Effects of sleep disruption on cognitive performance and mood in medical house officers

Citation for published version:

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
BMJ

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
Effects of sleep disruption on cognitive performance and mood in medical house officers

IAN J DEARY, ROSEMARY TAIT

Abstract
Twelve medical house officers were tested on a battery of memory, concentration, and work related tasks after three conditions: a night spent off duty; a night spent on call; and a night spent admitting emergency cases. Short term recall, but not digit span, concentration, or work related abilities, was impaired after a night of emergency admissions. A night spent on call had no effect on cognitive performance. Self reported mood scores showed that house officers were more deactivated (indicating a lack of vigour and drive) after nights of emergency admissions but not after nights on call. Significant between subject differences were found for five of the eight cognitive tests.

Though loss of sleep and long hours of work have an effect on memory and mood, the individual differences among doctors are the main source of the variance in performance of tasks.

Introduction
Junior house officers work long hours and experience sleep deprivation and sleep disruption. There is evidence from experiments on non-medical subjects that both conditions impair mental test performance and, by implication, work efficiency. Tasks that are boring and lengthy—for instance, card sorting and vigilance tasks—are more affected by sleep loss than are short tasks.

These effects appear to be related to fatigue rather than reduced motivation, as monetary reward for performance in vigilance tasks after loss of sleep has only a short lived improving effect. Sleep disruption, as opposed to total sleep loss, is more common and also has effects on cognitive performance. The intermittent disruption of sleep for two nights amounting to no more than one hour’s loss of sleep a night was found in one study to impair performance on reaction time, addition, and digit symbol substitution tasks to the same extent as one night of complete sleep loss.

Junior doctors’ subjective reports agree with these findings. In a survey of 2452 junior and senior house officers asked whether they thought their hours of duty so long as to impair their efficiency 37·3% replied “often” and 47·6% “occasionally.” There are few studies, however, of the effects of sleep loss and sleep disruption on the mental test and work related efficiency of house officers. House officers were studied after sleep loss (about 31-36 hours without sleep) and after a night off duty. Fatigued house officers made significantly more errors on a sustained attention detection task (diagnosing arrhythmias in electrocardiograms) and reported feeling more fatigued, more sad, less surjent, less elated, less egotistic, less vigorous, and of lower social affection on the Nowlis mood adjective check list. More recently a study of 27 interns examined sleep, mood states, and psychophysiological performance changes in a series of four sessions spread across the internship year. Of the mood states, only anger increased significantly over the year, the others (tension, depression, vigour, fatigue, and confusion) remaining unaltered. Interns who slept less became more angry, but both increasing anger and lack of sleep appeared to improve performance on choice reaction time and critical flicker fusion tasks.

The authors concluded that the internship year was not as “dramatically stressful as previously reported.” A study in Britain which examined the performance of two groups of 15 junior doctors on a three minute card sorting task that required grammatical reasoning found that scores became reliably worse in one group after a sleep debt of three to five hours but not in the second group. The second group also carried out a laboratory form sorting test and their performance in this test improved after a sleep debt of four to five hours but became worse after a sleep debt of eight hours. The authors concluded that compensation for sleep loss can be maintained for up to three minutes in challenging tasks.

We report a study aimed at improving on the few previous studies by (a) using a broader range of performance tests which are psychometrically sound and relevant to doctors’ duties; (b) testing a more homogeneous sample of British house officers (volunteer rates are often not reported but are as low as 31% in some series and may have included a disproportionate number who perform well when tired); (c) testing mood and performance of each house officer after each of three conditions—namely, a night off duty, a night on call, and a night spent admitting emergency cases (“waiting”); (d) describing the sleep, work, and leisure patterns of the group; and (e) examining the extent to which there are between subject differences

Department of Psychology, University of Edinburgh, Edinburgh EH8 9JZ
IAN J DEARY, BSc, MB, lecturer in psychology
ROSEMARY TAIT, MA, psychology student
Correspondence and requests for reprints to: Dr Deary.
in performance and the extent to which these are maintained across experimental conditions. This aspect is missing from other studies.

