



Original article

Scand J Work Environ Health [2003;29\(4\):261-269](#)

doi:10.5271/sjweh.730

Effects of the implementation of an 84-hour workweek on neurobehavioral test performance and cortisol responsiveness during testing

by [Persson R](#), [Ørbæk P](#), [Ursin H](#), [Kecklund G](#), [Österberg K](#), [Åkerstedt T](#)

Affiliation: Division of Occupational and Environmental Medicine, Lund University Hospital, SE-22185 Lund, Sweden. Roger.Persson@ymed.lu.se

Refers to the following texts of the Journal: [2000;26\(3\):219-226](#)
[1998;24 suppl 3:69-75](#)

The following articles refer to this text: [2006;32\(4\):318-327](#);
[2006;32\(5\):349-358](#)

Key terms: [84-hour workweek](#); [cognitive performance](#); [cortisol responsiveness](#); [long workhours](#); [neurobehavioral test performance](#); [shift work](#)

This article in PubMed: www.ncbi.nlm.nih.gov/pubmed/12934719



This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

Effects of the implementation of an 84-hour workweek on neurobehavioral test performance and cortisol responsiveness during testing

by Roger Persson, PhD,¹ Palle Ørbæk, PhD,¹ Holger Ursin, PhD,² Göran Kecklund, PhD,³ Kai Österberg, PhD,¹ Torbjörn Åkerstedt, PhD³

Persson R, Ørbæk P, Ursin H, Kecklund G, Österberg K, Åkerstedt T. Effects of the implementation of an 84-hour workweek on neurobehavioral test performance and cortisol responsiveness during testing. *Scand J Work Environ Health* 2003;29(4):261–269.

Objectives This study examined whether long workhours in combination with an extended workweek (12 hours/7 days), as requested by the workers, impaired attention and cognitive performance and whether the degree of hypothalamic-pituitary-adrenal (HPA) activation was related to the response to the performance tasks.

Methods A group of 41 male construction workers between 21 and 60 (mean 39) years of age who worked 84 hours a week, with alternate weeks off, was compared with a group of 23 male construction workers between 24 and 65 (mean 43) years of age who had a traditional 40-hour work schedule. Neurobehavioral test performance, self-ratings of fatigue and sleepiness, and salivary cortisol levels were evaluated in a counterbalanced repeated-measures design.

Results The 84-hour group did not show any signs of reduced test performance or elevated fatigue and sleepiness. The 84-hour group had faster reaction times on day 7 than on days 1 and 5. Although the expected activation of the HPA axis was only found in the total study sample when workdays 1 and 5 were collapsed, the HPA activation can be considered normal.

Conclusions The results suggest that an 84-hour work regimen in response to requests from workers does not induce more performance deficits than an ordinary 40-hour workweek. An extended work schedule of 84 hours cannot in the short-term be considered to affect basic mental capabilities negatively.

Key terms cognitive performance, long workhours, shift work.

The effects of long workhours on health and performance have been much debated. Potential benefits, which may be of a social, economic, or individually health promoting nature, are often contrasted against the assumed risks of negative effects on individual health and safety, as well as the possibility of negative social consequences (1–3). Even if working more than 48 to 56 hours a week is considered potentially harmful, current scientific evidence is inadequate to give any firm recommendations about long workhours (4–6). There is a need to supplement existing knowledge with detailed information regarding the multitude of alternative work schedules that are becoming increasingly implemented in a

variety of organizational and industrial settings. In the present case, construction workers building a bridge across the strait between Sweden and Denmark in the Malmö region suggested an 84-hour workweek (12 hours/7 days) with alternate weeks off. Working 84 hours a week with such a combined physically and mentally demanding job assignment was suspected to lead to accumulated fatigue and, therefore, compromise the workers' physical and mental status and jeopardize safety at work. For this reason, the work schedule was initially questioned by the local labor inspectorate, but was later accepted subject to medical supervision and evaluation. To investigate the possible adverse effects of this

¹ Division of Occupational and Environmental Medicine, Department of Laboratory Medicine, Lund University, Sweden.

² Department of Biological and Medical Psychology, University of Bergen, Norway.

³ National Institute for Psychosocial Factors and Health, Stockholm, Sweden.

Reprint requests to: Dr Roger Persson, Division of Occupational and Environmental Medicine, Lund University Hospital, SE-22185 Lund, Sweden. [E-mail: Roger.Persson@ymed.lu.se]

schedule, a multitude of methods and effect markers was used, for example, questionnaires, hormones in blood and saliva, heart rate variability, muscular fatigue, objective sleep monitoring, and neurobehavioral tests.

This paper presents the evaluation of the effects of a long workweek on neurobehavioral test performance and cortisol responsiveness during testing. One aim was to examine whether working long hours would lead to a decline in attention and cognitive performance. A supplementary aim was to investigate whether the hypothalamic-pituitary-adrenal (HPA) axis was properly activated during neurobehavioral testing and to determine whether HPA responsiveness was related to test performance. Since short-term (ie, 30 minutes) work with a visual display unit (VDU) with high demands on speed and accuracy has been reported to be associated with an increase in salivary cortisol levels (7), indicating a general stress response during such exposure, one would expect to observe a similar rise in cortisol levels during neurobehavioral testing if the participants are able to mobilize sufficient motivation and arousal across the week.

