Journal of Affective Disorders 164 (2014) 130–138

Contents lists available at ScienceDirect

Journal of Affective Disorders

journal homepage: www.elsevier.com/locate/jad

Research report

Longitudinal course of cognitive deficits in bipolar disorder: A meta-analytic study

Cecilia Samamé [a,b], Diego J. Martino [a,b], Sergio A. Strejilevich [a,c,*]

[a] Bipolar Disorders Program, Institute of Neurosciences, Favaloro University, Buenos Aires, Argentina
[b] National Scientific and Technical Research Council (CONICET), Buenos Aires, Argentina
[c] Institute of Cognitive Neurology (INECO), Buenos Aires, Argentina

ARTICLE INFO

Article history:
Received 27 March 2014
Accepted 11 April 2014
Available online 19 April 2014

Keywords:
Bipolar disorder
Cognitive impairment
Longitudinal
Meta-analysis

ABSTRACT

Objective: Persistent cognitive deficits in bipolar disorder represent a major impediment to functional adjustment, but their static or progressive nature remains to be ascertained. The aim of this study was to synthesize findings from longitudinal research in order to examine the trajectory of cognitive impairment in bipolar disorder.
Method: A literature search was conducted through online databases covering the period between January 1990 and February 2014. Two approaches were undertaken. First, the results of longitudinal studies including neuropsychological assessment of stable bipolar patients at baseline and after a follow-up period of at least one year were meta-analyzed so as to obtain overall test–retest effect sizes for neurocognitive domains. Second, meta-analysis was restricted to longitudinal studies of bipolar patients including a control group. Patients' and controls' overall test–retest effect sizes were compared.
Results: Bipolar patients' performance on 14 cognitive measures remained stable after a mean follow-up period of 4.62 years. When meta-analysis was restricted to controlled studies, no patient-control differences were found regarding longitudinal cognitive outcomes.
Limitations: Test–retest differences for medication variables and mood state could not be controlled. Sufficient data were not available to investigate a wider array of neuropsychological domains. Furthermore, most primary studies included relatively short test–restest intervals.
Conclusion: To date, the available evidence from longitudinal studies is not in accordance with the hypothesis of a progressive nature of cognitive deficits in BD. The implications of this finding for further research are discussed.

1. Introduction

Bipolar disorders (BDs) comprise a heterogeneous group of chronic and recurrent affective illnesses associated with impairments in different aspects of daily living (Gitlin et al., 1995; Huxley and Baldessarini, 2007; Jansen et al., 2012). Several studies have revealed that a considerable number of bipolar patients exhibit persistent cognitive dysfunctions, with medium-to-large effect sizes of impairment noted for attention/processing speed, verbal memory, and executive domains (Robinson et al., 2006; Torres et al., 2007; Mann-Wrobel et al., 2011). Flawed neuropsychological performance has been shown to be a strong predictor of functional maladjustment both in cross-sectional (Dickerson et al., 2004; Martino et al., 2008; Fulford et al., 2014) and longitudinal studies (Tabarés-Seisdedos et al., 2008; Martino et al., 2009; Bonnín et al., 2010). These considerations are particularly relevant to BDs, given that between one and two thirds of bipolar patients do not accomplish functional recovery even when syndromal recovery is evident (Tohen et al., 2000; Strejilevich et al., 2013b). Hence, neurocognitive dysfunctions are increasingly acknowledged as a target area for treatment and research on this group of disorders.

