

Neurocognition in individuals at high familial risk of mood disorders with or without subsequent onset of depression

M. Pappmeyer^{1*}, J. E. Sussmann¹, J. Hall², J. McKirdy¹, A. Peel¹, A. Macdonald¹, S. M. Lawrie¹, H. C. Whalley¹ and A. M. McIntosh¹

¹Division of Psychiatry, University of Edinburgh, Royal Edinburgh Hospital, Edinburgh, UK

²Neuroscience and Mental Health Research Institute, Cardiff University School of Medicine, Cardiff, UK

Background. Neurocognitive performance deficits have been observed in mood disorder patients and their unaffected relatives and may therefore qualify as endophenotypes. However, the precise time course of neurocognitive deficits has not been studied so that it is unknown whether neurocognitive abnormalities reflect the early effects of familial vulnerability to mood disorders or if they emerge at illness onset.

Method. A neuropsychological test battery was administered at baseline and after a 2-year follow-up interval in 111 initially unaffected young adults at high familial risk of mood disorders and 93 healthy controls (HC). During the follow-up period, 20 high-risk subjects developed major depressive disorder (HR-MDD), with the remainder remaining well (HR-well). Linear mixed-effects models were used to investigate differences and longitudinal changes in the domains of attentional processing, working memory, verbal learning and memory, and cognitive flexibility.

Results. Reduced long delay verbal memory and extradimensional set-shifting performance across both time points were found in the HR-well group relative to controls. The HR-MDD group displayed decreased extradimensional set-shifting abilities across both time points as compared with the HC group only. There were no significant performance differences between the two high-risk groups.

Conclusions. Reduced verbal memory and cognitive flexibility are familial trait markers for vulnerability to mood disorders in individuals with a close family history of bipolar disorder. Both neurocognitive performance deficits appear to be relatively stable over a 2-year time period and do not appear to be linked to the onset of MDD. These findings support their use as stable quantitative endophenotypes for mood disorders.

Received 23 February 2015; Revised 12 June 2015; Accepted 16 June 2015; First published online 20 July 2015

Key words: Bipolar disorder, endophenotypes, major depressive disorder, neuropsychology, longitudinal studies.

Introduction

Mood disorders including bipolar disorder (BD) and major depressive disorder (MDD) are among the most common mental disorders worldwide and a leading cause of disability (Kessler *et al.* 2005). They are known to aggregate in families, with first-degree relatives of BD patients having a 10-fold excess risk of BD compared with the general population, and a 3-fold increased risk of MDD (Smoller & Finn, 2003). Taking into account that the prevalence of MDD in the population is substantially greater than that of BD, with lifetime prevalence estimates for MDD being in the range of 16% (Kessler *et al.* 2005) versus

1% for BD (Merikangas *et al.* 2011), the absolute risk of developing MDD in first-degree BD relatives is much greater than their absolute risk of developing BD (Smoller & Finn, 2003). Moreover, the majority of first-degree BD relatives who go on to develop BD themselves are initially diagnosed with MDD since depressive episodes commonly emerge prior to the onset of manic episodes (Hillegers *et al.* 2005; Duffy, 2010). This finding together with moderate to high heritability estimates and evidence of a shared genetic architecture provides strong support for overlapping causal pathways in BD and MDD (McGuffin *et al.* 2003; Craddock, 2006; Cross-Disorder Group of the Psychiatric Genomics Consortium, 2013; Schulze *et al.* 2014).

To enhance the identification of susceptibility genes for complex, polygenetic disorders such as BD and MDD, the endophenotype approach has been applied. Endophenotypes are disease-associated traits that are

* Address for correspondence: M. Pappmeyer, Ph.D., University of Edinburgh, Royal Edinburgh Hospital, Morningside Park, Edinburgh EH10 5HF, UK.
(Email: Martina.Pappmeyer@ed.ac.uk)

more proximal to the molecular mechanisms than the clinical phenotype. Criteria for the identification of endophenotypes include that they must be: (a) associated with the illness; (b) heritable; (c) state-independent; (d) co-segregate with illness within families; and (e) found in unaffected relatives at higher rates than in the general population (Gottesman & Gould, 2003).

Several lines of evidence suggest that distinct neurocognitive deficits may qualify as endophenotypes for mood disorders. First of all, the majority of neurocognitive domains are strongly influenced by genetic factors and are highly heritable (Glahn *et al.* 2004, 2012). Second, neurocognitive deficits are known to be found in individuals with manifest mood disorder. A recent meta-analysis of neuropsychological performance in first-episode MDD patients observed neurocognitive impairments in attentional processing speed and cognitive flexibility (Lee *et al.* 2012). Similarly, meta-analyses of first-episode or euthymic BD patients reported reduced attentional processing speed and cognitive flexibility performance (Bourne *et al.* 2013; Lee *et al.* 2014). Moreover, these studies detected verbal learning and memory deficits as well as working memory impairments in BD patients (Bourne *et al.* 2013; Lee *et al.* 2014). These neuropsychological impairments were evident in both symptomatic and euthymic patients, hence suggesting that they may be relatively state-independent.

Third, neurocognitive deficits observed in affected patients have also been found in close unaffected relatives. In particular, first-degree relatives of BD patients have been shown to have deficits during tasks involving attentional processing (Glahn *et al.* 2010), verbal learning and memory (Arts *et al.* 2008; Balanza-Martinez *et al.* 2008; Bora *et al.* 2009), cognitive flexibility (Bora *et al.* 2009) and working memory (Balanza-Martinez *et al.* 2008; Glahn *et al.* 2010). However, the magnitude and consistency of these findings in first-degree relatives is generally reduced in comparison with affected individuals. Accordingly, it has been suggested that the more severe and widespread deficits observed in mood disorders patients may be related to disease-associated factors such as medication, illness progression or psychiatric co-morbidity (Balanza-Martinez *et al.* 2008). This hypothesis is in line with previous findings of associations between neuropsychological impairment in mood disorders with illness progression (Elgamil *et al.* 2010) or medication (Snyder, 2013).

