Effects of Burnout Treatment on Cognitive Functions and on Subjective Well-Being

Jasenko Dervisic
EFFECTS OF BURNOUT TREATMENT ON COGNITIVE FUNCTIONS AND ON SUBJECTIVE WELL-BEING

Jasenko Dervisic

Burnout syndrome is increasingly affecting more people in Western countries. The question of treatment is important. In this study, we recruited and investigated 27 burnout patients (16 women, 9 men) and 20 healthy controls (8 men and 12 women) before and after an intervention. The intervention consisted of a combination of both cognitive therapy, cognitive behavioral therapy, individual counseling, and a form of mindfulness group therapy (centered around own body awareness) to see whether the treatment had an effect on cognitive functions possibly affected by burnout and on subjective well-being. A battery of neuropsychological tests and questionnaires were administered to all participants, once before treatment, and once after. All test scores were z transformed and reduced to composite variables, measuring executive function, verbal memory function and psychomotor function. No significant interaction effects between group and test occasions were found. Treatment does not appear to influence cognitive functions affected in burnout. However, reports of subjective well-being as measured by questionnaires improved after treatment.

The present thesis focuses on the neuropsychological functions in patients with “burnout” syndrome, and how these functions may change from an intervention aimed to facilitate rehabilitation. There has been an increase in many Western countries in stress-related disease and long term-term absence from work due to illness among otherwise healthy, and high performing individuals (Ahola et al., 2006; Copertaro et al., 2007; Fernandez Torres et al., 2006; Rydmark et al., 2006; Bullmann et al. 2005; Henderson et al. 2005; Shanafelt et al., 2002). The symptoms of these individuals are generally characterized by memory and concentration problems, sleeplessness, diffuse aches, fatigue, irritability, and anxiety. The symptoms are often attributed to occupational stress, and have been characterized as "burnout” syndrome (Maslach et al., 2001, Melamed et al. 2006, Shirom and Melamed 2006). The concept of “burnout” was initially used to describe the frustration and emotional detachment seen in social and health workers that had developed as result of continuous stress exposure, including the conflict of individual values and organizational demands respectively (Melamed et al. 2006, Mason 1975; Maslach et al. 2001). There is no medical diagnostic definition of “burnout” (it is mainly used by researchers), and it should therefore not be used in a medical context. It should be noted that burnout is different from depression in the way that it refers to a person’s relationship to her or his work. (Maslach, 1996). Along with the symptoms of burnout mentioned above, patients report a feeling of being emotionally drained, which they often attribute to occupational stress. It has also been found that such stressed individuals can experience an acute phase with symptoms of hypertension, chest pain, dizziness and cognitive impairment (Sandström et al., 2005). Even though many individuals recover from the acute symptoms, the cognitive and emotional dysfunction as well as the increased sensitivity to stress often lasts for months, or years, forcing the affected individuals to work part-time, change jobs or retire early. Many professionals are still not accepting the described disabilities as a medical condition, and when accepted, they tend to be misdiagnosed as having depression. Only a minor portion of the
affected individuals respond to treatment with serotonin reuptake inhibitors or other antidepressants (Asberg et al, 2010).

One of the core symptoms of burnout is cognitive weariness, or problems with both concentration and memory. This is described by Melamed and co-workers (Shirom, 1989; Melamed, Kushner and Shirom, 1992). Exhaustion Disorder, or ED, as described by the Swedish Boarder of Health And Welfare, and in the Swedish version of the International Classification of Diseases, or ICD-10 (Swedish Board of Health and Welfare, 2003), is also characterized by cognitive impairments, including working memory difficulties, episodic memory difficulties, and difficulties relating to executive functions. One of the first studies to compare patient samples with healthy controls on measures of cognitive functioning was Sandström et al (2005). In this study, a broad range of verbal tests, non-verbal cognitive tests, memory tests, visual tests and auditory attention tests was used in a female sample. They found that patients and controls performed equally well on measures of general cognitive ability and verbal memory. They also found that patients performed significantly poorer in terms of immediate and delayed recall on tests of non-verbal memory, as well as auditory and visual attention. However, in Öhman et al (2007), no group differences were found in the domains of non-verbal memory, or auditory attention and visual attention. Patients did, however, perform poorer on such tests as letter fluency, trail making, digit symbol and prospective memory. Öhman and co-authors argue that their findings hint at the presence of deficits in executive control functioning and the prefrontal cortex. In another study, Rydmark and co-authors (2006), investigated a sample of female patients and controls and found that the former performed less well compared to the latter on tests of working memory and reaction time. Osterberg et al (2009) found that, compared to a sample of healthy controls, patients had significantly more subjective complaints of cognitive impairments. However, with exception for a small impairment on a cognitive speed test among patients, both the patient group and control group performed comparably well on objective tests. In line with this, Olsson, Roth and Melin (2010) found that patients responded more quickly but made more errors on a task involving vigilance and signal detection, perhaps due to impatience. In a more recent study, Sandström et al (2011) noted that patients performed significantly poorer on measures of attention and response control, as well as visuo-spatial memory ability. They also found that personality traits such as harm-avoidance, persistence and self-directedness was related to the patients' poorer performance on cognitive tasks (Sandström et al, 2011). In sum, the literature suggests executive control functioning, verbal working memory and psychomotor speed are impaired in burnout.

