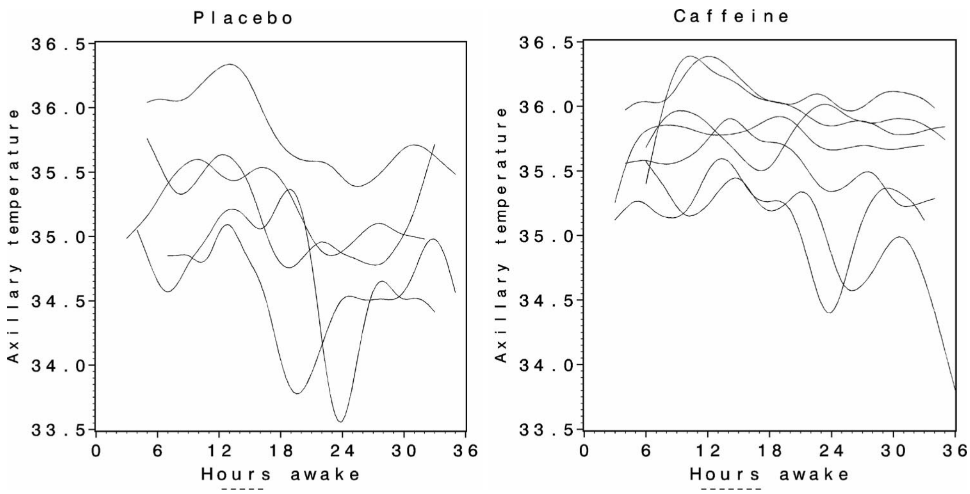

Fig. 9. Individual axillary temperatures in the placebo and caffeine groups.

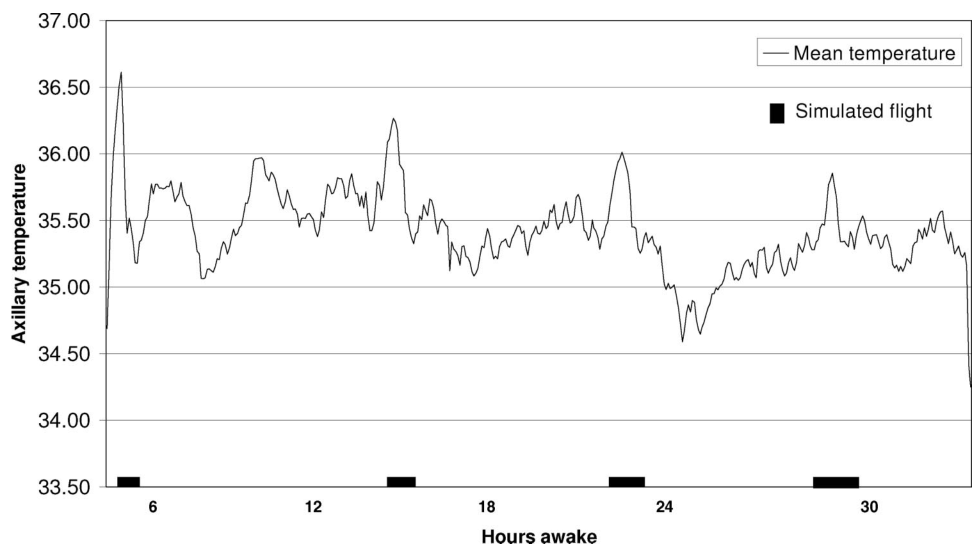

Fig. 10. Mean axillary temperatures during SD (n = 12).

The present study showed that a moderate 200-mg dose of caffeine does not have a significant effect on flight performance during a maximum of 30 hours of SD. The performance level of our young pilots was quite good even without taking any countermeasures against fatigue and, hence, possibilities for significant improvement in performance were somewhat limited. Still, the missing effect of caffeine intake on flight simulator performance was a little bit unexpected in view of notions that caffeine improves the driving capacity of a sleep-deprived car driver in a simulator.[18] Besides, caffeine intake is known to maintain the vigilance and target detection of soldiers in sustained simulated urban operations.[19–22]

It is known that the caffeine metabolism of individuals varies widely due to CYP1A2 alleles.[23] As we did not analyze caffeine levels in saliva samples or the urinary caffeine metabolic ratio of our subjects, the possible effect of inherited rapid or slow enzyme activity in the subjects randomized to the caffeine group remains unknown.

