Acta Psychiatrica Scandinavica

Meta-analysis

Neuropsychological testing of cognitive impairment in euthymic bipolar disorder: an individual patient data meta-analysis

Bourne C, Aydemir Ö, Balanzá-Martínez V, Bora E, Brissos S, Cavanagh JTO, Clark L, Cubukcuoglu Z, Videira Dias V, Dittmann S, Ferrier IN, Fleck DE, Frangou S, Gallagher P, Jones L, Kieseppä T, Martínez-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 SDR, Sundet K, Szöke 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. Neuropsychological testing of cognitive impairment in euthymic bipolar disorder: an individual patient data meta-analysis.

Objective: An association between bipolar disorder and cognitive impairment has repeatedly been described, even for euthymic patients. Findings are inconsistent both across primary studies and previous meta-analyses. This study reanalysed 31 primary data sets as a single large sample ($N = 2876$) to provide a more definitive view.
Method: Individual patient and control data were obtained from original authors for 11 measures from four common neuropsychological tests: California or Rey Verbal Learning Task (VLT), Trail Making Test (TMT), Digit Span and/or Wisconsin Card Sorting Task.
Results: Impairments were found for all 11 test-measures in the bipolar group after controlling for age, IQ and gender ($Ps \leq 0.001$, E.S. = 0.26–0.63). Residual mood symptoms confound this result but cannot account for the effect sizes found. Impairments also seem unrelated to drug treatment. Some test-measures were weakly correlated with illness severity measures suggesting that some impairments may track illness progression.
Conclusion: This reanalysis supports VLT, Digit Span and TMT as robust measures of cognitive impairments in bipolar disorder patients. The heterogeneity of some test results explains previous differences in meta-analyses. Better controlling for confounds suggests deficits may be smaller than previously reported but should be tracked longitudinally across illness progression and treatment.

C. Bourne[1,2,*], Ö. Aydemir[3], V. Balanzá-Martínez[4], E. Bora[5], S. Brissos[6,7], J. T. O. Cavanagh[8], L. Clark[9], Z. Cubukcuoglu[10], V. V. Dias[11], S. Dittmann[12], I. N. Ferrier[13], D. E. Fleck[14], S. Frangou[15], P. Gallagher[13], L. Jones[16], T. Kieseppä[17,18], A. Martínez-Aran[19,*], I. Melle[20,21], P. B. Moore[13], M. Mur[22], A. Pfennig[23,*], A. Raust[24,*], V. Senturk[25], C. Simonsen[20,26], D. J. Smith[27], D. S. Bio[28], M. G. Soeiro-de-Souza[28], S. D. R. Stoddart[16], K. Sundet[20,26], A. Szöke[24,29], J. M. Thompson[13], C. Torrent[19,*], T. Zalla[30], N. Craddock[27], O. A. Andreassen[20,21,*], M. Leboyer[24,29,*], E. Vieta[19,*], M. Bauer[23,*], P. D. Worhunsky[1], C. Tzagarakis[1], R. D. Rogers[1], J. R. Geddes[1,*], G. M. Goodwin[1,*]

Key words: bipolar disorder; cognitive impairment; review; neuropsychological tests

Corin Bourne, Department of Psychiatry, University of Oxford, Warneford Hospital, Oxford OX3 7JX, UK. E-mail: corin.bourne@psych.ox.ac.uk

*European Network of Bipolar Research Expert Centres.

Summations

- Cognitive deficits are present in euthymic bipolar patients, and although some confounds may explain part of the previously reported effect sizes, they cannot entirely explain the impairments.
- Individual patient data meta-analysis has important advantages over the use of published summary data for systematic review especially with regard to controlling for confounds.

Considerations

- The relative lack of drug effects on neuropsychological test performance should be treated with caution as this mega-analysis could not take into account duration or dosage of each drug treatment.
- Similarly, the correlational analysis suggesting that some impairments may track illness progression should also be treated with caution until longitudinal data supports the causality of this relationship.

[1]Department of Psychiatry, University of Oxford, Oxford, [2]Department of Psychology & Counselling, Newman University, Birmingham, UK, [3]Department of Psychiatry, Celal Bayar University, Manisa, Turkey, [4]Service of Psychiatry, University Hospital Doctor Peset, University of Valencia, CIBERSAM, Valencia, Spain, [5]Department of Psychiatry, University of Melbourne, Melbourne, Vic., Australia, [6]Lisbon's Psychiatric Hospitalar Centre, Lisbon, Portugal, [7]Janssen Pharmaceutical, Queluz de Baixo, Portugal, [8]Department of Psychological Medicine, University of Glasgow, Glasgow, [9]Department of Psychology, University of Cambridge, Cambridge, UK, [10]Department of Child and Adolescent Psychiatry, LVR-Kliniken Bedburg-Hau, Bedburg-Hau, Germany, [11]Bipolar Disorder Research Program, Faculty of Medicine, Hospital Santa Maria, University of Lisbon (FMUL), Lisbon, Portugal, [12]Department of Psychiatry, Ludwig-Maximilians-Universität, Munich, Germany, [13]Institute of Neuroscience, Newcastle University, Newcastle upon Tyne, UK, [14]Department of Psychiatry and Behavioral Neuroscience, University of Cincinnati Academic Health Center, Cincinnati, OH, USA, [15]Mount Sinai School of Medicine, NY, USA, [16]Department of Psychiatry, School of Clinical & Experimental Medicine, University of Birmingham, Birmingham, UK, [17]Department of Mental Health and Substance Abuse Services, National Institute for Health and Welfare, Helsinki, [18]Department of Psychiatry, Peijas Hospital, Helsinki University Central Hospital, Vantaa, Finland, [19]Institute of Neurosciences, Hospital Clinic, University of Barcelona, IDIBAPS, CIBERSAM, Barcelona, Catalonia, Spain, [20]Division of Mental Health and Addiction, Oslo University Hospital, Ulleval, Oslo, [21]Institute of Clinical Medicine, University of Oslo, Oslo, Norway, [22]Psychiatry Service, Santa Maria Hospital, University of Lleida, IRBLleida, Lleida, Spain, [23]Department of Psychiatry and Psychotherapy, University Hospital Carl Gustav Carus, Technische Universität Dresden, Dresden, Germany, [24]AP-HP, Service de psychiatrie, Hôpital H. Mondor — A. Chenevier, Créteil, France, [25]Department of Psychiatry, Ankara University School of Medicine, Ankara, Turkey, [26]Department of Psychology, University of Oslo, Oslo, Norway, [27]Department of Psychological Medicine and Neurology, Cardiff University School of Medicine, Cardiff, UK, [28]Mood Disorders Unit (GRUDA), School of Medicine, University of Sao Paulo (HC-FMUSP), Sao Paulo, Brazil, [29]Inserm, U955, Equipe 15, Créteil and [30]Institut Jean Nicod? CNRS, Ecole Normale Supérieure, Paris, France

Introduction

Bipolar disorder has been associated with cognitive impairment even in euthymia (1–4). Specific domains of impairments include the following: executive control (verbal and category fluency, mental manipulation, set shifting, response inhibition), verbal learning and memory, visual memory and attention (5–12). A subset of such deficits may also be present in first-degree relatives of bipolar patients indicating a possible endophenotype for bipolar disorder (6, 7) and a starting point for further genetic understanding of the disorder. Some studies (10) have reported a correlation between a subset of cognitive decrements and illness history, suggesting the competing possibility that clinical episodes may cause impairments in the neuronal function relevant to these domains. Such acquired effects might be more amenable to improved treatment. Given the important potential implications for the neurobiology of bipolar disorder and its treatment, it is disappointing that these claims rest on studies of modest size that allow little confidence in their individual conclusions. Replication is confused by the adoption of a too wide range of different neuropsychological tests with varying sensitivity to and specificity for particular cognitive domain impairments and their neural substrates.

Between 2006 and 2010, four papers have conducted meta-analyses on the cognitive deficits associated with bipolar disorder in purely euthymic patients: Arts et al. (6); Bora et al. (7); Robinson et al. (8) and Torres et al. (9). A fifth paper, Robinson and Ferrier (10), provided a narrative review of studies that considered the relationship between illness variables and cognitive deficits. Surprisingly, despite the similar aims, similar search terms and overlapping databases used across the five papers, they demonstrate wide variation in the primary studies chosen for inclusion and in their specific conclusions. An additional meta-analysis was published in 2011: Mann-Wrobel et al. (12). This paper differed from three of the earlier meta-analyses (6, 8, 9) by not supporting a differential impairment in verbal memory and executive function. Kurtz & Gerraty (13) provided a further meta-analysis although they considered clinical groups other than purely euthymic patients. If this study is restricted to euthymic patients, then their meta-analysis had similar sample sizes and effect size range to the meta-analyses reviewed in more detail here (see Table 1). However, the authors also suggested a greater impairment was present for verbal memory but also for non-verbal delayed memory in contrast to other meta-analyses (6, 7).