Participants and methods

Participants

Two groups of male construction workers were examined. The first group (N=41) had volunteered to work cycles of 84 hours a week (12 hours/7 days) followed by 7 days off. The daily workhours were between 0700 and 1900, including a sedentary 15-minute transport to and from an off-shore location in the middle of the strait. The second group (N=23) of construction workers worked a regular 40-hour week (8 hours/5 days) between 0700 and 1530 at an on-shore location in the harbor. The mean age of the 84-hour group was 39.2 (SD 10.4, range 21–60) years, and that of the 40-hour group was 42.5 (SD 13.9, range 24–65) years. The participants

were identified in collaboration with the employer, and inclusion in the groups was determined by their working under one of the five specified company foremen who were in charge of the workers on the different work-sites. It should be noted that the foremen in charge of the 84-hour group had a first-hand opportunity to select and hire workers when the project was started. Five participants in the 84-hour group were omitted from the statistical analysis since their participation rate was too low (ie, less than 2 of the 3 days of measurement). Their reasons for absence were vacation, common cold, working in key positions, and the like. Thirty-eight persons participated on all three workdays. In the 40-hour group, one participant was omitted from the statistical analysis due to poor cooperation. Altogether, 64 participants provided sufficient data for the analysis. When the testing was initiated, the 84-hour group had worked for approximately 6 months, whereas the 40-hour group had worked for approximately 8 months. Most of the participants were from the region and returned home after each workday. One-third of the participants in the 84-hour group commuted on a weekly basis, as they came from more distant parts of Sweden. During the workweek these long-distance commuters shared ordinary apartments, two by two, in the vicinity of the transport harbor. Questionnaire data showed that most of the participants in both groups slept well and were satisfied with the amount of sleep they got (table 1). Electrocardiographic data showed that the physical fitness level (estimated by heart rate during rest) of the two groups was similar (table 2).

Work conditions

Both groups were engaged in heavy construction work, involving different types of concrete work and steel reinforcement work. The main task for the 84-hour group was to build two H-pylons (rising 204 meters above sea level) to support the double decked cable-stayed highway and railroad bridge. The 40-hour group on

Table 1. Sleep data at baseline for the present investigation, during which the participants in each group had equivalent work conditions.

	84-hour group (N=40)		40-hour group (N=23)	
	Mean	95 % CI	Mean	95 % CI
Overall sleep quality (1=very poor; 5=very good)	4.18	3.96–4.39	4.09	3.77–4.40
Perceived satisfaction with the current amount of sleep (% satisfied)	88	74–95	78	58–90
Subjective sleep length during workdays (hour:minutes)	7:04	6:52–7:17	6:49	6:35–7:03
Signs of sleep disturbances during the preceding 6 months				
Awakening score ^a (1=almost everyday; 5=never)	4.17	3.97–4.36	4.17	3.86–4.49
Disturbed sleep ^a (1=almost every day; 5=never)	4.28	4.11–4.44	4.31	4.00–4.63

^a Cronbach's alpha for the awakening score (three items) reached 0.69 in the 84-hour group and 0.73 in the 40-hour group. For the disturbed sleep score (four items) the Cronbach's alpha reached 0.66 in the 84-hour group and 0.89 in the 40-hour group.

Table 2. Physical workload estimated by electrocardiographic recordings.^a

	84-hour group						40-hour group			
	Workday 1(N=31)		Workday 5(N=31)		Workday 7 (N=31)		Workday 1(N=24)		Workday 5(N=24)	
	Mean	95 % CI	Mean	95 % CI	Mean	95 % CI	Mean	95 % CI	Mean	95 % CI
Heart rate during rest (beats/min)	59	56–61	57	55–60	56	54–59	61	56–65	61	56–65
Average heart rate during work (beats/min)	84	80–87	84	81–86	83	81–85	95	88–101	92	88–101
Peak heart rate during work (beats/min)	143	139–148	143	137–149	141	137–146	134	125–142	130	125–142
Heart rate ratio (%)	21	19–22	21	20–23	21	20–22	28	24–32	26	23–29

^a The electrocardiographic data were obtained during different weeks (but with no known worktask variations) than the tests. The 10 youngest participants in the 84-hour group were not subjected to the recordings because of a lack of proper equipment. The maximum heart rate was calculated as $210 - (0.662 \times \text{age})$ (14). The percentage of the heart rate ratio was then calculated as $(\text{HR}_{\text{work}} - \text{HR}_{\text{rest}}) / (\text{HR}_{\text{max}} - \text{HR}_{\text{rest}}) \times 100$.

shore made the bridge components on which the railroad tracks rest (ie, the railroad trough). For both groups there were three scheduled breaks during the workday (ie, morning, lunch, and afternoon). The morning and afternoon breaks lasted 15 minutes. The lunch break lasted 30 minutes. Due to the location of the worksites, unplanned interruptions owing to disturbances in logistics or bad weather conditions occurred more frequently for the 84-hour group. Electrocardiographic data, which were obtained during different weeks but with no known worktask variations, showed that the 40-hour group had a higher workload despite the similar worktasks (table 2).