Most studies on neurocognition in mood disorders have assessed neuropsychological performance cross-sectionally in affected individuals or unaffected first-degree relatives. Accordingly, they only provide indirect evidence for the endophenotype criterion of state-independence. They do not, however, provide insights into the precise time course of the evolution

of cognitive deficits in mood disorders by assessing patients or unaffected relatives at various disease stages over time. Given the relative lack of longitudinal studies, it remains largely unknown if cognitive deficits remain relatively stable across time or whether they exhibit dynamic changes over time.

The Scottish Bipolar Family Study is a longitudinal cohort study based on familial risk which allows the examination of the time course of development of neurocognitive deficits in mood disorders and their relationship to familial risk and onset of illness. Based on the neurocognitive endophenotype literature reviewed above, we compared measures of attentional processing speed as assessed with the Digit Span forwards, working memory, verbal learning and memory, and cognitive flexibility. These were compared over a 2-year time interval between three groups of participants: high-risk of mood disorders individuals who were well at baseline but developed MDD during the follow-up period (HR-MDD), high-risk individuals who remained well over the same time period (HR-well) and unaffected healthy control (HC) subjects. We hypothesized that subtle neurocognitive performance deficits related to familial risk of mood disorders are present before the onset of illness. Furthermore, we hypothesized that these performance parameters worsen progressively in the 2-year period prior to illness and that an onset of MDD is associated with more pronounced neurocognitive deficits as compared with individuals who remain well.

Method

Participants

Participants were recruited as part of the Bipolar Family Study (Sprooten *et al.* 2011; Whalley *et al.* 2011). Individuals at high risk of mood disorders, because of a close family history of BD, were identified via affected relatives. They were considered at increased risk of mood disorders because of the known cross-over of risk between BD and MDD as outlined in the Introduction. In brief, it has been well established that first-degree relatives of BD patients are at increased risk of developing MDD or BD as compared with the general population (Smoller & Finn, 2003). Overall, they are more likely to develop MDD than BD as their absolute risk of developing MDD is at least twofold increased as compared with their absolute risk of BD (Smoller & Finn, 2003).

Psychiatrists across Scotland referred patients to the study with a primary diagnosis of BD, type I (BDI). The diagnosis of affected subjects was confirmed with the Operational Criteria Symptom Checklist (McGuffin *et al.* 1991) using information from clinical

case notes and the Structured Clinical Interview for DSM-IV Axis-I Disorders (SCID; First *et al.* 1996). The BDI patients were asked to identify close family members aged 16–25 years. Following informed consent, unaffected individuals with at least one first-degree, or two second-degree relatives with BDI were invited to participate. The participants at high risk of mood disorders were interviewed to confirm a lifetime absence of mood disorders or schizophrenia and to ensure that they did not fulfil any exclusion criteria outlined below.

Unaffected, unrelated control subjects with no personal or family history of BD were identified from the social contacts of the high-risk subjects and group-matched for age, sex and pre-morbid intelligence estimated with the National Adult Reading Test (Nelson, 1982). Comparison subjects were screened for Axis I disorders using the SCID. At baseline, exclusion criteria for all study groups included a personal history of MDD, mania or hypomania, psychosis, or any major neurological or psychiatric disorder, substance dependence, learning disability, head injury that included loss of consciousness and any contraindications to magnetic resonance imaging.

Approximately 2 years after the initial baseline examination, all participants were invited for a follow-up assessment. Written informed consent was acquired from all subjects and the study was approved by the Research Ethics Committee for Scotland.

Clinical assessment

Clinical assessments were conducted at the time of the first and second neuropsychological assessments. The mean interval between assessments was closely matched between the groups ($p \leq 0.820$), being 2.13 (s.d. 0.22), 2.15 (s.d. 0.22), 2.10 (s.d. 0.13) years for the HC, HR-well and HR-MDD groups. The diagnostic status of consenting subjects not returning for a second assessment was determined through written contact with the National Health Service (NHS). Clinical interviews were conducted by experienced psychiatrists (A.M.M., J.E.S.). Based on the follow-up clinical examination or information from case notes, high-risk subjects were grouped into those who remained well (HR-well) and those who subsequently developed MDD (HR-MDD). At both assessments, current manic and depressive symptoms were rated using the Young Mania Rating Scale (YMRS; Young *et al.* 2000) and Hamilton Depression Rating Scale (HAM-D; Hamilton, 1960).

Neuropsychological assessment

Neuropsychological tasks examining a broad range of neurocognitive domains that have been previously

found to be commonly impaired in mood disorders were administered at baseline and follow-up assessment by trained research assistants (A.M., A.P.). To assess attentional processing speed and working memory, the Digit Span of the Wechsler Adult Intelligence Scale (Wechsler, 1955) was administered. The number of correctly recalled strings of numbers during the forwards condition of the task was calculated to examine attentional processing speed. The number of correctly recalled strings of numbers during the backwards condition of the Digit Span served as an indicator of working memory performance.

To assess verbal learning and memory, the California Verbal Learning Test (CVLT; Delis *et al.* 2000) was administered. Here, the number of words correctly recalled during trials 1–5 (CVLT learning) served as an index of verbal learning ability. The number of words recalled during the free short delay recall (CVLT short delay) and free long delay recall (CVLT long delay) were calculated to estimate verbal memory performance.

Cognitive flexibility was examined using the Intra-/Extradimensional Set-Shifting Task (IED) of the Cambridge Neuropsychological Test Automated Battery (CANTAB; Roberts *et al.* 1988). For the IED, the number of trials needed to complete task stage 1 and the number of trials needed to complete task stages 2, 5 and 7 were calculated to assess simple discrimination learning and reversal learning performance, respectively. Furthermore, the number of trials needed to complete task stage 6 and the number of trials needed to complete task stage 8 were computed to extract intradimensional set-shifting and extradimensional set-shifting ability, respectively. The neuropsychological performance parameters extracted for analyses are presented in Table 1.

Statistical analyses

Statistical analyses of demographic and clinical data were conducted using one-way analysis of variance (ANOVA), χ^2 tests or Kruskal–Wallis tests where appropriate using SPSS, version 19 (<http://www.spss.com>; IBM Corp., USA). All statistical analyses regarding neuropsychological performance were computed in SPSS version 19, too, except for false discovery rate (FDR) corrections (Benjamini & Hochberg, 1995) which were conducted in R version 2.13.0 (<http://www.r-project.org>; R Foundation for Statistical Computing, Austria), using the ‘p.adjust(BH)’ function of the ‘stats’ package.