The present study provides an independent individual patient data meta-analysis (IPDMA) of the data from the most comparable studies included in the previous reviews, given the authors could provide original data for pooling. IPDMA has not been widely used in psychiatry although it is increasingly used in medical genetics (14–16) where it is often termed 'mega-analysis'. It has important

advantages over the use of published summary data for systematic review (17). In particular, IP-DMA allows the primary study effect sizes to be adjusted for confounding factors (i.e. factors such as age, education and IQ) prior to meta-analysis and for a large data set to be analysed for drug and illness severity effects. The latter having been previously restricted to primary studies of modest sample size or narrative review. The adjustment for confounding factors is especially valuable because, although some of the primary studies were very tightly matched case–control studies focusing on one or two neuropsychological tests, other included studies were more opportunistic samples running large neuropsychological test batteries with more sample variation. In a standard meta-analysis, the results from these two types of study are combined without adjustment.

Aims of the study

The main aim of the study was to synthesize data demonstrating cognitive deficits in euthymic bipolar patients in such a way as to be able to adjust for confounding factors to provide a more definitive estimate for effects sizes than in prior meta-analyses. A secondary aim was to create a large data set to provide a more definitive view of drug and illness severity effects on cognitive impairments than has been possible in relative small sample primary studies. We chose to include tests that had appeared consistently in the meta-analyses as showing impairment and for which data were actually available for the majority of individual patients.

Material and methods

Table 1 shows the results from the four existing meta-analyses as the rank of the neuropsychological tests showing the largest effects in each review. Effect sizes appear to be relatively large, but it is striking that sample numbers vary considerably due to the differences in criteria for study inclusion. The relative order of neuropsychological tests when ranked by effect size is variable from analysis to analysis partly due to the variation in study inclusion and probably partly due to noise.

Primary data were sought that tested both euthymic bipolar patients and healthy controls (aged 18–65) on at least one of four key neuropsychological tasks identified in Table 1: i) a verbal learning and memory task, that is, California Verbal Learning Task (CVLT) (18) or Rey Verbal Learning Task (RAVLT) (19); ii) the Trail Making Test (TMT) (20) as a measure of set shifting and pro-

cessing speed; iii) Digit Span [from WAIS-R Digit Span (21)] as a non-word working memory span task and iv) Wisconsin Card Sorting Task (WCST) (22) as a measure of set shifting and rule discovery. Verbal Learning Task (VLT), TMT and WCST all appear in the International Society for Bipolar Disorders recently recommended battery for neuropsychological assessment (23).

From the four selected neuropsychological tests, we focused on 11 specific outcome measures: VLT total score on trials 1–5 (Total1–5), VLT score on Short Delay (ShortDelay), VLT score on Long Delay (LongDelay), VLT score on Recognition (Recognition), VLT score for Recognition minus score for False Positives (Recog-FP); time to complete Trail Making Test A (TMTA), time to complete Trail Making Test B (TMTB); score on Forward Digit Span (FDS), score on Reverse Digit Span (RDS); number of categories achieved on Wisconsin Card Sorting Task (WCSTCats.) and number of perseverations on Wisconsin Card Sorting Task (WCSTPersev.).

Where possible, demographic and clinical variables were also collected for each primary data set including i) age; ii) IQ; iii) current mood; iv) age at onset; v) number of prior manic and depressed episodes; vi) number of prior manic and depressed hospitalizations and vii) drug treatment history.

Search strategy

Given the existence of five recent prior reviews in this area (each with similar but different search terms and inclusion/exclusion criteria), this study did not conduct an additional full systematic search under PRISMA (24) rules. Rather, in an attempt to include all the primary studies that had been in the prior reviews, all first authors of studies appearing in the five review papers that contained data on at least one of the four required neuropsychological tests were contacted. In addition, PsychInfo and PubMed databases were searched with the key concepts of bipolar disorder, euthymia and cognitive impairment to find any additional primary studies that met our criteria. These searches were restricted to articles published between 1 January 2007 and 30 June 2010 in English language peer-reviewed journals. In total, 45 primary studies were identified from 41 different authors (see Table S1). This number is smaller than may have first appeared from the literature search as some studies incorporated data sets used in other published studies and therefore did not constitute mutually exclusive data sets. Of the 45 eligible published studies, full data were provided by primary

Table 1. Summary of the effect sizes found for neuropsychological performance of bipolar patients relative to healthy controls. Top seven effect sizes in the meta-analysis by (a) Arts et al. (6), (b) Bora et al. (7), (c) Robinson et al. (8) and (d) Torres et al. (9)

	Neuropsychological test	Cognitive domain	N (bipolar)	N (control)	Effect size	P
(a)						
1	RDS	Executive	222	205	1.02	<0.0001
2	TMTB	Executive	309	306	0.99	<0.0001
3	WCST (Perseveration)	Resp. Inhib	268	288	0.88	<0.0001
4	Category Fluency	Executive	178	178	0.87	<0.0001
5	Rey/CVLT (Delayed Recall)	Verb. L + M	269	282	0.85	<0.0001
6	Digit Symbol Subtest	Attention	202	249	0.84	<0.0001
7	Rey/CVLT (Total Recall)	Verb. L + M	369	382	0.82	<0.0001
(b)						
1	TMTB	Executive	793	626	0.86	<0.0001
2	Rey/CVLT (Learning)	Verb. L + M	619	632	0.85	<0.0001
3	CPT Omission	Attention	303	279	0.83	<0.0001
4	Rey/CVLT (Delayed Recall)	Verb. L + M	578	612	0.77	<0.0001
5	Stroop	Resp. Inhib	746	707	0.76	<0.0001
6	Digit Symbol Subtest	Attention	381	479	0.75	<0.0001
7	RDS	Executive	375	487	0.75	<0.0001
(c)						
1	Category Fluency	Executive	149	135	1.09	<0.0001
2	RDS	Executive	222	209	0.98	0.0031
3	Rey/CVLT (Total Recall)	Verb. L + M	344	347	0.90	<0.0001
4	TMTB	Executive	418	355	0.78	<0.0001
5	WCST (Perseveration)	Resp. Inhib	195	216	0.76	<0.0001
6	Rey/CVLT (Short Free Recall)	Verb. L + M	345	349	0.73	<0.0001
7	Rey/CVLT (Long Free Recall)	Verb. L + M	365	368	0.71	<0.0001
(d)						
1	Rey/CVLT (Total Recall)	Verb. L + M	381	439	0.81	<0.0001
2	Digit Symbol Subtest	Attention	222	310	0.79	<0.0001
3	Rey/CVLT (Short Delay)	Verb. L + M	315	307	0.74	<0.0001
4	CPT Hits	Attention	188	208	0.74	<0.0001
5	Rey/CVLT (Long Delay)	Verb. L + M	361	441	0.72	<0.0001
6	Stroop	Resp. Inhib	346	329	0.71	<0.0001
7	WCST (Perseveration)	Resp. Inhib	244	229	0.69	<0.0001

CPT, continuous performance task; CVLT, California Verbal Learning Task; RDS, Reverse Digit Span; Resp. Inhib, Response Inhibition; TMTB, Trail Making Test B; Verb. L + M, Verbal Learning and Memory; WCST, Wisconsin Card Sorting Task.

authors in relation to 25 published papers (4, 25–48) with the data from the remaining 20 eligible studies unavailable and therefore not included in this reanalysis. Additionally, new primary data that met our criteria were also provided in relation to six unpublished data sets (49, 50) (A. Macritche, manuscript in preparation; A. Varma, manuscript in preparation; A. Pfennig, M. Alda, T. Young, G. MacQueen, J. Rybakowski, A. Suwalska, C. Simhandl, B. König, T. Hajek, C. O'Donovan, S. von Quillfeldt, D. Wittekind, J. Ploch, C. Sauer, M. Bauer, manuscript in preparation; M.G. Soeiro-de-Souza & D. Soares-Bio, manuscript in preparation), giving a total of 31 primary data sets for this reanalysis as shown in Table 2.

Where mood scores were available, euthymia was defined as ≤ 8 on Hamilton Depression Rating Scale (HDRS) (51) or ≤ 15 on Montgomery–Asberg Depression Rating Scale (MADRS) (52) or ≤ 11 on Inventory of Depressive Symptomatology (Clinician Rating; IDS-C) (53) and ≤ 8 on Young Mania Rating Scale (YMRS) (54) or ≤ 8 on Clini-

cian Administered Rating Scale for Mania Factor 1 CARS-M(F1) (55) or ≤ 20 on Manic State Rating Scale (MSRS) (56). If no mood ratings were available, then euthymia had been assessed by a qualified psychiatrist only.

The total sample size for the reanalysis was therefore 2876 participants: 1276 euthymic bipolar patients (54.7% female) and 1609 healthy controls (53.5% female). The bipolar patients were 83.5% Bipolar I, 12.3% Bipolar II, 2.7% Bipolar NOS, 1.4% Schizoaffective Disorder.

Statistical analyses

Parametric statistical tests were used to compare a variety of demographic variables between bipolar patients and healthy controls. Where appropriate, homogeneity of variance was checked using Levene's test. All continuous measures (including depression and mania scores) were converted to standardized z-scores within each study sample (patients plus controls) before further analysis.