Despite the growing awareness of the critical importance of neurocognitive functioning to BDs' outcome, data on the longitudinal trajectory of cognitive deficits across the course of the illness are scarce and inconsistent. Some studies found a negative association between the number of episodes, particularly manic ones, and neurocognitive functioning (Robinson and Ferrier, 2006; López-Jaramillo et al., 2010; Hellvin et al., 2012). These findings led authors to suggest that the experience of successive episodes might be related to progressive neurocognitive decline. The evidence supporting that cognitive impairments increase as a function of the number of previous episodes in bipolar patients is summarized in a recent review (Post et al., 2012). Moreover, this

* Corresponding author at: Congreso 2477 Dto. D (1428) Ciudad Autónoma de Buenos Aires, Argentina. Tel./fax: +54 11 4833 2424.
E-mail address: sstrejilevich@ffavaloro.org (S.A. Strejilevich).

http://dx.doi.org/10.1016/j.jad.2014.04.028
0165-0327/

association is usually considered as further evidence for illness progression in BDs (Robinson and Ferrier, 2006; Berk, 2009; Kapczinski et al., 2009; Post et al., 2012). However, almost all of these hypotheses are primarily based on cross-sectional studies, and the direction of causality is ambiguous (Martino et al., 2013b). As evident, the best approaches to understanding the trajectory of these deficits are longitudinal studies with serial neurocognitive assessments. To date, longitudinal studies have been scant and yielded mixed results: while some of them showed stable cognitive deficits over time (Balanzá-Martínez et al., 2005; Mur et al., 2008a; 2008b; Schouws et al., 2012; Gildengers et al., 2013), others revealed a pattern of progressive deterioration (Moorhead et al., 2007; Gildengers et al., 2009). Furthermore, most of them had high probabilities of type II error owing to small sample size.

Broadening our knowledge on the longitudinal course of cognition in BD is an indispensable step towards having a complete description of these disorders. It would contribute to better understanding of the pathophysiological mechanisms subserving BDs, to identify targets for treatment, to determine possible subtypes of the disorder, and to develop better therapeutic strategies. The aim of the current work was to pool the results of studies including cognitive measures of bipolar patients at different time points in order to overcome sample-size limitations and gain some insight into the longitudinal course of cognition in BDs.

2. Method

2.1. Search strategy

Articles were retrieved from the online databases Pubmed/PsychInfo using combinations of the following keywords: bipolar disorder, manic, cognition, neuropsychology, longitudinal/long term, prospective, follow-up, progression, stability, intelligence, IQ, attention, learning, memory, and executive. The reference lists of the studies identified for inclusion were also reviewed for additional relevant reports.

2.2. Primary study selection criteria

Reports were considered for the current meta-analysis if they met the following criteria: (I) Were published in a peer-reviewed English language journal between January 1990 and February 2014. (II) Included a patient group aged over 18 years, with the diagnosis of BD according to standardized diagnostic criteria (RDC, DSM-III, DSM-IV, ICD-10, etc.). (III) Involved longitudinal study design with neuropsychological assessment at baseline and after a follow-up period of at least one year. (IV) Patients were described as euthymic, stable or mildly symptomatic both at baseline and after the follow-up period. (V) Provided data to estimate effect sizes for patients' differences between test and re-test cognitive scores. (VI) Subjects were not given any specific treatment to enhance cognition. (VII) Included at least one cognitive measure that was examined in a minimum of three studies. (VIII) Included at least ten subjects at both time points.

Additionally, if there were studies with overlapping content based on the same patient sample, we considered the data from the study with the longest follow-up period. Two studies on the same patient group were only included if they reported different cognitive measures.

2.3. Meta-analytic procedure

Meta-analyses were performed using the Comprehensive Meta-Analysis software version 2.0 (Borenstein et al., 2005). Given that only a small number of studies provided data from healthy controls, and in order not to overlook the body of evidence provided by single group longitudinal studies, two different meta-analytic approaches were conducted. First, we included all longitudinal studies, regardless of whether or not they included a healthy control group. Subjects' test–retest effect sizes (d) for each cognitive measure were calculated by subtracting the average score after follow-up from the average score at baseline and dividing the result by the pooled standard deviations of both data sets. Given that correlations between pretest and posttest scores were not available, we used this approach in order to avoid overestimation of the magnitude of effect, as recommended by Dunlap et al. (1996) for the estimation of effect sizes in meta-analysis of repeated measures designs. When studies reported neuropsychological performance at more than two different time points, we only considered the scores reported at baseline and after the longest follow-up period, except in one case in which the subjects included after the longest period were less than ten (Yucel et al., 2007). Effect sizes were weighted using the inverse variance method. Whenever subjects performed better after the follow-up period we reported test–retest differences by positive effect sizes. Second, we performed meta-analyses based only on studies including a healthy control group both at baseline and after follow-up. Hence, we obtained overall test–retest effect sizes for both patients and controls.