Design

The 84-hour group was tested on days 1, 5, and 7, whereas the 40-hour group was tested on days 1 and 5. Since the design aimed to compare the 84-hour group with the 40-hour group, tests on days 1 and 5 were given in a counterbalanced order to minimize the influence of possible training and learning effects. For the 84-hour group, the test session on day 7 was always scheduled last.

Measures

Karolinska Sleepiness Scale (KSS). A slightly modified version of the KSS (8), in which the intermediate scale steps without verbal anchoring were removed, was responded to on a VDU. The subjects rated their current sleepiness on the following 5-point scale: (i) very alert, (ii) alert, (iii) neither alert nor sleepy, (iv) sleepy, no difficulty remaining awake, and (v) extremely sleepy, fighting sleep.

Fatigue symptom ratings. Eleven items, inspired by, and partially derived from, previous research (9), were used to assess aspects of physiological and mental fatigue. The items were presented on a VDU and responded to

on a 4-point scale indicating degree of compliance with a verbal expression concerning, for example, difficulties to concentrate, feeling stressed, overworked, and the like. A higher score indicated more symptoms of fatigue. In our present study, a global fatigue index comprised of all 11 items was used. The internal consistencies for this index were calculated as Cronbach's alpha and ranged between 0.74 and 0.85 across groups and workdays.

The neurobehavioral tests were selected for their empirically demonstrated high sensitivity to detect minor performance decrements caused by, for example, subtle brain dysfunction (10).

WAIS-R Digit Symbol: a test of perceptual and fine motor speed. With the WAIS-R Digit Symbol test the participants were instructed to enter code symbols as rapidly as possible in empty squares according to a code list that covered a series of numbers ranging from 1 to 9. The number of correctly substituted symbols within 90 seconds constituted the score (11). To reduce learning effects in the WAIS-R Digit Symbol substitution test, three different versions, one version for each specific workday, were used.

APT two-choice visual reaction time (APT RT-2). This simple reaction time test is part of the Automated Psychological Test (APT) system developed by Levander & Elithorn (12). The participants were instructed to react as fast as possible to a white square presented either to the left or to the right of a computer screen, and the responses were given with either of two corresponding keys. A fixation cross was presented in the center of the screen. Altogether, 50 stimuli were presented during approximately 5 minutes with an interstimulus interval of 2–6 seconds. The individual results were expressed as level (the mean of the 50 reaction-time responses) and variation (the standard deviation of the 50 reaction-time responses).

APT two-choice visual reaction time with sound inhibition (APT inhibition). This complex reaction-time test is similar to the APT RT-2, but the participants are supposed to inhibit a response if an auditory signal occurs at the same time as a visual stimulus (go/no go criteria). Half of the stimuli (ie, 25) are accompanied by the auditory stimulus, occurring at random. APT inhibition incorporates a preparatory stimulus (ie, the fixation cross-starts to flash before the squares are presented). The results were expressed in terms of (i) level, (ii) variation, and (iii) failed inhibition percentage.

APT k test: a test of selective attention. Each subject was told to scan a video screen in order to decide whether or not the letter *k* was present in a set of nontarget letters. If the letter *k* was present, the participants were instructed to respond by pressing a pair of buttons using the middle fingers. When the letter *k* was absent, the subject was encouraged to press another pair of buttons using the index fingers. One letter *k* was absent or present with equal probability (ratio 0.50) among a total of 10 letters. Approximately 100 sets of letters were presented during 5 minutes. The result was expressed as (i) reaction-time level of correct hits (covering all spatial locations on the screen) and (ii) error rates (missed and false hits).

Salivary cortisol

To investigate each participant's psychophysiological response to the neurobehavioral tests, salivary cortisol was sampled. Both cortisol and cortisone were analyzed by high-performance liquid chromatography (HPLC), and the total amount of corticosteroid was used in the statistical analyses. Saliva (2–5 ml) was collected in a 10-ml glass tube, frozen, and sent frozen for the HPLC analysis. Thawed samples were treated with 100% methanol, extracted (Bond Elut C18, Varian, Harbor City, CA, USA), separated (Supercosil LC8, Supelco, Bellefonte, PA, USA), and detected by ultraviolet light (spectra 100 at 240 nm). The data were recorded and integrated by WINer on Windows Labnet (Thermo Separation Products, San José, CA, USA). The coefficients of variance (CV) for within days were from 5.4% to 8.9%, and for between days they varied from 7.1% to 8.8%.

