Risk of Fatigue Among Airline Crew During 4 Consecutive Days of Flight Duty

Elisabeth M. Goffeng; Anthony Wagstaff; Karl-Christian Nordby; Anders Meland; Lars Ole Goffeng; Øivind Skare; Didrik Lilja; Jenny-Anne S. Lie

BACKGROUND: Airline crew are being exposed to extended workdays and compressed work periods, with quick returns between duties, implying a heightened physiological and psychological strain that may lead to sleep deprivation and fatigue. The aim of the study was assessment of the effect of an extended day of flight duty and a compressed work week with regard to recovery, cumulative fatigue, and neurobehavioral performance.

METHODS: We followed 18 pilots and 41 cabin crewmembers during four consecutive days of flight duty, comprising a total of ≥ 39 h, where the first day was ≥ 10 h. Information on demographics, work characteristics, health status, and physical activity was collected at baseline. Subjects completed logs for the first and fourth workday, including the Samn-Perelli Fatigue Checklist at three time points during these workdays. Two computer-based neurobehavioral tests were completed the evening prior to the first shift, and after the first and the fourth day of the work period.

RESULTS: Number of flight sectors during the work period was 10–20. Self-reported fatigue levels increased during the workdays. Neurobehavioral test-scores did not deteriorate. The effects of each additional flight sector during the work period was elevated reaction times (RT) both among cabin crewmembers (B = 5.05 ms, 95% CI 0.6, 9.5) and pilots (B = 4.95 ms, 95% CI 0.4, 9.5). Precision was unaffected.

DISCUSSION: Airline pilots and cabin crewmembers seem to obtain satisfactory sleep before and during the period of 4 consecutive days. The association between multiple flight sectors and increased fatigue supports previous findings.

KEYWORDS: airline crew, long working days, compressed working hours, fatigue.

Goffeng EM, Wagstaff A, Nordby K-C, Meland A, Goffeng LO, Skare Ø, Lilja D, Lie J-AS. *Risk of fatigue among airline crew during 4 consecutive days of flight duty*. Aerosp Med Hum Perform. 2019; 90(5):466–474.

Functional impairment associated with fatigue is recognized as a significant risk in commercial aviation.[18] Nonstandard working hours and shift work have been linked to both acute and chronic adverse effects, including a lack of sleep, disruption of circadian rhythms, and reduced alertness and performance. Airline crews are subject to long working hours, traditionally related to intercontinental flights. Recently, there has also been an increasing demand for extended daily work hours related to all types of operations, including domestic and short-haul flights. Airline crew schedules are typically irregular and involve early start times and long, compressed work periods, with short rests between work periods. All of these factors might contribute to fatigue and thus an increased risk of human error.[1] Fatigue is defined as a physiological state of reduced mental or physical performance capability resulting from sleep loss or extended wakefulness, circadian phase, or workload (mental or physical activity) that can impair a crewmember's alertness and ability to safely operate an aircraft or perform safety-related duties.[11] A substantial amount of evidence supports the finding that insufficient sleep may cause cognitive impairment.[13] However, the impact of fatigue is often underestimated. Compared to people who are well-rested, those who are sleep-deprived think and react more slowly, make more mistakes, and have memory difficulties.[4] This is especially problematic for aviation personnel (i.e., pilots and cabin crewmembers) performing critical safety duties that require long periods of sustained attention. The links between work schedules, sleep-wake behavior, and cognitive and task performance have been self-evident to researchers,

From the National Institute of Occupational Health, Oslo, and the Institute of Aviation Medicine, Norwegian Armed Forces Medical Services, Oslo, Norway.

This manuscript was received for review in August 2018. It was accepted for publication in February 2019.

Address correspondence to: Elisabeth Goffeng, Department of Occupational Medicine and Epidemiology, National Institute of Occupational Health, P.O. Box 8149 Dep, N-0033 Oslo, Norway; elisabeth.goffeng@stami.no.

DOI: https://doi.org/10.3357/AMHP.5236.2019

regulators, and safety professionals for decades. Although a number of studies on airline pilots support these assumptions,[2,7,16] there are fewer field studies measuring the effects of fatigue in cabin crewmembers.[5]

There is an increased awareness of the consequences of fatigue among legislative authorities, and European regulations now require that all airline companies have a fatigue risk management system. The impact of nonstandard working hours on fatigue and, thus, on accident risk is commonly acknowledged as a cause for concern; however, the mechanisms of that relationship remain unclear.[25] It is therefore important to differentiate between the effects of extended working hours per se, mental or physical workload, work characteristics, and sleep disruption due to the organization of the work hours.[12] The work pattern for airline crew flying short haul may involve multiple takeoffs and landings, resulting in a more demanding workload across the workday. Fatigue among pilots has been observed to be associated with the number of takeoffs and landings during a day of duty, where pilots with multisegment flights appeared more fatigued compared to those operating a single-segment flight duty.[2,10,19]

The aim of this study was twofold:

1. To evaluate whether a single long day of flight duty or a compressed work period with a varied number of flight sectors increases the risk of fatigue-related errors in airline pilots and cabin crewmembers, as measured by computerized neurobehavioral tests.
2. To assess the association between subjective reports on sleep and alertness and performance on neurobehavioral tests.

METHODS

The study protocol was approved in advance by the Regional Committee for Medical Research Ethics.

Subjects

Prior to beginning the study, the subjects received written detailed information and provided written informed consent. The study sample consisted of 18 pilots and 41 cabin crewmembers in a medium-sized European airline, all based in Norway. According to current regulations, a normal workday may be scheduled to last up to 13 h, with the possibility of one extended workday of up to 14 h per 7 d. A work period should not exceed 47.5 h. In unforeseen circumstances, a workday may be extended up to 16 h and a work period up to 60 h. The same limitations apply to both pilots and cabin crewmembers. In the present study, we followed the crew during 4 scheduled consecutive days of flight duty, comprising a total of ≥ 39 h, where the flight duty period of the first day was ≥ 10 h.

Subjects were recruited through e-mails sent by the management of flight and cabin operations at a national level, in coordination with pilot and cabin crew union representatives. In addition, handouts were distributed at the crew base. For operational reasons, the airline did not schedule any crewmember with rosters that met our criteria for inclusion in the study during the winter program, which is a reason for the rather long period of data collection, from April 2015 to September 2017.

