Eur Arch Psychiatry Clin Neurosci (2005) 255 : 373–380 DOI 10.1007/s00406-005-0577-7

ORIGINAL PAPER

Eva Biringer · Astri Lundervold · Kirsten Stordal · Arnstein Mykletun · Jens Egeland · Ronald Bottlender · Anders Lund

Executive function improvement upon remission of recurrent unipolar depression

Received: 1 April 2004 / Accepted: 13 January 2005 / Published online: 1 April 2005

Abstract The aim of the study was to investigate the improvement of executive function measures upon recovery from unipolar depression. Thirty patients who suffered from recurrent major unipolar depression were retested with regard to their executive function approximately two years after an initial baseline examination. At baseline, patients were depressed (average 17-item HAM-D score 21.8), at retesting they were partially or totally recovered (average HAM-D score 8.2). There was a significant positive association between improvement on the HAM-D and improvement of executive function. In those with complete recovery, overall executive function and most examined executive function measures were no longer different from the baseline performance of healthy controls (with the possible exception of semantic fluency and Stroop Colour-word). In conclusion, recovery from major unipolar depression was accompanied by a recovery of many aspects of executive function to a normal level. Our findings support previous studies that have shown that neuropsychological impairment associated with long-standing depressive symptomatology is reversible (i. e. state-related) in recurrent unipolar depression.

Key words depression · longitudinal · executive function

E. Biringer · K. Stordal · A. Lund
Institute of Clinical Medicine
Section of Psychiatry
University of Bergen
Bergen, Norway

E. Biringer (✉)
Section of Psychiatry
Haukeland University Hospital
5021 Bergen, Norway
Fax: +4755974419
E-Mail: eva.biringer@psyk.uib.no

A. Lundervold
Dept. of Biological and Medical Psychology
University of Bergen
Bergen, Norway

A. Mykletun
Research Centre for Health Promotion
University of Bergen
Bergen, Norway

J. Egeland
Vestfold Mental Health Care Trust
Tønsberg, Norway

R. Bottlender
Dept. of Psychiatry
Ludwig-Maximilian University (LMU)
Munich, Germany

Introduction

Major depression is a common and severe disorder that traditionally has been seen as a disturbance in mood and vegetative functioning. Since the beginning of the 1980s, a number of studies have shown that major depression can also be associated with impairment of cognitive function. One of the neuropsychological domains that have been shown to be particularly impaired in depression, is that of 'executive function' [4, 12, 20, 21, 24, 29, 33, 35, 44, 49]. However, it is yet not clear whether this depression related impairment persists after remission from the depressive episode, i. e. whether it is 'trait' or 'state' dependent [4, 27].

Executive functions are higher-level cognitive functions that control and regulate lower cognitive operations. They are most likely linked to functional neuronal circuits involving the frontal lobe [50, 51]. Pennington and Ozonoff have conceptualised executive function by including the following components: set-shifting, planning, inhibition, working memory, and fluency [43].

Depression is regarded as a chronic disorder with a release-relapse course in which patients experience recurrence of episodes after having partially or fully recovered, and in which sub-threshold symptoms might be present between the episodes [3, 30]. The symptomatology and cognitive deficits in depression have been linked to a neurobiological dysfunction that involves the frontal-subcortical neuronal circuitries [23, 34, 46].

Functional brain imaging studies (fMRI) have shown that the disease is associated with pathological changes of blood flow and glucose metabolism in the prefrontal cortex, striatum, pallidum, and thalamus during depression, and that some of these disturbances improve upon recovery [15, 54]. Cognitive activation studies with positron emission tomography (PET) have indicated that depression is associated with an activation level in frontal and prefrontal regions that is different from that found in normal controls, and that post-treatment activation is improved compared to baseline activation [9, 31]. It has also been shown that a portion of depressed patients suffer from a dysregulation in the HPA-axis that results in excessive and dysfunctional cortisol secretion [36, 37, 41]. It is possible that such a long-lasting hypercortisolaemia may exert a detrimental effect upon the neuronal functioning, and that it may even lead to neuronal loss [41, 48]. If this holds true, one could expect that neuropsychological test performance would fail to improve to a level comparable to that of healthy controls after recovery in subjects with long disease duration [5].

However, several studies have suggested that the changes in executive function in long-lasting depression are state-related (i. e. reversible), although these studies are by no means consistent [4]. Beblo et al. (1999) and Neu et al. (2001) found a significant improvement of the Verbal Fluency Test upon recovery from recurrent depression [8, 39]. However Reischies et al. (2000) found that there was no significant difference of improvement between retested remitted depressives and retested controls on this test [45]. Beats et al. (1996) found that whereas performances on a set-shifting task normalised after recovery in a group of remitted elderly patients with long disease duration, performance on a planning task was slower, and number of words generated on the Verbal Fluency Test fewer when compared to healthy controls [7]. Paradiso et al. (1997) showed that remitted unipolar depressives with long disease duration were impaired on the Trail Making Test, and on the Stroop Colour-word Test [42]. Possibly, test performance in remitted patients could be a function of time of follow-up and level of psychopathology at the time of follow-up, as well as of task-specific characteristics. Perhaps some aspects of executive function return to normal earlier in the course of remission, whereas others remain impaired longer, or do not return to normal at all, i. e. they are 'trait' markers that could possibly be related to the underlying pathobiological mechanisms. Therefore, one should examine the patients after a sufficiently long test-retest interval in order to avoid the effects of depression lag and of persistent sub-threshold symptomatology.