Group effect size of cognitive deficits. To investigate group (patient vs. control) effects on neuropsychological performance, group, age, IQ and gender were regressed on to each of the 11 neuropsychological test outcome measures within each of the 31 studies. For the eight studies that did not use an explicit measure of IQ, years of education was used as a proxy ($r_p = 0.50$, $P < 0.001$). The regression coefficient and standard error for group within each study were then entered for meta-analysis for each outcome variable. Thus, the meta-analysis was effectively performed on study group effect sizes adjusted *a priori* for the confounds of age, IQ and gender. The meta-analyses were conducted on both fixed and random effects assumptions, but results did not differ materially. This analysis did not use the more standard IPDMA technique of mixed model regression (with fixed and random effects) as the between-study heterogeneity for group effect size was considered too high for at least some of the outcome measures (see Table 4).

Residual mood effects. Residual mood symptoms (both depression and mania) could not be added to the above analysis because they were confounded with group. However, in an attempt to understand how much of the group effect on performance might be attributable to residual confounding by mood, two further analyses were conducted. The first approach used meta-regression, with each of the studies ascribed a factor relating to the relative level of residual mood symptoms in the patient group. The second method considered mood effects within the patient group only using mixed model regression with data collapsed across studies. Depression scores and mania scores along with age, IQ, gender (all fixed effects) and study (random effect) were regressed on to each of the 11 neuropsychological test outcome measures.

Drug effects within patient group. To investigate potential drug effects within the patient group, mixed model linear regression was used. Patients were coded for five binary (yes/no) drug status variables: lithium, anticonvulsants, antipsychotics, antidepressants and drug free. Each drug status variable (fixed effect) together with age, IQ, gender (fixed effects) and study (random effect) was regressed on to each of the 11 neuropsychological test outcome measures.

Relationship between illness variables and cognitive deficits. Mixed model linear regression was also used to investigate potential relationships between illness severity measures and neuropsychological test performance within the patient group. Num-

ber of depressed episodes, number of manic episodes, total number of episodes, number of depressed hospitalizations, number of manic hospitalizations, total number of hospitalizations and illness duration were each fitted separately into the regression model with age, IQ and gender as universal confounders (fixed effect) and study (random effect) for each of the 11 neuropsychological test outcome measures.

Statistical analysis was conducted in R 2.12.2 (The R Foundation for Statistical Computing, Vienna, Austria) except for the meta-analysis which was conducted in STATA IC Version 11 (StataCorp LP., College Station, TX, USA). All statistical tests were two-tailed.

Table 2. List of studies in reanalysis data set

	Study	N	N_{bp}	N_{cont}
1	Balanza-Martinez et al. (26)	41	15	26
2	Bora et al. (27)	95	65	30
3	Cavanagh et al. (28)	39	19	20
4	Clark et al. (29)	60	30	30
5	Cubukcuoglu & Aydemir (49)	101	51	50
6	Dias et al. (46)	115	65	50
7	Dittmann et al. (30)	116	74	42
8	El-Badri et al. (31)	57	30	27
9	Fleck et al. (32)	51	11	40
10	Fleck et al. (33)	70	22	48
11	Frangou et al. (34)	86	42	44
12	Goswami et al. (35)	74	37	37
13	Hellvin et al. (50)*	228	63	165
14	Kaya et al. (48)	62	43	19
15	Kieseppa et al. (36)	140	26	114
16	A. Macritche (manuscript in preparation)	56	28	28
17	Martinez-Aran et al. (4)	69	39	30
18	Martinez-Aran et al. (37)	112	77	35
19	Mur et al. (38)	89	43	46
20	A. Pfennig, M. Alda, T. Young, et al. (manuscript in preparation)	54	33	21
21	Senturk et al. (39)	56	27	29
22	Simonsen et al. (25)	146	29	117
23	Simonsen et al. (47)†	204	31	173
24	Smith et al. (40)	54	21	33
25	M.G. Soeiro-de-Souza & D. Soares-Bio (manuscript in preparation)	134	38	96
26	Stoddart et al. (41)	59	19	40
27	Szoke et al. (42)	145	97	48
28	Thompson et al. (43)	126	63	63
29	Torrent et al. (44)	73	38	35
30	A. Varma (manuscript in preparation)	106	53	53
31	Zalla et al. (45)	58	38	20
	Grand total	2876	1267	1609

*Data set reduced from that published to exclude participants already included in Simonsen et al. (25, 47).

†Data set reduced from that published to exclude participants already included in Simonsen et al. (25).

Results

Table 3 shows the demographic profile of the patient and control groups. Overall, the groups were well matched for gender ($\chi^2_1 = 0.71$, $P = 0.40$) but showed a significant difference in age ($t_{2866} = 5.51$, $P < 0.001$, $d = 0.21$; 95% CI, 1.57–3.30) with bipolar patients being, on average, 2.4 years older. The bipolar group also had, on average, 0.6 fewer years of education ($t_{2714} = 5.14$, $P < 0.001$, $d = 0.20$; 95% CI, −0.88 to −0.40) and showed a difference in premorbid IQ on the two IQ measures with substantial sample sizes: National Adult Reading Test (NART) (57)/WAIS-R (21) ($t_{985} = 3.87$, $P < 0.001$, $d = 0.25$; 95% CI, −3.86 to −1.26) and Wechsler Abbreviated Scale of Intelligence (WASI) (58) ($t_{959} = 6.99$, $P < 0.001$, $d = 0.48$; 95% CI, −6.61 to −3.71). The groups did not differ on IQ for those studies that used the WAIS Vocabulary Subtest (18) ($t_{179} = 1.2$, $P = 0.23$; 95% CI, −6.65 to 1.66) or WAIS Information Subtest (21): $t_{54} = 1.7$, $P = 0.10$; 95% CI, −4.41 to 0.41). One study used the Wechsler Memory Scale (WMS-R) (59) as an IQ measure which showed a group difference ($t_{98} = 2.31$, $P = 0.02$, $d = 0.46$; 95% CI, −14.1 to −1.1) but as this is a memory measure and not a measure of premorbid IQ, this difference is not surprising. It should be noted that the last three measures were only used in relatively small sample subsets. Overall, the data set showed significant group differences in a range of confounding variables reinforcing the need to covary for these factors in any combined analysis. This can only be done convincingly using IPDMA.

Group effect size of cognitive deficits

The patient group had large reductions in performance on all 11 outcome variables relative to controls when controlling for the effect of age, IQ and gender. The overall effect size for group varied between 0.63 on TMTB to 0.26 on WCSTCats. (Table 4). The sample sizes (n in Table 4) were substantially larger than for the meta-analyses in Table 1a,c,d and comparable to or larger than Table 1b. Figures 1–3 show forest plots for the meta-analysis of the confound-adjusted group effect sizes associated with VLT, TMT and WCST neuropsychological tests.

The studies showed a wide range of between-study heterogeneity across the 11 outcome measures, ranging from 0% to 84% (Table 4). The I^2 measure of heterogeneity provides an indication of the proportion of total variation in effect size estimates attributable to between-study heterogeneity. I^2 values of 8% for TMTA, 12% for WCSTCats.

Table 3. Demographics of patient and control groups

$N = 2876$	Patients ($N = 1267$) M(SD)	Controls ($N = 1609$) M(SD)
Age ($n = 2868$)	38.8 (11.7)	36.4 (11.8)
Male	569	748
Female	698	861
Years of education ($n = 2716$)	12.9 (3.4)	13.6 (3.0)
IQ measures		
NART/WAIS ($n = 1103$)	112.4 (11.6)	114.7 (10.6)
WASI ($n = 961$)	107.5 (10.5)	112.6 (10.8)
WMS-R ($n = 100$)	97.4 (17.8)	105.0 (14.8)
WAIS Vocab. Subtest ($n = 181$)	44.6 (11.5)	47.1 (12.0)
WAIS Info. Subtest ($n = 56$)	19.4 (5.0)	21.4 (3.9)

NART, National Adult Reading Test; WAIS, Wechsler Adult Intelligence Scale; WASI, Wechsler Abbreviated Scale of Intelligence; WMS-R, Wechsler Memory Scale.

and 15% for VLT Recog-FP can be considered minor; values of 39% (VLT ShortDelay), 42% (VLTLlongDelay FDS) and 45% (WCSTPersev.) can be considered moderate; whilst between-study heterogeneity on VLT Total1–5 (Fig. 1), TMTB (Fig. 2) and both FDS and RDS with $I^2 = 61\%$, 69%, 71% and 84%, respectively, was substantial (60). Magnitude of effect size was associated with increased heterogeneity.