Every month, when new rosters were published, only volunteers with the compressed work period of interest were included. The study was restricted to short-haul flights in order to avoid confounding effects from night work and jetlag. All flights were operated by Boeing 737 aircrafts. Characteristics of the included subjects are listed in **Table I**.

Procedure

We applied a cross-shift/cross-week design in which the subjects served as their own controls. Data collection started at enrollment on the evening before the first day of flight duty, was repeated at the time of check-out on the first and fourth workday, and was completed again after subjects had been off-duty for 2 d. Both the neurobehavioral tests were performed in the evening after the first and fourth workday, and as far as possible at a similar hour on days off.

At enrollment, the subjects completed a questionnaire that included items on demographics, work characteristics, health history, sleep, and physical activity. The questions were based on validated questionnaires, including the QPS Nordic[14] and the Bergen Insomnia Scale.[17] In addition, questions were developed specifically for this study. The subjects reported their subjective perceptions of work characteristics, including work-related stress indicators, control/demand issues, and possible work-family conflict.

During the first and fourth days of the actual work period, the subjects completed logs developed for the study. The logs included questions about subjects' length of time sleeping and pattern of sleeping from the previous night, commuting time, check-in and check-out times, physical activity, and self-reported health issues. The subjects also rated their level of fatigue on the Samn-Perelli Fatigue Checklist (SP)[22] at check-in time for duty, after 8 h, and at check-out time after 10–14 h of flight duty on both the first and the fourth day of their actual work period. The checklist is a seven-point scale and fatigue is treated as an ordinal variable with the following categories: 1) fully alert, wide awake; 2) lively, responsive, not at peak; 3) OK, somewhat fresh; 4) a little tired, less than fresh; 5) moderately tired, let-down; 6) very tired, difficulty concentrating; and 7) completely exhausted.

To measure possible detrimental fatigue effects, we used two computer-based neurobehavioral tests, both addressing attention: the Attentional Capture Task (ACT)[24] and the Sustained Attention to Response Task (SART).[21]

The ACT procedure, as described by Theeuwes and Chen,[24] is a sensitive test of attention and was chosen because fatigue reduces arousal level and vigilance and could, therefore, influence attention and vulnerability to distractions. The ACT measures vulnerability to automatic and reflexive distractions due to attentional capture. To reduce the burden on the subjects, we chose a shortened version of the ACT, lasting 15 min, previously used by Meland et al.[15] Attentional capture is assessed by a task-irrelevant peripheral stimulus (i.e., a red blink) that flashes for 60 ms equally often at one of six positions at various

Table I. Subject Characteristics.

	CABIN CREWMEMBERS ($N = 41$)	PILOTS ($N = 17$*)
Men	6	15
Women	35	2
Mean age	40 yr (SD 12.3)	52 yr (SD 7.4)
Current smokers	1	1
Mean BMI (kg · m^{-2})	24 (SD 3.6)	25 (SD 2.8)
Married/cohabiting	27 (66%)	11 (65%)
Children <18 yr living at home	12 (29%)	8 (47%)
Commuting time (one way)	57 min	70 min
Hours of sleep before workday 1 (mean)	5 h 56 min	6 h 49 min
Hours of sleep before workday 4 (mean)	6 h 57 min	7 h 1 min
Reported sleep disturbances		
Last year	21 (51%)	9 (50%)
Last month	12 (12%)	5 (27%)
Self-reported chronotype:		
Morning type	14 (34%)	2 (12%)
Evening type	9 (22%)	5 (29%)
Neither/nor	18 (41%)	10 (59%)
Physical activity (1-3 hours per week)	24 (59%)	11 (65%)
Physical activity (> 3 hours per week)	17 (41%)	6 (35%)
Mean years of employment	17 years (SD 11.4)	26 years (SD 7.7)
Content with work pattern	27 (66%)	8 (44%)
Reported work/family conflict	18 (44%)	7 (39%)

* Questionnaires were returned by 17 of 18 pilots.

distances prior to the target appearance (**Fig. 1**). The following variables were derived from the ACT: reaction time (RT), number of correct responses, and precision [number of correct responses/(number of correct responses + number of errors)] in the following four conditions: 1) overall; 2) "cued", red blink at the target position; 3) "large distance", red blink at the three largest distances; and 4) "short distance", red blink closest to the target. Prior to their first test, the subjects performed a habituation trial lasting for 8 min. This trial was shortened to 2 min in the remaining test sessions, followed by the actual 15-min test.

The SART is sensitive to transitory reductions in attention and particularly the response inhibition element of our executive functioning.[21] In the procedure, 225 single digits, 25 of each digit from 1 to 9, are presented centrally on a laptop. Each digit is presented for 250 ms and is then replaced by an encircled X mask for 900 ms. Presentation is regularly paced at an onset-to-onset interval of 1150 ms. Both digits and mask are white against a black background. Subjects are instructed to respond to the go stimuli (1, 2, 4, 5, 6, 7, 8, and 9) with a key press and to withhold this response when the no-go stimulus (3) is presented. Presentation is regularly paced in order to create a habitual response. Three scores were derived from the SART: 1) number of errors of commission (responses to the rare no-go digit); 2) precision [number of correct responses/(number of correct responses + number of errors of commission)]; and 3) RT to frequent go stimuli > 100 ms. We regarded RT > 100 ms to frequent go stimuli as anticipation, and these were omitted from the analysis. The SART procedure takes 5 min to complete.

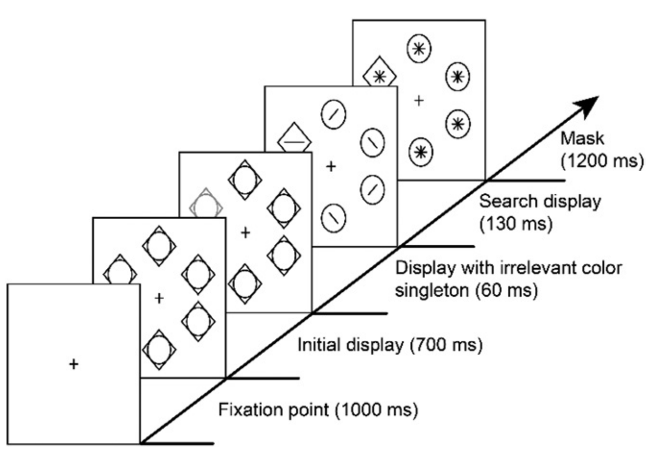

Fig. 1. Time and sequence of the Attentional Capture Task.