The Q-test for heterogeneity was used to test the homogeneity of the resulting mean weighted effect size for each variable and to compare patients' and controls' overall test–retest effect sizes. The I^2 index was calculated to describe the percentage of total variation across reports due to heterogeneity rather than chance. I^2 values of 25%, 50%, and 75% indicate low, moderate, and high heterogeneity respectively. Based on the small sample size and the presence of heterogeneity in some of the analyses, we chose a random effects model. A significance level of $p < 0.05$ was set for the random effects model and homogeneity analyses.

2.4. Cognitive variables

For the purposes of this study, the results of reports utilizing the same test or tapping approximately the same neuropsychological construct were combined into a single summary measure. Fourteen overall neuropsychological measures were obtained. Crystallized Intelligence was explored using the full-scale National Adult Reading Test (Nelson, 1982) and the revised Wechsler Adult Intelligence Scale –WAIS- vocabulary/information scores (Wechsler, 1955; 1997). Two distinct attention summary measures were calculated using results of the Trail Making Test part A – TMT_A – (Reitan, 1958) and variants of the Continuous Performance Test – CPT- (Conners and Staff, 2000). The test parameters considered were 'seconds employed to conclude the task' and 'target detection' respectively. Immediate verbal memory was assessed by means of word list learning (trials 1–5) of the California Verbal Learning Test – CVLT- (Delis et al., 1987) and the Rey Auditory Verbal Learning Test – RAVLT- (Rey, 1964). The results of these tests were combined into a list learning overall score. Delayed list learning was assessed by combining free delayed recall measures of the CVLT and RAVLT. Verbal fluency was assessed by means of tasks requiring either the naming of words corresponding to a common category (animals) or words beginning with a certain letter (Benton et al., 1983). Meta-analyses for categorical and phonemic scores were conducted separately. Processing speed was assessed using latencies (ms) on reaction time tests. Overall measures for digit span were obtained by combining the results of studies utilizing the WAIS Digit Span scores. Forward and backward digit spans were meta-analyzed separately. A measure of cognitive flexibility was obtained by

combining the results of studies including the Wisconsin Card Sorting Test – WCST- (Heaton, 1981). The parameter used was the number of perseverative errors. Two different measures of executive control were obtained using the results of the Stroop test (Golden, 1978) and the Trail Making Test part B (Reitan, 1958). The parameters considered were 'interference score' and 'time employed to conclude the task'. Finally, a global cognitive score was also computed by either pooling all domain effect sizes within a study assessing multiple neuropsychological domains or using global cognition scores when studies reported them.