Procedure

Testing was carried out (for both groups) in the morning between 0800 and 1130 in stationary work barracks placed in the immediate proximity to the worksites. There was no systematic difference in the timing of the tests between the groups. Each session lasted approximately 30 minutes. Typically, four subjects were tested simultaneously under the supervision of two, or three,

psychologists. The first sampling of salivary cortisol preceded the tests. To stimulate saliva excretion, the participants chewed on a piece of paraffin. The session then started by having them fill out a computer-administered questionnaire concerning current state of sleepiness and fatigue. The tests were then given in the following order: WAIS-R Digit Symbol, APT RT-2, APT inhibition, and finally the *k* test. The second sampling of salivary cortisol followed the tests. Since the participants had been awake for at least 2 hours before the testing, the sampling of salivary cortisol occurred in a phase in which cortisol levels are expected to show physiologically a slight decline during a 30-minute period.

Ethics

All of the participants gave their written informed consent to participate. The Ethics Committee of Lund University approved the study (LU 63–98).

Data management

Descriptive data: questionnaire and electrocardiographic data. The awakening score was calculated as the mean score of three items that assessed (i) ease of awakening, (ii) whether the sleep was refreshing, and (iii) exhaustion at awakening. The score ranged from one to five. Higher scores represented a more refreshing sleep that was easily terminated. The overall sleep quality was assessed with one question. The score ranged from one to five, and a higher score represented a better quality of sleep. The disturbed sleep score was calculated as the mean score of four items that assessed whether the subject had (i) difficulties falling asleep, (ii) disturbed or restless sleep, (iii) repeated awakenings, and (iv) premature awakenings. The score ranged from one to five. Higher scores represented less disturbed sleep. The sleep time during workdays was calculated as the difference between the reported time for night sleeping and reported time for awakening. The perceived satisfaction with the attained amount of sleep item was dichotomized according to yes and no answers. The results were expressed as the percentage satisfied (yes answers) within each group. Confidence intervals for the percentage were calculated according to Altman et al (13). Maximum heart rate was calculated as $210 - (0.662 \times \text{age})$ (14). The percentage of the heart rate ratio (HRR%) was then calculated as $(\text{HR}_{\text{work}} - \text{HR}_{\text{rest}}) / (\text{HR}_{\text{max}} - \text{HR}_{\text{rest}}) \times 100$.

Outcome measures: self-ratings and neurobehavioral tests. The neuropsychological test scores and the KSS scores were calculated as raw scores if nothing else was indicated. The ratings of the individual fatigue items

were averaged to form a fatigue index score. The distribution of every test variable and rating scale score was plotted and visually inspected for deviations from normality. Positively skewed deviations from normality were found for the two APT RT-2 variables and the three APT RT-inhibition variables. The APT *k* error variable and the fatigue index score were also found to be positively skewed. To comply with normality assumptions, the positively skewed variables in which only positive values were achievable were logarithmically transformed with the base of 10. The log-transformed variables were antilogged and converted back into the original scale after the statistical testing. Hence the log-transformed variables are presented as geometric means with accompanying confidence intervals. To lessen the impact of extreme scores for the APT RT-inhibition and APT *k* test error variables, which incorporated possibilities to achieve nonpositive values (ie, zero), the data were subjected to square root transformation.

Statistical analysis

Most of the statistical computations were made with SPSS 11.0.1 (15). P-values below 0.05 were considered statistically significant. With the use of the General Linear Model module, univariate two-way repeated-measures analyses with group as a between-participant factor and age as a covariate evaluated the group differences in the mean scores for the neuropsychological tests. If a regression with $P < 0.10$ was found between the covariate age and a test score, the former was retained in the final model. As a consequence, nearly all the group computations involving neurobehavioral test scores were adjusted for age (ie, all except the APT RT failed inhibition variable). Age did not yield any statistically significant effects on the cortisol measures.

The main question of the study was tested by the interaction effect between workday (day 1 and day 5) and group (84-hour versus 40-hour). For the 84-hour group only, the possible effect of all workdays (ie, day 1 versus day 5 versus day 7) on the tests and ratings was also evaluated. Analyses were performed on the participants that had a complete set of data for the particular items or tests. In the 84-hour group the homogeneity of variance assumption was tested with Mauchly's test of sphericity. If the sphericity assumption was violated, the Geisser-Greenhouse F-test was used to adjust the degrees of freedom in order to increase the critical F-value; this procedure lessened the risk for type I errors. Within-group post-hoc comparisons were made with the Bonferroni-adjusted paired t-test. Given that the within-subject factor represents ordered levels, polynomial contrasts were used to evaluate the linear and quadratic effects. The polynomial contrasts were also screened for

higher order contrasts. No higher order contrast fitted better than any linear contrast.

A univariate three-way repeated-measures analysis was used to evaluate the possible differences in the cortisol levels for the 30 and 15 participants from the 84-hour and 40-hour groups, respectively, who left a complete set of data for workdays 1 and 5. There were no differences in the mean cortisol levels between the participants with a complete set of data and those with an incomplete dataset. Specifically, the analysis involved two factors with repeated measurements with two levels each, that is, workday (day 1 and 5) and pre-post (pretest and posttest levels of cortisol), as well as one between group factor (84-hour versus 40-hour group). The main question to be evaluated was tested by the three-way interaction between workday, group, and pre-post.