Statistical Analysis

The scores on the neurobehavioral tests were analyzed using linear mixed models. We selected three variables measured in four conditions for the ACT procedure and three variables for the SART procedure. For each score, we evaluated the difference between days, where baseline was day 0 and day 6 combined. All subjects were off duty 2 d preceding day 0 and day 6. Having day 0 and day 6 as a combined baseline category allowed us to separate the effect of work period (day 1 and day 4 vs. baseline) from the practice effect. We adjusted for possible practice effects by including the logarithm of the test number (1 = day 0, 2 = day 2, 3 = day 4, 4 = day 6) as a covariate, assuming the practice effect to follow a logarithmic shape, meaning the practice effect was largest at the first test session. Visual inspections of the plots confirmed this finding. We also investigated how workload, as measured by the number of flight sectors, affected the outcome scores of both the ACT and the SART. Here, we completely adjusted for any possible practice effect by including test number as a categorical variable. All analyses as additionally adjusted for sex and age. A random intercept for subjects was included and restricted maximum likelihood was used to estimate variance parameters. We used Statistical Package for Windows 24.0 (SPSS Inc., Chicago, IL, United States) for descriptive analyses, while the lme4 package (version 1.1-12) in R (version 3.3.3) (R-project.org) was used for the multivariate analyses. A critical *P*-value of 0.05 was chosen.

RESULTS

Table I shows the characteristics of the cabin crewmembers and pilots. Compared to the cabin crewmembers, pilots were older, had longer commuting time, reported more sleep disturbances, and were less content with their work pattern. Both groups reported being physically active. We followed the subjects from baseline, the evening before the start of their work period, to the evening after they were off-duty for 2 d (**Fig. 2**).

There were 18 pilots and 41 cabin crewmembers who completed the 2 first test sessions (at baseline and after the first day of flight duty). These numbers were reduced to 25 and 16 after the fourth workday, due to rescheduling, delays, diversions, and sick leave. Start of holidays, long commuting time, and other private issues all influenced the further reductions in the number of subjects completing the second baseline after 2 d off. In addition, some subjects were excluded from the second baseline because they had a 5-d working period, which made their new baseline after 2 d off not comparable with the new baseline of subjects with a 4-d work period.

According to the logs and the subject's rosters, the work periods ranged from 39 to 41 h in both professional groups. The night before the fourth workday, 14 out of 15 pilots and 13 out of 25 cabin crewmembers had a hotel layover. The average duty time on the first workday among the cabin crewmembers was 11.2 h (SD 1.1) and 10.6 h (SD 1.8) on the fourth day. Mean check-in time for duty was 07:45 (SD 2.0) on the first workday and 09:19 (SD 2.6) on the fourth workday. The duty time among the pilots was 10.9 h (SD 1.3) on the first workday and 10.1 (SD 1.6) on the fourth workday. The mean check-in time for the pilots on the first workday was 08:13 (SD 2:47), and 10:17 (SD 3.3) on the fourth workday. The total number of flight sectors during the work period ranged from 10 to 20, with a mean of 15.3 among both cabin crewmembers and pilots. The mean number of flight sectors among the cabin crewmembers on the first workday was 3.8 (range 2–6), and on the fourth 3.5 (range 2–5). Among the pilots the number of flight sectors was 3.4 (range 2–6) on the first workday and 3.9 (range 3–5) on the fourth workday. Reported sleep quality among both groups improved the night preceding the fourth workday (**Fig. 3**).

Neither cabin crewmembers nor pilots reported fatigue, as measured by the SP, at the start of the first or the last day of the work period. Cabin crewmembers had a mean score of 2.1 at the start of the first workday and 2.4 at the start of the fourth workday, while the pilots had a mean score of 2.2 on both days. The cabin crewmembers reported a mean SP score of 4.2 after the first day and 4.4 after the last day of the work period. Using a *t*-test, we did not observe any significant difference in the SP score between cabin crew who spent the night preceding the fourth workday in hotels and those who spent this night at home, neither at the start of the fourth workday ($t = 0.2$, df $= 19.3$, $P = 0.90$), nor at the end ($t = 1.5$, df $= 20$, $P = 0.16$). The pilots scored a mean of 4.6 on the SP after the first workday and 4.1 after the fourth workday (**Fig. 4**).

The main observation was that overall cognitive performance, as measured by the ACT, did not deteriorate in the 4 d compared to baseline performance, neither among the pilots nor among the cabin crewmembers, when adjusting for age, sex, and practice effects (**Table II** and **Table III**). However, the effects of each additional flight sector during the work period was elevated RT both among cabin crewmembers (B = 5.05 ms, 95% CI 0.6, 9.5) and pilots (B = 4.95 ms, 95% CI 0.4, 9.5). Check-in times for duty and duration of the workdays did not influence the overall RT (not shown).

Overall cognitive performance as measured by the SART did not deteriorate during the 4 d, neither among the pilots nor among the cabin crewmembers. Precision among cabin crewmembers increased after the first workday compared to baseline (B = 4.6, 95% CI 0.3, 8.9), while precision among pilots increased after the fourth workday compared to baseline (B = 9.3, 95% CI 1.8, 16.9) (**Table IV**).

