SUPPLEMENTARY MATERIAL

Meta-analyses of cognitive functioning in euthymic bipolar patients and their first-degree relatives

B. Arts[1]*, N. Jabben[1], L. Krabbendam[1] and J. van Os[1,2]

[1] *Department of Psychiatry and Neuropsychology, South Limburg Mental Health Research and Teaching Network, EURON, Maastricht University, PO Box 616 (KAP2), 6200 MD Maastricht, The Netherlands*
[2] *Division of Psychological Medicine, Institute of Psychiatry, London, UK*

Results

Bipolar patients

Twenty-eight studies were included in the meta-analysis (Table 1; the tables appear at end of text). Four of these stratified their samples by a third variable (van Gorp *et al.* 1998; Ferrier *et al.* 1999; Nehra *et al.* 2006; Torrent *et al.* 2006). For reasons of homogeneity, in the case of stratification, only one study group was included, with bias to the less severe patients or those with a better established diagnosis. Thus, Ferrier *et al.* (1999) stratified by outcome, contrasting a good outcome *versus* a poor outcome group; for the purpose of the current meta-analysis only the good outcome group was included. The study by Nehra *et al.* (2006) used first- and multiple-episode patients, only established bipolar patients with multiple episodes were included in the current analysis. Van Gorp *et al.* (1998) included patients with and without prior alcohol dependence, only the group without alcohol dependence was used in the analysis. Finally, Torrent *et al.* (2006) used bipolar I and bipolar II patients, only bipolar I patients were included.

Neuropsychological domains

The neuropsychological tests used in these studies were divided into 11 categories measuring approximately the same cognitive construct (adapted from Krabbendam *et al.* 2005). A neuropsychological test was included by the *a priori* criterion of having been used in at least four different studies. Immediate verbal memory was assessed using word list learning [California Verbal Learning Test (CVLT; Delis, 1987); Rey Auditory Verbal Learning Test (RAVLT; Rey, 1964); Auditory Verbal Learning Test (AVLT; Brand & Jolles, 1985)]. For the purposes of the analysis, results of these comparable tests were included together. Delayed verbal memory was assessed using the delayed recall version of the CVLT, RAVLT, and AVLT. Delayed visual memory was measured using the delayed recall version of the Rey Osterrieth Complex Figure (Rey, 1941). Working memory was assessed using the Digit Span (Wechsler, 1955). Verbal fluency was measured using either words from a certain category or beginning with a certain letter (FAS; Benton, 1978). Concept formation and shifting was assessed with the Wisconsin Card Sorting Test (WCST; Heaton, 1981); number of perseverative errors and categories achieved were separately analysed. Executive control was measured using the Stroop Color-Word interference (Stroop, 1935) and Trailmaking Test part B (Reitan, 1958). Sustained attention was assessed using a variant of the Continuous Performance Test (Kurtz, 2001). The test parameter used was number and/or percentage correct response. Mental Speed was measured using the Digit Symbol Substitution Test (DSST; Wechsler, 1955) and the Trailmaking Test part A (Reitan, 1958). Visuoperception was assessed using the copy version of the Rey Osterrieth Complex Figure (Rey, 1941). Intelligence was measured using the full-scale NART (Grober, 1991) or the WAIS-R vocabulary score (Wechsler, 1981), both good estimates of premorbid intelligence.

Meta-analytical results: patients

All effect sizes were in the same direction (Table 2), suggesting worse performance in euthymic bipolar patients compared to healthy controls.

In all instances, with the exception of visuoconstruction (Rey copy) and intelligence, bipolar patients displayed significantly poorer performance compared to controls. The largest effect sizes were evident for working memory (Digit Span backward), delayed and

* Author for correspondence (b.arts@np.unimaas.nl).

immediate verbal recall (CVLT) and mental speed (DSST) (effect sizes >0.8). Medium effect sizes ($0.5<d<0.8$) were observed for fluency (categories), executive control (Trail B, Stroop), concept shifting (WCST perseverative errors), delayed visual memory (Rey figure), sustained attention (CPT) and mental speed (Trail A). A small effect size ($0.2<d<0.5$) was noted for concept shifting (WCST categories), fluency (FAS) and working memory (Digit Span forward).