Axillary temperatures decreased in both groups studied, with a larger decrease in the placebo group (0.6°C) compared with the caffeine group (0.3°C). However, the main effect of caffeine on axillary temperature was only moderate, falling in the range of normal individual circadian variation. Body temperature was, on average, slightly higher in the caffeine group than in the placebo group, probably due to a higher metabolism and level of activation caused by caffeine intake. As expected, a negative correlation was found between the SSS score and axillary temperature in our study, showing that the students were most tired when their body temperature was in its batyphase. The effect of nicotine on the temperature results may raise some questions. The higher proportion of nicotine users (67%) in the placebo group compared with the caffeine group (29%) may be associated with axillary temperatures decreasing somewhat more in the placebo group, as nicotine is known to act as a vasoconstrictor in the skin.

In this study, some individuals in the caffeine group had slightly too optimistic self-perception of their flight performance after being awake for 20 hours (Fig. 7). This phenomenon of so-called "overconfidence" has previously been observed with modafinil use.[24] Caffeine use in sustained flight operations might lead to a flight safety problem due to decreased self-criticism. Our study sample was very small, being comprised of only military pilot students with the same flight experience available at the same time in Finland. The generalizability of our results is therefore not very high. However, it seems that young military pilot students are quite capable of performing during overnight SD and no clear decrement was seen in their flight performance. Additionally, the results suggest that moderate doses of caffeine do not significantly improve the flight performance of sleep-deprived military pilot students during simulated flight missions where the pilots are facing mainly cognitive, not physical, challenges.

In conclusion, the flight performance of the young pilot students did not change significantly during the SD period. Caffeine did not improve their flight performance. To avoid unwanted side effects and possibly decreased self-perception of the pilots, we do not recommend the use of caffeine pills in military pilot operations. However, caffeine users can get a rebound headache if they do not get caffeine. We therefore suggest that conventional portions of coffee be made available to regular caffeine users.

Acknowledgments

We acknowledge Hannu Rintamäki and Tero Mäkinen for their help with the temperature recordings and Jani Määttä, Matti Mustanoja, Toni Ovaskainen, and the personnel of the Training Air Wing, Finnish Air Force for their help in data collection. We also thank Heikki Lyytinen, Martti Sorri, and Pentti Kuronen for their advice in planning the study and for supporting the authors in many ways.

This work was supported by grants from the Ministry of Defence/Scientific Advisory Board of Defence (2004 and 2005) and The Finnish Cultural Foundation. The Finnish Air Force has also financially supported this study.