Few burnout patients respond to treatment with serotonin reuptake inhibitors and antidepressants (Asberg et al 2010). However, research has shown that mindfulness meditation (Cohen-Katz et al, 2004), cognitive behavioral therapy and yoga (Granath et al, 2006), qigong (Stenlund et al, 2009) and peer-support groups (Peterson et al, 2011), all have positive effects in terms of alleviating burnout symptoms. At Stressmottagningen, a stress clinic in Stockholm, Sweden, burnout patients receive a form of stress rehabilitation which includes a combination of both cognitive therapy, cognitive behavioral therapy, individual counseling, and a form of mindfulness group therapy centered around own body awareness (http://www.stressmottagningen.nu/behandling-och-rehabilitering/individuell-stresshantering/).

**Aim**

To test the efficacy of the burnout rehabilitation program offered at Stressmottagningen, we designed a longitudinal study of burnout patients, who received intervention between Time 1 and Time 2, and healthy controls, who did not. All participants were investigated with regards to performance on a battery of neuropsychological tests selected specifically to measure aspects
of executive functioning, verbal memory function, psychomotor speed, as well as word fluency and visuo-spatial ability. These are some of the cognitive functions which according to the literature are primarily affected by burnout. In addition, we wanted to investigate self-reported levels of occupational stress and burnout using a set of questionnaires which were administered both before and after treatment. Although research on the matter is inconclusive, several studies, including a meta-analysis of gender differences in burnout (Puranova, 2010), suggests that men and women are affected differently by burnout: women report more emotional exhaustion whereas men appear to be more depersonalized. We thus also wanted to investigate whether male and female burnout patients differ in terms of test performance, on self-reported levels of stress, and finally whether they respond differently to the burnout rehabilitation offered at Stressmottagningen compared to healthy controls who did not receive any intervention.

Methods

Participants
Burnout patients were recruited from Stressmottagningen in Stockholm, Sweden. They were tested on one occasion before treatment during the year 2011-2012, and once again after treatment during 2013-2014, at Karolinska Hospital, more specifically at Astrid Lindgren’s Children’s Hospital. Healthy controls were recruited online. Most controls were students at the time at Karolinska Institute in Stockholm, Sweden. The controls were also recruited for a second visit in 2013-2014. In total, more than 50 patients, about 50 healthy male controls, as well as 50 healthy female controls were recruited for the first visit. For the second visit, 27 right-handed non-smoking burnout patients returned for testings, 54% (17 women and 10 men, mean age 38.44 ± 5.21, range 26 – 46 years, education 16.96 ± 1.70), as well as 20 healthy right-handed non-smoking controls, about 20% of the original sample (12 women, and 8 men, mean age 30.9 ± 7.79, range 20 – 44 years, education 16.42 ± 3.22). The same 27 patients and 20 controls who participated in the second visit were selected from the first round of testings. All others were subsequently omitted from this study. (See table on next page).

Table 1. Demographics and baseline data for Maslach Burnout Inventory – General Survey (MBI-GS), including sub scores for Cynicism, Professional Efficacy and Exhaustion, as well as Montgomery Asberg Depression Rating Scale (MADRS).