For five out of 17 analyses there was evidence for significant heterogeneity between the results of the different studies. The largest heterogeneity was found for working memory (Digit Span backward), executive control (Trail B) and concept shifting (WCST categories). Two studies were largely responsible for this heterogeneity, namely the studies of Balanza-Martinez *et al*. (2005) and Goswami *et al*. (2006). Both showed larger effect sizes. In a sensitivity analysis of working memory (Digit Span backward) excluding the study of Goswami *et al*. (2006), the observed heterogeneity largely disappeared (before exclusion: $\chi^2=30.50$, $p=0.000$; after exclusion: $\chi^2=5.70$, $p=0.223$). The effect size reflecting bipolar–control differences remained significant ($d=0.73$, $p=0.000$). The study of Goswami *et al*. (2006) also caused most of the heterogeneity in the analysis on executive control (Trail B). Leaving this study out resulted in non-significant heterogeneity (before exclusion: $\chi^2=33.93$, $p=0.000$; after exclusion: $\chi^2=10.78$, $p=0.214$). The effect size reflecting bipolar–control differences remained significant ($d=0.58$, $p=0.000$). In the case of concept shifting (WCST categories), heterogeneity was largely caused by the study of Balanza-Martinez *et al*. (2005) (before exclusion: $\chi^2=19.76$, $p=0.019$; after exclusion: $\chi^2=11.21$, $p=0.190$). The effect size reflecting bipolar–control differences remained significant ($d=0.39$, $p=0.000$). The fact that the most significant heterogeneity was due to only two studies suggests that certain characteristics of these studies may be responsible for this finding. The study by Balanza-Martinez *et al*. (2005) was relatively small and used a bipolar population with rather low educational level and no specification of characteristics of disease (duration, number of episodes, etc.). One could speculate that they described a rather severely ill population. Goswami *et al.* (2006) used a rather young population with a relatively long duration of illness and early illness onset. This study also probably included a rather severely ill group of patients.

Meta-regression revealed a significant effect of sex ratio on the concept formation and shifting case-control difference (WCST) ($p=0.001$, $B=-2.63$, 95% CI -4.156 to -1.11). This finding indicates that studies with higher male/female ratios showed smaller effect sizes. Age had a significant effect on the case-control difference of concept formation (WCST) and working memory (Digit Span backward) ($p=0.000$, $B=-32.51$, 95% CI -43.6 to -21.4, and $p=0.029$, $B=-11.18$, 95% CI -21.2 to -1.18). Thus, studies with higher mean age showed smaller effect sizes. Finally, educational level had a significant effect on the working memory case-control difference (Digit Span backward), fluency (FAS) and concept formation (WCST) ($p=0.03$, $B=-0.014$, 95% CI -0.027 to -0.001; $p=0.007$, $B=36.93$; 95% CI 8.38–52.24; $p=0.014$, $B=2.53$, 95% CI 0.52–4.55). This points in the direction of larger effect sizes in studies with higher educated participants.

In conclusion, part of heterogeneity may be due to differences between the various studies in these independent variables.

Meta-analytical results: first-degree relatives

A total of 14 studies were included (Table 3). Two of these studies used more than one family group (Sobczak *et al.* 2003; McIntosh *et al.* 2005). In the study by McIntosh *et al.* (2005), a group of unaffected relatives from bipolar families and a group from 'mixed' families was used; only the group from bipolar families was included in the analyses. Sobczak *et al.* (2003) used a group of first-degree relatives of bipolar I patients and a group of relatives of bipolar II patients; only the group of family-members of bipolar I patients was used in the meta-analysis.