Given the longitudinal study design and the fact that the data consist of non-uniform numbers of repeated measurements, linear mixed-effects models were applied to investigate neuropsychological performance

Table 1. Neuropsychological performance parameters

Task parameter	Measure	Neuropsychological domain
Digit Span forwards	Number of correctly recalled strings of numbers	Attentional processing speed
Digit Span backwards	Number of correctly recalled strings of numbers in reversed order	Working memory
CVLT learning	Number of words correctly recalled during trials 1–5	Verbal learning
CVLT short delay	Number of words correctly recalled during free short delay recall	Verbal memory
CVLT long delay	Number of words correctly recalled during free long delay recall	Verbal memory
IED SDL	Number of trials needed to complete stage 1	Simple discrimination learning
IED RL	Number of trials needed to complete stages 2, 5, 7	Reversal learning
IED IDS	Number of trials needed to complete stage 6	Intradimensional set-shifting
IED EDS	Number of trials needed to complete stage 8	Extradimensional set-shifting

CVLT, California Verbal Learning Test; IED, Intra-/Extradimensional Set-Shifting Task; SDL, simple discrimination learning; RL, reversal learning; IDS, intradimensional set-shifting; EDS, extradimensional set-shifting.

over time. Linear mixed-effects modelling has several advantages over the commonly applied repeated-measures ANOVA as casewise deletion of missing data is not necessary which allows the analysis of all available data. Moreover, it handles the correlation structures of repeated measurements nested within participants. In the linear mixed-effects model used, the intercept term is treated as a random effect that varies by individual so that intraindividual correlation among the neuropsychological performance measures of a particular individual is taken into account. The following independent variables were used as predictors of neurocognitive function: group; time (baseline *v.* follow-up assessment); group \times time interaction. Age and sex served as covariates. Accordingly, significant group effects represent differences in neuropsychological performance between the

groups across both time points. Time effects represent differences in neurocognitive function between baseline and follow-up examination. Group \times time interactions represent differences in neuropsychological performance over time between groups.

A statistical significance level of $p_{\text{FDR}} \leq 0.05$ was chosen, fully corrected for multiple comparisons using the Benjamini and Hochberg FDR procedure (Benjamini & Hochberg, 1995). For ease of comparison of our results with future findings, we report the original uncorrected p values ($p_{\text{uncorrected}}$) and whether or not they survived the FDR procedure. Wherever significant between-group differences or interaction effects were found, pairwise comparisons were performed between the three groups, for which p values were corrected according to Tukey's 'honestly significant difference' (HSD) method ($p_{\text{HSD}} \leq 0.05$).

Wherever significant between-group differences were found in the longitudinal analysis, an additional analysis of covariance (ANCOVA) was conducted between the groups for neuropsychological performance at baseline, adjusted for age and sex. This analysis was intended to assess whether observed longitudinal neuropsychological impairments were also predictive at baseline for an onset of MDD.

To assess the relationship between severity of depressive symptoms and neuropsychological performance, we calculated the Spearman correlation coefficient between the HAM-D total score and the neuropsychological performance parameters for each group. In each case, p values were corrected using the FDR procedure and considered significant when $p_{\text{FDR}} \leq 0.05$.

To examine potentially confounding effects of exposure to medication and relatedness of subjects on neuropsychological performance, we performed the following additional analyses for significant findings: We first repeated our analyses excluding medicated HR-MDD subjects ($n=4$), followed by randomly excluding related subjects from the same pedigree ($n=2$ HC; $n=17$ HR-well; $n=2$ HR-MDD).

Ethical standards

All procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008.

Results

Sociodemographic and clinical characteristics

In total, 114 high-risk individuals provided neuropsychological data along with clinical information at baseline, at which time point none of them had a

clinical mood disorder. Overall, 20 high-risk participants received a diagnosis of MDD within the 2-year period, but one individual had to be excluded from baseline analysis since he did not complete the neuropsychological test battery. Two of the 114 high-risk individuals developed BD during the 2-year follow-up period and were excluded from all analyses due to the small sample size. Accordingly, our analyses included 92 HR-well and 19 HR-MDD subjects at baseline. Of the HC individuals, 96 provided neuropsychological data along with clinical information at baseline. Three developed MDD in the follow-up period and were therefore excluded from all analyses, leading to a sample size of 93 HC subjects. At follow-up, 63 HR-well, 20 HR-MDD and 62 HC subjects provided suitable data. Four HR-MDD participants were prescribed antidepressant medication at follow-up. Three subjects were taking selective serotonin reuptake inhibitors (one fluoxetine, one citalopram, one sertraline) and one participant was on a tricyclic antidepressant (lofepramine). The remaining 16 HR-MDD subjects were unmedicated.

There were no significant differences between the groups in terms of age, sex, handedness, verbal intelligence and YMRS sum score at any assessment point (see Table 2). There were, however, significant group differences at baseline ($p_{\text{uncorrected}} \leq 0.007$) and follow-up ($p_{\text{uncorrected}} \leq 0.023$) for clinical measures of depression from the HAM-D. At baseline, HR-well and HR-MDD subjects had significantly higher depression scores ($p_{\text{HSD}} \leq 0.047$ and $p_{\text{HSD}} \leq 0.003$, respectively) than HC individuals, with no significant differences between the high-risk groups. At follow-up, HR-MDD subjects had higher depression scores than HC and HR-well individuals ($p_{\text{HSD}} \leq 0.013$ and $p_{\text{HSD}} \leq 0.010$, respectively) as expected, with no significant differences between HC and HR-well individuals.

Neurocognitive performance over time

Table 3 provides the results of the linear mixed-effects model analyses. A significant group effect was found for the long delay free recall of the CVLT ($p_{\text{uncorrected}} \leq 0.003$) and extradimensional set-shifting of the CANTAB ($p_{\text{uncorrected}} \leq 0.004$) that passed the FDR procedure (see Fig. 1). *Post-hoc* analyses revealed that HC subjects recalled significantly more words over both assessment time points during the long delay free recall than HR-well participants ($p_{\text{HSD}} \leq 0.002$), with no significant differences between the HC and HR-MDD groups ($p_{\text{HSD}} \leq 0.136$) or the HR-well and HR-MDD groups ($p_{\text{HSD}} \leq 0.485$). For the extradimensional set-shifting performance, *post-hoc* analyses showed that HC subjects had a significantly superior task

performance across both time points as compared with HR-well ($p_{\text{HSD}} \leq 0.031$) and HR-MDD subjects ($p_{\text{HSD}} \leq 0.004$), with no significant difference between the two high-risk groups ($p_{\text{HSD}} \leq 0.077$). These findings are in line with our hypothesis that subtle neurocognitive performance deficits related to familial risk of mood disorders are present before the onset of illness.