<table>
<thead>
<tr>
<th>Patients.</th>
<th>Controls</th>
<th>Pat Female.</th>
<th>Pat Male.</th>
<th>Control F.</th>
<th>Control Male</th>
</tr>
</thead>
<tbody>
<tr>
<td>(N=27) M(SD)</td>
<td>(N =20) M(SD)</td>
<td>(N=17) M(SD)</td>
<td>(N=10) M(SD)</td>
<td>(N = 12) M(SD)</td>
<td>(Control = 8) M(SD)</td>
</tr>
<tr>
<td>Age</td>
<td>38.44 (5.21) 30.9 (7.79)</td>
<td>36.94 (5.66) 41 (3.16)</td>
<td>30.83 (8.08) 31 (7.86)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education</td>
<td>16.96 (1.70) 16.4 (3.22)</td>
<td>16.71 (1.86) 17.4 (1.57)</td>
<td>16.66 (3.59) 15.75 (1.73)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MBI-GS Av</td>
<td>3.97 (0.57) 2.59 (0.33)</td>
<td>3.92 (0.64) 4.05 (0.42)</td>
<td>2.57 (0.39) 2.63 (0.24)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MBI-GS Cyn</td>
<td>3.25 (1.26) 1.03 (0.93)</td>
<td>3.42 (1.32) 2.96 (1.17)</td>
<td>0.82 (0.82) 1.35 (1.07)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MBI-GS Pro.Ef</td>
<td>4.16 (1.09) 4.98 (0.78)</td>
<td>3.91 (1.14) 4.59 (0.89)</td>
<td>4.93 (0.87) 5.06 (0.68)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MBI-GS Exh.</td>
<td>4.47 (0.96) 1.25 (0.73)</td>
<td>4.34 (1.1) 4.7 (0.64)</td>
<td>1.42 (0.67) 1.0 (0.78)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MADRS</td>
<td>17.55 (5.42) 4.15 (3.23)</td>
<td>18.58 (5.16) 15.8 (3.65)</td>
<td>4.16 (2.94) 4.15 (3.83)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The study was approved by the Ethics Committee at the Karolinska Institute, and written informed consent was received from each participant. Once enrolled in the study, all subjects were first screened for burnout using the Maslach Stress-Burnout Inventory – General Survey (MBI-GS) – (Schaufeli and Van Dierendonck, 1995). The MBI-GS is a seven point rating scale,
ranging from 0 (never) to 6 (daily), and it consists of three subscales: Exhaustion (five items), cynicism (five items) and professional efficacy (six items). When rating perceived stress, subjects were asked to take into consideration the previous six months, not the actual time-point. In Scandinavia, average scores for MBI-GS are around 2. Participants were also assessed with Shirom-Melamed Burnout Questionnaire (SMBM). (Melamed et al., 1999). MBI-GS and SMBM scores are highly correlated (Grossi et al., 2003). They were screened for depression using the Montgomery-Asberg Depression Scale (Montgomery et al., 1979).

Additionally, all participants were administered a shortened version of the Edinburgh Handedness Inventory (Oldfield, 1971). The Revised NEO Personality Inventory (NEO-PI-R; Costa & McCrae, 1992), which was designed to measure the Five-Factor Model of personality, was administered to each participant and used to assess self-reported personality traits. The NEO-PI-R contains 240 items which are answered on a five-point Likert scale ranging from strongly disagree to strongly agree. The NEO-PI-R divides each of the five dimensions of personality (Neuroticism, Extraversion, Openness, Agreeableness, Conscientiousness) into six smaller facets, meaning that a total of 30 facets are assessed by the questionnaire.

**Tests and Material**

The test battery which was then administered to all participants during first- and second visit of the study covered three major cognitive domains: executive functioning, verbal memory function, and psychomotor function.

**Executive functioning**

This denotes a set of different cognitive abilities that are involved in complex, goal-directed thought and behaviour (Anderson, et al. 2001). The executive dimensions assessed here were (a) focused and sustained attention, (b) set-shifting, and (c) working memory function. Focused attention was tested with the dichotic listening task, (Hugdahl 2003, Kompus, 2011), and with the digit span task forward (Weschler Adult Intelligence Scale-Revised, [WAIS-R]). Sustained attention and response control (inhibition) was tested with a version of Go/No-go and stop test (McNab et al., 2008). In the Go/No-go test the participants were asked to press the space bar in reply to the presentation of a yellow square on the computer screen. The measure parameters were the percent of commission errors. In the stop task, participants were asked to press the space bar on a computer keyboard in response to the presentation of a yellow horizontal arrow, and to inhibit this response during certain conditions. Stop signal reaction time (SSRT) was used as measure parameter. It was estimated by first calculating the average stop signal duration for each participant (the average duration of the first arrow in the stop trial after stabilization, i.e. in stop trials 20-38; see Appendix A, Supplementary material). The stop signal reaction time (SSRT) was then assessed by subtracting the mean stop signal duration from the median reaction time for the oddball trials (McNab et al., 2008). Set-shifting or cognitive flexibility was assessed using the Modified Wisconsin Card Sorting Test (MWCST; Heaton, et al, 1993), which was conducted on a portable laptop computer. It requires the participant to match a response card with a stimulus card according to a rule. The criterion variable chosen was the number of perseverative errors, i.e. how many times the subject failed to match a new card to the correct stimulus card due to a rule change.