The neuropsychological tests used in the studies were divided in the same categories as described earlier and included only if used in at least four different studies. This resulted in less cognitive domains analysed than in the bipolar studies. These domains were immediate and delayed verbal memory, working memory, concept formation and shifting, verbal fluency, executive control, mental speed, and intelligence. The Visual Verbal Learning Test used in the study by Sobczak *et al*. (2003), measuring immediate and delayed verbal memory and resembling the CVLT and RAVLT (Lezak, 1995), was added to the analysis.

Meta-analysis of the neuropsychological domains indicated that all meta-analytical effect sizes were in the direction of worse performance in the first-degree relatives compared to the healthy controls (Table 4). Effect sizes, however, were much smaller than in the bipolar–control comparisons (<0.5), and only significantly different for executive control (Stroop and Trail B).

There was evidence of significant heterogeneity for three out of 12 analyses, namely for the domains of intelligence and working memory (Digit Span).

Heterogeneity may be due to the small number of studies with small, heterogeneous groups of

first-degree relatives with different family histories and genetic load. The study by Gourovitch *et al.* (1999), for example, using monozygotic twins, showed relatively large but differential effect sizes for working memory and verbal memory, contributing to heterogeneity.

Meta-regression revealed no significant effects of the independent variables examined.

Table 1. *Studies with bipolar patients included in the meta-analysis*

Author (year)	Subjects (*n*)		Definition of euthymia[a]	Neuropsychological test parameters	*d*[b]
	Patients	Controls			
Altshuler *et al*. (2004)	40	22	HAMD < 6 YMRS < 7 prospectively for 3 months	CVLT immediate recall CVLT delayed recall Rey figure delayed FAS WCST perseverative errors WCST category Stroop time Trail A Trail B IQ Rey figure copy	0.75 0.78 0.57 0.16 0.77 0.89 0.41 0.38 0.40 0.20 0.30
Balanza-Martinez *et al.* (2005)	>15	>26	HAMD < 8 HAMD 3.4 (2.9) CARS < 8 CARS 1.3 (1.8) 2 months euthymia	FAS Fluency cat. WCST perseverative errors WCST category Stroop time Trail A Trail B DSST	1.28 1.79 1.67 1.48 1.62 0.68 0.89 1.05
Blumberg *et al*. (2003)	15	20	HAMD < 8 HAMD 7.3 (7.1) CARS < 8 CARS 4.1 (5.0)	Stroop time	0.74
Bozikas *et al*. (2005)	19	30	MADRS < 9 MADRS 1.53 (2.61) YMRS < 9 YMRS 3.16 (2.48)	CPT	0.10
Cavanagh *et al*. (2002)	20	20	HAMD < 8 1.0 (2.9) MMS < 3 MMS 0.5 (1.5)	CVLT delayed recall FAS Stroop correct	0.96 0.29 0.61
Clark *et al.* (2002)	30	30	HAMD < 9 HAMD 2.07 (2.26) YMRS < 9 YMRS 1.67 (2.22)	CVLT immediate recall CVLT delayed recall CPT	0.48 0.95 0.96
Clark *et al*. (2005*a*)	15	15	HAMD < 9 HAMD 3.2 (2.5) YMRS < 9 YMRS 1.9 (2.5)	CPT	1.00
Deckersbach *et al.* (2004*b*)	30	30	HAMD 3.4 (2.6) YMRS 1.0 (1.6)	CVLT immediate recall CVLT delayed recall	1.40 1.67

Table 1 (*cont.*)