Moreover, a significant effect of time was observed for the Digit Span forwards ($p_{\text{uncorrected}} \leq 0.009$), verbal learning ($p_{\text{uncorrected}} \leq 0.003$), free short delay recall ($p_{\text{uncorrected}} \leq 0.006$), simple discrimination learning ($p_{\text{uncorrected}} \leq 0.003$) and extradimensional set-shifting ($p_{\text{uncorrected}} \leq 0.001$) that passed the FDR procedure. For all neurocognitive measures except for simple discrimination learning, the effect was driven by all participant groups displaying enhanced performance during the follow-up assessment as compared with the baseline assessment. By contrast, the time effect for simple discrimination learning ability was in the opposite direction, with all participant groups showing worse performance at the follow-up appointment as compared with the baseline appointment.

The analyses revealed no significant group \times time interactions. This finding is in contrast to our hypothesis that neurocognitive performance parameters worsen progressively in the 2-year period prior to illness onset in HR-MDD as compared with HR-well and HC individuals.

Neurocognitive performance as a predictor of illness onset

To assess whether the observed longitudinal group effects for the long delay free recall of the CVLT and extradimensional set-shifting were also predictive at baseline assessment for a subsequent onset of MDD, additional ANCOVAs were performed for these two neurocognitive measure at baseline (see online Supplementary Table S1). There was a significant group effect for extradimensional set-shifting ($p_{\text{uncorrected}} \leq 0.021$). *Post-hoc* tests indicated that the HR-MDD group ($p_{\text{HSD}} \leq 0.019$) and the HR-well group ($p_{\text{HSD}} \leq 0.038$) needed more trials to successfully complete the extradimensional set shifting stage of the IED as compared with HC participants, with no performance differences between the high-risk groups ($p_{\text{HSD}} \leq 0.232$). There were no significant differences between the groups at baseline for performance during the long delay free recall of the CVLT ($p_{\text{uncorrected}} \leq 0.231$).

Correlation analysis

There were no significant FDR-adjusted correlations between neurocognitive performance measures and

Table 2. Demographic and clinical characteristics

	Baseline			Statistics		Follow-up			Statistics	
	HC (<i>n</i> = 93)	HR-well (<i>n</i> = 92)	HR-MDD (<i>n</i> = 19)	F/χ^2	<i>p</i>	HC (<i>n</i> = 62)	HR-well (<i>n</i> = 63)	HR-MDD (<i>n</i> = 20)	F/χ^2	<i>p</i>
Age, years	21.01 (2.45)	21.20 (2.88)	21.10 (2.82)	0.13	0.88	22.82 (2.73)	23.71 (2.84)	23.33 (2.98)	1.77	0.17
Sex, <i>n</i>				0.46	0.80				2.12	0.35
Male	40	44	9			21	29	7		
Female	53	48	10			41	34	13		
Handedness, <i>n</i>				5.54	0.24				5.43	0.07
Right	88	81	19			61	57	20		
Other	5	11	0			1	6	0		
NART IQ	110.31 (8.00)	108.39 (9.37)	107.26 (6.80)	1.64	0.20	–	–	–	–	–
Time, years	2.13 (0.22)	2.15 (0.22)	2.10 (0.13)	0.20	0.82	–	–	–	–	–
HAM-D ^a	0 (1)	0 (2)	1 (5)	9.79	0.01*	1 (3)	1 (2)	5 (12)	7.59	0.02*
YMRS ^a	0 (0)	0 (0)	0 (0)	3.48	0.18	0 (0)	0 (1)	0 (0)	0.79	0.68

Data are given as mean (standard deviation) unless otherwise indicated.

HC, unaffected healthy control subjects; HR-well, individuals at high risk of mood disorders who were well at baseline and remained well during the follow-up period; HR-MDD, individuals at high risk for mood disorders who were well at baseline but developed major depressive disorder during the follow-up period; NART, National Adult Reading Test; IQ, intelligence quotient; Time, time between baseline and follow-up assessment; HAM-D, Hamilton Depression Rating Scale; YMRS, Young Mania Rating Scale.

^aKruskal–Wallis test, median and interquartile range presented for skewed variables.

*Significant effect.

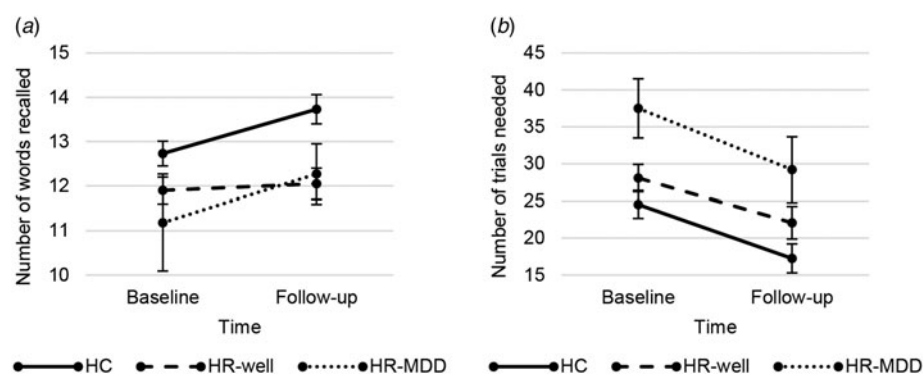


Fig. 1. Performance on (a) the long delay free recall of the California Verbal Learning Test and (b) the extradimensional set-shifting stage of the Cambridge Neuropsychological Test Automated Battery. Values are means, with standard errors represented by vertical bars. Significant group effects were observed ($p < 0.05$). HC, Healthy controls; HR-well, individuals at high risk of mood disorders who were well at baseline and remained well during the follow-up period; HR-MDD, individuals at high risk for mood disorders who were well at baseline but developed major depressive disorder during the follow-up period.

depressive symptom severity as measured with the HAM-D total score (see online Supplementary Table S2).