Residual mood effects

The meta-regression showed that the factor relating to a study's ability to minimize residual mood within the patient group significantly explained the between-study heterogeneity for two of the 11 outcome variables: TMTA regression coefficient = −0.05 ($t = 2.18$, $P = 0.047$, Adj.$R^2 = 100\%$, 95% CI, −0.10 to −0.001) and WCSTCats. regression coefficient = −0.07 ($t = 2.78$, $P = 0.018$, Adj.$R^2 = 100\%$, 95% CI, −0.12 to −0.02). None of the other nine outcome variables were associated with significant meta-regression coefficients (VLT Total1–5: $t = 0.70$, $P = 0.50$, 95% CI, −0.08 to 0.04; VLT ShortDelay: $t = 1.10$, $P = 0.29$, 95% CI, −0.07 to 0.02; VLT LongDelay: $t = 0.82$, $P = 0.42$, 95% CI, −0.07 to 0.03; VLT Recognition: $t = 0.18$, $P = 0.86$, 95% CI, −0.04 to 0.05; VLT Recog-FP: $t = 1.43$, $P = 0.19$, 95% CI, −0.08 to 0.02; TMTB: $t = 0.72$, $P = 0.48$, 95% CI, −0.12 to 0.06; FDS: $t = 0.07$, $P = 0.94$, 95% CI, −0.09 to 0.08; RDS: $t = 0.20$, $P = 0.84$, 95% CI, −0.89 to 0.11; WCSTPersev.: $t = 1.56$, $P = 0.15$, 95% CI, −0.02 to 0.13).

The second approach to understand residual mood effects was to consider the effect of depression score and mania score on neuropsychological performance within the patient group only. Depression score showed an overall main effect on

Table 4. Overall effect size of group for the 11 outcome variables

Test	Outcome variable	N (bipolar)	N (control)	Overall effect size (95% CI)	P	I^2 (%)
VLT	Total1–5	624	661	0.51 (0.42–0.60)	<0.001	61
VLT	Short Delay	667	680	0.48 (0.39–0.57)	<0.001	39
VLT	Long Delay	667	680	0.55 (0.47–0.64)	<0.001	42
VLT	Recognition	576	590	0.46 (0.36–0.57)	<0.001	0
VLT	Recog-FP	333	404	0.38 (0.26–0.50)	<0.001	15
TMT	A	879	752	−0.49 (−0.58 to −0.40)	<0.001	8
TMT	B	903	778	−0.63 (−0.72 to −0.55)	<0.001	69
Digit Span	Forward	533	650	0.30 (0.20–0.40)	<0.001	71
Digit Span	Reverse	533	650	0.60 (0.51–0.69)	<0.001	84
WCST	Categories	605	639	0.26 (0.15–0.37)	<0.001	12
WCST	Perseverations	606	639	−0.29 (−0.40 to −0.17)	<0.001	45

Recog-FP, recognition minus false positives; TMT, Trail Making Test; VLT, Verbal Learning Task; WCST, Wisconsin Card Sorting Task.

just three of 11 outcome measures (when accounting for the effect of mania, age, IQ and gender), typically on measures of memory, speed and executive function: VLT Total1–5 effect size = −0.09, t_{652} = 2.68, P = 0.008, 95% CI, −0.16 to −0.03; VLT Recognition effect size = −0.13, t_{605} = 3.32, P = 0.001, 95% CI, −0.02 to −0.05; and TMTA effect size = 0.09, t_{682} = 2.62, P = 0.009, 95% CI, 0.02–0.16. Higher depression scores were related to worse cognitive performance but the effect size was considerably smaller than the relevant effect size for group (see Table 4). There was no overall main effect of mania score on any of the 11 outcome measures (when accounting for the effect of depression, age, IQ and gender).

Drug effects within patient group

Within the patient sample, there was full information on drug treatment for 952 patients (75%) and information on lithium status for 1122 (89%). Thus, for comparative analysis, 652 patients were on lithium with 470 lithium free, 337 were on anticonvulsants with 409 anticonvulsant free, 209 were on antidepressants with 537 antidepressant free, 209 were on antipsychotics with 537 antipsychotic free and 72 were drug free compared to 880 on at least one drug type. The mixed model regression analysis within the patient group suggested that neither lithium (given effects of study, age, IQ and gender) nor antidepressants (given effects of study, age, IQ and gender) affected performance on any of the 11 outcome measures (Ps > 0.1 for all effect sizes of lithium or antidepressant status). Similarly, anticonvulsants showed no effect on performance (given effects of study, age, IQ and gender) on any of the 11 outcome measures (Ps > 0.1 for all effect sizes of anticonvulsants except for WCST Cats. with P = 0.08). Antipsychotics (given effects of study, age, IQ and gender) showed a reduced performance on VLT Total1–5 only (effect

size = −0.29, P = 0.006, 95% CI, −0.49 to −0.08) of the 11 outcome measures (Ps > 0.1 for all other effect sizes of antipsychotic status except for VLT ShortDelay and VLT LongDelay both with P = 0.08 and WCSTPersev. with P = 0.09). Being drug free improved performance (given effects of study, age, IQ and gender) relative to any drug on two of the 11 outcome measures: VLT Total1–5 (effect size = −0.39, P = 0.010, 95% CI, −0.69 to −0.09) and VLT LongDelay (effect size = −0.35, P = 0.017, 95% CI, −0.64 to −0.06; Ps > 0.1 for all other effect sizes of drug-free status).