A univariate two-way repeated-measures analysis was used to evaluate the possible differences in the cortisol levels for the 25 subjects in the 84-hour group who had a complete dataset for workdays 1, 5, and 7. There were no differences in the mean cortisol levels between the participants with or without complete sets of data.

Results

Comparisons between the 84-hour and 40-hour work schedules

Self-rated sleepiness and fatigue symptoms. There was no group difference for sleepiness or fatigue development between days 1 and 5. The 40-hour group had higher KSS and fatigue index scores than the 84-hour group (table 3).

Neurobehavioral test performance. There was typically no group difference for performance development between days 1 and 5. Only one interaction effect between workday and group, involving the APT *k*-test reaction-time variable, was detected (table 1). This interaction was due to the slower reaction times observed in the 40-hour group on day 1, which subsequently became faster and comparable with those observed in the 84-hour group on day 5. Furthermore, the 40-hour group exhibited larger variation scores than the 84-hour group in the APT RT-2 and APT RT-inhibition tests.

Salivary cortisol. No interaction indicating group differences in physiological activation between the days was observed. The levels of cortisol were, on the average, higher after the testing than before it [$F = 6.15$, $df = 1$, $P = 0.017$, partial $\eta^2 = 0.13$] (table 4).

Table 3. Descriptive data and P-values and partial eta squares from the univariate two-way repeated-measures analyses of variance with group as a between-participants factor. (WAIS = Wechsler Adult Intelligence Scale, APT = Automated Psychological Test, RT = reaction time)

	84-hour group (N=41)				40-hour group (N=23)				Two-way repeated measures analysis					
	Workday 1		Workday 5		Workday 1		Workday 5		Main effect workday		Between-participants effect: group		Interaction effect group × workday	
	Mean	95% CI	Mean	95% CI	Mean	95% CI	Mean	95% CI	P	partial η^2	P	partial η^2	P	partial η^2
Self-rating measures														
Karolinska Sleepiness Scale	2.02	1.77–2.02	2.17	1.94–2.40	2.30	1.90–2.70	2.52	2.16–2.89	0.913	..	0.018	0.09	0.813	..
Fatigue symptom index ^a	1.17	1.11–1.23	1.23	1.16–1.31	1.30	1.19–1.42	1.50	1.35–1.68	0.183	..	0.001	0.18	0.069	..
Neurobehavioral tests														
WAIS-R Digit Symbol	51.3	48.2–54.4	51.0	47.2–54.9	45.6	40.8–50.4	46.7	42.4–51.0	0.405	..	0.142	..	0.416	..
APT RT-2, level (ms) ^a	330	316–345	321	308–334	334	310–360	336	311–364	0.990	..	0.721	..	0.185	..
APT RT-2, variation (ms) ^a	63	58–67	63	56–70	74	64–85	72	63–83	0.534	..	0.042	0.07	0.763	..
APT RT inhibition, level (ms) ^{a, b, c}	420	387–456	402	375–430	426	387–469	430	397–465	0.149	..	0.863	..	0.219	..
APT RT inhibition, variation (ms) ^{a, b, c}	84	71–99	78	69–88	99	85–115	104	84–128	0.188	..	0.030	0.08	0.189	..
Neurobehavioral tests														
APT RT inhibition, failed inhibition ratio (%) ^b	7.5	5.4–9.6	7.8	5.7–9.9	8.3	4.4–12.3	7.5	4.1–10.8	0.969	..	0.761	..	0.509	..
APT k test, RT level for correct responses (ms)	1445	1369–1520	1448	1378–1517	1602	1470–1733	1467	1352–1583	0.609	..	0.223	..	0.022	0.08
APT k test, error responses (%)	3.1	2.2–3.9	3.0	1.3–4.7	2.8	1.8–3.9	5.6	0.4–10.9	0.035	0.07	0.670	..	0.239	..

^a Results expressed as geometric means with accompanying confidence intervals.

^b For the 84-hour group N = 40.

^c For the 40-hour group N = 22.

Table 4. Mean salivary cortisol levels (ng/l) prior to and after the neurobehavioral tests in the 84-hour and 40-hour groups.

	84-hour group (N=41)				40-hour group (N=23)			
	Pretest		Posttest		Pretest		Posttest	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Workday 1	14.35	4.37	16.23	4.47	15.22	8.17	17.84	6.10
Workday 5	13.51	6.04	14.38	6.92	12.37	8.16	14.33	11.29

^a Computations performed on the 30 and 15 participants from the 84-hour and 40-hour groups, respectively, who had left a complete set of data. This analysis showed a significant main effect indicating higher cortisol levels after the testing than before the testing [$P = 0.017$, partial $\eta^2 = 0.13$]. No interaction effects were found.