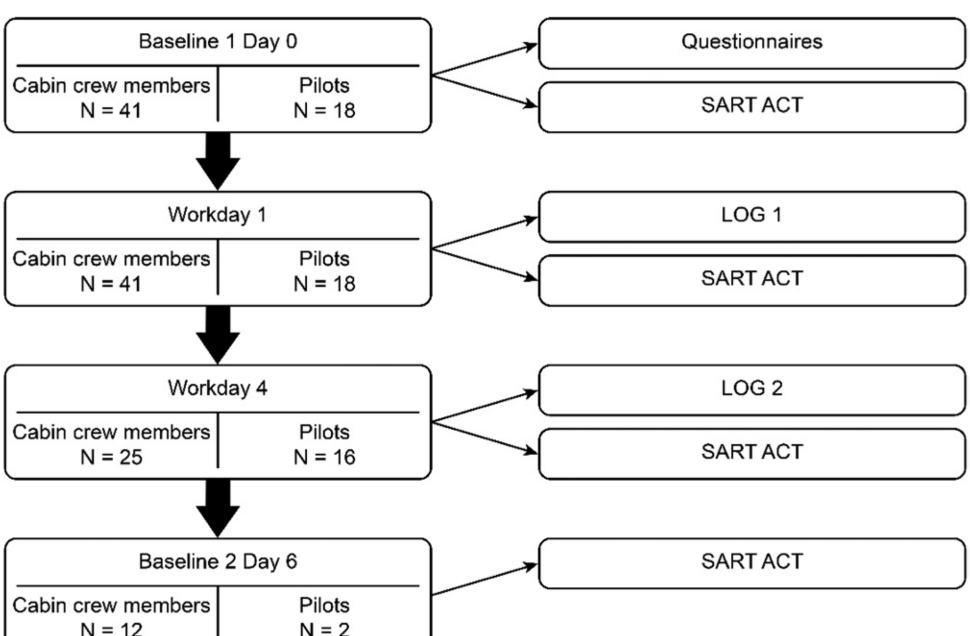

Fig. 2. Flowchart of the data collection among cabin crewmembers and pilots.

DISCUSSION

As expected, there was a clear increase in self-reported fatigue levels from the start to the end of the work period, in both cabin crewmembers and pilots. Contrary to our hypothesis, we found no evidence of a decrease in objective performance, as measured by the ACT and SART, in any of the groups looking at work hours only. However, we found evidence of longer RT, as measured by the ACT, when the number of flight sectors increased during the 4 workdays in both pilots and cabin crewmembers.

The safety risk associated with fatigue is widely recognized in commercial aviation operations.[4]

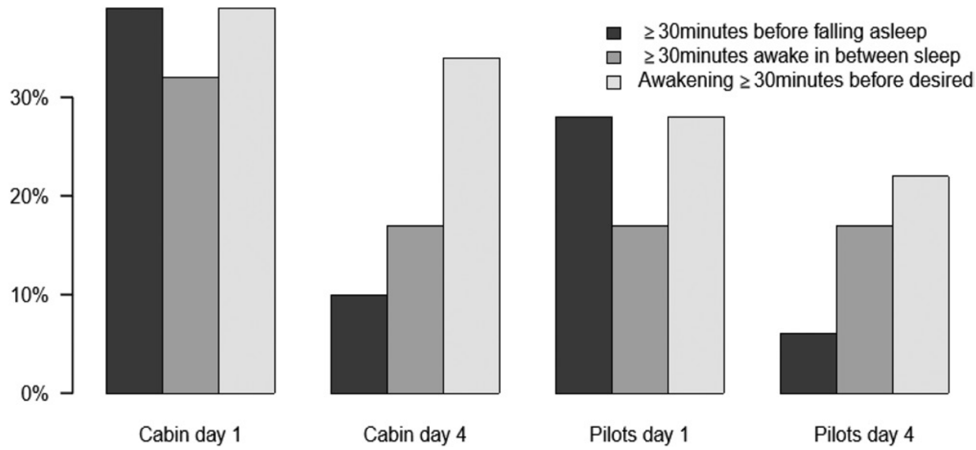

Fig. 3. Percentage of subjects reporting sleep parameters during nights preceding day 1 and day 4, among cabin crewmembers (*N* = 41 d 1, *N* = 26 d 4) and pilots (*N* = 18 d 1, *N* = 16 d 4).

The findings in the present study show that the majority of both cabin crewmembers and pilots managed to obtain satisfactory length of sleep[9] before and during their work periods. Both professional groups reported sleep disturbances the nights preceding the workdays; however, most prominently the night before the first workday. Of the pilots, 47% had children under 18 yr at home and 56% were not content with their work pattern, which could have influenced sleep disturbances the night before the first workday. The hotel layovers before the fourth workday may have made it easier for both groups of subjects to unwind and sleep undisturbed during that night.

The cabin crewmembers obtained 1 h less sleep than the pilots the night before the first day of flight duty. This finding could be due to the cabin crewmembers having earlier start times. Other questionnaire-based studies of short-haul pilots in the commercial aviation industry indicate that duty periods with early-morning starts is one of the major causes of fatigue.[2] The earlier workers must start work, the less sleep they will obtain, and the more tired or fatigued they will be.[20] A study of short-haul Australian pilots found that self-rated fatigue among pilots at the start of duty was highest for duty periods that commenced between 04:00 and 05:00 and lowest for duty periods that commenced between 09:00 and 10:00.[20] In the present study, none of the cabin crewmembers started their duty before 05:00, and only one pilot had to check in for duty at 04:55 on one of his workdays. This finding may explain why most of the subjects were able to obtain at least 6 h of sleep and reported that they were rested before both days of their flight duty.

According to the cabin crewmembers' and the pilots' scores on the SP, they appeared to be well-rested when commencing their work in the morning of both the first and the fourth day of flight duty. SP score was elevated toward the end of both workdays in both professional groups. However, among the pilots, the elevation was less pronounced on the fourth day, which may be due to the fact that their mean working hours were 48 min shorter on this day compared to the first, and that the majority of them had a hotel layover with reduced sleep disturbances. The cabin crewmembers' mean SP score at the end of the fourth workday was elevated compared to the first day. This in spite of the fourth workday being 24 min shorter and the mean check-in time 1.5 h later than on the first workday.

We did not observe any significant differences between the SP scores when comparing the cabin crewmembers with hotel layovers on the fourth workday with the ones who started their duty from the home base. Cabin crewmembers reported smaller variations than pilots regarding sleep duration the nights preceding both workdays, and less sleep disturbances the night preceding the fourth workday. A possible explanation may be the very different work content.