Relationship between illness variables and cognitive deficits

Table 5 shows the illness characteristics of the patient sample. The mixed model regression analysis within the patient group suggested that some of these illness variables correlated at better than chance with some of the 11 outcome variables (eight out of 66) but effects were generally small. Thus, number of manic episodes affected performance on three of the outcome measures (given effects of study, age, IQ and gender): VLT Short-Delay (effect size = −0.07, P = 0.03, 95% CI, −0.14 to −0.01); VLT LongDelay (effect size = −0.09, P = 0.007, 95% CI, −0.16 to −0.03); and TMTA (effect size = 0.09, P = 0.03, 95% CI, 0.01–0.17). Number of total episodes only affected performance on TMTA (effect size = 0.08, P = 0.03, 95% CI, 0.01–0.15). Number of depressive episodes had no main effects. Number of depressive hospitalizations also only affected performance on TMTA (effect size = 0.26, P = 0.003, 95% CI, 0.09–0.42) whilst number of total hospitalizations affected performance on TMTA (effect size = 0.12, P = 0.008, 95% CI, 0.03–0.21), TMTB (effect size = 0.13, P = 0.005, 95% CI, 0.04–0.21) and WCSTCats. (effect size = −0.12, P = 0.01, 95% CI, −0.21 to −0.03). Number of manic hospitalizations had no main effects. Thus, of the four
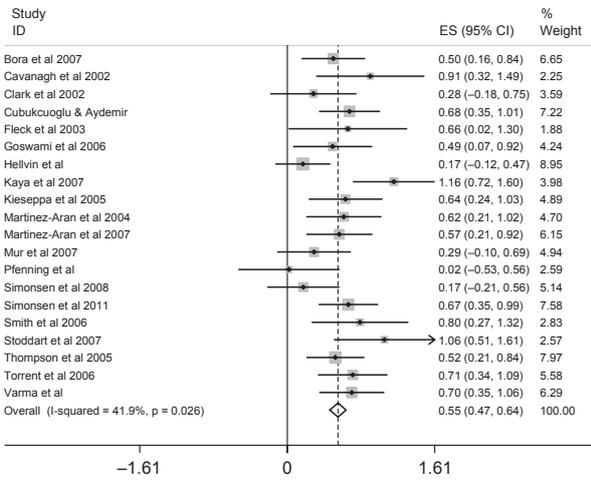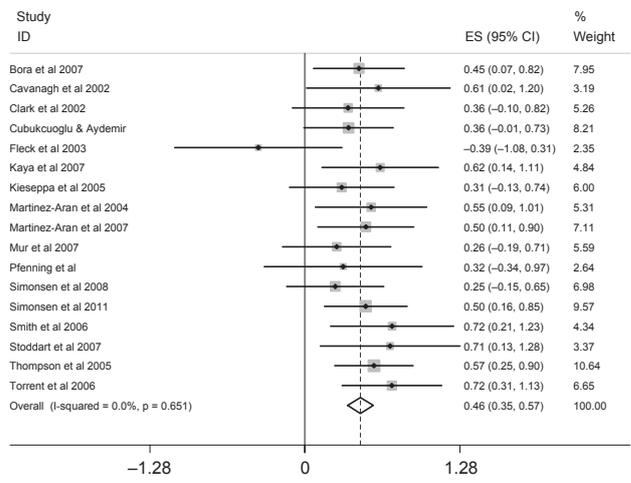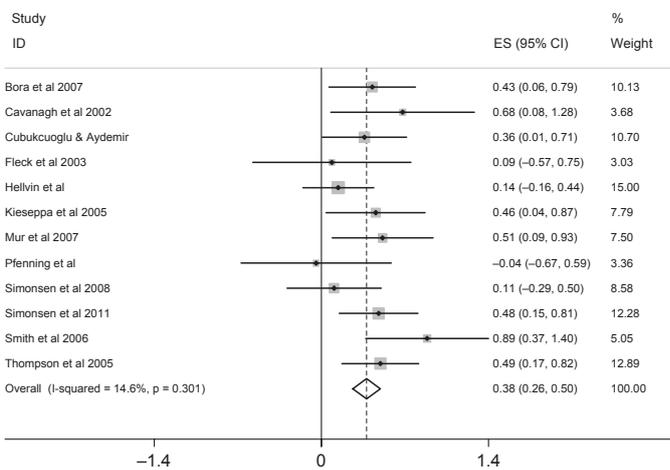
Bourne et al.
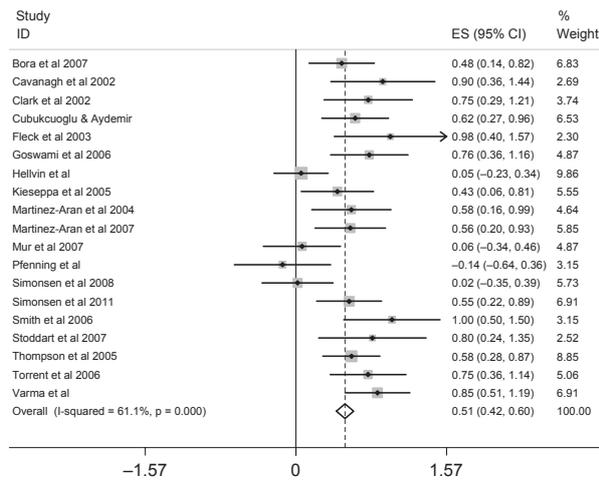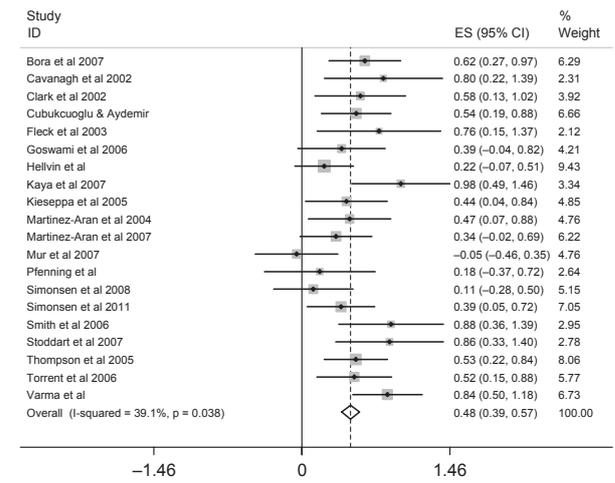
Fig. 1. Forest plots showing the main effect of group (accounting for effect of age, IQ and gender) for the five outcome variables associated with Verbal Learning Task (VLT).
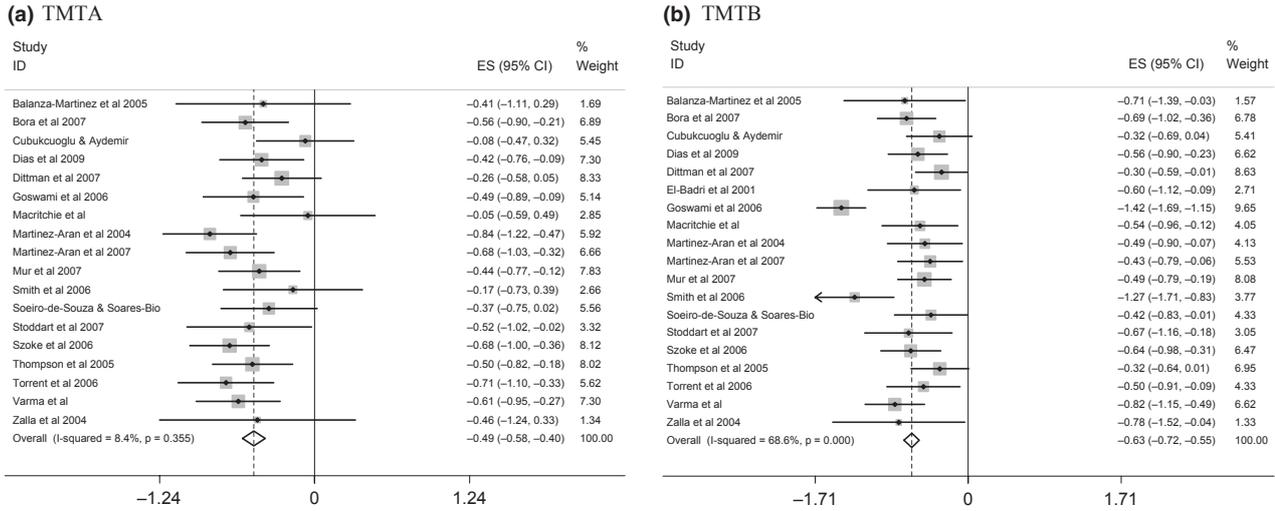
(a) TMTA

(b) TMTB

Fig. 2. Forest plots showing the main effect of group (accounting for effect of age, IQ and gender) for the two outcome variables associated with Trail Making Test (TMTA and TMTB).
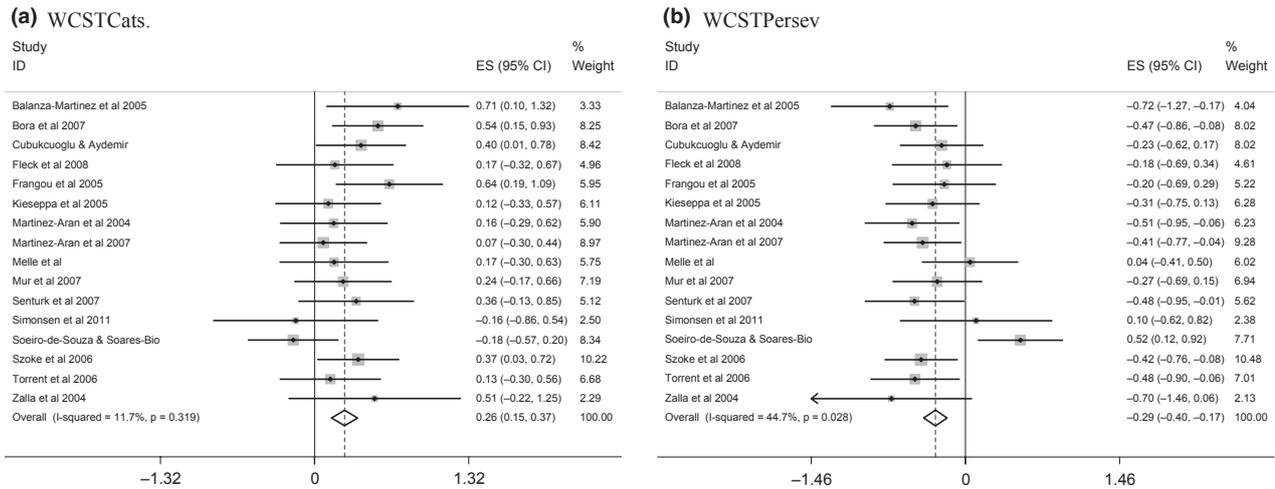
(a) WCSTCats.

(b) WCSTPersev

Fig. 3. Forest plots showing the main effect of group (accounting for effect of age, IQ and gender) for the two outcome variables associated with Wisconsin Card Sorting Task (WCSTCats. and WCSTPersev.).

illness variables that affected cognitive performance, TMTA was affected by all four.

Discussion

This analysis of individual patient data across the 31 studies provides further evidence that euthymic bipolar patients exhibit moderate cognitive impairments on a range of standard neuropsychological tests. Cognitive deficits remain significant even after controlling for key baseline factors such as age, IQ and gender that are known to affect neuropsychological test performance. The current level of minor depressive symptoms and the effects of some drug treatments may contribute to these effects but cannot explain them. Thus, there is significant residual cognitive impairment associated

Table 5. Clinical indices of the patient group

	Patients M (SD)	Range
Age at onset (n = 1129)	25.0 (8.7)	6–60
Illness duration (n = 1104)	13.8 (9.9)	0–51
No. of depressive episodes (n = 992)	5.6 (10.7)	0–100
No. of manic episodes (n = 989)	3.4 (4.5)	0–88
Total no. of episodes (n = 1115)	11.6 (19.8)	0–200
No. of depressive hospitals (n = 271)	0.6 (1.4)	0–10
No. of manic hospitals (n = 271)	1.4 (2.3)	0–15
Total no. of hospitalisations (n = 806)	2.9 (3.8)	0–40

with bipolar disorder over and above the known confounding factors.

The effect sizes for such deficits were lower (0.26–0.63) than those reported in prior meta-anal-

yses (6–8, 10) (ds = 0.5–1.0). This reduction in observed effect sizes is in part due to controlling better for the effect of age, IQ and gender. However, we were also able to include unpublished studies which often had the lowest effect sizes [e.g. Hellvin et al. (50) and A. Pfennig, M. Alda, T. Young, et al. (manuscript in preparation) for VLT Total1–5, LongDelay and Recog-FP; Cubukcuoglu & Aydemir (49) and A. Macritche (manuscript in preparation) for TMTA and TMTB; A. Varma (manuscript in preparation) for FDS and RDS; M.G. Soeiro-de-Souza & D. Soares-Bio (manuscript in preparation) for WCSTCats.; and Hellvin et al. (50) and M.G. Soeiro-de-Souza & D. Soares-Bio (manuscript in preparation) for WCSTPersev.]. This suggests the field has had some impact from publication bias, which perhaps is unsurprising.