Evaluation of performance and cortisol excretion across workdays in the 84-hour group

Self-rated sleepiness and fatigue symptoms. The number of days worked influenced the feelings of fatigue but not sleepiness (table 5). The fatigue index scores increased across the days. Differences were mainly observed when day 7 was compared with day 1 or 5.

Neurobehavioral test performance. The number of days worked influenced performance with respect to the

following tests and variables: WAIS-R Digit Symbol, APT RT-2 variation, APT RT inhibition level, and APT k-test level for correct responses (table 5). These differences, which emerged either between days 1 and 7 or between days 5 and 7, showed that performance became faster and less variable after working 7 days. The evaluation with polynomial contrasts suggested a gradual and linear progress for some performance variables.

Salivary cortisol measurement. There were no overall differences in the cortisol levels after the testing compared with before the testing [$F=1.57$, $df=1$, $P=0.222$], and no interaction occurred between workday and the factor pre-post [$F=1.69$, $df=2$, $P=0.195$]. The cortisol levels generally decreased as a consequence of the number of days worked [$F=4.32$, $df=1.53$, $P=0.029$, partial $\eta^2=0.15$]. Specifically, the pretest levels on day 1 were higher than the pretest levels on day 7, and all the posttest levels of salivary cortisol deviated significantly from each other on all workdays (table 6). Polynomial contrast indicated a linear trend for the salivary cortisol levels, the means decreasing over the workweek [$F=17.40$, $df=1$, $P<0.001$, partial $\eta^2=0.42$].

Table 5. Descriptive data and results from the univariate two-way repeated-measures analyses. (WAIS = Wechsler Adult Intelligence Scale, APT = Automated Psychological Test, RT = reaction time)

	84-hour group (N= 38)						One-way repeated-measures analysis	
	Workday 1		Workday 5		Workday 7		Main effect workday	Polynomial contrast (linear effect)
	Mean	95 % CI	Mean	95 % CI	Mean	95 % CI		
Self-rating measures								
Karolinska Sleepiness Scale	2.05	1.79–2.32	2.21	1.97–2.45	2.24	1.96–2.52	F=0.792, df=2, P=0.457	F=1.27, df=1, P=0.268
Fatigue symptom index ^a	1.18	1.12–1.25	1.24	1.16–1.32	1.39 ^{b,c}	1.29–1.50	F=12.34, df=2, P<0.001 (partial $\eta^2=0.25$)	F=17.38, df=1, P<0.001 (partial $\eta^2=0.32$)
Neurobehavioral tests								
WAIS-R Digit Symbol	51.3	48.0–54.5	51.4	47.4–55.4	56.0 ^{b,c}	52.4–59.6	F=14.66, df=2, P<0.001, (partial $\eta^2=0.28$)	F=40.54, df=1, P<0.001, (partial $\eta^2=0.52$)
APT RT-2, level (ms) ^a	324	315–347	320	309–333	324	304–339	F=2.56, df=2, P=0.084	F=2.20, df=1, P=0.146
APT RT-2, variation (ms) ^a	62	57–67	63	55–71	54 ^{b,c}	49–59	F=4.85, df=2, P=0.010, (partial $\eta^2=0.12$)	F=11.46, df=1, P=0.002, (partial $\eta^2=0.23$)
APT RT inhibition, level (ms) ^{a,d}	410	382–441	402	373–432	367 ^{b,c}	345–392	F=8.42, df=2, P=0.001, (partial $\eta^2=0.19$)	F=14.24, df=1, P=0.001, (partial $\eta^2=0.28$)
APT RT inhibition, variation (ms) ^{a,d}	80	69–93	78	68–89	68	59–80	F=2.58, df=2, P=0.082	F=4.45, df=1, P=0.042, (partial $\eta^2=0.11$)
APT RT inhibition, failed inhibition ratio (%) ^d	7.9	5.7–10.1	8.1	5.8–10.4	5.3	4.0–6.7	F=2.22, df=2, P=0.116	F=2.83, df=1, P=0.101
APT k test, RT level for correct responses (ms) ^e	1465	1389–1542	1451	1381–1521	1320 ^{b,c}	1260–1381	F=14.87, df=2, P<0.001, (partial $\eta^2=0.29$)	F=29.50, df=1, P<0.001, (partial $\eta^2=0.45$)
APT k test, error responses (%) ^e	2.7	1.9–3.5	2.9	1.1–4.7	2.6	1.8–3.4	F=0.34, df=1.69, P=0.674	F=0.46, df=1, P=0.502

^a Results expressed as geometric means with accompanying confidence intervals.

^b Post hoc paired samples t-test, P<0.05, workday 1 versus workday 7.

^c Post hoc paired samples t-test, P<0.05, workday 5 versus workday 7.

^d APT RT inhibition: N=36.

^e APT k test: N=37.

Table 6. Pre- and posttest mean levels of salivary cortisol across the workweek for the 84-hour group.