The present study investigated improvement in performance on measures of executive function after a long test-retest interval (26 months) in a well defined group of unipolar patients with long disease duration. Their executive function performance was impaired while they were in the depressed state [17, 49]. The aim of the study was to explore the association between symptom relief and improvement on tests related to frontal lobe functioning. Specifically, we tested the hypotheses that 1) improvement of executive function follows improvement in depressive symptomatology, and 2) after complete recovery, executive function is not significantly different from that of healthy controls.

Methods

Sample

Thirty subjects from a baseline sample of 50 with a DSM-IV diagnosis of recurrent major unipolar depression [1, 19] were re-examined with psychiatric and neuropsychological measures after a mean test-retest interval of 26.3 months (standard deviation SD = 6.1). Inclusion criteria at baseline (T1) were a minimum score of 18 on both the 17-item Hamilton Depression Rating Scale (HAM-D) [26] and the Montgomery Aasberg Depression Rating Scale (MADRS) [38], and a lifetime minimum of two major depressive episodes including the baseline episode. At follow-up (T2), level of depressive psychopathology was assessed by the HAM-D and the subjects were examined by the Mini International Neuropsychiatric Interview (version 5.0.0) [1, 47] in order to detect if any changes of primary diagnosis had occurred. Out of the 50 patients included at T1, three were deceased at T2, two had developed a bipolar disorder, and fifteen subjects did not respond to the invitation to participate in the follow-up examination. Sample characteristics and a drop-out analysis are presented in Table 1. There were no significant group differences with respect to sociodemographic and clinical variables, level of functioning as measured by the Global Assessment of Functioning Scale (GAF) [1], general intellectual abilities as measured by the Similarities and the Picture Completion subtests from the Wechsler Adult Intelligence Scale-revised (WAIS-R) [55], or medication (medicated/non-medicated) at T1 between the retested (N = 30) and the retest non-attendees (N = 20). However, the retested had a longer disease duration (P = 0.01, two-tailed, Mann-Whitney U test) and a higher number of depressive episodes than the retest non-attendees (P = 0.009, two-tailed, Mann-Whitney U test). At T2, the sample was divided into recovered (N = 17) and non-recovered (N = 13) (T2 cut-off HAM-D ≤ 7), clinical characteristics of these subgroups are presented in Table 2. There were no significant differences in sociodemographic or clinical variables between the recovered and the non-recovered.

The study was approved by the Regional Committee for Medical Ethics and it was performed in accordance with the Helsinki Declaration of the World Medical Association Assembly. All participants provided written informed consent to participate in the study at inclusion and at follow-up. The study was funded by the Norwegian Research Council.

Selected EF tests

Ten subtests were used to measure executive function (EF). These were selected from a neuropsychological test battery according to the theoretically and empirically based model of executive function outlined by Pennington and Ozonoff in 1996 [43]. Seven of these measures had significantly separated the depressed sample from the control sample (N = 50) at T1 after medication and psychomotor speed had been controlled for [49]. From the Wisconsin Card Sorting Test (WCST) [28], four measures were included: the Number of categories completed (WCST Catc), the Number of perseverative errors (WCST Perr), the Failure to maintain set (WCST Ftms), and the Number of nonperseverative errors (WCST Nperr). WCST Catc and WCST Perr were used as measures of set-shifting. The PASAT 2 and the PASAT 3 seconds subtests from the Paced Auditory Serial Addition Test (PASAT) [25] and the Backward Digit Span (WAIS-R Dsb) subtask from the Wechsler Adult Intelligence Scale-revised (WAIS-R) [55] were included as measures of working memory. The Colour-word (Stroop C/W) subtask from the Stroop Colour and Word Test (Stroop) [22] was included as a measure of inhibition. The Number of words

Table 1 T1 sample characteristics of the retest attendees and the retest non-attendees

	Retest attendees (N = 30)			Retest non-attendees (N = 20)		
	Mean	SD	Range	Mean	SD	Range
Age, years	35.8	8.4	20–50	33.8	9.3	19–51
HAM-D	21.8	3.2	18–30	22.5	5.6	18–42
Education, years	13.8	3.0	9–19	14.0	2.8	10–19
Intellectual abilities						
WAIS-R Picture completion	14.5	3.4	6–19	14.7	2.9	9–19
WAIS-R Similarities	19.1	6.4	4–27	21.1	4.1	10–27
Age at first episode, years	23.7	9.6	7–44	27.9	10.5	15–47
Lifetime number of episodes*	3.8	1.2	2–5	2.4	0.5	2–3
Total disease duration, years*	13.6	9.3	1–32	6.1	5.1	1–22
GAF	44.6	8.0	25–55	47.9	9.5	33–70
Sex, male/female	12/18			9/11		
Handedness, right/left	28/2			18/2		
Work status	10 employed/4 students/ 14 sick leave or disability pension/2 information missing			6 employed/5 students/6 sick leave or disability pension/ 2 out of work/1 information missing		
Medication	26 on medication, 2 not on medication, 2 information missing			18 on medication, 2 not on medication		