Specifically, the following effect sizes were found (compared to prior studies) in the following cognitive domains: i) verbal memory – Total Score effect size = 0.51 (prior studies = 0.90–0.81), Short Delay effect size = 0.48 (prior studies = 0.85–0.73), Long Delay effect size = 0.55 (prior studies = 0.85–0.71), Recognition effect size = 0.46 (prior study = 0.43), Recog-FP effect size = 0.38; ii) visual scanning speed – TMTA effect size = 0.49 (prior studies = 0.82–0.60); iii) working memory capacity – FDS effect size = 0.30 (prior studies = 0.47–0.37); iv) executive function – TMTB effect size = 0.63 (prior studies = 0.99–0.55), RDS effect size = 0.60 (prior studies = 1.02–0.54), WCSTCats. effect size = 0.26 (prior studies = 0.69–0.52); v) response inhibition = WCST-Persev. = 0.29 (prior studies = 0.88–0.70).

The high heterogeneity of some tests appears to underlie the differences in the results of prior meta-analyses. The variation in effect sizes between the previously published meta-analysis (Table 1) is likely to have been due to variations in the studies included. In turn, the range of effect sizes produced by including a different subset of studies can be directly explained by the relatively high level of heterogeneity revealed in this sample by our analysis (typically 39–84%; see Table 4) especially for some tests. The test with the most heterogeneity in this analysis was TMTB. TMTB is known to have considerable variability across test sites (61), thus there appears to be a strong case for trying to refine the operationalization of TMTB as well as VLT (encoding and short term recall) and Digit Span (Forward and Reverse). Each test taps domains of function markedly impaired in bipolar patients as shown by the large average effect sizes. One important possibility would be to present them in more standardized computerized formats locally or even on line.

Nevertheless, the group effect sizes allow confidence that a substantial average effect is present for the domains of attention/working memory, verbal memory, speed and executive function. It is somewhat easier to say what cannot explain these effects, than to say what can. Residual mood symptoms within the patient group were understandably confounded with group. However, our analysis suggests that residual symptom scores in the patient group cannot explain much of the difference found between the groups across the various tests. Cognitive deficits are also not simply explained as side-effects of drug therapy. This has previously been the subject of debate; some studies suggesting that antipsychotic drugs may cause some cognitive impairment (62, 63) and others suggesting no drug effect on cognitive performance (64). The present analysis suggests that most neuropsychological tests do not exhibit any significant effect attributable to drug treatment. The only possible exception is on measures of verbal memory with antipsychotics having an impairing effect on VLT Total1–5 and drug-free status being associated with improved performance on VLT Total1–5 and LongDelay (relative to any drug). However, any potential implied drug effects must be treated with caution due to the potential for confounding by indication. For example, a history of psychosis may be related to specific working memory impairments (65–67), and those with a history of psychosis are also likely to be those currently taking antipsychotics (68). We could not analyse the effect of polypharmacy, which is common in clinical samples, but not in these research samples. It is likely that there was a deliberate effort to exclude symptomatic and heavily medicated patients from these studies given the intention was usually to reduce the confounds between the patient and control groups.

If illness course had had a negative impact on cognition, it would potentially be a key finding; it could imply that neuropsychological outcome measures are sensitive to treatment. In a partial support of this hypothesis, some of the neuropsychological measures correlated with illness intensity variables, for example number of manic episodes appears to affect performance on certain VLT measures, whilst TMTA appears to be especially sensitive to potential illness progression effects. However, the magnitude of these associations may be unreliable for various reasons. First, the impact of illness may not be simply cumulative, and the largest effects may occur early in the illness course, as appears likely in schizophrenia (69).

Second, measures of illness severity that depend on counting episodes in mature samples of patients are of uncertain validity. Quantifying depressive episodes when so much of the depressive burden of bipolar disorder is chronic, subsyndromal and poorly recalled is questionable; indeed, we found no associations with number of depressive episodes. Positive findings for more memorable events, like manic episodes and numbers of hospitalizations, appear more likely to be valid and did produce some significant results in this analysis. The hypothesis that much of the apparent cognitive impairment of bipolar disorder is attributable to the accumulated impact of the illness course remains plausible but not proven by the present study. Only adequately powered prospective studies in early stages of illness will establish the effect beyond doubt.

Although the range in effect sizes reported here appears to support previous suggestions that executive function and memory may be especially affected in bipolar disorder (6, 8, 9), it is also notable that all of the effect sizes reported here could be considered to be small to medium (70) in magnitude across all the cognitive domains investigated. Our results could therefore also be interpreted as being consistent with the notion of cognitive impairment in bipolar disorder being a relatively non-specific effect on multiple functional brain networks. This can be related to similarly non-specific imaging findings suggesting lateral ventricle enlargement (effect size = 0.39) and increased rates of deep white matter hyperintensities without grey matter volume decrements (71) in the many imaging studies conducted in bipolar patients. Although these structural abnormalities can be greater in older patients they are also found in samples of similar mean age as the sample in this study (71). The evolving evidence for widely distributed disturbances in white matter structure from diffusion tensor imaging is also supportive of an underlying functional neuropathology (72). Although its aetiology remains poorly understood, a contribution from intracellular mechanisms regulating oxidative stress is one hypothesis that is assuming increasing importance (73). Given the putative neuroprotective effects of lithium (74, 75), an improved cognitive performance for those patients taking lithium relative to those lithium free might have been expected. However, no such effect was found; either because lithium does not enhance cognitive performance or because any neuroprotective effect is dependent upon factors, such as chronic use, which could not be estimated in this dataset. In support of the former 'ineffective hypothesis', two recent longitudinal cohort studies

indicate that deficits are stable despite long-term lithium therapy (76, 77).

As with all analyses of neuropsychological performance, this study's findings and conclusions are limited by the reliability, validity and psychometric properties of the individual neuropsychological tests. The high levels of heterogeneity found in this study and the previous standard meta-analyses (6–9, 12, 13) for some measures highlight the need for standardization in test presentation to try and meet this limitation. Indeed, the high levels of heterogeneity consistently found for some measures raises the question as to whether it is meaningful to combine them in a meta-analysis at all. This study is also limited by the response bias of authors allowing access to their primary data sets. Furthermore, it is acknowledged that this study considered outcome measures from a relatively small number of neuropsychological tests. However, despite being limited to those primary studies that consented to provide data, and partly because the analysis was limited to the most frequently used neuropsychological tests, this study contained sample sizes substantially greater than many of the prior standard meta-analyses and thus represents a major data synthesis. Furthermore, by using IPDMA (rather than standard meta-analysis) this study was both able to i) provide the least confounded estimates of the effect size relating to cognitive impairment in euthymic bipolar patients and ii) provide the first analysis of potential medication and illness severity effects on neuropsychological performance in a statistically valuable sample size.

In summary, this reanalysis provides further evidence that euthymic bipolar patients exhibit significant cognitive impairment on a range of neuropsychological tests. These impairments remain substantial but less than previous work (including previous meta-analyses) has suggested (1–4, 6–10). The advantage of IPDMA in controlling for a greater range of confounding factors and the inclusion of unpublished studies accounts for this. The impairment effect appears largely independent of drug treatment. Performance on some neuropsychological tests appears to have deteriorated further as illness progressed (i.e. number of episodes increased) but longitudinal data from earlier in the illness course are needed to show that the relationship is causal and clinically important. Finally, this review and reanalysis has highlighted the variability and heterogeneity between individual primary studies. This means the field remains polarized between the certainty that cognitive impairment is a feature of bipolar disorder and uncertainty, for example about its heritability, specificity or the impact of illness

intensity. Specific and correct findings on the latter may be reasonably based on studies that are well conducted but too small for confidence and too subtle to be replicated in cohorts of convenience. On the other hand, small studies can always generate false positives findings, and this is too often forgotten in the field (78). The present result, from a study sample larger than the samples reported in three of the previous meta-analyses of published data sets, may well be giving us the true picture. A clear goal for future research is operationally to refine all test procedures and variables being measured to reduce heterogeneity and combine data prospectively across centres to obtain the necessary power essential to statistical confidence.