Present study

SUBJECTS

We obtained permission from the physicians’ committee and junior doctors’ committee to approach all medical house officers in a large Edinburgh teaching hospital. Twelve of the 13 house officers who were approached agreed to take part (seven men, five women). All had completed a house officer post in surgery and all testing was done between mid-February and mid-April.

DESIGN

We used a repeated measures design. We chose this design because it provides an estimate of the test performance variance that is due to differences between subjects as well as that which is due to the experimental conditions. Each subject was tested three times. To control for practice and sequence effects the order of conditions for each subject was determined by using a Latin square. Thus practice and sequence effects were distributed across the cells of the design and into the error variance term. When alternative forms of tests were used each form was spread across experimental conditions and order by using a Latin square.

TEST BATTERY

The Novel mood adjective check list contains 46 adjectives, each of which is rated on a four point scale from zero (subject “definitely does not” feel the particular mood) to 3 (subject “definitely does” feel the mood indicated). These give scores on 11 mood factors (see below) which may be condensed to two higher order factors which have various labels but are often referred to as ‘general tone’ and ‘vigor’.10

Digit span is a standard test of short term memory from the Wechsler adult intelligence scale—revised.11 Both forward and backward forms of the test were given. The test resembles the setting in which a house officer receives verbal reports of laboratory results, short instructions, telephone numbers, and so on.

The serial 13s test is an adaptation of the standard serial 7s test of concentration. House officers were required to count backwards accurately from 200 in steps of 13. Total time to reach the end point (5) was recorded. This was used in attempt to mimic the type of arithmetical skill and concentration required to calculate drug dosages and infusion rates.

Logical memory—immediate—entails being read a passage of prose which contains 23 facts and attempting to recall as many facts as possible when the reading stops. Three passages were selected from the Wechsler memory scale12 and were intended to simulate the setting in which a doctor receives a verbal report of a patient’s case history.

Logical memory—delay—After the immediate logical memory test the same passage was repeated and the doctor’s recall of the passage delayed until the next two tests had been completed. This is a modified form of the delayed memory test from the Wechsler memory scale and tests memory for a prose passage after a period of distraction.

Information processing was indexed by using a laboratory report sorting task adapted from that used elsewhere.9 House officers were presented with 50 clinical chemistry reports which contained patients’ results and normal ranges for various values. Subjects were required to sort the reports as quickly and accurately as possible into those with any abnormal results and those with none. The reports often contained data from more than one date, whereupon house officers were required to make a judgment based on the most recent results. Measures of sorting time and sorting errors were taken.

Electrocardiogram assessments—Twenty seven electrocardiographic records were collected over two months before the study. These were chosen to represent patterns corresponding to the main myocardial infarctions and arrhythmias and various printout formats from a series of emergency and elective electrocardiographs in the hospital. These were then made into three parallel nine item tests which, when examined independently by a senior electrocardiography technician and two lecturers in clinical medicine, appeared to be of equal diagnostic difficulty. House officers were allowed 10 minutes to diagnose the tracings in each test. These were scored at a later date by a lecturer in clinical medicine who was a specialist in cardiology. Each item was marked on a three point scale as 0 (unacceptable or no answer), 1 (acceptable answer, main diagnosis not exactly correct), and 2 (main diagnosis correct). House officers are often the first on the scene when an electrocardiogram has been recorded and this test was an attempt to measure their acumen in this aspect.

Questionnaire—We designed a sleep, work, and leisure questionnaire, which was completed at the conclusion of each test session.

PROCEDURE

Subjects were tested three times. Testing took 45 minutes to one hour. To eliminate effects of time of day all testing was performed between 2 pm and 5 pm on the day after each of the following conditions: (a) off duty—the previous evening and night were spent at home; (b) on call—the previous evening and night were spent attending calls to usually two or three general medical wards, but no new admissions were accepted or expected; (c) without—the previous evening and night might have spent both admittance and emergency cases to wards and looking after patients already in two or three general medical wards.