* Significant mean group difference, P = 0.009 and P = 0.01, Mann-Whitney U test, two-tailed

Table 2 Sample characteristics of the recovered and the non-recovered

	Recovered (N = 17)			Non-recovered (N = 13)		
	Mean	SD	Range	Mean	SD	Range
Age at inclusion, years	34.1	9.4	20–47	37.9	6.6	29–50
HAM-D						
T1	20.1	1.6	18–23	23.9	3.6	18–30
T2	2.7	2.1	0–7	15.4	5.8	8–27
Education, years	14.0	3.0	9–19	13.5	3.0	9–18
Intellectual abilities						
WAIS-R Picture completion	15.0	3.2	7–19	13.8	3.7	6–18
WAIS-R Similarities	18.4	6.9	5–27	20.1	5.8	4–25
T1-T2 interval, months	27.4	6.9	15.3–40.0	24.7	4.5	17.3–33.1
Age at first episode, years	25.1	11.0	7–44	22.2	7.7	14–37
Lifetime number of episodes	3.8	1.3	2–5	3.7	1.11	2–5
Total disease duration, years	10.8	7.4	1–23	16.8	10.5	1–32
GAF						
T1	43.9	8.9	25–55	45.5	68.0	31–55
T2	82.2	8.6	70–100	58.1	11.0	45–80
Sex, male/female	8/9			4/9		
Handedness, right/left	17/0			11/2		
Work status	T1: 9 employed/1 student/7 sick leave or disability pension T2: 11 employed/2 students/4 sick leave or disability pension			T1: 3 employed/3 students/7 sick leave or disability pension T2: 2 employed/10 sick leave or disability pension/1 no income		
Antidepressant use*	T1: 11 T2: 11			T1: 12 T2: 12		
SSRI	T1: 10 T2: 11			T1: 11 T2: 9		
other antidepressants	T1: 1 mianserin/1 venlafaxine/ 1 nefazodone T2: 2 MAOH			T1: 1 MAOI/2 venlafaxine/ 1 nefazodone T2: 2 MAOI/1 tricyclic/1 lithium		
Neuroleptics	T1: 4 T2: 1			T1: 3 T2: 3		
Sedatives	T1: 8 T2: 2			T1: 6 T2: 7		

* Some patients were on more than one antidepressant

in phonemic categories (COWAT Phon) and Number of words in semantic categories (COWAT Sem) from the Controlled Oral Word Association Test (COWAT) [10] were included as measures of fluency. More details about the tests are given in Stordal et al. [49].

EF operationalisation

The Stroop C/W subtask, the WCST Perr, the WCST Nperr, and the WCST Ftms subtasks were reverse-scaled so that higher scores indicated better performance. Scores of change between T1 and T2 for raw scores and for standardised scores (Z-scores) of each of the ten EF measures were computed by subtracting T1 scores from T2 scores. A Cronbach's alpha of 0.83 was found at T1 for raw scores of the ten selected EF measures in a N = 100 sample consisting of the total baseline sample and the sample of healthy controls previously mentioned. This high internal consistency of the EF measures indicated that a composite scale of change of EF from T1 to T2 could be computed. Weighted Z-scores of EF change were, therefore, added together in order to produce a composite scale. The weighted Z-scores of change were computed by multiplying Z-scores of change for each EF measure with their respective weights derived from a one-factor Principal Components Analysis (PCA) on T1 raw scores in the N = 100 sample.

Data analysis

The HAM-D total score was reverse scaled and change variables were computed by subtracting T1 scores from T2 scores. In order to evaluate the hypothesis that improvement of depressive symptomatology correlated with EF improvement, Pearson's correlation coefficients r were calculated for the association between the HAM-D change scores and the composite scale of EF change as well as between the HAM-D change scores and change scores of each of the ten selected EF measures in the total sample. Then, the subjects were divided into two groups: recovered and non-recovered (T2 HAM-D cut-off ≤ 7). Chi-square tests and independent-samples t tests were conducted to assess differences of sociodemographic and clinical variables between the two groups. A Mann-Whitney U test was conducted to assess differences between the two groups on the composite scale of EF change. Independent-samples t tests or Mann-Whitney U tests were conducted to assess between-group differences of change of EF raw scores.

In order to assess to what degree EF returned to normal levels after depression recovery, Z-scores of the recovered group's test scores at T2 were compared to Z-scores from the T1 control group by independent samples t tests in a N = 67 data set consisting of T1 values from the healthy control group (N = 50) and T2 values from the recovered group (N = 17). An independent-samples t test was conducted to evaluate if the recovered group's score on a PCA-weighted sumscale of Z-scores of the ten selected EF measures in this data set was different from the healthy controls' score.