3. Results

Our search strategy identified 35 studies exploring cognitive functioning in bipolar patients at different time points. Twenty-one of them were excluded as they did not meet the inclusion criteria. Finally, 14 reports met all the inclusion criteria required for this meta-analysis (Balanzá-Martínez et al., 2005; Burdick et al., 2006; Moorhead et al., 2007; Yucel et al., 2007; Depp et al., 2008; Mur et al., 2008a; 2008b; Tabarés-Seisdedos et al., 2008; Delaloye et al., 2011; Schouws et al., 2012; Torrent et al., 2012; Braw et al., 2013; Gildengers et al., 2013; Mora et al., 2013). Two of them were excluded as they were based on the same sample used in other study (Mur et al., 2008a; 2008b). The studies by Tabarés-Seisdedos et al. (2008) and Balanzá-Martínez et al. (2005) were based on the same sample but reported some different neuropsychological data and were therefore not meta-analyzed together. Twelve reports comparing the neurocognitive performance of at least 357 BD patients at baseline with that exhibited after a follow-up period (weighted mean=3.33 years) were included in the current review (Fig. 1, Table 1). Meta-analyses were performed for 14 cognitive variables including, on average, 152 patients followed for a mean period of 4.62 years. Finally, six reports (Depp et al., 2008; Tabarés-Seisdedos et al., 2008; Delaloye et al., 2011; Schouws et al., 2012; Gildengers et al., 2013; Mora et al., 2013) enabled to compare the longitudinal course of cognition of 233 patients (mean follow-up period=2.20 years) with that of 165 healthy controls (mean follow-up period=2.18 years).

3.1. Meta-analytic findings

Patients' test–retest effect sizes for cognitive variables, together with their confidence intervals, significance tests, and homogeneity statistics are reported in Table 2. No significant effect sizes were observed for any of the variables analyzed. The distributions of effect sizes were highly homogeneous (Q-test $p > 0.05$, $I^2 < 25\%$) for ten variables, namely crystallized IQ, TMT_A, phonemic fluency, continuous performance test, list learning, delayed list recall, forward digit span, backward digit span, Stroop interference score, psychomotor speed, and global cognitive score. In the presence of high homogeneity, the same result could be obtained using either a fixed or a random effects model. The hypothesis of homogeneity was not rejected in any of the analyses performed, though a trend towards significance of the Q-test was observed in the category fluency analysis.

When restricting meta-analysis only to those studies including a healthy control group, four summary measures were obtained: backward digit span, Stroop (interference score), phonemic fluency, and global cognitive score. No patient-control differences were found for age in any of the four analyses. Significant differences favouring controls were found for years of education in three analyses, namely phonemic fluency ($d=0.59$, CI=0.05 to 1.14, $p=0.03$), Stroop ($d=0.70$, CI=0.28 to 1.12, $p<0.01$), and backward digit span ($d=0.59$, CI=0.05 to 1.14, $p=0.03$). The within-group effect sizes for cognitive variables, together with their confidence intervals, significance tests, and homogeneity statistics are reported in Table 3. No significant test–retest effect sizes were observed neither for patients nor controls. The hypothesis of homogeneity was not rejected in any of the analyses