To test verbal working memory function we administered the N-Back test, which was conducted on a portable laptop computer. In the test, each participant had to monitor, and give an answer in reply to, letters presented one by one in a sequence on a computer screen. The measure parameters used were the average RT (in ms) and score (%) in the 2-back and 3-back condition. In addition, working memory number storage capacity was assessed with the Digit Span Task (forward and backward). The measure parameter included the longest digit span the
participant was able to obtain in the forward and backward directions (Wechsler Adult Intelligence Scale-Revised [WAIS-R]).

**Verbal Memory Function**
To assess verbal memory we used The Claesson Dahl inventory of learning and memory (revised CD), as described in Jovanovic et al., 2010; Ciumas, 2010. This test measures efficacy in the ability to immediately recall auditory presented verbal materials, and the ability to recollect the words attached to the learning situation (Nyman, 1998). The Claeson-Dahl inventory of learning and memory-revised (CD) consists of two phases. The first, when the subject learns a list of 10 words to be immediately recalled after presentation, measures the efficacy in the ability for immediate recall of auditory presented verbal materials. The second phase, where the subject 30 min after presentation is asked to recall the 10 words, is designed to test the ability to recollect the words attached to the learning situation. The output parameters used is the number of times the word list is repeated until the subject can repeat all the words in the right order, and the weighted points.

**Psychomotor Function**
Psychomotor function was investigated using the Grooved Pegboard Test (Klove 1963) as described in Ciumas et al., 2010, and motor speed with Trail Making Test, Trace (Lezak 2004). In addition, we tested verbal function with the word fluency test (Spreen & Strauss 1998), and visuospatial function with Vandenbergs test of mental rotation (Vandenberg & Kruse 1978). See appendix A (Supplementary data) for detailed descriptions of all the neuropsychological tests.

**Analysis**
All relevant output variables were first adjusted so that higher scores indicated better performance. They were then transformed into z scores. For the first visit, we calculated the mean and standard deviation for each score from each test. Then, we subtracted the mean from each score and divided it by the standard deviation. For the second visit, we subtracted the mean scores from the first visit, and divided it by the standard deviations from the first visit. Furthermore, using the taxonomy described in the method, a composite variable was created in SPSS for executive function, verbal memory function and psychomotor function, respectively. Each new composite variable was created using the mean of several output variables from the different tests described in methods. This method allows for a reduction of data according to the various cognitive categories subjects’ scores fell into.

**Results**
A repeated measure ANOVA with factors time (before/after) and group (patients, controls) and sex (male/female) revealed a main effect of time on verbal function \((F(1, 30) = 18.924, p < .001)\) and psychomotor function \((F(1, 30) = 13.823, p = .003)\), where subjects received higher composite scores during the second visit, but not on executive function \((F(1, 30) < 1, p > .5)\).

However, there was no interaction between group and time (and between group, time and sex in a three-way ANOVA) on executive function \((F(1, 30) = 0.796, p = .380)\) verbal function \((F(1, 30) = 0.506, p = .483)\), and psychomotor function \((F(1, 30) = 2.123, p = .156)\). There appears to be no treatment effect. We do have effects of time, but they are the same for both groups (Figure 1).
Overall, healthy controls score significantly better than patients on tasks measuring verbal function ($F(1, 30) = 9.016, p = .015$), and there is a trend in the same direction for tests measuring psychomotor function ($F(1, 30) = 6.135, p = .057$). No difference between groups has been observed for executive function ($F(1, 30) = 0.037, p = .848$).

Furthermore, male subjects across all participants and in all groups perform better than female subjects on tests measuring executive function ($F(1, 30) = 7.694, p = .027$), but there are no significant differences in tests measuring verbal ($F(1, 30) = 5.009, p = .099$) and psychomotor functions ($F(1, 30) = 1.549, p = .669$), and no significant interactions.