The Bonferroni approach was used in order to control for Type I error across the analyses with the ten EF measures as dependent variables. The tests were two-tailed with an alpha level of 0.05, except for tests done on the associations between improvement on EF variables and improvement on the HAM-D, which were one-tailed, because here only positive associations were expected. Statistical procedures were performed using the SPSS 11.5.

Results

Relationship between depression recovery and EF improvement

The relationship between HAM-D change scores and the composite scale of EF change is shown as a scatterplot in Fig. 1. A significant positive Pearson's correlation coefficient was found between the HAM-D change and the composite scale of EF change (r = 0.33, P = 0.04, one-tailed). Depression recovery as measured by the HAM-D explained 11 % of the variance of EF improvement from T1 to T2 ($R^2 = 0.11$). Pearson's correlation coefficients for the relationship between HAM-D change and change of raw scores from T1 to T2 on each of the ten selected EF measures in the total sample are presented in Fig. 2. Change from T1 to T2 was positively correlated with HAM-D change on eight out of the ten EF measures. The only statistically significant positive correlation was found between HAM-D change and change of the COWAT Sem (r = 0.33, P = 0.04, one-tailed). The positive correlations between change of HAM-D and change of the Stroop C/W, the WCST Ftms, and the PASAT 2 were moderate, yet noteworthy (r = 0.28, r = 0.27, and r = 0.25, respectively), but not of statistical significance.

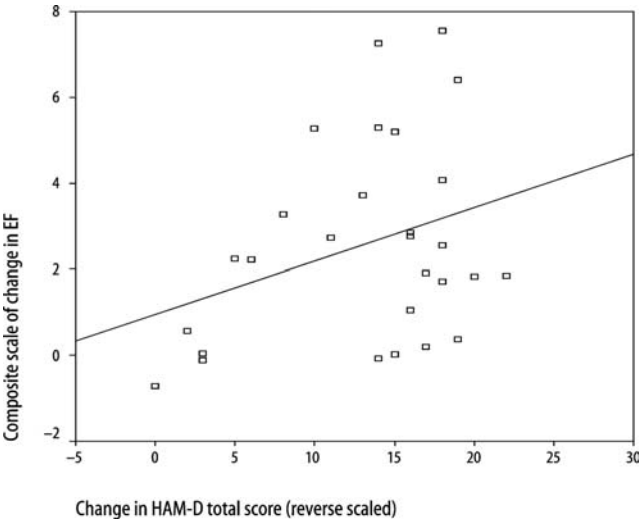

Fig. 1 Scatterplot of HAM-D total score improvement from T1 to T2 versus the composite scale of EF improvement

The correlations between change of the HAM-D total score and change of the PASAT 3 and of the WCST Perr

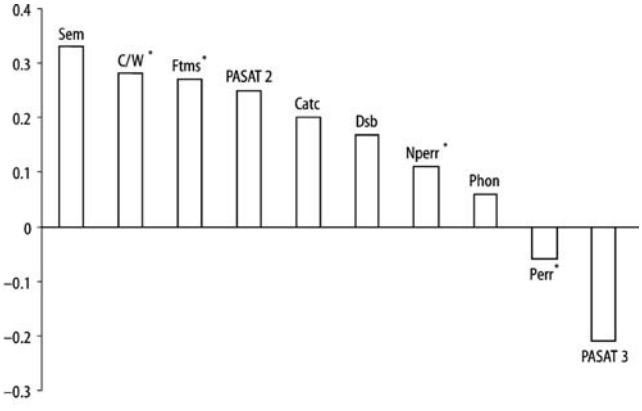

Fig. 2 Pearson's correlation coefficients between HAM-D total score improvement and change of EF measures from T1 to T2 in the total sample (N = 30)
* test scores are reverse scaled

were negative. After an alpha correction was applied for the single test correlations in Fig. 2, none of the correlations were significant.

EF improvement in the recovered and the non-recovered groups

Means and standard deviations of raw scores, change scores from T1 to T2, and Z-scores of change for each of the ten selected EF measures for the recovered and the non-recovered groups are shown in Table 3. Eight out of the ten EF measures showed greater mean improvements in the group that recovered than in the group that did not. An independent-samples t test conducted to evaluate group differences between the recovered and the non-recovered groups' change from T1 to T2 showed that the recovered group had a significantly greater performance improvement on the PASAT 2 compared to the non-recovered group (P = 0.02, one-tailed, independent-samples t test). After an alpha-correction was performed, however, this difference was not significant. No other measure of change was significantly different between the two groups. In addition, the difference between the two groups with regard to improvement on the composite scale of EF change from T1 to T2 was not significant.

Test performance in the recovered group compared to the healthy controls

Mean Z-score differences (recovered-controls) are presented in Fig. 3. After recovery, mean Z-scores in the recovered group were in the range from + 0.30 to –0.61 SD above/below the group test means of the control group.