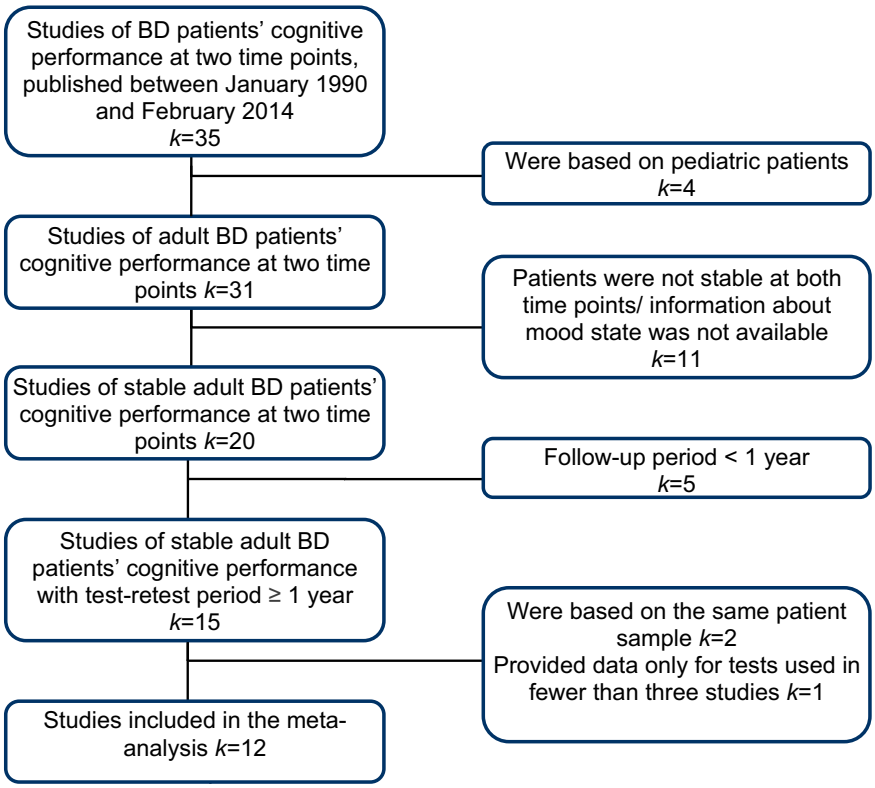

Fig. 1. Selection process for the primary reports included in the meta-analysis.

Table 1
Studies included in the current meta-analysis.

Primary study	Sample BD (type)/HC	Mean age at baseline (SD)	Follow-up (years)	Mood state (baseline)	Mood state (follow-up)	Cognitive variables	BDs' Test–retest ES	HCs' Test–retest ES
Balanzá-Martínez et al. (2005)	15(I)/26	41.5 (11.1)	3	Stable HDRS: 3.4 (2.9) CARS-M factor1: 1.3 (1.8)	Stable HDRS: 3.4 (2.9) CARS-M factor1: 1.3 (1.8)	TMT_A TMT_B WCST (perseverative errors)	0.47 0.47 0.14	NA
Braw et al. (2013)	31(I–II)/31	41.10 (13.43)	2	Stable, minimally symptomatic HDRS: 11.0 (4.4) YMRS: 2.5 (1.7)	Stable, minimally symptomatic HDRS: 11.1 (8.4) YMRS: 3.7 (3.0)	Processing speed CPT (target detection)	0.3 0.01	NA
Burdick et al. (2006)	16(?)	38.1 (5.7)	5	Euthymic (criteria NA) SADS(D): 2.5 (1.6) SADS(P): 1.4 (0.7)	Euthymic (criteria NA) SADS(D): 2.6 (1.1) SADS(P): 1.3 (0.6)	Crystallized IQ TMT_A TMT_B Phonemic fluency List learning Delayed list recall WCST (perseverative errors) Global score	-0.22 0.33 -0.2 0.52 0.53 0.52 0.63 0.26	NA
Delaloye et al. (2011)	15(I–II)/15	67.93 (5.18)	2	Euthymic (GDS < 5 YMRS < 5)	Euthymic (GDS < 5, YMRS < 5)	Processing speed Stroop (interference score)	−0.1 −0.09	−0.2 −0.17
Depp et al. (2008)	35(I–II)/35	57.7 (10.0)	1.3	Stable, minimally symptomatic	Stable	Global score	0.16	0.26
Gildengers et al. (2013)	47(I–II)/22	68.0 (9.3)	2	Euthymic (HDRS < 10, YMRS < 10)	Euthymic (HDRS < 10, YMRS < 10)	Global score	−0.3	−0.08
Moorhead et al. (2007)	20(I)/21	41.5 (8.9)	4.1	Euthymic (HDRS < 6, YMRS < 6)	Euthymic (HDRS < 6, YMRS < 6)	Crystallized IQ	−0.1	0.48
Mora et al. (2013)	28(I–II)/26	41.71 (12.4)	6	Euthymic (HDRS < 8, YMRS < 6) HDRS: 1.46 (1.7) YMRS: 1.64 (1.9)	Euthymic (HDRS < 8, YMRS < 6) HDRS: 2.11 (2.2) YMRS: 1 (1.3)	Crystallized IQ TMT_A TMT_B Phonemic Fluency Digit backwards Digit span forward Delayed list recall List learning WCST (perseverative errors) Stroop (interference score) CPT (target detection) Psychomotor speed Global score	0.37 −0.09 −0.49 −0.09 −0.11 0 0.05 −0.19 0.27 0.25 −0.22 −0.2 −0.06	0.38 0.11 0.01 −0.04 0.06 −0.15 0.97 0.43 0.55 0.48 0.51 −0.03 0.27
Schouws et al. (2012)	65(I–II)/42	68.35 (6.5)	2	Euthymic CES-D: 9.86 (7.5) YMRS: 1.06 (1.7)	Euthymic CES-D: 12.51 (5.5) YMRS: 1.06 (1.1)	TMT_A TMT_B Digit span forward Digit span backwards List learning Delayed list recall Category fluency Phonemic fluency Global score	0.05 −0.06 0.17 −0.16 0.05 0.02 −0.12 0.13 0.06	−0.22 −0.24 0.02 −0.25 −0.4 −0.23 −0.07 −0.21 0.06
Tabarés-Seisdedos et al. (2008)	43(I)/25	41.2 (11.5)	1	Stable 77% euthymic (HDRS ≤ 8, CARS-M factor 1 ≤ 7). HDRS: 4.2 (4.1) CARS-M Factor 1: 12.2 (4.2)	Stable 63% euthymic (HDRS ≤ 8, CARS-M factor 1 ≤ 7). HDRS: 4.7 (5.7) CARS-M Factor 1: 1.3 (2.3)	Digit backwards Category fluency Phonemic fluency Stroop (interference score) CPT (target detection) Global score	0.1 0.19 0.25 0.17 0.15 0.19	0.34 0.39 0.35 0.33 0.22 0.3
Torrent et al. (2012)	45(I–II)/45	39.31 (12.04)	8.9	Euthymic (HDRS ≤ 8, YMRS ≤ 6) HDRS: 2.81 (2.34) YMRS: 1.28 (1.65)	Euthymic (HDRS ≤ 8, YMR ≤ 6) HDRS: 3.96 (3.06) YMRS: 1.89 (1.76)	Crystallized IQ TMT_A TMT_B Digits backwards Digits forward Category fluency Phonemic fluency List learning Delayed list recall WCST (perseverative errors) Stroop (interference score) Global score	0.19 −0.15 −0.44 −0.01 0.35 −0.49 −0.27 −0.19 −0.09 −0.32 −0.15 −0.11	NA