Analysis of potential confounders

All results remained significant after FDR correction when randomly excluding related subjects. A significant group effect was found for the long delay free recall ($p_{\text{uncorrected}} \leq 0.003$) and extradimensional

set-shifting ($p_{\text{uncorrected}} \leq 0.003$). Moreover, a significant effect of time was observed as in the original analysis for the Digit Span forwards ($p_{\text{uncorrected}} \leq 0.012$), verbal learning ($p_{\text{uncorrected}} \leq 0.005$), free short delay recall ($p_{\text{uncorrected}} \leq 0.012$), simple discrimination learning ($p_{\text{uncorrected}} \leq 0.004$) and extradimensional set-shifting ($p_{\text{uncorrected}} \leq 0.001$).

When excluding medicated HR-MDD subjects from the analysis, the significant group effect for the long

Table 3. Longitudinal analysis of neuropsychological performance

							Statistics					
	HC		HR-well		HR-MDD		Group effect		Time effect		Group × time	
	Baseline	Follow-up	Baseline	Follow-up	Baseline	Follow-up	F	p	F	p	F	p
Digit Span forwards	6.77 (1.28)	7.50 (1.48)	6.87 (1.21)	6.98 (1.34)	6.68 (1.53)	7.10 (1.41)	0.36	0.70	7.02	0.01*	1.35	0.26
Digit Span backwards	5.18 (1.24)	5.46 (1.36)	4.80 (1.14)	5.00 (1.11)	5.22 (1.22)	5.29 (1.71)	2.57	0.08	0.78	0.38	0.20	0.82
CVLT learning	54.55 (10.87)	60.48 (9.47)	54.35 (9.45)	56.57 (10.34)	50.47 (17.72)	56.23 (10.79)	1.63	0.20	9.32	<0.01*	2.05	0.13
CVLT short delay	11.94 (2.74)	13.18 (2.70)	11.35 (2.78)	11.75 (3.00)	10.18 (4.76)	12.00 (3.18)	4.09	0.02	7.77	0.01*	1.38	0.25
CVLT long delay	12.73 (2.67)	13.73 (2.59)	11.90 (2.88)	12.05 (2.77)	11.18 (4.75)	12.27 (3.04)	6.11	<0.01*	1.74	0.19	1.47	0.23
IED SDL	6.61 (0.80)	6.98 (1.60)	7.00 (1.87)	7.31 (2.37)	6.75 (0.86)	7.30 (1.53)	0.54	0.58	8.89	<0.01*	1.28	0.28
IED RL	22.29 (4.47)	22.30 (2.15)	21.97 (2.71)	24.36 (11.95)	24.63 (8.78)	22.70 (2.58)	0.61	0.54	3.59	0.06	0.21	0.81
IED IDS	6.50 (1.33)	6.83 (1.67)	6.50 (1.03)	7.10 (2.79)	6.50 (0.89)	6.25 (0.55)	0.76	0.47	0.80	0.37	1.64	0.20
IED EDS	24.53 (18.03)	17.25 (15.42)	28.07 (17.74)	22.07 (17.47)	37.50 (17.49)	29.20 (19.88)	5.80	<0.01*	19.23	<0.01*	0.22	0.80

Data are given as mean (standard deviation).

HC, unaffected healthy control subjects; HR-well, individuals at high risk of mood disorders who were well at baseline and remained well during the follow-up period; HR-MDD, individuals at high risk for mood disorders who were well at baseline but developed major depressive disorder during the follow-up period; CVLT, California Verbal Learning Test; IED, Intra-/Extradimensional Set-Shifting Task; SDL, simple discrimination learning; RL, reversal learning; IDS, intradimensional set-shifting; EDS, extradimensional set-shifting.

* Significant effect after false discovery rate correction.

delay free recall ($p_{\text{uncorrected}} \leq 0.003$) remained significant after FDR correction. The group effect for extradimensional set-shifting ($p_{\text{uncorrected}} \leq 0.035$) remained significant at nominal level only. The significant effects of time for simple discrimination learning ($p_{\text{uncorrected}} \leq 0.003$) and extradimensional set-shifting ($p_{\text{uncorrected}} \leq 0.001$) remained significant after the FDR procedure. The observed time effects for the Digit Span forwards ($p_{\text{uncorrected}} \leq 0.019$), verbal learning ($p_{\text{uncorrected}} \leq 0.029$) and free short delay recall ($p_{\text{uncorrected}} \leq 0.045$) remained significant at a nominal level only. All results are shown in online Supplementary Tables S3 and S4.

Discussion

This is, to the best of our knowledge, the first prospective longitudinal study examining neurocognitive performance in high-risk of mood disorders individuals who were unaffected at initial assessment and either developed MDD or remained well during the follow-up period. Reduced long delay verbal memory and extradimensional set-shifting performance across the two time points were found in the HR-well group relative to controls, with the HR-MDD group displaying decreased extradimensional set-shifting abilities as compared with the HC group only. An additional analysis of these two neurocognitive domains revealed no significant differences between the two high-risk groups at baseline so that they do not appear to be predictive for a subsequent onset of disease.

In line with our hypothesis, the finding of reduced long delay verbal memory and extradimensional set-shifting performance in the HR-well group across time suggests that neurocognitive deficits in these domains constitute a familial trait marker for vulnerability to mood disorders in close relatives of BD patients. It cannot be determined from the data if the decreased task performance is a consequence of shared genetic and/or environmental effects. Given that they are already present in early adulthood, they are unlikely to be of degenerative origin but probably represent disturbances of normal brain development predisposing to illness. Since no significant differences between the two high-risk groups and no group \times time interaction emerged, the results do not speak towards reduced verbal memory and extradimensional set-shifting abilities to be directly linked to an onset of MDD. These findings are in contrast to our hypothesis that neurocognitive performance parameters worsen progressively in the 2-year period prior to illness onset in HR-MDD as compared with HR-well and HC individuals. It should be highlighted, however, that visual inspection of extradimensional set-shifting performance (Fig. 1) shows that the HR-MDD group

performed worse than the HR-well group across time. Accordingly, it cannot be ruled out that the small sample size of the HR-MDD group did not allow for detection of significant effects due to a lack of power. Indeed, a meta-analysis by Bora *et al.* (2009) showed that effect sizes for cognitive flexibility in healthy relatives of BD patients were small. It appears therefore important to investigate neurocognition further in a larger sample size of high-risk subjects who go on to develop MDD.