The independent-samples t test conducted with the sumscale of EF Z-scores as dependent variable in order to evaluate if the recovered group's test performance at T2 differed from that of the healthy controls' at T1 showed that there was no significant group difference between the recovered at T2 and the healthy controls (P = 0.190, two-tailed). In the analysis with the ten single EF test measures as dependent variables, the recovered group did not perform significantly lower than the control group on eight out of the ten measures. Their performance was still significantly lower on the COWAT Sem and on the Stroop C/W (P = 0.030 and P = 0.033, two-tailed). After alpha correction, no EF-measure was significantly lower in the recovered group compared to the control group.

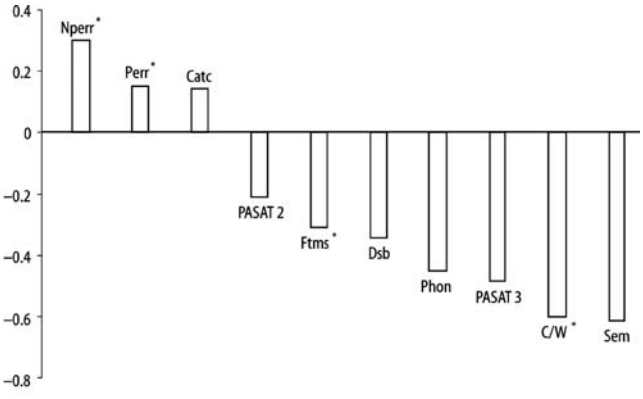

Fig. 3 Mean Z-score difference between the recovered (N = 17) and the controls (N = 50)
* test scores are reverse scaled

Discussion

The results demonstrated that improvement of depressive symptomatology is followed by improvement on measures of neuropsychological test measures regarded as indicators of executive function. A significant medium-sized positive correlation between improve-

Table 3 HAM-D total scores and EF scores at T1 and at T2, change scores and Z-scores of change (T1 to T2) for the recovered and the non-recovered

	Recovered				Non-recovered			
	T1 Mean (SD)	T2 Mean (SD)	Change (T2–T1)		T1 Mean (SD)	T2 Mean (SD)	Change (T2–T1)	
			Mean (SD)	Mean Z-score (SD)			Mean (SD)	Mean Z-score (SD)
PASAT 3	44.4 (13.6)	47.1 (14.1)	2.7 (10.8)	0.2 (0.9)	42.9 (12.9)	48.4 (10.7)	5.5 (6.5)	0.4 (0.5)
PASAT 2	35.6 (12.6)	41.4 (10.8)	5.9 (6.0)	0.5 (0.5)	38.5 (13.9)	39.4 (10.8)	0.9 (6.4)	0.1 (0.5)
Stroop C/W*	57.9 (16.4)	49.7 (10.1)	8.3 (11.5)	0.6 (0.8)	55.4 (10.6)	53.2 (17.8)	2.2 (14.2)	0.2 (1.0)
WAIS Dsb	5.2 (1.6)	6.1 (2.3)	0.9 (2.5)	0.5 (1.3)	5.1 (1.4)	5.9 (2.0)	0.8 (1.6)	0.4 (0.9)
COWAT Phon	24.2 (8.3)	26.8 (9.8)	2.6 (4.7)	0.3 (0.6)	22.8 (6.8)	24.5 (8.1)	1.7 (5.1)	0.2 (0.6)
COWAT Sem	39.5 (8.6)	41.0 (9.1)	1.5 (6.4)	0.2 (0.6)	40.9 (12.2)	39.5 (11.0)	1.4 (9.2)	0.1 (0.9)
WCST Catc	4.8 (1.8)	5.5 (1.2)	0.8 (1.5)	0.5 (1.1)	5.6 (0.9)	5.9 (0.4)	0.2 (0.9)	0.2 (0.7)
WCST Perr*	14.6 (11.2)	9.6 (8.8)	4.9 (8.2)	0.6 (0.9)	14.3 (7.7)	8.0 (4.1)	6.3 (7.9)	0.7 (0.9)
WCST Nperr*	15.6 (9.5)	8.9 (7.5)	6.7 (9.0)	0.9 (1.2)	12.4 (6.9)	8.3 (4.1)	4.1 (8.1)	0.5 (1.0)
WCST Ftms*	2.1 (2.1)	1.1 (1.3)	0.9 (1.9)	0.5 (1.1)	1.2 (1.3)	0.8 (1.3)	0.4 (1.5)	0.2 (0.9)

* lower score indicating better performance

ment of depressive symptomatology and improvement on a composite score of change of EF over a period of approximately two years was found in this sample with long depression duration.

In our study, there are several arguments in favour of the view that long-lasting depression does not lead to cognitive impairment: 1) The significant positive correlation between improvement of depressive symptomatology and improvement of EF. 2) No significant group difference between the recovered at T2 and the controls. 3) The improvement of the HAM-D accounted for 11 % of the total variance of the composite scale of EF change. The amount of variance explained by the group difference depressed-controls at T1 was 17.9 % on the composite scale of EF. Because it was not to be expected that the improvement in EF upon recovery should exceed the amount of variance explained by the group difference at baseline, this 11 % gain was a substantial one. At baseline, the depressed group's mean test performance was 0.55 SD below that of the healthy controls, and at T2, it had improved to merely 0.24 SD below it.