All subjects were assured of complete confidentiality and at the completion of the study were informed of the main findings in a group setting.

ANALYSIS

There were no missing values in the data. Analysis of cognitive and mood changes across the three experimental conditions was by one way analysis of variance for repeated measures. Analysis of variance is valid when populations are symmetrical or skewed in a similar fashion and when the largest variance is less than four or five times that of the smallest.13 Our data met these requirements. Analysis of variance for repeated measures was a suitable test in our study because it allowed us to separate the variance in cognitive and mood scores into three portions: variance due to our three conditions; variance due to differences among subjects; and error variance. Thus the present test with three conditions and mood score. One indicates whether the three sleep conditions contributed variance and the other whether subjects differed significantly in their cognitive abilities.

Between subject variance is often left unreported in repeated measures experiments but we thought it important to discover the various sources of differential test performance. When F values were not significant pairwise comparisons between conditions were carried out. This was a legitimate procedure as two of the conditions studied (on call and waiting) entail a degree of sleep disruption and might lead to similar results. To avoid type I errors we used Dunn’s test for a priori pairwise comparisons.14 Dunn’s tests were set at p<0.05 for each pair of comparisons (off duty versus on call, off duty versus waiting) within each cognitive test or mood scale.

Results

House officers slept a mean of 39.2 ± 1 hours a week. House officers slept a median of 7 hours (range 6-10) when off duty, 5 hours (0-7) when on duty, and 1-5 hours (0-7) when waiting. Pairwise comparisons showed that, as predicted, off duty nights allowed more sleep than either on call nights (t=3.91, df=11, p<0.005; one tailed) or waiting nights (t=5.45, df=11, p<0.0005; one tailed). On call nights allowed more sleep than waiting nights (t=3.27, df=11, p<0.005; one tailed). When on call house officers’ sleep was disturbed a median of once (t=0.3 times). Of these, about half required the house officers to leave their beds, and these calls lasted for a median of 1 hour (range 0-5). House officers worked for a mean of 98.2 hours (SD 31.6; range 30-144) a week (determined by the number of hours that a sleep was carried, which included time spent sleeping while on call and waiting). Five subjects were on a 1:2 rota, two on a 1:2:5 rota, and five on a 1:3 rota. House officers drank a mean of 9-2 units (SD 6.5; range 0-24 units) of alcohol a week. One house officer had seen a general practitioner for anxiety related problems and none had seen a general practitioner for depression related problems.

Table I gives summary data for the eight cognitive tests. One way analysis of variance for repeated measures resulted in no significant F values. Pairwise comparisons of the off duty condition versus both on call and waiting showed that logical memory (immediate) was impaired after waiting (p<0.05).

Using one way analysis of variance for repeated measures we found significant between subject effects for the serial 13s test (F=17.46; df=11,22; p<0.001), logical memory (immediate) (F=5.37; df=11,22; p<0.01), logical memory (delay) (F=4.49; df=11,22; p<0.01), laboratory reports (sorting time) (F=9.78; df=11,22; p<0.01), and laboratory reports (sorting errors) (F=2.35; df=11,22; p<0.05). Table II shows the Pearson product moment correlation coefficients for those cognitive test results that showed significant between subject differences. Ten of the 15 correlations were
TABLE I—Mean (SD) cognitive test scores across three experimental conditions (n=12)