All $p$-values have been corrected for three comparisons using Bonferroni corrections.

**Questionnaires – Subjective Well-Being**

A significant interaction between time (before/after) and group (patients, controls) was found for MBI-GS Exhaustion ($F(1, 41) = 26.448, p < .01$), MBI-GS Cynicism ($F(1, 41) = 6.901, p = .048$), and for Montgomery Asberg Depression Rating Scale (MADRS), ($F(1, 41) = 20.441, p < .01$), but not for MBI-GS Professional Efficacy ($p > .1$).

Within patients, a significant improvement over time was found on MBI-GS Exhaustion ($F(1, 23) = 30.934, p < .001$), MBI-GS Cynicism ($F(1, 23) = 6.922, p = .045$), and MADRS, ($F(1, 23) = 36.238, p < .001$), but not for MBI-GS Professional Efficacy ($p > .1$). Within the control group, no changes over time have been observed for any of the questionnaires (all $p > .1$). This indicates that patients are recovering in subjective exhaustion and cynicism.

All $p$-values have been corrected for four comparisons using Bonferroni corrections.
We also investigated whether the patients who declined to return, and the controls who declined to return, were any different in terms of performance on the neuropsychological tests compared to the patients and controls who returned for the second visit, which is, after the intervention. No differences were found in Verbal Memory Function \( (F(1, 116) = .693, p = .407) \), Psycho Motor Function \( (F(1, 116) = .231, p = .632) \) and Executive Function \( (F(1, 116) = .2.172, p = .143) \) in the first visit between the twenty seven patients and twenty controls who returned after the intervention and the rest, who did not. Additionally, there were no significant group interactions for Verbal Function \( (F(1, 116) = .171, p = .680) \), Psycho Motor Function \( (F(1, 116) = 1.383, p = .242) \), and Executive Function \( (F(1, 116) = .455, p = .501) \).

We performed the same analysis on the questionnaire data, to see whether the returning 27 patients and 20 controls differed in terms of subjective well-being measured by MADRS, and MBI-GS (including the sub scores for Exhaustion, Cynicism and Professional Efficacy) during the first visit compared to the patients and controls who did not return after the intervention for second rounds of tests. No differences were found in MADRS \( (F(1, 130) = 1.123, p = .291) \), MBI-GS Exhaustion \( (F(1, 130) = .394, p = .531) \), MBI-GS Cynicism \( (F(1, 130) = .281, p = .597) \) and MBI-GS Professional Efficacy \( (F(1, 130) = .139, p = .710) \) between the patients and healthy controls who returned after treatment and those who chose not to return. Moreover, there were no significant group interactions for MADRS \( (F(1, 130) = .577, p = .449) \), MBI-GS Exhaustion \( (F(1, 130) = .726, p = .396) \), MBI-GS Cynicism \( (F(1, 130) = .544, p = .462) \) and MBI-GS Professional Efficacy \( (F(1, 130) = .651, p = .421) \).

**Discussion**

The purpose of this study was to test whether the rehabilitation program offered to burnout patients at Stressmottagningen was effective, in terms of performance on neuropsychological tests designed to measure aspects of executive functioning, verbal working memory and psychomotor function – cognitive functions which, according to the literature, are affected in burnout (Öhman et al, 2007; Sandström et al, 2005) –, and on self-reported measures of perceived stress.

It appears that the composite scores for executive function and verbal memory function only improve over time, but there are no apparent differences in improvement between patients and
controls. It does however appear that female patients and male patients differ in terms of improvement on the tests which were designed to measure aspects of executive function, but not on verbal memory function and psychomotor function – but this effect is present in males across all groups. Thus, based on our results, it appears that the treatment offered at Stressmottagningen has no effect on the cognitive functions affected in burnout, and that there are no apparent gender differences in the patient group.

It should be mentioned that had they been examined independently, the individual tests which were used to calculate the composite variables would have perhaps painted a different picture regarding the effects of the intervention on burnout patients.

However, according to the scores for exhaustion and cynicism obtained from analysis of the questionnaires Maslach Burnout Inventory – General Survey (MBI-GS) and Montgomery Asberg Depression Rating Scale (MADRS), burnout patients are less exhausted and cynical after treatment. Also, their depression scores have significantly decreased after treatment, which could signify an improvement due to intervention.