However, the failure of the measures that were mostly sensitive to changes in depressive symptomatology, the Semantic fluency measure and the inhibition measure from the Stroop Colour and Word Test, to improve completely to the level of the controls, is an argument in favour of the 'trait' hypothesis. Because these measures were still about 0.60 SD below the controls after recovery, they could be markers of permanent depression-associated changes. Thus, our study gives support to the model in which changes are 'state' related, but it cannot completely reject the 'trait' hypothesis.

The pattern of test improvement in relation to improvement of depressive symptomatology was heterogeneous with respect to task characteristics, and because of the low statistical power in our study, we are careful about drawing conclusions with regard to the improvement profile. However, it seems like speeded attention-demanding tasks, as represented by the COWAT Sem, the Stroop C/W, and the PASAT, were more sensitive to depression than the non-speeded problem solving tasks. This is in line with previous studies that have claimed that speeded attention is impaired in depression [11, 13, 16], and that it might return to normal to a lesser extent than the non-speeded complex set-forming tasks of the WCST [7, 53].

This is also supported by Paradiso et al. (1997), who found that remitted patients performed impaired compared to healthy controls on the Stroop C/W, and indirectly by Trichard et al. (1995) who found that patients improved their scores on the Stroop C/W interference score (which takes height for the speed component) at retesting to a level where they did not differ significantly from controls [42, 53].

As in our study, the COWAT Sem was found to correlate with depression improvement in the studies of Beblo et al. (1999) and Trichard et al. (1995) [8, 53]. Interestingly, improvement in their studies was found after one month in remission only, which suggests that improvement of fluency happens early in the course of recovery. The Semantic fluency task can also be seen as a measure of long-term episodic memory, and because depression has been shown to be associated with impaired memory performance [27, 39, 45], the improvement of this task could be due to an effect of improvement of memory functioning along with depression recovery.

In our study, there might be several other reasons than the hypothesised detrimental effect of long-standing pathology for the failure of the EF to remit totally to the level of the controls. Firstly, the premorbid level of EF was not known in the patient sample; therefore we do not know to what extent it returned to the level it was at before the onset of the first depressive episode. Subjects who suffer from depression might already have a dysfunction or vulnerability prior to the index episode which make their baseline performance different from and lower than that of healthy controls. In a recent review, Clark et al. referred to the 'mediating vulnerability factor', a model in which a trait impairment is combined with state modulation [11, 40], and possibly, this model is relevant in unipolar depression as well. Secondly, the retested sample in our study was a relatively chronified one, i. e. with a high number of depressive episodes and a long total lifetime disease duration, and we have not been able to control for the probable negative effect of having had a reduced level of general functioning due to illness over many years.

Factors that can affect neuropsychological test results in studies like ours are participants' features such as baseline performance, medication, motivation, and test features such as test-retest reliability, practice effects, floor-ceiling effects and duration of the test-retest interval [6, 14, 18, 32, 44, 52]. When practice effects are combined with regression to the mean, low scores tend to show a marked improvement on retesting [2]. A large correlation ($r = 0.52$) between the average of the initial and the final measurement and the observed change indicated that this was the case for the WCST Catc. Another weakness of our study is the lack of retesting of the healthy controls, which might have led to an overestimation of the improvement of the recovered patients, particularly for the problem solving task, which is easily subjected to ceiling-effects at retesting [6]. The strength of our study is the homogeneity of the sample (purely younger unipolar subjects with recurrent subtype) and the EF operationalisation. Because the number of construct indicators was increased, we presumed that EF reliability increased when our selected EF measures were summarised into a composite scale. This increase of reliability could be the reason why the composite scale of EF improvement correlated more with the HAM-D improvement than what the single EF measures did.

The correlation between the HAM-D change and the composite scale of EF change was significant in the continuous approach, but there was not a significant group difference in the categorical approach. This was to be expected, as variance in the HAM-D was lost when the

variable was dichotomised by the introduction of a cut-off value. We therefore emphasise the significant finding from the continuous approach in our study more than we emphasise the finding from the categorical approach.

In conclusion, our study is an argument for the model in which depression-related changes in executive function are reversible upon recovery. This is important knowledge for the patients who experience cognitive problems while they are depressed, as well as for the clinicians who are responsible for their treatment. However, our study fails to completely reject the possibility of persistent changes. Therefore, the heterogeneity of cognitive changes displayed in this and other studies should be investigated further in future studies. In order to avoid the effect of rest-symptomatology upon cognitive performance, this should be done with follow-up intervals of a long enough duration to ensure that complete recovery has been reached, and with subsamples that should be homogeneous and well-characterised with respect to diagnosis and clinical characteristics.