<table>
<thead>
<tr>
<th></th>
<th>Off duty</th>
<th>On call</th>
<th>Waiting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Digit span (forward)</td>
<td>6-67 (0.99)</td>
<td>6-75 (1.06)</td>
<td>6-42 (0.90)</td>
</tr>
<tr>
<td>Digit span (backward)</td>
<td>5-00 (1-13)</td>
<td>5-17 (0.94)</td>
<td>5-08 (1.38)</td>
</tr>
<tr>
<td>Serial 13s</td>
<td>57-6 (20-6)</td>
<td>56-0 (22-8)</td>
<td>57-4 (21-4)</td>
</tr>
<tr>
<td>Logical memory (immediate)</td>
<td>8-67 (2-90)</td>
<td>7-33 (3-92)</td>
<td>6-67 (2.81)</td>
</tr>
<tr>
<td>Logical memory (delay)</td>
<td>10-67 (4-91)</td>
<td>10-08 (4-48)</td>
<td>9-33 (3-60)</td>
</tr>
<tr>
<td>Laboratory reports (sorting time in seconds)</td>
<td>330-1 (117-3)</td>
<td>392-7 (217-17)</td>
<td>363-5 (170-6)</td>
</tr>
<tr>
<td>Laboratory reports (errors)</td>
<td>4-33 (2-10)</td>
<td>4-58 (1-62)</td>
<td>4-50 (2-81)</td>
</tr>
<tr>
<td>Electrocardiographic diagnosis</td>
<td>12-58 (2.23)</td>
<td>12-17 (2.95)</td>
<td>11-59 (3.73)</td>
</tr>
</tbody>
</table>

TABLE II—Pearson product moment correlation coefficients for house officers’ scores on cognitive tests across experimental conditions (n=12)

<table>
<thead>
<tr>
<th></th>
<th>Off duty</th>
<th>On call</th>
<th>Waiting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Serial 13s</td>
<td>0.96; p&lt;0.01</td>
<td>0.84; p&lt;0.01</td>
<td>0.81; p&lt;0.01</td>
</tr>
<tr>
<td>Logical memory (immediate)</td>
<td>0.65; p&lt;0.05</td>
<td>0.56; p&lt;0.05</td>
<td>0.58; p&lt;0.05</td>
</tr>
<tr>
<td>Logical memory (delay)</td>
<td>0.42</td>
<td>0.74; p&lt;0.01</td>
<td>0.45</td>
</tr>
<tr>
<td>Laboratory reports (time to sort)</td>
<td>0.70; p&lt;0.05</td>
<td>0.84; p&lt;0.01</td>
<td>0.78; p&lt;0.01</td>
</tr>
<tr>
<td>Laboratory reports (errors)</td>
<td>0.39</td>
<td>0.52; p&gt;0.05</td>
<td>0.05</td>
</tr>
</tbody>
</table>

TABLE III—Median (range) Novlis mood adjective check list factor scores across experimental conditions (n=12)

<table>
<thead>
<tr>
<th></th>
<th>Off duty</th>
<th>On call</th>
<th>Waiting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concentration</td>
<td>7-0 (0-11)</td>
<td>6-0 (2-8)</td>
<td>4-0 (1-12)</td>
</tr>
<tr>
<td>Aggression</td>
<td>1-0 (0-4)</td>
<td>0-5 (0-6)</td>
<td>1-5 (0-6)</td>
</tr>
<tr>
<td>Depression*</td>
<td>2-0 (0-9)</td>
<td>3-5 (0-8)</td>
<td>7-5 (2-9)</td>
</tr>
<tr>
<td>Social affection</td>
<td>6-5 (1-11)</td>
<td>6-5 (3-9)</td>
<td>6-5 (0-11)</td>
</tr>
<tr>
<td>Anxiety</td>
<td>0-5 (0-4)</td>
<td>0-5 (0-7)</td>
<td>0-5 (0-6)</td>
</tr>
<tr>
<td>Depression</td>
<td>1-0 (0-3)</td>
<td>0-5 (0-7)</td>
<td>1-5 (0-6)</td>
</tr>
<tr>
<td>Egotism</td>
<td>0-0 (0-5)</td>
<td>1-0 (0-7)</td>
<td>0-0 (0-7)</td>
</tr>
<tr>
<td>Pleasants</td>
<td>3-0 (0-10)</td>
<td>4-0 (1-10)</td>
<td>2-0 (0-8)</td>
</tr>
<tr>
<td>Activation</td>
<td>5-5 (0-9)</td>
<td>5-5 (0-7)</td>
<td>1-0 (0-9)</td>
</tr>
<tr>
<td>Nonchalance</td>
<td>2-5 (0-6)</td>
<td>2-0 (0-7)</td>
<td>1-5 (0-3)</td>
</tr>
<tr>
<td>Seclusivity</td>
<td>1-0 (0-4)</td>
<td>2-0 (0-5)</td>
<td>0-5 (0-3)</td>
</tr>
</tbody>
</table>