Our finding of decreased verbal memory and extradimensional set-shifting in relatives of BD patients are in line with findings from recent meta-analyses (Arts *et al.* 2008; Balanza-Martinez *et al.* 2008; Bora *et al.* 2009). However, it should be noted that the analyses presented here do not confirm previous meta-analyses of attentional processing, verbal learning and working memory deficits in high-risk BD subjects (Arts *et al.* 2008; Balanza-Martinez *et al.* 2008; Bora *et al.* 2009). As outlined previously, one potential reason for the absence of significant findings may be the large sample size that can only be obtained using meta-analyses to detect subtle differences in relatives in these domains (Arts *et al.* 2008). Another putative reason relates to the heterogeneity in neurocognitive tasks that have been employed to study these neurocognitive domains.

From a neuroanatomical point of view, both short-term and long-term storage and retrieval of verbal information have been linked to a bilateral frontal and parietal network of brain regions, including the posterior inferior frontal, anterior middle frontal, anterior cingulate and supramarginal gyrus (Andreassen *et al.* 1995; Henson *et al.* 2000; Dupont *et al.* 2002). Moreover, it has been shown that enhanced performance during the CVLT is associated with higher engagement of the right hippocampus and right frontal lobe (Johnson *et al.* 2001). Dysfunction of this network of brain regions may well be in line with studies suggesting that there may be diminished prefrontal modulation of various brain regions including the anterior cingulate that results in dysregulation of mood as evident in BD (Strakowski *et al.* 2005).

While complex tasks involving extradimensional set-shifting undoubtedly rely on the interplay of various brain regions including lateral prefrontal, orbital and parietal brain areas that may serve as a supervisory attentional network, the ventrolateral prefrontal cortex in particular has been hypothesized to be functionally specialized for extradimensional set-shifting (Hampshire & Owen, 2006). Reduced extradimensional set-shifting performance may thus be in line with a hypothesis of malfunction of the ventral brain system to be underlying the pathogenesis of BD which is thought to be essential for affective processing

and modulation, with the ventrolateral prefrontal cortex playing a central role (Phillips *et al.* 2003a, b). In line with this, Dias *et al.* (1996) showed that lesions in the ventrolateral prefrontal cortex of marmoset monkeys selectively impair extradimensional set-shifting performance as assessed with the same task that we employed in this study.

Importantly, extradimensional set-shifting deficits have also been documented in obsessive-compulsive disorder patients as well as their unaffected first-degree relatives (Chamberlain *et al.* 2007). Accordingly, extradimensional set-shifting deficits appear to form a neurocognitive endophenotype that is not specific to mood disorders such as BD and MDD but rather appears to be a risk marker for a variety of psychiatric conditions. The pathophysiological mechanisms underlying this overlap remain currently unknown. One potential explanation may be related to the function of the ventrolateral prefrontal cortex which is thought to be central to extradimensional set-shifting as described before (Hampshire & Owen, 2006). Ventral aspects of the prefrontal cortex have not only been associated with extradimensional set-shifting performance but also with emotional regulation (Phillips *et al.* 2003b).

The finding of significant time effects for attentional processing speed as assessed with the Digit Span forwards, verbal learning, short delay verbal memory and extradimensional set-shifting in the direction of superior performance during follow-up as compared with baseline assessment across groups most probably reflects the effects of repeated task presentation. These practice effects during neuropsychological examinations have been well documented in the literature (Bartels *et al.* 2010). There was also a significant effect of time for simple discrimination learning due to decreased task performance at follow-up as compared with baseline assessment across all groups. One plausible explanation for this conflicting result may be that participants remembered the task from the baseline assessment and already shifted their attention to the currently irrelevant stimulus as they were expecting the reversal of the rule to occur at follow-up examination.

The strengths of this study are its longitudinal nature, the assessment of subjects prior to illness onset, the relatively young age of the participants and the comparatively large sample size of high-risk subjects and controls. In addition, all subjects underwent careful clinical assessment at both time points and the effects of medication and relatedness of subjects were ruled out.

Nevertheless, some limitations need to be addressed. First, it remains unknown whether currently unaffected HR-well subjects may develop a mood disorder in the future. Second, previous longitudinal studies have reported that the majority of the high-risk

subjects who developed BD themselves experienced depressive episodes years before conversion (Hillegers *et al.* 2005; Duffy, 2010) so that it appears likely that some of our HR-MDD subjects may develop BD in the future. The follow-up assessments of our study cohort will clarify if some of the HR-MDD participants will convert to BD and if some of our HR-well subjects go on to develop a mood disorder. Third, our study groups differed with respect to depression symptom severity at baseline. However, the median of the HAM-D total score was only 1 in the HR-MDD group, suggesting only subsyndromal depression symptoms. Moreover, our correlation analysis revealed no relationship between depression symptom severity and our neurocognitive performance measures. Therefore, it appears unlikely that general mood differences at baseline between the groups have influenced our findings. Fourth, as already highlighted previously, the relatively small HR-MDD sample size might have resulted in a lack of power to detect significant effects. Last, we only assessed neurocognitive function 2 years after the initial baseline assessment and can thus not draw any conclusions if neurocognitive function fluctuated in between the two assessment points. It should be highlighted, however, that the assessment of neurocognitive function using narrower time intervals would have probably caused extensive training effects which are already evident at the follow-up assessment and which would have probably hampered the interpretation of results.

In summary, our findings suggest that reduced long delay verbal memory and extradimensional set-shifting performance across time constitute familial trait markers for vulnerability to mood disorders in close relatives of BD patients. Both neurocognitive performance deficits appear to be relatively stable over a 2-year time period and do not appear to be linked to an onset of MDD. Accordingly, verbal memory and extradimensional set-shifting appear to be relatively state-independent, which supports their potential as putative endophenotypes for mood disorders. These findings add important information for the identification of neurocognitive endophenotypes for mood disorders. Future longitudinal studies should particularly examine the time course of neurocognitive performance before and after the onset of depression using longer time intervals and larger sample sizes.