Author (year)	Subjects (*n*)		Definition of euthymia[a]	Neuropsychological test parameters	*d*[b]
	Patients	Controls			
Deckersbach *et al.* (2004*a*)	25	25	HAMD 3.3 (2.5) YMRS 1.2 (1.5)	Rey figure delayed Rey figure copy	0.70 0.06
Dixon *et al.* (2004)	15	30	BDI 6.5 (4.3) YMRS 2.7 (2.2)	FAS Fluency category Stroop correct IQ	0.17 0.30 0.82 −0.32
Ferrier *et al.* (1999)	20	20	HAMD 2.7 (2.1) MSS 4.1 (1.9)	RAVLT immediate recall Rey figure delayed Digit Span backward FAS Trail A Trail B DSST Digit Span forward Rey figure copy	0.93 0.92 1.11 0.40 0.81 0.92 0.81 0.28 0.64
Fleck *et al.* (2003)	14	40	HAMD < 10 HAMD 3.7 (2.8) YMRS < 10	CVLT immediate recall CVLT delayed recall	1.01 0.77
Frangou *et al.* (2005*a*)	10	43	HAMD < 6 HAMD 3.0 (1.2) YMRS < 6 YMRS 1.1 (0.5) at least 1 month	WCST perseverative errors WCST category	0.55 0.04
Frangou *et al.* (2005*b*)	44	44	HAMD < 10 HAMD 7 MRS < 10 MRS 0	FAS WCST perseverative errors WCST category Stroop correct IQ	0.88 0.38 0.25 0.57 0.31
Goswami *et al.* (2006)	37	37	Euthymia > 1 month HAMD 2.35 (1.48) MSRS 7.91 (4.88)	RAVLT immediate recall Digit Span backward Trail A Trail B DSST Digit Span forward	0.69 2.28 0.54 1.99 0.19 0.50
Krabbendam *et al.* (2000)	21	22	HAMD 3.4 (3.0) YMRS 0.77 (1.5)	AVLT immediate recall AVLT delayed recall Fluency category Stroop time DSST	0.94 0.93 0.54 0.67 1.12
Larson *et al.* (2005)	18	18	HAMD 3 (3) YMRS 2 (3) Follow-up for 4–8 weeks	IQ	0.12
Malhi *et al.* (2005)	12	12	HAMD < 7 HAMD 4.3 (1.1) YMRS < 7 YMRS 0.9 (0.5)	Stroop time	1.02

Table 1 (*cont.*)

Author (year)	Subjects (*n*)		Definition of euthymia[a]	Neuropsychological test parameters	*d*[b]
	Patients	Controls			
Martinez-Aran *et al*. (2004)	44	30	HAMD < 9	CVLT immediate recall	0.84
			HAMD 3.6 (2.6)	CVLT delayed recall	0.96
			YMRS < 7	Digit Span backward	0.86
			YMRS 1.4 (1.8)	FAS	0.56
			6 months remission	Fluency category	0.83
				WCST perseverative errors	0.62
				WCST category	0.38
				Stroop correct	0.59
				Trail A	0.90
				Trail B	0.57
				Digit Span forward	0.56
				IQ	0.75
McIntosh *et al.* (2005)	27	50	HAMD 5	FAS	0.71
			YMRS 2	DSST	1.34
				IQ	−0.07
Nehra *et al*. (2006)	30	20	HAMD < 8	FAS	0.45
			HAMD 2.67 (0.92)	Fluency category	0.46
			YMRS < 8	WCST perseverative errors	0.37
			YMRS 1.47 (1.25)	WCST category	0.07
				Trail A	0.41
				Trail B	0.69
Strakowski *et al*. (2004)	10	10	HAMD < 8	CPT	0.21
			HAMD 3.1 (2.5)		
			YMRS < 6		
			YMRS 1.6 (1.8)		
Thompson *et al*. (2005)	63	63	HAMD < 8	RAVLT immediate recall	0.59
			HAMD 2.1 (1.7)	Digit Span backward	0.37
			YMRS < 8	FAS	0.36
			YMRS 1.4 (2.0)	Stroop correct	0.58
			Prospectively verified for 1 month	Trail A	0.47
				Trail B	0.23
				DSST	0.91
				Digit Span forward	0.05
Thompson *et al.* (2006)	20	20	HAMD < 8	Digit Span backward	0.75
			HAMD 1.90 (2.38)	Digit Span forward	0.25
			YMRS < 8		
			YMRS 1.40 (2.08)		
Torrent *et al.* (2006)	38	35	HAMD < 9	CVLT immediate recall	0.58
			HAMD 4.29 (2.51)	CVLT delayed recall	0.80
			YMRS < 7	Digit Span backward	0.86
			YMRS 0.79 (1.19)	FAS	0.41
				Fluency category	0.76
				WCST perseverative errors	0.56
				WCST category	0.23
				Stroop correct	0.58
				Trail A	0.80
				Trail B	0.57
				Digit Span forward	0.70