*One way analysis of variance: F=6.99; df=2.22; p<0.01. F Statistics for other mood factors not significant.

Many reports have indicated that there are two main higher order mood factors contained in the Novlis mood adjective check list.8 These have been termed factor A (hedonic tone; a combination of the aggression, anxiety, and depression factors from the Novlis check list) and factor B (vigour; a combination of the activation and deactivation factors from the check list). Table IV shows the mood scores for these higher order factors in the three experimental conditions. The one way analysis of variance for repeated measures showed a significant effect across conditions for vigour (F=4.21; df=2.22; p<0.05). Using Dunn’s test for pairwise comparisons we found that house officers felt more vigorous after being off duty when compared with after waiting (p<0.05) but not when compared with after being on call.

Discussion

We found that short term memory was impaired in house officers after the sleep loss and disruption caused by a night spent waiting. This adds to previous reports of decreased vigilance in sustained electrocardiogram monitoring and variable reports of decreased card sorting ability associated with logical reasoning.9 The single study that reported improvements in abilities8 was recording more elementary processes (and reporting them in the form of significant but very low correlation coefficients that accounted for between 2-6% and 6-1% of the variance in test performance), and it may be concluded that the combination of long hours of work and loss of sleep does affect some mental abilities that are related to work efficiency. A night spent on call had neither the cognitive impairing effects nor mood impairing effects of a night spent waiting. We attempted to include more directly work related tasks in our study but none of these was significantly affected by the experimental conditions. It may be that the impairment after sleep loss affects basic psychological processes such as memory rather than specific diagnostic skills. Also the impairment of a general process like short term memory may be enough to account for occasional slips made by doctors when tired, such as missing clinical signs or writing a wrong prescription.

Despite its importance the attention that has been given to cognitive change after long hours of work has overshadowed other determinants of house officers’ decision making and wellbeing. In this and other studies significant changes in mood have been found after sleep disruption. We did not simply record the mood changes in a wide range of factors but also related our scores to those mood factors which are currently thought to be most central—activation and hedonic tone. Finding that a night spent waiting significantly impairs activation has two implications. Firstly, it is important to consider how mood and cognition interact. Deactivated mood, which our house officers tended to have after a night spent waiting, might be the signal of an information processing system under stress. Thus house officers may compensate for tiredness by drawing on reserve mental capacity, but at a cost. Secondly, mood states induced by sleep loss and long hours of work may beget pathologically low moods and result in the high rates of depressive episodes reported elsewhere.9 10 Discussions on the efficiency of house officers focus on cognitive changes; if, however, there are detectable impairments in one of the two main dimensions of mood then we must ask two questions: Are there lapses in clinical judgment arising from these mood states? and Is the alteration in mood by itself reason enough to question the current work practices?