	84-hour group (N=25)											
	Workday 1				Workday 5				Workday 7			
	Pretest		Posttest		Pretest		Posttest		Pretest		Posttest	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Salivary cortisol (ng/l)	14.82 ^b	4.32	16.66 ^{a,b}	4.47	13.49	5.64	14.42 ^{b,c}	7.45	12.56	3.55	12.48	3.71

^a Post hoc paired samples t-test P<0.05, workday 1 versus workday 5.

^b Post hoc paired samples t-test P<0.05, workday 1 versus workday 7.

^c Post hoc paired samples t-test P<0.05, workday 5 versus workday 7.

Discussion

Our study disclosed no evidence of performance decrement in neurobehavioral tests or of increasing symptoms of fatigue or sleepiness as a result of an 84-hour workweek in comparison with corresponding results for an ordinary 40-hour workweek. Contrary to expectations, the 40-hour group reported more sleepiness and fatigue and had occasionally a more variable and slower performance than the 84-hour group. Interestingly, the 84-hour group had the fastest reaction times and the smallest variation scores on day 7. At first glance, this finding might be suspected to indicate training or learning

effects, as day 7 was not balanced out in the study design. However, a previously repeated administration of the same neurobehavioral test has never suggested any signs of learning, in either the short term (ie, 4 times within 2 hours) or the intermediate term (the same procedure repeated 1 week later) (13). On the other hand, training or learning effects cannot be ruled out for the WAIS-R Digit Symbol test (16, 17). An alternative interpretation of the lower and less variable reaction times is excessive arousal on day 7. This position is supported by the self-reports of low levels of fatigue and sleepiness prior to the testing, as well as the fact that unpublished observations have shown that the 84-hour group

required at least two major sleep episodes (ie, two consecutive nights' sleep) in order to reach baseline levels of fatigue and sleepiness. Following this line of reasoning, a central question becomes whether an increase in arousal is detrimental or beneficial. From a traditional perspective, one might argue that the shorter reaction times reflect improved performance, and are desirable, whereas a more unconventional approach would be to suggest that the central nervous system may accelerate towards a state of dysfunction. Yet another possibility is that the faster and less variable performance on day 7 in the 84-hour group was due to more motivation, as the subjects were likely to have positive expectations of the forthcoming leave. It is also imaginable that these subjects, in order to retain the current work schedule and the benefits there of, have had an unspoken ambition to perform well the last workday. Regardless of the reason, a motivation hypothesis finds some support in the fact that the decreases in reaction times seemed to become more prominent as a function of the complexity of the tests. For example, between days 5 and 7, no difference in mean reaction time was observed for the APT RT-2 test, whereas the mean reaction time of the slightly more cognitively demanding APT inhibition test decreased by 40 milliseconds ($Z=-0.42$). In the APT k test, which involved a markedly more complex visual search component, the mean reaction time decreased by 130 milliseconds ($Z=-0.62$). Possibly more complex tasks leave more room for motivation to become an issue; thus trying harder pays off.

The observed rise in cortisol levels in the total study sample during the neurobehavioral testing agrees fairly well with previous findings concerning changes in saliva cortisol levels in response to short-term stress among VDU operators (7). Even if mean levels indicate diminishing cortisol responsiveness during testing in the 84-hour group, the intraindividual variability between days was large, and no workday by pre-post interaction was observed. Similarly, the general rise in cortisol during the testing was not observed when the development across all 7 days was studied. Remarkably, the 84-hour group displayed a general decrease in absolute levels of cortisol across the week. At a glance this finding may give the impression of the start of a fatigue reaction in the HPA axis, as a response to prolonged demands and too little sleep. This impression is reinforced by the agreement with falling morning serum cortisol and serum testosterone levels during the workweek, found in the same group during another workweek (18). On the other hand, when the data are considered in combination with the observation that the melatonin level of the 84-hour group was markedly higher on day 1 than on days 5 and 7 (18), another explanation may be that the drop in cortisol merely reflects

the normalization process after the phase shift that occurred when the participants entered the work schedule and were forced to wake up earlier.

At this point, some methodological issues need to be addressed. The present design has more than sufficient power to detect a difference between the means of the size of one standard deviation, both regarding main effects and interactions. Clinical experience indicates that smaller effects are hardly worth detecting. Hence, failures to detect workday by group interactions are not primarily due to a lack of statistical power. It is more likely that practical circumstances may have reduced the possibilities of finding the expected negative effects of the 84-hour workweek. One should, for example, note that our study focused on the potential effects of accumulated fatigue on cognitive performance. It is possible that the 84-hour group would have reported more fatigue and performed worse if the measures had been obtained at the end of the shifts. Moreover, despite great effort to find similar worktasks, the electrocardiographic recordings showed that the 40-hour group worked at a higher percentage of their cardiovascular capacity. Since it seems reasonable to assume that workers who feel out of shape, or have a tendency to become easily fatigued, do not likely volunteer for this type of extreme work scheduling, one might think that this observation reflects selection bias due to fitness. However, in view of the group's similar heart rates during rest, this seems not to be a plausible selection mechanism. It is more likely that other incentives have affected the allocation of the participants, for example, economic or social reasons. Nonetheless, even if major construction work is likely to be associated with a considerable degree of self-selection that may be caused by a wide range of factors that do not necessarily have medical relevance, one should note that the participants first and foremost were likely to be representative of their own, and similar, lines of work. As with our findings, however, many studies of compressed work scheduling have failed to find effects on performance (19). Our results stress the fact that long workhours are not automatically associated with performance decrements. It seems apparent that work content and context and employed strategies to execute work (individual as well as organizational) are important determinants of the impact of a work schedule on the individual worker's performance and health.