But, since we couldn’t control what people did during treatment, it is difficult to draw any conclusions as to whether patients were less exhausted, less cynical and less depressed due to the intervention, or whether it was due to some other, confounding factor. Although the research suggests that burnout is resistant to phenomena like spontaneous recovery (Melamed et al, 2006), lack of control over the treatment and of what patients did in their spare time could have affected the results. For example, in one instance, a patient reported practicing yoga during their spare time. Moreover, there was no clinical control group (i.e. untreated patients), which could have further controlled for spontaneous recovery.

Additionally, many patients declined to return after treatment for the second round of neuropsychological testing, which could also have affected the outcome of the comparisons between cognitive functions, or could have rendered them inconclusive in terms of evaluating the effects of the intervention on cognitive functions in burnout. One of the major methodological problems, in particular in a longitudinal study, is attrition, or the loss of participants (Gustavson et al, 2012). Psychological distress, socio-economic status, unemployment can influence drop-out rate and affect the data in some way, deteriorating the generalizability. In order to investigate whether the patients and controls who returned for second round of testing differed in any way from those who did not, we compared the results obtained from them during the first visit on the neuropsychological tests and on the questionnaires. No significant differences were found. However, the sample of patients and healthy controls who returned could still, in theory, have been different from the ones who dropped out based on some other factors which were not addressed by the questionnaires, such as their quality of sleep, family responsibilities, or other uncontrollable factors.

Future studies should aim for better control over the treatment, and better control over what people did during their treatment, in their spare time and in between sessions of treatment. Additionally, there should be a clinical control group of patients that do not receive treatment, or placebo treatment. Furthermore, the control group in the current study did not control for anything but the neuropsychological tests and stress questionnaires. In order to draw any conclusion on the effects of the intervention, perhaps they should have had a form of placebo treatment as well.
References


Supplementary material

Below is a detailed description of the neuropsychological tests.

**Executive functioning**

**Dichotic Listening** During the dichotic listening task (1), the participants hear two auditory stimuli simultaneously, with one sound presented to each ear, and are asked to report which stimulus they heard first or best. The auditory stimuli used in this study consisted of consonant-vowel (CV) syllables. Each syllable contained one of the six stop consonants /b/, /d/, /g/, /p/, /t/ or /k/, combined with the vowel /a/. Thirty-six stimulus-pairs such as /ba/-/da/, /ga/-/pa/, etc., were formed whereby one of the syllables was presented to the right, and the other one simultaneously to the left ear. Each CV syllable was presented for 350 ms and they were presented with an inter-stimulus interval of 4000 ms. Thirty of the 36 stimulus pairs were dichotic, that is, the syllable presented to the right and left ear were different, and served as basis for the laterality analysis. The 6 homonymous pairs were used to confirm that the subject was able to discriminate the syllables. The test had three conditions; non-forced (NF), forced right (FR), and forced left (FL). In the non-forced trial, participants were asked to listen to the sounds and repeat. The participants were told that it could seem as if there were two sounds at the same time, but that they could only give one response at each trial (the syllable they heard best, most clearly). In the forced right trial, participants were asked to listen to the sounds in the right ear and report the syllable they heard best. In the forced left trial, participants were
asked to listen to the sounds in the left ear and report the syllable they heard best, or most clearly. The percentage of correctly identified left- and right- ear syllables was recorded and used as dependent variable in the statistical analysis.

**Digits span test** was used to measure the number storage capacity of the working memory. During the test, a series of digits (e.g., ‘5, 7, 4’) was read aloud, and the participant was asked to repeat the digits. If they did this successfully, they were given a longer list (e.g., ’3, 2, 7, 0’). The length of the longest list a person can remember is that person's digit span. **Digits Span Forward** requires the participant to repeat the digits in the order they were read. It is a measure of auditory memory span and attention. **Digits Span Backward** requires the participant to repeat the digits in a reversed order. It is a measure of both working memory and attention. These tasks were administered and scored according to standard procedures. Points were given for correctly repeated series of digits. The tester stops when the participant fails to correctly repeat two series of digits in a row. (2)

**Focused and sustained attention**

**The Go/No-Go test** (3) was used to measure the participants’ capacity for sustained attention and response inhibition. It was conducted on a portable laptop computer. During the GNG test, 50% of trials involved the presentation of a yellow square for 1300 ms, followed by a blank screen for 400 ms, and a fixation cross for 300 ms (go trials). Each participant was asked to press the space bar as quickly as possible in reply to the presentation of a yellow square. For 25% of the trials a yellow triangle was presented, which indicated that participants should not make a response (no-go trials, inhibited response). For the remaining 25% of trials a blue square was presented on the computer screen, according to the same presentation sequence, indicating that participants should not make a response. The measure parameter was the number of commission errors.