Our finding of a significantly longer RT, as measured by the ACT, in both professional groups with an increasing number of flight segments during the 4-d on-duty period supports earlier research on short-haul flight operations.[2,10,19] Cognitive slowness is known to be one indicator of fatigue.[3] The subjects' precision, however, was unchanged, which may be a result of a speed/accuracy trade-off.[8] It may seem that they did their best to be precise, but they took more time because they were tired. Most short-haul days of duty involve

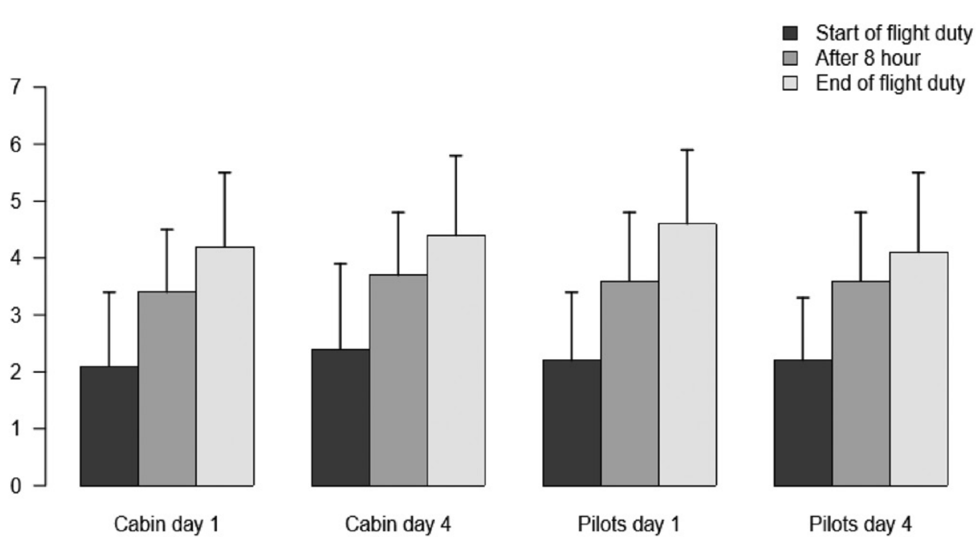

Fig. 4. Mean scores and SD, according to the Samn-Perelli Fatigue Check List, during day 1 and day 4 among cabin crewmembers (*N* = 41 d 1, *N* = 26 d 4) and pilots (*N* = 18 d 1, *N* = 16 d 4).

Table II. ACT: Reaction Time (RT) in ms, with Confidence Intervals (95% CI), in Cabin Crewmembers and Pilots at Baseline (BL), After Workday 1, and After Workday 4.

					UNADJUSTED				ADJUSTED			
					DAY 1 vs. BL		DAY 4 vs. BL		DAY 1 vs. BL		DAY 4 vs. BL	
VARIABLE	PROFESSION	BL	DAY 1	DAY 4	B (95% CI)	*P*	B (95% CI)	*P*	B (95% CI)	*P*	B (95% CI)	*P*
Overall	Cabin	797.0	790.7	761.4	−6.4 (−24.4, 11.7)	0.49	−35.7 (−57.4, −13.9)	0.002	10.5 (−9.8, 30.9)	0.31	−5.8 (−33.4, 21.7)	0.68
	Pilots	821.4	809.3	783.9	−12.1 (−39.7, 15.4)	0.39	−37.5 (−67.0, −8.0)	0.014	7.1 (−22.9, 37.2)	0.64	−3.7 (−39.2, 31.8)	0.84
	Diff*	24.4	18.6	22.5	−9.6 (−28.5, 9.2)	0.32	−10.6 (−39.3, 18.1)	0.47	6.7 (−14.6, 28.0)	0.54	8.9 (−22.7, 40.4)	0.58
Cued	Cabin	727.9	725.4	698.0	−2.6 (−24.8, 19.7)	0.82	−30.0 (−56.7, −3.2)	0.03	13.6 (−11.8, 38.9)	0.3	−0.5 (−34.9, 33.9)	0.98
	Pilots	755.6	734.1	689.2	−21.4 (−55.4, 12.5)	0.22	−66.4 (−102.6, −30.1)	0.001	−5.0 (−42.6, 32.6)	0.79	−34.6 (−78.9, 9.8)	0.13
	Diff*	27.7	8.7	−8.8	−2.6 (−24.8, 19.7)	0.82	−21.4 (−55.4, 12.5)	0.22	13.6 (−11.8, 38.9)	0.3	−5.0 (−42.6, 32.6)	0.79
Large distance	Cabin	810.0	800.3	773.0	−9.6 (−28.5, 9.2)	0.32	−37.0 (−59.7, −14.3)	0.002	6.7 (−14.6, 28.0)	0.54	−8.1 (−37.1, 20.8)	0.58
	Pilots	831.9	821.3	801.4	−10.6 (−39.3, 18.1)	0.47	−30.5 (−61.2, 0.2)	0.054	8.9 (−22.7, 40.4)	0.58	2.5 (−34.9, 39.8)	0.9
	Diff*	21.9	21.0	28.4	−30.0 (−56.7, −3.2)	0.03	−66.4 (−102.6, −30.1)	0.001	−0.5 (−34.9, 33.9)	0.98	−34.6 (−78.9, 9.8)	0.13
Small distance	Cabin	806.2	807.6	773.6	1.4 (−19.5, 22.3)	0.89	−32.6 (−57.8, −7.4)	0.013	14.2 (−9.9, 38.4)	0.25	−9.8 (−42.5, 23.0)	0.56
	Pilots	835.0	828.4	804.3	−6.6 (−38.5, 25.3)	0.69	−30.7 (−64.8, 3.4)	0.08	7.4 (−28.3, 43.2)	0.68	−5.3 (−47.5, 36.9)	0.81
	Diff*	28.8	20.8	30.7	−37.0 (−59.7,−14.3)	0.002	−30.5 (−61.2, 0.2)	0.054	−8.1 (−37.1, 20.8)	0.58	2.5 (−34.9, 39.8)	0.9

Linear mixed models analyses with random intercept for subjects and adjustments for age, sex, and practice effects.
B: Estimate of difference between day 4 and 1.
* Difference between pilots and cabin crewmembers.