Table 1 (*cont.*)

Author (year)	Subjects (*n*)		Definition of euthymia[a]	Neuropsychological test parameters	*d*[b]
	Patients	Controls			
Van Gorp *et al.* (1998)	13	22	HAMD < 7	CVLT immediate recall	0.70
			YMRS < 6	CVLT delayed recall	0.52
				Rey figure delayed	0.25
				FAS	−0.11
				WCST perseverative errors	0.95
				WCST category	1.00
				Stroop time	0.08
				Trail A	0.32
				Trail B	0.24
				Rey figure copy	−0.09
Varga *et al.* (2006)	19	31	MADRS 2.26 (3.69)	AVLT immediate recall	1.52
			MRS 2.32 (4.10)	AVLT delayed recall	1.04
				WCST perseverative errors	0.55
				WCST category	0.15
				Stroop correct	0.80
				Trail A	0.53
				Trail B	1.24
				DSST	0.54
				IQ	0.57
Zubieta *et al.* (2001)	15	15	HAMD < 6	Fluency category	0.77
			HAMD 3.4 (2.1)	WCST perseverative errors	1.52
			YMRS < 4	WCST category	0.84
			YMRS 0.4 (0.6)	Stroop correct	1.12
			At least 6 months' euthymia		

HAMD, Hamilton Depression Rating Scale; YMRS, Young Mania Rating Scale; CARS, Clinician Administered Rating Scale for Mania; MADRS, Montgomery–Asberg Depressive Rating Scale; MMS, Modified Manic Scale; BDI, Beck Depression Inventory; MSS, Manic State Scale; MRS, Manic Rating Scale; MSRS, Manic State Rating Scale; CVLT, California Verbal Learning Test; FAS, Verbal fluency test; WCST, Wisconsin Card Sorting Test; DSST, Digit Symbol Substitution Test; CPT, Continuous Performance Test; RAVLT, Rey Auditory Verbal Learning Test; AVLT, Auditory Verbal Learning Test.

[a] Values in parentheses are standard deviations.

[b] Effect size, positive values indicate better performance in controls.

Table 2. *Results of meta-analyses of cognitive test performance differences between bipolar patients versus normal controls*