**The Stop Test** (3) was used to measure response inhibition. It was conducted on a portable laptop computer. In the stop task, 50% of the trials consisted of the presentation of a yellow horizontal arrow which was displayed for 1500 ms, followed by a blank screen for 500 ms, and a fixation cross for 300 ms. In response to the presentation of a yellow horizontal arrow participants were urged to press the space key as quickly as possible. In 25% of trials the yellow horizontal arrow was followed by a vertical yellow arrow, signaling to participants that they should inhibit the response (stop trials). The presentation times for some arrows was adapted according to the following procedure: During the first trial in the inhibition condition, the horizontal arrow was displayed for 250 ms and the vertical arrow for 12,500 ms. Following this, 50 ms was added to the duration of the first arrow if the participant had been successful during the previous inhibition trial. If, however, the participant had failed to inhibit their response, the duration was reduced by 50 ms. These limits were imposed so that the minimum duration was 50 ms and the maximum was 1000 ms. The duration of the second arrow was adapted so that the total duration for the two arrows was always 1500 ms, in keeping with the control condition. This adaptive duration procedure was used to achieve approximately 50% accuracy, and to reduce the likelihood of participants anticipating the onset time of the second arrow, and delaying their response to the first arrow accordingly. Moreover, in the remaining 25% of trials (oddball trials) the initial horizontal yellow arrow was followed by a blue horizontal arrow, upon which a button press was still required. In such oddball trials the duration of exposure of the first arrow was determined by the accuracy and exposure duration in the previous stop trial, the same as for stop trials. However, unlike stop trials, the accuracy of the oddball trials did not influence any subsequent exposure duration. In the stop test an
estimate of the stop signal reaction time was made by first calculating the average stop signal duration for each participant (the average duration of the first arrow in the stop trials after stabilization) i.e. in stop trials 20-38). The stop signal reaction time (SSRT) was then determined by subtracting the mean stop signal duration from the median reaction time for the oddball trials (the median was used because of standard positive skewing of RT data). (3)

**Set-shifting**

**The Modified Wisconsin Card Sorting Test (MWCST)** was used to measure set-shifting, or cognitive flexibility (Heaton, et al, 1993). It was conducted on a portable laptop computer. During the test, four stimulus cards were continuously present on the upper half of the computer screen. A single response card was subsequently presented in the lower half of the screen. The task was to match the response card with one of the stimulus cards according to a specific rule. To sort a response card to a matching stimulus card the participants pressed a number key between 1-4 (in the left-to-right order these keys corresponded to the four stimulus cards). A trial is considered correct if a response card is matched with a specific stimulus card according to the principles of the rule. For each trial, the software would tell the participant whether a particular match was right or wrong. If the participant successfully sorting a response card to a stimulus card a specific number of trials, the rule would change. In total, the test consisted of 128 trials or until all rule changes have been accomplished. The outcome variable was the number of perseverative errors, i.e. how many times the subject failed to match a new card to the correct stimulus card due to a rule change. (12).

**Working Memory Function**

**The N-back test** was conducted on a portable laptop computer. In the test, each participant was tasked with monitoring a sequence of letters presented, one at a time, in a fixed position on a computer screen (onset times?), separated by a blank screen (onset times?) and a fixation cross (onset times?). During the test, the software would shift between four conditions, indicated by the presentation of the words “0back”, “1back”, “2back”, or “3back”, respectively (onset times?).

The participant was asked to give an answer in response to each letter in each condition using the keys 1 and 2 (which corresponded to “yes” and “no”). In the 0-back condition, the participant had to give a response to any letter that matched a pre-specified letter (i.e., “z”). In the 1-back condition, the target was any letter identical to the letter presented one trial back. In the 2-back condition, the target was any letter that was identical to the one presented two trials back. In the 3-back condition, the target was any letter that was identical to the one presented three trials back. The criterion variables used were the RT and score (%) for 2-back and 3-back condition.