multiple flight sectors and the number of takeoffs and landings per pilot during the work period varied between 10 and 20. For cabin crewmembers, multiple flight sectors involve a potential increase in workload due to the need to monitor a higher number of passengers and perform multiple safety and service tasks for each flight. In the course of a day with 5–6 flight sectors, cabin crewmembers may attend to up to 700 passengers. For airline pilots, the multiple takeoffs and landings involve the highest cognitive demand and represent increased workload.[19] Harsh weather conditions, heavy traffic, delays, and other unforeseen irregularities could exaggerate the strain of multiple takeoffs and landings. In the Nordic countries, harsh weather conditions and less daylight during the winter months represent an additional operational challenge. Because of this recognized hazard, the airline from which the subjects in the current study were recruited strive to avoid the longest work schedules during the winter months and, consequently, none of the evaluated work periods occurred during the months of November, December, and January. This eliminated certain operational conditions, which probably would have influenced the workload, and thus the results in the present field study.

RT and number of lapses have also been assessed in other studies of flight personnel. Thomas et al.[25] administered the Psychomotor Vigilance Test and found a significant difference between rested and fatigued airline pilots, with the fatigued pilots showing a faster mean RT, while the number of lapses remained unchanged. The authors argue that their tests may not be sufficiently sensitive to fatigue.[25] In our study, we observed that the number of errors as measured by the SART remained statistically unchanged from baseline to the subsequent tests. This finding could be due to the test not being sufficiently sensitive to the effects of fatigue, or that the actual work periods were not sufficiently strenuous to produce a risk of making errors. It could also be explained by inadequate adjustments for practice effects.

Although the SP scores on workdays indicate somewhat strenuous days of work, questionnaire-based registrations reflect subjective perceptions, and a variety of factors of both psychological and situational origins may influence questionnaire responses.[27] The focus on fatigue in aviation, the effect of being studied, or the interpretation of the consequences of the research may explain some of the discrepancies between the results based on self-report and the neurobehavioral results.

Other factors that may play a role include the conditions under which one performs the task, the time of the day the test is performed, and whether one is tired or hungry.[6] The neurobehavioral tests were all performed in the evening. However, due to the slight variations in subjects' schedules and irregularities of arrival times, it was not possible to carry out the tests at exactly the same time, which may have influenced the results of these tests.

A strength of the current study is the use of a case-crossover design, which minimizes confounding effects.[23] In addition the study was conducted in a real-life situation with detailed exposure information, thus increasing its external validity. A limitation is the relatively small number of subjects. Another

Table III. ACT: Number of Correct Responses (*N*) and Precision (P), with Confidence Intervals (95% CI), in Cabin Crewmembers and Pilots at Baseline (BL), After Workday 1 and After Workday 4.

					UNADJUSTED				ADJUSTED			
VARIABLE	**PROF.**	**BL**	**DAY 1**	**DAY 4**	**B (95% CI)**	***P***	**B (95% CI)**	***P***	**B (95% CI)**	***P***	**B (95% CI)**	***P***
N cued	Cabin	28.0	28.5	29.3	0.5 (−1.0, 2.0)	0.48	1.3 (−0.5, 3.1)	0.15	−1.1 (−2.7, 0.5)	0.19	−1.6 (−3.8, 0.6)	0.15
	Pilots	28.7	29.7	32.5	1.0 (−1.2, 3.3)	0.37	3.8 (1.4, 6.2)	0.003	−0.8 (−3.2, 1.6)	0.49	0.5 (−2.4, 3.3)	0.75
	Diff†	0.7	1.2	3.2	−6.4 (−24.4, 11.7)	0.49	−12.1 (−39.7, 15.4)	0.39	10.5 (−9.8, 30.9)	0.31	7.1 (−22.9, 37.2)	0.64
N LDist*	Cabin	62.8	65.7	72.8	2.9 (−1.1, 6.9)	0.16	10.0 (5.3, 14.8)	7.6e-05	−2.5 (−6.7, 1.7)	0.25	0.3 (−5.4, 5.9)	0.93
	Pilots	62.9	69.7	74.4	6.8 (0.7, 12.8)	0.032	11.5 (5.0, 18.0)	0.0008	−0.2 (−6.4, 6.0)	0.94	0.1 (−7.2, 7.4)	0.98
	Diff†	0.1	4.0	1.6	−35.7 (−57.4, −13.9)	0.002	−37.5 (−67.0, −8.0)	0.014	−5.8 (−33.4, 21.7)	0.68	−3.7 (−39.2, 31.8)	0.84
N overall	Cabin	154.6	162.7	174.1	8.1 (0.2, 15.9)	0.047	19.5 (10.0, 29.0)	1e-04	−4.6 (−12.5, 3.2)	0.25	−3.2 (−13.8, 7.4)	0.56
	Pilots	156.4	170.6	182.1	14.2 (2.2, 26.2)	0.022	25.7 (12.9, 38.5)	0.0002	−1.8 (−13.4, 9.8)	0.76	−0.8 (−14.5, 12.9)	0.9
	Diff†	1.8	7.9	8.0	1.4 (−19.5, 22.3)	0.89	−6.6 (−38.5, 25.3)	0.69	14.2 (−9.9, 38.4)	0.25	7.4 (−28.3, 43.2)	0.68
N S Dist**	Cabin	41.6	44.5	46.7	2.9 (0.3, 5.5)	0.033	5.1 (2.0, 8.2)	0.002	−0.9 (−3.6, 1.8)	0.51	−1.7 (−5.4, 1.9)	0.36
	Pilots	41.9	45.8	47.6	3.9 (−0.1, 7.9)	0.058	5.7 (1.4, 9.9)	0.01	−0.8 (−4.8, 3.2)	0.69	−2.2 (−6.9, 2.5)	0.37
	Diff†	0.3	1.3	0.9	−32.6 (−57.8, −7.4)	0.013	−30.7 (−64.8, 3.4)	0.08	−9.8 (−42.5, 23.0)	0.56	−5.3 (−47.5, 36.9)	0.81
P Cued	Cabin	77.8	79.4	81.4	1.6 (−2.6, 5.7)	0.46	3.6 (−1.4, 8.5)	0.16	−3.0 (−7.4, 1.5)	0.2	−4.6 (−10.6, 1.5)	0.14
	Pilots	79.7	82.6	90.3	2.9 (−3.4, 9.2)	0.37	10.6 (3.9, 17.3)	0.003	−2.3 (−8.9, 4.3)	0.49	1.3 (−6.5, 9.1)	0.75
	Diff†	1.9	3.2	8.9	0.5 (−1.0, 2.0)	0.48	1.0 (−1.2, 3.3)	0.37	−1.1 (−2.7, 0.5)	0.19	−0.8 (−3.2, 1.6)	0.49
P LDist*	Cabin	58.2	60.9	67.4	2.7 (−1.0, 6.4)	0.15	9.2 (4.8, 13.7)	8e-05	−2.3 (−6.1, 1.6)	0.26	0.2 (−5.0, 5.4)	0.94
	Pilots	58.2	64.5	68.9	6.3 (0.6, 11.9)	0.031	10.6 (4.7, 16.6)	0.001	−0.2 (−5.9, 5.5)	0.94	0.1 (−6.7, 6.8)	0.98
	Diff†	0.0	3.6	1.5	1.3 (−0.5, 3.1)	0.15	3.8 (1.4, 6.2)	0.003	−1.6 (−3.8, 0.6)	0.15	0.5 (−2.4, 3.3)	0.75
P overall	Cabin	61.4	64.6	69.1	3.2 (0.1, 6.4)	0.043	7.7 (3.9, 11.4)	0.0001	−1.8 (−4.9, 1.3)	0.26	−1.3 (−5.5, 2.9)	0.54
	Pilots	62.1	67.7	72.3	5.6 (0.9, 10.4)	0.022	10.2 (5.1, 15.3)	0.0001	−0.7 (−5.3, 3.9)	0.76	−0.3 (−5.7, 5.1)	0.91
	Diff†	0.7	3.1	3.2	2.9 (−1.1, 6.9)	0.16	6.8 (0.7, 12.8)	0.032	−2.5 (−6.7, 1.7)	0.25	−0.2 (−6.4, 6.0)	0.94
P S Dist**	Cabin	57.8	61.8	64.8	4.0 (0.4, 7.7)	0.031	7.0 (2.7, 11.4)	0.002	−1.2 (−4.9, 2.5)	0.52	−2.4 (−7.5, 2.6)	0.35
	Pilots	58.2	63.6	66.1	5.4 (−0.1, 10.9)	0.058	7.9 (2.0, 13.8)	0.001	−1.1 (−6.6, 4.4)	0.69	−3.0 (−9.5, 3.5)	0.37
	Diff†	0.4	1.8	1.3	10.0 (5.3, 14.8)	7.6e-05	11.5 (5.0, 18.0)	0.001	0.3 (−5.4, 5.9)	0.93	0.1 (−7.2, 7.4)	0.98