Acknowledgements Liv Heldal and the Department of Biological and Medical Psychology, University of Bergen; Kjetil Sundet and Bjørn Rund in the Bergen-Oslo Project group; Åsa Hammar, Atle Roness, and Ketil Ødegaard at the University of Bergen; Norma Mjellem; and T. Sato at the LMU, Munich.

References

1. APA (1994) Diagnostic and Statistical Manual of Mental Disorders (DSM-IV). Fourth Edition. American Psychiatric Association, Washington DC
2. Altman D (1991) Practical statistics for medical research. Chapman and Hall, London
3. Angst J, Merikangas K (1997) The depressive spectrum: diagnostic classification and course. J Affect Disord 45:31–40
4. Austin M, Mitchell P, Goodwin G (2001) Cognitive deficits in depression. Possible implications for functional neuropathology. Br J Psychiatry 178:200–206
5. Basso M, Bornstein R (1999) Relative memory deficits in recurrent versus first-episode major depression on a word-list learning task. Neuropsychology 13:557–563
6. Basso M, Bornstein R, Lang J (1999) Practice effects on commonly used measures of executive function across twelve months. Clin Neuropsychol 13:283–292
7. Beats B, Sahakian B, Levy R (1996) Cognitive performance in tests sensitive to frontal lobe dysfunction in the elderly depressed. Psychol Med 26:591–603
8. Beblo T, Baumann B, Bogerts B, Wallesch C, Herrmann M (1999) Neuropsychological correlates of major depression: a short-term follow-up. Cognitive Neuropsychiatry 4:333–341
9. Bench C, Frackowiak R, Dolan R (1995) Changes in regional cerebral blood flow on recovery from depression. Psychol Med 25:247–251
10. Benton A, Hamsher K (1989) Multilingual aphasia examination. AJA Associates, Iowa
11. Clark L, Goodwin G (2004) State- and trait-related deficits in sustained attention in bipolar disorder. Eur Arch Psychiatry Clin Neurosci 254:61–68
12. Degl'Innocenti A, Agren H, Backman L (1998) Executive deficits in major depression. Acta Psychiatr Scand 97:182–188
13. Den Hartog H, Derix M, Van Bemmel A, Kremer B, Jolles J (2003) Cognitive functioning in young and middle-aged nonmedicated out-patients with major depression: testing the effort and cognitive speed hypotheses. Psychol Med 33:1443–1451
14. Dikmen S, Heaton R, Grant I, Temkin N (1998) Test-retest reliability and practice effects of the Expanded Halstead-Reitan Neuropsychological Test Battery. J Int Neuropsychol Soc 5:346–356
15. Drevets W (2000) Neuroimaging studies of mood disorders. Biol Psychiatry 48:813–829
16. Egeland J, Rund BR, Sundet K, Landrø NI, Asbjørnsen A, Lund A, Roness A, Stordal K, Hugdahl K (2003) Attention profile in schizophrenia compared with depression: differential effects of processing speed, selective attention and vigilance. Acta Psychiatr Scand 108:276–284
17. Egeland J, Sundet K, Asbjørnsen A, Hugdahl K, Rund B, Landro N, Lund A, Roness A, Stordal K (2003) Sensitivity and specificity of memory dysfunction in schizophrenia: A comparison with major depression. J Clin Exp Neuropsychol 25:79–93
18. Elliott R, Sahakian B, McKay A, Herrod J, Robbins T, Paykel E (1996) Neuropsychological impairments in unipolar depression: the influence of perceived failure on subsequent performance. Psychol Med 26:975–989
19. First M, Spitzer R, Gibbon M, Williams J (1995) Structured clinical interview for DSM-IV axis I disorders- patient edition (SCID I/P, version 2.0). Biometrics Research Department, New York State Psychiatric Institute, New York
20. Fossati P, Amar G, Raoux N, Ergis A, Allilaire J (1999) Executive functioning and verbal memory in young patients with unipolar depression and schizophrenia. Psychiatry Res 89:171–187
21. Franke P, Maier W, Hardt J, Frieboes R, Lichtermann D, Hain C (1993) Assessment of frontal lobe functioning in schizophrenia and unipolar major depression. Psychopathology 26:76–84
22. Golden C (1978) Stroop Color and Word Test. Manual. Psychological Assessment Resources, Odessa
23. Goodwin G (1997) Neuropsychological and neuroimaging evidence for the involvement of the frontal lobes in depression. J Psychopharmacol 11:115–122
24. Grant M, Thase M, Sweeney J (2001) Cognitive disturbance in outpatient depressed younger adults: evidence of modest impairment. Biol Psychiatry 50:35–43
25. Gronwall D (1977) Paced auditory serial-addition task: a measure of recovery from concussion. Percept Mot Skills 44:367–373
26. Hamilton M (1960) A rating scale for depression. J Neurol Neurosurg Psychiatry 23:56–62
27. Harrison J, Owen A (2002) Cognitive Deficits in Brain Disorders. Martin Dunitz Ltd., London
28. Heaton R, Chelune G, Talley J, Kay G, Curtiss G (1993) Wisconsin Card Sorting Test. Manual. Psychological Assessment Resources Inc
29. Kaiser S, Unger J, Kiefer M, Markela J, Mundt C, Weisbrod M (2003) Executive control deficit in depression: event-related potentials in a Go/Nogo task. Psychiatry Res: Neuroimaging 122: 169–184
30. Keller M (2003) Past, present, and future directions for defining optimal treatment outcome in depression: remission and beyond. J Am Med Assoc 289:3152–3160
31. Kennedy S, Javanmard M, Vaccarion F (1997) A review of functional neuroimaging in mood disorders: positron emission tomography and depression. Can J Psychiatry 42:467–475
32. Kuny S, Stassen H (1995) Cognitive performance in patients recovering from depression. Psychopathology 28:190–207
33. Landrø N, Stiles T, Sletvold H (2001) Neuropsychological function in nonpsychotic unipolar major depression. Neuropsychiatry Neuropsychol Behav Neurol 14:233–240
34. Mega M, Cummings J (1994) Frontal-subcortical circuits and neuropsychiatric disorders. J Neuropsychiatry Clin Neurosci 7:271–272
35. Merriam E, Thase M, Haas G, Keshavan M, Sweeney J (1999) Prefrontal cortical dysfunction in depression determined by Wisconsin Card Sorting Test Performance. Am J Psychiatry 156: 780–782
36. Mitchell A (1998) The role of corticotropin releasing factor in depressive illness: a critical review. Neurosci Biobehav Rev 22: 636–651