References

1. Lagarde D: Introduction–overview: the sleep-wakefulness cycle and the SUSOPS/CONOPS. In: Sleep/Wakefulness Management in Continuous/Sustained Operations. RTO Lecture Series 223 (RTO-EN-016 AC/323(HFM-064)TP/39). Neuilly-Sur-Seine, France, North Atlantic Treaty Organisation, Research and Technology Organisation, 2002.
2. Killgore WD, Balkin TJ, Wesensten NJ: Impaired decision making following 49 h of sleep deprivation. J Sleep Res 2006; 15: 7–13.
3. Russo MB, Kendall AP, Johnson DE, et al; Visual perception, psychomotor performance, and complex motor performance during an overnight air refuelling simulated flight. Aviat Space Environ Med 2005; 76(Suppl 7): C92–103.
4. Åkerstedt T, Folkard S, Portin C: Predictions from the three-process model of alertness. Aviat Space Environ Med 2004; 75(Suppl 3): A75–83.
5. Rouch I, Wild P, Ansiau D, Marquie JC: Shiftwork experience, age and cognitive performance. Ergonomics 2005; 48: 1282–93.
6. Åkerstedt T, Landström U: Work place countermeasures of night shift fatigue. J Ergonomics 1998; 21: 167–78.
7. Brice CF, Smith AP: Effects of caffeine on mood and performance: a study of realistic consumption. Psychopharmacology (Berl) 2002; 164: 188–92.
8. McLellan TM, Bell DG, Kamimori GH: Caffeine improves physical performance during 24 h of active wakefullness. Aviat Space Environ Med 2004; 75: 666–72.
9. Wesensten NJ, Belenky G, Thorne DR, Kautz MA, Balkin TJ: Modafinil vs. caffeine: effects on fatigue during sleep deprivation. Aviat Space Environ Med 2004; 75: 520–5.
10. Kamimori GH, Johnson D, Thorne D, Belenky G: Multiple caffeine doses maintain vigilance during early morning operations. Aviat Space Environ Med 2005; 76: 1046–50.
11. Lieberman HR, Tharion WJ, Shukitt-Hale B, Speckman KL, Tulley R: Effects of caffeine, sleep loss, and stress on cognitive performance and mood during U.S. Navy SEAL training, Sea-Air-Land. Psychopharmacology (Berl) 2002; 164: 250–61.
12. Wyatt JK, Cajochen C, Ritz-De Cecco A, Czeisler CA, Dijk DJ: Low-dose repeated caffeine administration for circadian-phase-dependent performance degradation during extended wakefulness. Sleep 2004; 27: 374–81.
13. Doan BK, Hickey PA, Lieberman HR, Fischer JR: Caffeinated tube food effect on pilot performance during a 9-hour, simulated night time U-2 mission. Aviat Space Environ Med 2006; 77: 1034–40.
14. Svensson E, Angelborg-Thanderz M, Sjöberg L, Olsson S: Information complexity–mental workload and performance in combat aircraft. Ergonomics 1997; 40: 362–80.
15. Hoddes E, Zarcone V, Smythe H, Philips R, Dement WC: Quantification of sleepiness: a new approach. Psychophysiology 1973; 10: 431–6.
16. Lieberman HR, Wurtman RJ, Emde GG, Coviella ILG: The effects of low doses of caffeine on human performance and mood. Psychopharmacology (Berl) 1987; 92: 315–20.
17. Shann F, Mackenzie A: Comparison of rectal, axillary, and forehead temperatures. Arch Pediatr Adolesc Med 1996; 150: 74–8.
18. Baker WJ, Theologus GC: Effects of caffeine on visual monitoring. J Appl Physiol 1972; 56: 422–7.
19. Gillingham RL, Keefe AA, Tikuisis P: Acute caffeine intake before and after fatiguing exercise improves target shooting engagement time. Aviat Space Environ Med 2004; 75: 865–71.
20. McLellan TM, Kamimori GH, Bell DG, Smith IF, Johnson D, Belenky G: Caffeine maintains vigilance and marksmanship in simulated urban operations with sleep deprivation. Aviat Space Environ Med 2005; 76: 39–45.
21. McLellan TM, Kamimori GH, Voss DM, Bell DG, Cole KG, Johnson D: Caffeine maintains vigilance and improves run times during night operations for special forces. Aviat Space Environ Med 2005; 76: 647–54.
22. Tharion WJ, Shukitt-Hale B, Lieberman HR: Caffeine effects on marksmanship during high-stress military training with 72 hour sleep deprivation. Aviat Space Environ Med 2003; 74: 309–14.
23. Woolridge H, Williams J, Cronin A, Evans N, Steventon GB: CYP1A2 in a smoking and a non-smoking population; correlation of urinary and salivary phenotypic ratios. Drug Metabol Drug Interact 2004; 20: 247–61.
24. Baranski JV, Pigeau RA; Self-monitoring cognitive performance during sleep deprivation: effects of modafinil, D-amphetamine and placebo. J Sleep Res 1997; 6: 84–9.