Test	K[a]	Subjects (n)		d[b]	95% CI	Z[c]	p	χ^2[d]	p
		Bipolar	Control						
Digit backward	6	222	205	1.02	0.49 to 1.54	3.85	0.000	30.50	0.000
CVLT delayed recall	10	269	282	0.85	0.60 to 1.09	6.83	0.000	16.27	0.061
DSST	7	202	249	0.84	0.53 to 1.14	5.32	0.000	13.76	0.032
CVLT immediate recall	12	369	382	0.82	0.65 to 0.99	9.25	0.000	13.96	0.235
Fluency categories	7	178	178	0.75	0.44 to 1.04	4.83	0.000	10.91	0.091
Trail B	10	319	306	0.75	0.42 to 1.1	4.47	0.000	33.93	0.000
Stroop time	6	116	124	0.73	0.32 to 1.13	3.49	0.000	11.00	0.051
WCST perseverative errors	10	268	288	0.72	0.48 to 0.95	5.90	0.000	15.24	0.085
Stroop correct	8	258	268	0.65	0.47 to 0.83	7.17	0.000	2.37	0.937
Rey figure recall	4	98	89	0.62	0.32 to 0.92	4.04	0.000	2.01	0.570
CPT correct	4	74	85	0.58	0.09 to 1.08	2.31	0.021	6.52	0.089
Trail A	10	319	306	0.58	0.42 to 0.75	7.02	0.000	4.88	0.845
WCST categories	10	268	288	0.49	0.22 to 0.76	3.59	0.000	19.76	0.019
FAS	12	369	382	0.47	0.30 to 0.65	5.14	0.000	15.54	0.159
Digit forward	6	222	205	0.37	0.15 to 0.59	3.33	0.001	6.19	0.288
Rey copy	4	103	94	0.22	−0.06 to 0.51	1.52	0.129	2.89	0.409
IQ	8	237	247	0.16	−0.11 to 0.44	1.15	0.250	15.36	0.032

CI, Confidence interval; CVLT, California Verbal Learning Test; DSST, Digit Symbol Substitution Test; WCST, Wisconsin Card Sorting Test; CPT, Continuous Performance Test; FAS, Verbal fluency test.

[a] Number of studies included in the analysis.

[b] Mean, weighted effect size Cohen's d.

[c] Test of significance of effect size (p).

[d] Test of within category heterogeneity between studies (p).

Table 3. *Studies with first-degree family members included in the meta-analysis*

Author (year)	Subjects (n)		Sample characteristics[a]	Neuropsychological test parameters	d[b]
	Family	Controls			
Christensen *et al.* (2006)	7	36	MZ twins discordant for bipolar disorder	Stroop Trail A Trail B	0.37 0.20 0.63
Christensen *et al.* (2006)	14	52	DZ twins discordant for bipolar disorder	Stroop Trail A Trail B	0.45 −0.10 0.25
Clark *et al.* (2005*b*)	27	47	10 parents, 12 siblings, 5 children HAMD 1.2 (1.9) YMRS 0.4 (1.1)	CVLT immediate recall CVLT delayed recall	0.20 0.12
Ferrier *et al.* (2004)	17	17	First-degree relatives HAMD 0.82 (1.01) YMRS 0.47 (1.28) Controls HAMD 0.35 (0.86) YMRS 0.18 (0.53)	RAVLT immediate recall Digit Span backward FAS Stroop Trail A Trail B DSST Digit Span forward	0.18 0.99 −0.12 0.00 −0.07 0.37 0.24 0.39
Frangou *et al.* (2005*b*)	15	43	Unaffected offspring of bipolar probands	WCST perseverative errors WCST category WAIS-R IQ	−0.42 −0.53 −0.09

Table 3 (*cont.*)