Declaration of interest

Drs. Bora, Bourne, Craddock, Cubukcuoglu, Dittmann, Fleck, Gallagher, Geddes, Jones, Kieseppä, Leboyer, Martínez-Aran, Melle, Moore, Mur, Raust, Rogers, Senturk, Simonsen, Soares-Bio, Smith, Soeiro-de-Souza, Sundet, Szöke, Thompson, Torrent, Tzagarakis, Worhunsky and Zalla declare that they have no conflicts of interest over the past 2 years. Dr. Andreassen has received speakers's honorarium from Lilly, Lundbeck and GSK. Dr. Clark is a consultant for Cambridge Cognition Ltd. Dr. Aydemir has participated in a clinical trial sponsored by AstraZeneca, received speaker honoraria from Lundbeck, AstraZeneca, Janssen-Cilag and Pfizer and consultant for Servier. Dr Balanzá-Martínez has received grants and served as consultant, advisor or CME speaker from Angelini, AstraZeneca, Bristol-Myers-Squibb, Grunenthal, Janssen, Juste, the Spanish Ministry of Science and Innovation (CIBERSAM) and 'Fundación Alicia Koplowitz'. Dr. Bauer has received grant/research support from The Stanley Medical Research Institute, NARSAD, Deutsche Forschungsgemeinschaft and the European Commission (FP7). He is a consultant for Alkermes, AstraZeneca, BristolMyers Squibb, Ferrer Internacional, Janssen, Lilly, Lundbeck, Otsuka, Servier, Takeda. Dr. Bauer has received speaker honoraria from AstraZeneca, BristolMyers Squibb, GlaxoSmithKline, Lilly, Lundbeck, Otsuka. Pfizer. Dr. Brissos has been working full time as Medical Affairs Manager for Janssen Pharmaceutical. Dr. Cavanagh has received investigator-originated research grant funding from Pfizer and Biogen IDEC. Dr. Dias is consultant for Angelini Pharmaceutical, Portugal and has received educational grants from Lundbeck, Sanofi-Aventis, AstraZeneca and Bristol-Myers Squibb. Dr. Ferrier has received speaker honoraria for lectures given at educational meetings sponsored by Astra Zeneca and Organon. Dr. Frangou has participated in advisory boards for Janssen-Cilag and Ferrer Grupo and has been a speaker for Janssen-Cilag. Dr. Goodwin has received grants/research support, consulting fees and honoraria from AstraZeneca, Bristol-Myers Squibb, Eisai, Eli Lilly, Lundbeck, P1Vital, Servier, Takeda and Teva. Dr. Pfennig has received research support and speaker honoraria from AstraZeneca. Dr. Stoddart currently works for a consultancy firm that has pharmaceutical companies among its clients. Dr. Vieta has received grants and served as consultant, advisor or CME speaker for the following entities: Adamed, Alexza, Almirall, AstraZeneca, Bial, Bristol-Myers Squibb, Elan, Eli Lilly, Ferrer, Forest Research Institute, Gedeon Richter, Glaxo-Smith-Kline, Janssen-Cilag, Jazz, Johnson & Johnson, Lundbeck, Merck, Novartis, Organon, Otsuka, Pfizer, Pierre-Fabre, Qualigen, Roche, Sanofi-Aventis, Servier, Shering-Plough, Shire, Solvay, Sunovion, Takeda, Teva, the Spanish Ministry of Science and Innovation (CIBERSAM), the Seventh European Framework Programme (ENBREC), the Stanley Medical Research Institute, United Biosource Corporation and Wyeth.

References

1. BEARDEN CE, HOFFMAN KM, CANNON TD. The neuropsychology and neuroanatomy of bipolar affective disorder: A critical review. Bipolar Disord 2001;**3**:106–150.
2. QURAISHI S, FRANGOU S. Neurospychology of bipolar disorder. J Affect Disord 2002;**72**:209–226.
3. SAVITZ J, SOLMS M, RAMESAR RS. Neuropsychological dysfunction in bipolar affective disorder: a critical opinion. Bipolar Disord 2005;**7**:216–235.
4. MARTINEZ-ARAN A, VIETA E, REINARES M et al. Cognitive function across manic or hypomanic, depressed, and euthymic states in bipolar disorder. Am J Psychiatry 2004;**161**:262–270.
5. GOODWIN GM, ANDERSON I, ARANGO C et al. ECNP consensus meeting. Bipolar depression. Nice, March 2007. Eur Neuropsychopharmacol 2008;**18**:535–549.
6. ARTS B, JABBEN N, KRABBENDAM L, van OS J. Meta-analyses of cognitive functioning in euthymic bipolar patients and their first-degree relatives. Psychol Med 2007;**38**:771–785.
7. BORA E, YUCEL M, PANTELIS C. Cognitive endophenotypes of bipolar disorder: a meta-analysis of neuropsychological deficits in euthymic patients and their first-degree relatives. J Affect Disord 2009;**113**:1–20.
8. ROBINSON LJ, THOMPSON JM, GALLAGHER P et al. A meta-analysis of cognitive deficits in euthymic patients with bipolar disorder. J Affect Disord 2006;**93**:105–115.
9. TORRES IJ, BOUDREAU VG, YATHAM LN. Neuropsychological functioning in euthymic bipolar disorder: a meta-analysis. Acta Psychiatr Scand 2007;**116**:17–26.
10. ROBINSON LJ, FERRIER IN. Evolution of cognitive impairment in bipolar disorder: a systematic review of cross-sectional evidence. Bipolar Disord 2006;**8**:103–116.
11. KRABBENDAM L, ARTS B, van OS J, ALEMAN A. Cognitive functioning in patients with schizophrenia and bipolar disorder: a quantitative review. Schizophr Res 2005;**80**: 137–149.
12. MANN-WROBEL MC, CARRENO JT, DICKINSON D. Meta-analysis of neuropsychological functioning in euthymic bipolar disorder: an update and investigation of moderator variables. Bipolar Disord 2011;**13**:334–342.
13. KURTZ MM, GERRATY RT. A meta-analytic investigation of neurocognitive deficits in bipolar illness: profile and effects of clinical state. Neuropsychology 2009;**5**:551–562.
14. The Psychiatric GWAS Consortium Steering Committee. A framework for interpreting genome-wide association studies of psychiatric disorders. Mol Psychiatry 2009;**14**: 10–17.
15. RIVAS MA, BEAUDOIN M, GARDET A et al. Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. Nat Genet 2011;**43**:1066–1075.

16. Belmonte Mahon P, Pirooznia M, Goes FS et al. Genome-wide association analysis of age at onset and psychotic symptoms in bipolar disorder. Am J Med Genet B Neuropsychiatr Genet 2011;**156**:370–378.

17. Clarke MJ. Individual patient data meta-analyses. Best Pract Res Clin Obstet Gynaecol 2005;**19**:47–55.

18. Delis DC, Kramer JH, Kaplan E, Ober BA. California verbal learning test: adult version. San Antonio, TX: The Psychological Corporation, 1987.

19. Rey A. L'examen clinique en psychologie. Paris: Presses Universitaires de France, 1964.

20. Reitan RM. Validity of trail making test as an indication of organic brain disease. Percept Mot Skills 1958;**8**: 271–276.

21. Wechsler D. Wechsler adult intelligence scale – revised: manual. New York: Psychological Corporation, 1981.

22. Heaton RK. Wisconsin card sorting test manual. Odessa, FL: Psychological Assessment Resources Inc., 1981.

23. Yatham LN, Torres IJ, Malhi GS et al. The International Society for Bipolar Disorders – battery for assessment of neurocognition (ISBD-BANC). Bipolar Disord 2010;**12**: 351–363.

24. Moher D, Liberati A, Tetzlaff J, Altman DG, Group TP. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. J Clin Epidemiol 2009;**62**:1006–1012.

25. Simonsen C, Sundet K, Vaskinn A et al. Neurocognitive profiles in bipolar I and bipolar II disorder: differences in pattern and magnitude of dysfunction. Bipolar Disord 2008;**10**:245–255.

26. Balanza-Martinez V, Tabares-Seisdedos R, Selva-Vera G et al. Persistent cognitive dysfunctions in bipolar I disorder and schizophrenic patients: a 3-year follow-up study. Psychother Psychosom 2005;**74**:113–119.

27. Bora E, Vahip S, Akdeniz F et al. The effect of previous psychotic mood episodes on cognitive impairment in euthymic bipolar patients. Bipolar Disord 2007;**9**:468–477.

28. Cavanagh JTO, van Beck M, Muir M, Blackwood DHR. Case-control study of neurocognitive function in euthymic patients with bipolar disorder: an association with mania. Br J Psychiatry 2002;**180**:320–326.

29. Clark L, Iversen SD, Goodwin GM. Sustained attention deficit in bipolar disorder. Br J Psychiatry 2002;**180**: 313–319.

30. Dittmann S, Seemuller F, Schwarz MJ et al. Association of cognitive deficits with elevated homocysteine levels in euthymic bipolar patients and its impact on psychosocial functioning: preliminary results. Bipolar Disord 2007;**9**:63–70.

31. El-Badri SM, Ashton CH, Moore PB, Marsh VR, Ferrier IN. Electrophysiological and cognitive function in young euthymic patients with bipolar affective disorder. Bipolar Disord 2001;**3**:79–87.

32. Fleck DE, Shear PK, Zimmerman ME et al. Verbal memory in mania: effects of clinical state and task requirements. Bipolar Disord 2003;**5**:375–380.

33. Fleck DE, Shear PK, Madore M, Strakowski SM. Wisconsin card sorting test performance in bipolar disorder: effects of mood state and early course. Bipolar Disord 2008;**10**:539–545.