Supplementary material

For supplementary material accompanying this paper visit <http://dx.doi.org/10.1017/S0033291715001324>

Acknowledgements

The University of Edinburgh is supported by the Wellcome Trust through a Strategic Award (104036/Z/14/Z). The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007–2013) under grant no. 602450 through the Imaging Genetics for Mental Disorders (IMAGEMEND) project. The investigators also acknowledge the financial support of NHS Research Scotland through the Scottish Mental Health Research Network (<http://www.smhrn.org.uk>) who provided assistance with subject recruitment and cognitive assessments. M.P. has been supported by a studentship from the Medical Research Council. J.E.S. is supported by a Clinical Research Training Fellowship from the Wellcome Trust. J.H. is supported by a Senior Clinical Fellowship from the Chief Scientists Office in Scotland. H.C.W. is supported by a College Fellowship from the University of Edinburgh and a John, Margaret, Alfred and Stewart *Sim* (JMAS *Sim*) fellowship from the Royal College of Physicians of Edinburgh. A.M.M. was supported by the Health Foundation through a Clinician Scientist Fellowship (reference 2268/4295), by the Brain and Behaviour Research Foundation through a National Alliance for Research on Schizophrenia and Depression (NARSAD) Independent Investigator Award, and currently by a Scottish Funding Council Senior Clinical Fellowship. The funders had no role in study design, data collection, data analysis, data interpretation, or writing of the report. We thank all participants and their families for taking part in this study.

Declaration of Interest

S.M.L. has received research funding from Abbvie, Roche and Sunovion for therapeutic studies of people with schizophrenia, and has been paid by Janssen, Roche and Sunovion to speak at chair educational meetings about schizophrenia, as well as to contribute to advisory boards about new antipsychotic treatments. S.M.L., J.H., H.C.W. and A.M.M. have received financial support from Pfizer (formerly Wyeth) in relation to imaging studies of people with schizophrenia and BP. M.P., J.E.S., J.M., A.P. and A.M. report no potential financial conflicts of interest.

References

Andreasen NC, O'Leary DS, Arndt S, Cizadlo T, Hurtig R, Rezaei K, Watkins GL, Ponto LL, Hichwa RD (1995). Short-term and long-term verbal memory: a positron emission

tomography study. *Proceedings of the National Academy of Sciences of the United States of America* **92**, 5111–5115.

- Arts B, Jabben N, Krabbendam L, van Os J (2008). Meta-analyses of cognitive functioning in euthymic bipolar patients and their first-degree relatives. *Psychological Medicine* **38**, 771–785.
- Balanza-Martinez V, Rubio C, Selva-Vera G, Martinez-Aran A, Sanchez-Moreno J, Salazar-Fraile J, Vieta E, Tabares-Seisdedos R (2008). Neurocognitive endophenotypes (endophenocognitypes) from studies of relatives of bipolar disorder subjects: a systematic review. *Neuroscience and Biobehavioral Reviews* **32**, 1426–1438.
- Bartels C, Wegrzyn M, Wiedl A, Ackermann V, Ehrenreich H (2010). Practice effects in healthy adults: a longitudinal study on frequent repetitive cognitive testing. *BMC Neuroscience* **11**, 118.
- Benjamini Y, Hochberg Y (1995). Controlling the false discovery rate – a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B-Methodological* **57**, 289–300.
- Bora E, Yucel M, Pantelis C (2009). Cognitive endophenotypes of bipolar disorder: a meta-analysis of neuropsychological deficits in euthymic patients and their first-degree relatives. *Journal of Affective Disorders* **113**, 1–20.
- Bourne C, Aydemir O, Balanza-Martinez V, Bora E, Brissos S, Cavanagh JT, Clark L, Cubukcuoglu Z, Dias VV, Dittmann S, Ferrier IN, Fleck DE, Frangou S, Gallagher P, Jones L, Kieseppa T, Martinez-Aran A, Melle I, Moore PB, Mur M, Pfennig A, Raust A, Senturk V, Simonsen C, Smith DJ, Bio DS, Soeiro-de-Souza MG, Stoddart SD, Sundet K, Szoke A, Thompson JM, Torrent C, Zalla T, Craddock N, Andreassen OA, Leboyer M, Vieta E, Bauer M, Worhunsky PD, Tzagarakis C, Rogers RD, Geddes JR, Goodwin GM (2013). Neuropsychological testing of cognitive impairment in euthymic bipolar disorder: an individual patient data meta-analysis. *Acta Psychiatrica Scandinavica* **128**, 149–162.
- Chamberlain SR, Fineberg NA, Menzies LA, Blackwell AD, Bullmore ET, Robbins TW, Sahakian BJ (2007). Impaired cognitive flexibility and motor inhibition in unaffected first-degree relatives of patients with obsessive-compulsive disorder. *American Journal of Psychiatry* **164**, 335–338.
- Craddock N (2006). Genetics of mood disorders. *Psychiatry* **5**, 170–174.
- Cross-Disorder Group of the Psychiatric Genomics Consortium (2013). Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet* **381**, 1371–1379.
- Delis DC, Kramer JH, Kaplan E, Ober BA (2000). *California Verbal Learning Test: Second Edition*. Psychological Corporation: San Antonio, TX.
- Dias R, Robbins TW, Roberts AC (1996). Dissociation in prefrontal cortex of affective and attentional shifts. *Nature* **380**, 69–72.
- Duffy A (2010). From predisposition to illness: genetically sensitive intermediate pathways to mood disorders. *British Journal of Psychiatry* **197**, 341–342.