Author (year)	Subjects (*n*)		Sample characteristics[a]	Neuropsychological test parameters	*d*[b]
	Family	Controls			
Gourovitch *et al.* (1999)	7	15	MZ twins	CVLT immediate recall	0.33
				CVLT delayed recall	0.80
				Digit Span backward	0.97
				FAS	0.28
				WCST perseverative errors	0.52
				Trail A	−0.10
				Trail B	0.01
				Digit Span forward	1.16
				WAIS-R IQ	0.40
Kéri *et al.* (2001)	20	20	Unaffected siblings BP-I probands	Digit Span backward	−0.18
				FAS	0.12
				WCST perseverative errors	0.10
				WCST category	0.11
				Digit Span forward	−0.33
Kieseppa *et al.* (2005)	19	114	Twins discordant for BP-I	CVLT delayed recall	0.08
				Digit Span backward	−0.18
				DSST	−0.12
Kremen *et al.* (1998)	14	44	Relatives of psychotic bipolar probands	WCST perseverative errors	0.09
				WCST category	0.45
				Trail A	−0.28
				Trail B	−0.11
				DSST	−0.05
				WAIS-R IQ	−0.58
McIntosh *et al.* (2005)	24	50	Unaffected relatives with >1 first- or second-degree BP proband HAMD 1.5 (median) YMRS 0 Controls HAMD 0 YMRS 0	FAS	0.58
				DSST	0.50
Pirkola *et al.* (2005)	16	100	Unaffected co-twins 3 MZ, 13 DZ	Digit Span backward	−0.30
				Digit Span forward	−0.78
Sobczak *et al.* (2003)	22	15	First-degree relatives BP-I	VVLT immediate recall	0.25
				VVLT delayed recall	0.34
Szoke *et al.* (2006)	51	50	First-degree relatives of BP-I patients	WCST perseverative errors	0.22
				Trail A	0.41
				Trail B	0.54
Toulopoulou *et al.* (2006)	50	69	17 parents, 23 siblings 10 children	WAIS-R IQ	0.42
Zalla *et al.* (2004)	33	20	11 parents, 22 siblings MADRS < 16 MAS < 7	WCST perseverative errors	0.57
				WCST category	0.12
				Stroop	1.03
				Trail A	0.31
				Trail B	0.60
				WAIS-R IQ	0.79

MZ, Monozygotic; DZ, dizygotic; HAMD, Hamilton Depression Rating Scale; YMRS, Young Mania Rating Scale; MADRS, Montgomery–Asberg Depressive Rating Scale; MAS, Beck–Rafaelsen Mania Scale; CVLT, California Verbal Learning Test; RAVLT, Rey Auditory verbal Learning Test; FAS, Verbal fluency test; DSST, Digit Symbol Substitution Test; WCST, Wisconsin Card Sorting Test.

[a] Values in parentheses are standard deviations.

[b] Effect size, positive values indicate better performance in controls.

Table 4. *Results of meta-analyses of cognitive test performance differences between first-degree relatives versus normal controls*

		Subjects (*n*)							
Test	*K*[a]	Family	Control	*d*[b]	95% CI	*z*[c]	*p*	χ^{2}[d]	*p*
Stroop	4	71	125	0.49	0.045 to 0.93	2.16	0.031	5.35	0.148
Trail B	7	143	234	0.37	0.15 to 0.60	3.27	0.001	4.98	0.546
FAS	4	68	102	0.27	−0.04 to 0.59	1.70	0.090	3.01	0.391
CVLT immediate	4	73	94	0.22	−0.09 to 0.53	1.38	0.167	0.08	0.994
CVLT delayed recall	4	75	191	0.21	−0.07 to 0.50	1.45	0.146	2.11	0.550
IQ	5	119	191	0.19	−0.27 to 0.65	0.82	0.414	12.77	0.012
Digit span backward	5	79	266	0.18	−0.33 to 0.69	0.69	0.490	13.29	0.010
WCST perseverative errors	6	140	192	0.17	−0.09 to 0.43	1.26	0.207	6.26	0.282
DSST	4	74	225	0.14	−0.16 to 0.45	0.91	0.361	3.66	0.300
Trail A	7	143	234	0.13	−0.09 to 0.35	1.14	0.256	5.28	0.508
Digit span forward	4	60	152	0.04	−0.72 to 0.81	0.11	0.911	15.23	0.002
WCST categories	4	82	127	0.04	−0.36 to 0.43	0.18	0.861	5.35	0.148

CI, Confidence interval; FAS, Verbal fluency test; CVLT, California Verbal Learning Test; IQ, intelligence quotient; WCST, Wisconsin Card Sorting Test; DSST, Digit Symbol Substitution Test.

[a] Number of studies included in the analysis.

[b] Mean, weighted effect size Cohen's *d*.

[c] Test of significance of effect size (*p*).

[d] Test of within category heterogeneity between studies (*p*).