34. Frangou S, Donaldson S, Hadjulis M, Landau S, Goldstein LH. The Maudsley bipolar disorder project: executive dysfunction in bipolar I disorder and its clinical correlates. Biol Psychiatry 2005;**58**:859–864.

35. Goswami U, Sharma A, Khastigir U et al. Neuropsychological dysfunction, soft neurological signs and social disability in euthymic patients with bipolar disorder. Br J Psychiatry 2006;**188**:366–373.

36. Kieseppa T, Tuulio-Henriksson A, Haukka J et al. Memory and verbal learning functions in twins with bipolar-I disorder and the role of information-processing speed. Psychol Med 2005;**35**:205–215.

37. Martinez-Aran A, Vieta E, Torrent C et al. Functional outcome in bipolar disorder: the role of clinical and cognitive factors. Bipolar Disord 2007;**9**:103–113.

38. Mur M, Portella MJ, Martinez-Aran A, Pifarre J, Vieta E. Persistant neuropsychological deficit in euthymic bipolar patients: executive function as a core deficit. J Clin Psychiatry 2007;**68**:1078–1086.

39. Senturk V, Goker C, Bilgic A et al. Impaired verbal memory and otherwise spared cognition in remitted bipolar patients on monotherapy with lithium or valproate. Bipolar Disord 2007;**9**:136–144.

40. Smith DJ, Muir WJ, Blackwood DHR. Neurocognitive impairment in euthymic young adults with bipolar spectrum disorder and recurrent major depressive disorder. Bipolar Disord 2006;**8**:40–46.

41. Stoddart SDR, Craddock NJ, Jones LA. Differentiation of executive function and attention impairments in affective illness. Psychol Med 2007;**37**:1613–1623.

42. Szoke A, Schurhoff F, Golmard J et al. Familial resemblance for executive functions in families of schizophrenic and bipolar patients. Psychiatry Res 2006;**144**: 131–138.

43. Thompson JM, Gallagher P, Hughes JH et al. Neurocognitive impairment in euthymic patients with bipolar affective disorder. Br J Psychiatry 2005;**186**:32–40.

44. Torrent C, Martinez-Aran A, Daban C et al. Cognitive impairment in bipolar II disorder. Br J Psychiatry 2006;**189**:254–259.

45. Zalla T, Joyce C, Szoke A et al. Executive dysfunctions as potential markers of familial vulnerability to bipolar disorder and schizophrenia. Psychiatry Res 2004;**121**: 207–217.

46. Dias VV, Brissos S, Frey BN, Andreazza AC, Cardoso C, Kapczinski F. Cognitive function and serum levels of brain-derived neurotrophic factor in patients with bipolar disorder. Bipolar Disord 2009;**11**:663–671.

47. Simonsen C, Sundet K, Vaskinn A et al. Neurocognitive dysfunction in bipolar and schizophrenia spectrum disorders depends on history of psychosis rather than diagnostic group. Schizophr Bull 2011;**37**:73–83.

48. Kaya E, Aydemir O, Selcuki D. Residual symptoms in bipolar disorder: the effect of the last episode after remission. Prog Neuropsychopharmacol Biol Psychiatry 2007;**31**:1387–1392.

49. Cubukcuoglu Z, Aydemir O. Specifity and sensitivity of neurocognitive tests in bipolar disorder: A comparative study. Unpublished thesis, 2011.

50. Hellvin T, Sundet K, Simonsen C et al. Neurocognitive functioning in patients recently diagnosed with bipolar disorder. Bipolar Disord 2012;**14**:227–238.

51. Hamilton M. A rating scale for depression. J Neurol 1960;**23**:55–62.

52. Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. Br J Psychiatry 1979;**134**:382–389.

53. Rush AJ, Carmody T, Reimitz PE. The inventory of depressive sympotmatology (IDS): clinician (IDS-C) and self-report (IDS-SR) ratings of depressive symptoms. Int J Methods Psychiatr Res 2000;**9**:45–59.

54. Young RC, Briggs JT, Ziegler VE, Meyer DA. A rating scale for mania: reliability, validity, and sensitivity. Br J Psychiatry 1978;**133**:429–435.

55. ALTMAN EG, HEDEKER DR, JANICAK PG, PETERSON JL, DAVIS JM. The clinician-administered rating scale for mania (CARS-M): development, reliability, and validity. Biol Psychiatry 1994;**36**:124–134.

56. BECH P, BOEWIG TG, DEIN E, JACOBSEN O, GRAM LF. Quantitative rating of manic states. Acta Psychiatr Scand 1975;**59**:1–6.

57. NELSON HE, WILLISON J. National adult reading test manual. 2nd ed. Windsor: NFER-Nelson, 1991.

58. WECHSLER D. Wechsler abbreviated scale of intelligence (WASI). San Antonio, TX: Harcourt Assessment, 1999.

59. WECHSLER D. Wechsler memory scale – revised. New York: Psychological Corporation, 1987.

60. HIGGINS JPT, THOMPSON SG. Quantifying heterogeneity in a meta-analysis. Stat Med 2002;**21**:1539–1558.

61. NUECHTERLEIN KH, GREEN MF, KERN RS et al. The MATRICS consensus cognitive battery, part 1: test selection, reliability, and validity. Am J Psychiatry 2008;**165**: 203–213.

62. DONALDSON S, GOLDSTEIN LH, LANDAU S, RAYMONT V, FRANGOU S. The Maudsley Bipolar Disorder Project: the effect of medication, family history, and duration of illness on IQ and memory in bipolar I disorder. J Clin Psychiatry 2003;**64**:86–93.

63. JAMROZINSKI K, GRUBER O, KEMMER C, FALKAI P, SCHERK H. Neurocognitive functions in euthymic bipolar patients. Acta Psychiatr Scand 2009;**119**:365–374.

64. GOSWAMI U, SHARMA A, VARMA A et al. The neurocognitive performance of drug-free and medicated euthymic patients do not differ. Acta Psychiatr Scand 2009;**120**: 456–463.

65. GLAHN DC, BEARDEN CE, BARGUIL M et al. The neurocognitive signature of psychotic bipolar disorder. Biol Psychiatry 2007;**62**:910–916.

66. GLAHN DC, THERMAN S, MANNINEN M et al. Spatial working memory as an endophenotype for schizophrenia. Biol Psychiatry 2003;**53**:624–626.

67. GLAHN DC, BEARDEN CE, CAKIRA S et al. Differential working memory impairment in bipolar disorder and schizophrenia: effects of lifetime history of psychosis. Bipolar Disord 2006;**8**:117–123.

68. TORRENT C, MARTINEZ-ARÁN A, DABAN C et al. Effects of atypical antipsychotics on neurocognition in euthymic bipolar patients. Compr Psychiatry 2011;**52**:203–213.

69. MOHAMED S, PAULSEN JS, O'LEARY D, ARNDT S, ANDREASEN N. Generalized cognitive deficits in schizophrenia: a study of first-episode patients. Arch Gen Psychiatry 1999;**56**: 749–754.

70. COHEN JW. Statistical power analysis for the behavioural sciences. 2nd ed. Hillsdale NJ: Lawrence Erlbaum Associates, 1988.

71. KEMPTON MJ, GEDDES J, ETTINGER U, WILLIAMS SCR, GRASBY PM. Meta-analysis, database, and meta-regression of 98 structural imaging studies in bipolar disorder. Arch Gen Psychiatry 2008;**65**:1017–1032.

72. VEDERINE FE, WESSA M, LEBOYER M, HOUENOU J. A meta-analysis of whole-brain diffusion tensor imaging studies in bipolar disorder. Prog Neuropsychopharmacol Biol Psychiatry 2011;**35**:1820–1826.

73. ANDREAZZA AC, KAUER-SANT'ANNA M, FREY BN et al. Oxidative stress markers in bipolar disorder: a meta-analysis. J Affect Disord 2008;**111**:135–144.

74. SHAO L, YOUNG LT, WANG J. Chronic treatment with mood stabilizers lithium and valproate prevents excitotoxicity by inhibiting oxidative stress in rat cerebral cortical cells. Biol Psychiatry 2005;**58**:879–884.

75. TAN H, YOUNG LT, SHAO L, CHE Y, HONER WG, WANG J. Mood stabilizer lithium inhibits amphetamine-increased 4-hydroxynonenal-protein adducts in rat frontal cortex. Int J Neuropsychopharmacol 2012;**15**:1275–1285.

76. MORA E, PORTELLA MJ, FORCADA I, VIETA E, MUR M. Persistence of cognitive impairment and its negative impact on psychosocial functioning in lithium-treated, euthymic bipolar patients: a 6-year follow-up study. Psychol Med 2013;**43**:1187–1196.

77. TORRENT C, MARTINEZ-ARÁN A, DEL MAR BONNIN C et al. Long-term outcome of cognitive impairment in bipolar disorder. J Clin Psychiatry 2012;**73**:e899–e905.

78. IOANNIDIS JPA. Why most published research findings are false. PLoS Med 2005;**2**:e124.

Supporting information

Additional Supporting information may be found in the online version of this article:

Table S1. List of 45 identified eligible studies.