- Dupont S, Samson Y, Le Bihan D, Baulac M (2002). Anatomy of verbal memory: a functional MRI study. *Surgical and Radiologic Anatomy* 24, 57–63.
- Elgamal S, Denburg S, Marriott M, MacQueen G (2010). Clinical factors that predict cognitive function in patients with major depression. *Canadian Journal of Psychiatry* 55, 653–661.
- First MB, Spitzer RL, Gibbon M, Williams JBW (1996). *Structured Clinical Interview for DSM-IV Axis I Disorders, Clinical Version (SCID-CV)*. American Psychiatric Press: Washington, DC.
- Glahn DC, Almasy L, Barguil M, Hare E, Peralta JM, Kent JW Jr, Dassori A, Contreras J, Pacheco A, Lanzagorta N, Nicolini H, Raventos H, Escamilla MA (2010). Neurocognitive endophenotypes for bipolar disorder identified in multiplex multigenerational families. *Archives of General Psychiatry* 67, 168–177.
- Glahn DC, Bearden CE, Niendam TA, Escamilla MA (2004). The feasibility of neuropsychological endophenotypes in the search for genes associated with bipolar affective disorder. *Bipolar Disorders* 6, 171–182.
- Glahn DC, Curran JE, Winkler AM, Carless MA, Kent JW Jr, Charlesworth JC, Johnson MP, Goring HH, Cole SA, Dyer TD, Moses EK, Olvera RL, Kochunov P, Duggirala R, Fox PT, Almasy L, Blangero J (2012). High dimensional endophenotype ranking in the search for major depression risk genes. *Biological Psychiatry* 71, 6–14.
- Gottesman II, Gould TD (2003). The endophenotype concept in psychiatry: etymology and strategic intentions. *American Journal of Psychiatry* 160, 636–645.
- Hamilton M (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery and Psychiatry* 23, 56–62.
- Hampshire A, Owen AM (2006). Fractionating attentional control using event-related fMRI. *Cerebral Cortex* 16, 1679–1689.
- Henson RN, Burgess N, Frith CD (2000). Recoding, storage, rehearsal and grouping in verbal short-term memory: an fMRI study. *Neuropsychologia* 38, 426–440.
- Hillegers MH, Reichart CG, Wals M, Verhulst FC, Ormel J, Nolen WA (2005). Five-year prospective outcome of psychopathology in the adolescent offspring of bipolar parents. *Bipolar Disorders* 7, 344–350.
- Johnson SC, Saykin AJ, Flashman LA, McAllister TW, Sparling MB (2001). Brain activation on fMRI and verbal memory ability: functional neuroanatomic correlates of CVLT performance. *Journal of the International Neuropsychological Society* 7, 55–62.
- Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry* 62, 593–602.
- Lee RS, Hermens DF, Porter MA, Redoblado-Hodge MA (2012). A meta-analysis of cognitive deficits in first-episode major depressive disorder. *Journal of Affective Disorders* 140, 113–124.
- Lee RS, Hermens DF, Scott J, Redoblado-Hodge MA, Naismith SL, Lagopoulos J, Griffiths KR, Porter MA, Hickie IB (2014). A meta-analysis of neuropsychological functioning in first-episode bipolar disorders. *Journal of Psychiatric Research* 57, 1–11.
- McGuffin P, Farmer A, Harvey I (1991). A polydiagnostic application of operational criteria in studies of psychotic illness. Development and reliability of the OPCRIT system. *Archives of General Psychiatry* 48, 764–770.
- McGuffin P, Rijsdijk F, Andrew M, Sham P, Katz R, Cardno A (2003). The heritability of bipolar affective disorder and the genetic relationship to unipolar depression. *Archives of General Psychiatry* 60, 497–502.
- Merikangas KR, Jin R, He JP, Kessler RC, Lee S, Sampson NA, Viana MC, Andrade LH, Hu C, Karam EG, Ladea M, Medina-Mora ME, Ono Y, Posada-Villa J, Sagar R, Wells JE, Zarkov Z (2011). Prevalence and correlates of bipolar spectrum disorder in the world mental health survey initiative. *Archives of General Psychiatry* 68, 241–251.
- Nelson H (1982). *The National Adult Reading Test Manual*. NFER-Nelson: Windsor.
- Phillips ML, Drevets WC, Rauch SL, Lane R (2003a). Neurobiology of emotion perception I: the neural basis of normal emotion perception. *Biological Psychiatry* 54, 504–514.
- Phillips ML, Drevets WC, Rauch SL, Lane R (2003b). Neurobiology of emotion perception II: implications for major psychiatric disorders. *Biological Psychiatry* 54, 515–528.
- Roberts AC, Robbins TW, Everitt BJ (1988). The effects of intradimensional and extradimensional shifts on visual discrimination learning in humans and non-human primates. *Quarterly Journal of Experimental Psychology Section B – Comparative and Physiological Psychology* 40, 321–341.
- Schulze TG, Akula N, Breuer R, Steele J, Nalls MA, Singleton AB, Degenhardt FA, Nöthen MM, Cichon S, Rietschel M; Bipolar Genome Study, McMahon FJ (2014). Molecular genetic overlap in bipolar disorder, schizophrenia, and major depressive disorder. *World Journal of Biological Psychiatry* 15, 200–208.
- Smoller JW, Finn CT (2003). Family, twin, and adoption studies of bipolar disorder. *American Journal of Medical Genetics Part C: Seminars in Medical Genetics* 123C, 48–58.
- Snyder HR (2013). Major depressive disorder is associated with broad impairments on neuropsychological measures of executive function: a meta-analysis and review. *Psychological Bulletin* 139, 81–132.
- Sprooten E, Sussmann JE, Clugston A, Peel A, McKirdy J, Moorhead TW, Anderson S, Shand AJ, Giles S, Bastin ME, Hall J, Johnstone EC, Lawrie SM, McIntosh AM (2011). White matter integrity in individuals at high genetic risk of bipolar disorder. *Biological Psychiatry* 70, 350–356.
- Strakowski SM, Delbello MP, Adler CM (2005). The functional neuroanatomy of bipolar disorder: a review of neuroimaging findings. *Molecular Psychiatry* 10, 105–116.
- Wechsler D (1955). *Wechsler Adult Intelligence Scale (Manual)*. Psychological Corporation: New York.
- Whalley HC, Sussmann JE, Chakirova G, Mukerjee P, Peel A, McKirdy J, Hall J, Johnstone EC, Lawrie SM, McIntosh AM (2011). The neural basis of familial risk and temperamental variation in individuals at high risk of bipolar disorder. *Biological Psychiatry* 70, 343–349.
- Young RC, Biggs JT, Ziegler VE, Meyer DA (2000). Young Mania Rating Scale. In *Handbook of Psychiatric Measures* (ed. American Psychiatric Association), pp. 540–542. American Psychiatric Association: Washington, DC.