Our results suggest that there is a factor that accounts for more variation in cognitive performance levels than the effects of being on call—namely, the differences between individual doctors. In five of our eight tests the between subject effects were significant and, for the most part, reliable when tested by correlation analysis. An able doctor after a night of waiting is likely to be better than a poor doctor after a night at home (on our group of tests). Despite our finding that sleep loss impaired short term memory our raw data also show that at best one doctor recalled four facts out of 23, whereas another doctor at worst recalled nine items. In the laboratory report sorting task one house officer’s worst time over the three conditions was three minutes 10 seconds, whereas another house officer, making no fewer errors, took seven minutes 24 seconds in the fastest of the three conditions. A complete attempt at investigating the effects of doctors will have to address the question of individual differences in performance. Until now variation in efficiency has been looked for in settings alone. In a serious attempt to account for most of the variance in cognitive performance we must bring the person into the equation and acknowledge that some doctors are better at these tasks than are others. Further, we must ask whether there are specific person setting interactions rather than assume a general effect on all personalities. A starting point for this future research
Abstract

Data from the Tucson epidemiological study of airways obstructive disease on smoking of non-tobacco cigarettes such as marijuana were analysed to determine the effect of such smoking on respiratory symptoms and pulmonary function. Among adults aged under 40, 14% had smoked non-tobacco cigarettes at some time and 9% were current users. The prevalence of respiratory symptoms was increased in smokers of non-tobacco cigarettes. After tobacco smoking had been controlled for men who smoked non-tobacco cigarettes showed significant decreases in expiratory flow rates at low lung volumes and in the ratio of the forced expiratory volume in one second to the vital capacity. This effect on pulmonary function in male non-tobacco cigarette smokers was greater than the effect of tobacco cigarette smoking.

These data suggest that non-tobacco cigarette smoking may be an important risk factor in young adults with respiratory symptoms or evidence of airways obstruction.

Introduction

The adverse effects of tobacco cigarette smoking have been shown consistently in population studies. The effects of non-tobacco cigarettes have not, however, been examined in a general population. Various illicit drugs are smoked as cigarettes, but by far the most widely used in the United States is marijuana. Data from a representative population sample would provide information on the effects of "usual" non-tobacco cigarette smoking.

Subjects participating in the Tucson epidemiological study of airways obstructive disease frequently inquired whether non-tobacco cigarette smoking (specifically marijuana) should be included in the responses to the smoking questions. For this reason, questions about non-tobacco cigarette smoking were included in this survey (1981-3). We report on the results of this survey as a cross-sectional study of the effects of non-tobacco cigarettes on respiratory symptoms and pulmonary function.

Methods

The methods of selection for the study population have been described. In brief, the population is a random stratified cluster sample of households in Tucson, Arizona, enrolled in 1972-3. Details of the study questionnaire and spirometry methods have been reported.

Questionnaire and spirometric data from the seventh survey (1981-3) were available for analysis on 2251 white non-Mexican-American subjects aged over 14 years. The survey questionnaire contained questions about the duration and intensity of non-tobacco cigarette smoking and the depth of inhalation. The questions referred to "non-tobacco cigarette" smoking because of the illegality of marijuana use.

Values for the forced expiratory volume in one second (FEV1) and the forced vital capacity (FVC) were the best of at least three attempts. The flow rates at 50% (Vmax50) and 75% (Vmax75) of the expired forced vital capacity were derived from the best sum curve, FEV1+FVC. Values were expressed as percentage of predicted. Predicted values were based on the subject's age, sex, and height using prediction equations derived from asymptomatic, non-diseased, non-smoking subjects in this population.

Preliminary analyses showed that there were only two current and six ex-smokers of non-tobacco cigarettes aged over 40 years. Analyses were therefore confined to the 15-40 year age group. There were 990 subjects in this age range with questionnaire and spirometric data. Subjects were grouped according to their tobacco and non-tobacco cigarette smoking habits as current smokers, ex-smokers, and those who had never smoked. In some analyses current and ex-smokers of non-tobacco cigarettes were grouped together and referred to as "ever" smokers of non-tobacco cigarettes.

Data were processed on the DEC-10/Cyber 175 computer system of the University of Arizona. Statistical techniques included cross tabulation with χ² tests, analysis of variance (ANOVA), and logistic analysis using the statistical package for the social sciences routines.