Linear mixed models analyses with random intercept for subjects and adjustments for age, sex, and practice effects.
B: Estimate of difference between workday 4 and 1.
* Large distance; **small distance; †difference between pilots and cabin crewmembers.

Table IV. SART: Number of Errors of Commission (NOC), Precision (Prec), and Reaction Time (RT) in ms, with Confidence Intervals, in Cabin Crewmembers and Pilots at Baseline (BL), After Workday 1, and After Workday 4.

					UNADJUSTED				ADJUSTED			
VARIABLE	**PROFESSION**	**BL**	**DAY 1**	**DAY 4**	**B (95% CI)**	***P***	**B (95% CI)**	***P***	**B (95% CI)**	***P***	**B (95% CI)**	***P***
NOC	Cabin	13.5	14.1	14.8	0.6 (−0.8, 2.1)	0.38	1.3 (−0.4, 3.0)	0.14	0.2 (−1.4, 1.9)	0.79	0.5 (-1.7, 2.8)	0.65
	Pilots	11.9	11.7	12.1	−0.2 (−2.4, 2.0)	0.85	0.1 (−2.2, 2.4)	0.91	−0.6 (−3.1, 1.8)	0.61	-0.7 (-3.6, 2.2)	0.65
	Diff*	-1.6	-2.4	-2.7	0.6 (−0.8, 2.1)	0.38	−0.2 (−2.4, 2.0)	0.85	0.2 (−1.4, 1.9)	0.79	-0.6 (-3.1, 1.8)	0.61
Prec	Cabin	87.8	88.8	86.0	1.0 (−2.8, 4.9)	0.61	−1.9 (−6.4, 2.7)	0.43	4.6 (0.3, 8.9)	0.038	4.7 (-1.1,10.5)	0.11
	Pilots	91.1	93.0	93.2	1.9 (−4.0, 7.8)	0.53	2.0 (−4.2, 8.3)	0.52	6.1 (−0.3, 12.5)	0.065	9.3 (1.8,16.9)	0.017
	Diff*	3.3	4.2	7.2	1.3 (−0.4, 3.0)	0.14	0.1 (−2.2, 2.4)	0.91	0.5 (−1.7, 2.8)	0.65	-0.7 (-3.6, 2.2)	0.65
RT	Cabin	347.0	330.3	323.7	−16.7 (−27.8, −5.6)	0.00	−23.2 (−36.6, −9.8)	0.001	−12.3 (−25.2, 0.7)	0.067	-15.3 (-32.9, 2.3)	0.092
	Pilots	367.8	362.1	356.3	-5.7 (−22.7, 11.3)	0.52	−11.4 (−29.6, 6.8)	0.22	0.3 (−18.9, 19.6)	0.97	-2.0 (-24.8,20.8)	0.86
	Diff*	20.8	31.8	32.6	1.0 (−2.8, 4.9)	0.61	1.9 (−4.0, 7.8)	0.53	4.6 (0.3, 8.9)	0.038	6.1 (-0.3,12.5)	0.065

Linear mixed models analyses with random intercept for subjects and adjustments for age, sex, and practice effects.
B: Estimate of difference between day 4 and 1.
* Difference between pilots and cabin crewmembers.

weakness is that the practice effect was not fully adjusted for when analyzing the change in ACT and SART outcomes from baseline to day 4. Furthermore, the few individuals at the second baseline (day 6) influence the precision of the estimated practice effect. However, when studying the impact of the number of flight sectors, check-in times for duties, and duration of the workdays with respect to the neurobehavioral outcomes, the models did fully adjust for the practice effect.