37. Moffoot A, O'Carroll R, Bennie J, Carroll S, Dick H, Ebmeier K, Goodwin G (1994) Diurnal variation of mood and neuropsychological function in major depression with melancholia. J Affective Disord 32:257–269
38. Montgomery S, Aasberg M (1979) A new depression scale designed to be sensitive to change. Br J Psychiatry 134:382–389
39. Neu P, Kiesslinger U, Schlattmann P, Reischies F (2001) Time-related cognitive deficiency in four different types of depression. Psychiatry Res 103:237–247
40. Nuechterlein K, Dawson M, Gitlin M, Ventura J, Goldstein M, Snyder K, Yee C, Mintz J (1992) Developmental processes in schizophrenic disorders: longitudinal studies of vulnerability and stress. Schizophr Bull 18:387–425
41. O'Brien J (1997) The 'glucocorticoid cascade' hypothesis in man. Prolonged stress may cause permanent brain damage. Br J Psychiatry 170:199–201
42. Paradiso S, Lamberty G, Garvey M, Robinson R (1997) Cognitive impairment in the euthymic phase of chronic unipolar depression. J Nerv Mental Dis 185:748–754
43. Pennington B, Ozonoff S (1996) Executive functions and developmental psychopathology. J Child Psychol Psychiatry 37:51–87
44. Porter R, Gallagher P, Thompson J, Young A (2003) Neurocognitive impairment in drug-free patients with major depressive disorder. Br J Psychiatry 182:214–220
45. Reischies F, Neu P (2000) Comorbidity of mild cognitive disorder and depression – a neuropsychological analysis. Eur Arch Psychiatry Clin Neurosci 4:186–193
46. Royall D (1999) Frontal systems impairment in major depression. Seminars in Clinical Neuropsychiatry 4:13–23
47. Sheehan D, Lecrubier Y, Harnett-Sheehan K, Janavs J, Weiller E, Bonara L, Keskiner A, Schinka J, Knapp E, Sheehan M, Dunbar G (1997) Reliability and Validity of the MINI International Neuropsychiatric Interview (M. I. N. I.): According to the SCID-P and its reliability. Eur Psychiatry 12:232–241
48. Sheline Y (2000) 3D MRI studies of neuroanatomic changes in unipolar major depression: the role of stress and medical comorbidity. Biol Psychiatry 48:791–800
49. Stordal K, Lundervold A, Egeland J, Mykletun A, Asbjørnsen A, Landrø N, Roness A, Rund B, Sundet K, Oedegaard K, Lund A (2004) Impairment across executive functions in recurrent major depression. Nord J Psychiatry 2004:41–47
50. Stuss D, Alexander M (2000) Executive functions and the frontal lobes: a conceptual view. Psychol Res 63:289–298
51. Stuss D, Levine B (2002) Adult clinical neuropsychology: lessons from studies of the frontal lobes. Annu Rev Psychol 53:401–433
52. Temkin N, Heaton R, Grant I, Dikmen S (1999) Detecting significant change in neuropsychological test performance: A comparison of four models. J Int Neuropsychol Soc 5:357–369
53. Trichard C, Martinot J, Alagille M, Masure M, Hardy P, Ginestet D, Feline A (1995) Time course of prefrontal lobe dysfunction in severely depressed in-patients: a longitudinal neuropsychological study. Psychol Med 25:79–85
54. Videbech P (2000) PET measurements of brain glucose metabolism and blood flow in major depressive disorder: a critical review. Acta Psychiatr Scand 101:11–20
55. Wechsler D (1981) Wechsler adult intelligence scale-revised. The Psychological Corporation, New York