**Verbal Memory Function**

**Claeson-Dahl**: Immediate and delayed recall of words (episodic memory) was assessed by the Claeson-Dahl Verbal Learning and Retention Test (4). It is a two-part, standardized, auditory, verbal learning and retention task consisting of a list of ten common Swedish words, two syllables in length. The examiner reads the list to the participant (at the rate of 2 seconds per word). After 15 seconds the participant is asked to reproduce as many words as possible. The procedure is repeated until the participant has produced the entire list twice correctly, during maximally ten trials. The first part of the test, when the subject learns a list of 10 words to be immediately recalled after presentation, measures the efficacy in the ability for immediate recall of auditory presented verbal materials (Nyman, 1998). The second phase, where the subject 30 min after presentation is asked to recall the 10 words, is designed to test the ability to recollect
the words attached to the learning situation. Recall was tested again after 60 minutes, as well as recognition memory, and memory for word order. The output parameters used was the number of times the word list is repeated until the subject can repeat all the words in the right order, and the weighted points.

**Psychomotor speed and attention**

**The Trail Making Test A (TMT-A) and Tracing A** are tasks involving visual, conceptual and visuo-motor tracking. TMT-A (5) requires primarily perceptual-motor speed, visual scanning, attention and numeric sequencing. TMT-A also measures working memory. During TMT-A, the subject is required to, as quickly as possible and in a numerical order, draw a line between 25 circles using a pen. The output measure was the number of seconds needed to complete the task and the number of correct connections.

**Tracing A** is a task that requires the participant to draw a line as quickly as possible between 25 circles on a paper. The criterion variable was the number of seconds needed to complete the task. (5).

**Psychomotor speed**

**Grooved Pegboard test** (6) requires participants to place 25 pegs, one at a time, into 25 keyhole-shaped holes with varying orientations that are randomly positioned on a five-by-five matrix. The pegs had to be rotated into position to match the hole before they were inserted, which was to be done as quickly as possible and in a prescribed order. Subjects were required to place the pegs from left to right when using both hands. The outcome measure was the time needed to complete the task, number of inserted pegs L/R hand, and the number of drops.

**Dual task condition** (7) required the participants to tap with one hand and insert pegs in the board with the other and vice versa. The participants were instructed to perform both tasks simultaneously, as quickly as possible, for 30 seconds, and the task was performed twice. The outcome measure was the ratio between the number of pegs and taps during 30 seconds period for the right hand, and for the left hand.

**Executive functions**

**The Trail Making Test B (TMT-B)** is task of visual, conceptual, and visuo-motor tracking which requires primarily perceptual-motor speed, visual scanning, attention, working memory and secondarily the ability to shift conceptually between numerical and alphabetical order (4). TMT-B was administered according to standard procedures (8). It requires the participant to draw a line, as quickly as possible, between 25 circles (of which thirteen are labeled numerically from 1-13, and the remainder are labeled alphabetically from A-L). If the participant made an error, it was pointed out for correction and patient was asked to return to and continue from the correct location while the time remained running. Criterion variables were number of seconds needed to complete TMT-B.

**Visuospatial function**

**The Mental Rotation Test** is comprised of ten items, five in each part of the test. Each item consisted of a geometrical target figure in the left-most position on a paper followed by four response-choice figures: two rotated reproductions of the target figure and two distractor figures. The task was to indicate which two of the four response-choice figures are rotated reproductions of the target figure, allowing free rotation in three-dimensional space. In each item there are always two correct figures and two incorrect distractor figures. Points were given
for correctly identified response-choice figures. Point reductions were made for wrong answers. The participants were urged not to guess. The time limit for the test is six minutes (9). The outcome measure is the number of correctly identified response-choice figures.

Verbal tests

The Word Fluency Test (FAS) was used to measure word fluency. Participants were asked to say as many words they could think of starting with the letters F, A, and S (60s/letter). Points were given for the total number of words produced. No points were given for inflections of previously mentioned words, names or numbers. The outcome measure was the total amount of words produced (in total) (10).

The Letter-Digit Substitution Task (LDST). Contains a key in which the numbers 1 to 9 are paired with a different letter. Beneath the key the test items are printed. Participants are required to replace the randomized letters with the appropriate digit given by the key. The first 10 items are used as practice items, to ensure that participants understand the test instructions. After completion of these items, participants are instructed to replace the remaining items as quickly as possible. The number of correct substitutions made in 60 seconds is the dependent variable (11).