An increasing number of flight sectors during the 4 d was associated with a significantly longer RT among both pilots and cabin crewmembers. However, subjects' precision remained unchanged. While subjective fatigue increased during the 4-d work period, the overall objective performance, as measured by the neurobehavioral tests, did not reveal any deterioration when focusing on work hours only. This emphasizes the importance of looking at both work content and work hours when studying the effect of long workdays. The present study illustrates the complexity of factors that contribute to the development of fatigue among airline crew. Further research with a larger sample size would strengthen the validity of our findings.

ACKNOWLEDGMENTS

Authors and affiliations: Elisabeth M. Goffeng, M.Sc., Karl-Christian Nordby, M.D., Ph.D., Lars Ole Goeffeng, Ph.D., Øivind Skare, Ph.D., Didrik Lilja, M.Sc., and Jenny-Anne S. Lie, Ph.D., National Institute of Occupational Health, Oslo, Norway, and Anthony Wagstaff, M.D., Ph.D., and Anders Meland, Ph.D., Norwegian Armed Forces Medical Services, Institute of Aviation Medicine, Oslo, Norway.

REFERENCES

1. Avers K, Johnson WB. A review of Federal Aviation Administration fatigue research: transitioning scientific results to the aviation industry. Aviat Psychol Appl Hum Factors. 2011; 1(2):87–98.
2. Bourgeois-Bougrine S, Cabon P, Mollard R, Coblentz A, Speyer J-J. Fatigue on aircrew from short-haul flights in civil aviation: the effects of work schedules. Human Factors and Aerospace Safety. 2003; 3(2):177–188.
3. Brezonakova A. Pilot burnout as a human factor limitation. Transportation Research Procedia. 2017; 28:11–15.
4. Caldwell JA. Fatigue in aviation. Travel Med Infect Dis. 2005; 3(2):85–96.
5. Castro M, Carvalhais J, Teles J. Irregular working hours and fatigue of cabin crew. Work. 2015; 51(3):505–511.
6. Deary IJ, Liewald D, Nissan J. A free, easy-to-use, computer-based simple and four-choice reaction time programme: the Deary-Liewald reaction time task. Behav Res Methods. 2011; 43(1):258–268.
7. Goode JH. Are pilots at risk of accidents due to fatigue? J Safety Res. 2003; 34(3):309–313.
8. Heitz RP. The speed-accuracy tradeoff: history, physiology, methodology, and behavior. Front Neurosci. 2014; 8:150.
9. Hirshkowitz M, Whiton K, Albert SM, Alessi C, Bruni O, et al. National Sleep Foundation's sleep time duration recommendations: methodology and results summary. Sleep Health. 2015; 1(1):40–43.
10. Honn KA, Satterfield BC, Mccauley P, Caldwell JL, Van Dongen H. Fatiguing effect of multiple take-offs and landings in regional airline operations. Accid Anal Prev. 2016; 86:199–208.
11. ICAO. Fatigue management. [Accessed May 2018]. Available from https://icao.int/safety/fatiguemanagement.

12. Knauth P. Extended work periods. Ind Health. 2007; 45(1):125–136.
13. Lim J, Dinges DF. A meta-analysis of the impact of short-term sleep deprivation on cognitive variables. Psychol Bull. 2010; 136(3):375–389.
14. Lindström K, Elo A, Skogstad A, Dallner M, Gamberale F, et al. QPS Nordic: general Nordic questionnaire for psychological and social factors at work: user's guide. Copenhagen: Nordic Council of Ministers; 2000.
15. Meland A, Ishimatsu K, Pensgaard AM, Wagstaff A, Fonne V, et al. Impact of mindfulness training on physiological measures of stress and objective measures of attention control in a military hnelicopter unit. Int J Aviat Psychol. 2015; 25(3−4):191–208.
16. O'Hagan AD, Issartel J, Fletcher R, Warrington G. Duty hours and incidents in flight among commercial airline pilots. Int J Occup Saf Ergon. 2016; 22(2):165–172.
17. Pallesen S, Bjorvatn B, Nordhus IH, Sivertsen B, Hjørnevik M, Morin CM. A new scale for measuring insomnia: the Bergen Insomnia Scale I. Percept Mot Skills. 2008; 107(3):691–706.
18. Petrilli RM, Roach GD, Dawson D, Lamond N. The sleep, subjective fatigue, and sustained attention of commercial airline pilots during an international pattern. Chronobiol Int. 2006; 23(6):1357–1362.
19. Powell DM, Spencer MB, Holland D, Broadbent E, Petrie KJ. Pilot fatigue in short-haul operations: effects of number of sectors, duty length, and time of day. Aviat Space Environ Med. 2007; 78(7):698–701.
20. Roach GD, Sargent C, Darwent D, Dawson D. Duty periods with early start times restrict the amount of sleep obtained by short-haul airline pilots. Accid Anal Prev. 2012; 45:22–26.
21. Robertson IH, Manly T, Andrade J, Baddeley BT, Yiend J. Oops!': Performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. Neuropsychologia. 1997; 35(6): 747–758.
22. Samn S, Pirelli L. Estimating aircrew fatigue: a technique with application to airlift operations. Brooks AFB (TX): USAF School of Medicine; 1982. Technical Report SAM-TR-82-21.
23. Silva IDS. Cancer epidemiology. Lyon (France): IARC; 1999.
24. Theeuwes J, Chen CYD. Attentional capture and inhibition (of return): the effect on perceptual sensitivity. Percept Psychophys. 2005; 67(8): 1305–1312.
25. Thomas LC, Gast C, Grube R, Craig K. Fatigue detection in commercial flight operations: results using physiological measures. Procedia Manufacturing. 2015; 3:2357–2364.
26. Tucker P. What constitutes an unsafe work schedule? Scan J Work Environ Health. 2011; 37(3):169–171.
27. Williamson A. Using self-report measures in neurobehavioural toxicology: can they be trusted? Neurotoxicology. 2007; 28(2):227–234.