In conclusion, the expected consequences of fatigue build-up were not observed. The results showed that implementing a demanding 84-hour work regimen in response to the workers' requests does not necessarily create poorer cognitive performance than an ordinary 40-hour workweek. Working an extended work schedule of 84 hours cannot in the short term be considered to affect basic mental capabilities negatively.

Acknowledgments

Lena Seger, MSc, is acknowledged for her supervision of the collection of the data. Emma Karlsson, MSc, and Sofia Hansson, MSc, gave valuable assistance during the collection of the data. Randi Espelid provided skillful technical assistance during the analysis of the salivary samples.

AFA insurances, Stockholm, Sweden, supported this study financially.

References

1. Rosa RR. Performance, alertness, and sleep after 3.5 years of 15-hour shifts: a follow-up study. *Work Stress* 1991;5:107–16.
2. Gillespie A. A comparison of a 12-hour and eight-hour shift system. *Nurs Times* 1996;39:36–9.
3. Tucker P, Barton J, Folkard S. Comparison of eight and 12 hour shifts: impacts on health, wellbeing and alertness during the shift. *Occup Environ Med* 1996;53:767–72.
4. Harrington JM. Shift work and health: a critical review of the literature on working hours. *Ann Acad Med Singapore* 1994;23:699–705.
5. Rosa RR. Extended workshift and excessive fatigue. *J Sleep Res* 1995;4 Suppl 3:52–6.
6. Spurgeon A, Harrington JM, Cooper CL. Health and safety problems associated with long working hours: a review of the current position. *Occup Environ Med* 1997;54:367–75.
7. Schreinicke G, Hinz A, Kratzsch J, Huber B, Voigt G. Stress-related changes of saliva cortisol in VDU operators. *Int Arch Occup Environ Health* 1990;6:319–21.
8. Åkerstedt T, Gillberg M. Subjective and objective sleepiness in the active individual. *Int J Neurosci* 1990;52:29–37.
9. Bjorvatn B, Kecklund G, Åkerstedt T. Bright light treatment used for adaptation to night work and re-adaptation back to day life: a field study at an oil platform in the North Sea. *J Sleep Res* 1999;8:105–12.
10. Österberg K, Ørbæk P, Karlson B, Bergendorf U, Seger L. A comparison of neuropsychological tests for the assessment of chronic toxic encephalopathy. *Am J Ind Med* 2000;6:666–80.
11. Wechsler D. WAIS-R . [Wechsler adult intelligence scale-revised]. Stockholm: Psykologiförlaget; 1992. [Manual in Swedish]
12. Levander S, Elithorn A. The automated psychological test system: manual. Trondheim (Norway): Department of Psychiatry and Behavioral Medicine, University of Trondheim; 1987.
13. Altman DG, Machin D, Bryant TN, Gardner MJ, editors. *Statistics with confidence*. 2nd ed. Bristol: BMJ Books; 2000.
14. Bruce RA, Fischer LD, Cooper MN, Gey GO. Separation of effects of cardiovascular disease and age on ventricular function with maximal exercise. *Am J Cardiol* 1974;34:757–63.
15. SPSS Inc. *SPSS Base 9.0 for windows: user's guide*. Chicago (IL): SPSS Inc; 1998.
16. Österberg K, Ørbæk P, Karlson B, Seger L, Åkesson B, Bergendorf U. Psychological test performance during experimental challenge to toluene and n-butyl acetate in cases of solvent-induced toxic encephalopathy. *Scand J Work Environ Health* 2000;26:219–26.
17. Otto DA, Hudnell HK, House DE, Molhave L, Counts W. Exposure of humans to a volatile organic mixture, I: behavioral assessment. *Arch Environ Health* 1992;47:23–30.
18. Ørbæk P, Kecklund G, Seger L, Åkerstedt T. Arbetsvecka på 84 timmar: Trötthet, sömnmönster och stressreaktioner [84-hour workweek: fatigue, sleep patterns and stress reactions]. Lund: Department of Occupational and Environmental Medicine, Lund University, and National Institute for Psychosocial Medicine, Stockholm, Sweden; 2000.
20. Lowden A, Kecklund G, Axelsson J, Åkerstedt T. Change from an 8-hour shift to a 12-hour shift, attitudes, sleep, sleepiness and performance. *Scand J Work Environ Health* 1998;24 Suppl 3:69–75.

Received for publication: 31 December 2002