Table 1 (*continued*)

Primary study	Sample BD (type)/HC	Mean age at baseline (SD)	Follow-up (years)	Mood state (baseline)	Mood state (follow-up)	Cognitive variables	BDs' Test–retest ES	HCs' Test–retest ES
Yucel et al. (2007)	12(I–II)	28.4 (10.7)	2	Minimally symptomatic HDRS: 8.1 (6.3) YMRS: 2.4 (3.8)	Euthymic HDRS: 4.3 (5.1) YMRS: 0.2 (0.6)	List learning	0.23	NA

ES=effect size; BD=bipolar disorder; CARS-M=clinician-administered rating scale for mania; CES-D=center for epidemiologic studies depression scale; CPT=continuous performance test; GDS=geriatric depression scale; HC=healthy controls; HDRS=Hamilton depression rating scale; SADS(D)=schedule for affective disorders and schizophrenia-depression; SADS(P)=schedule for affective disorders and schizophrenia-psychosis; SD=standard deviation; TMT_A=trail making test–part A; TMT_B=trail making test–Part B; WCST=Wisconsin card sorting test; YMRS=Young mania rating scale; NA=not available.

Table 2
Mean weighted effect sizes for BDs' neurocognitive performance at two different time points.

Variable	Studies (*k*)	Subjects	Follow-up	ES[a]	95% CI	*Z*[b]	*P*	*Q*-test (*P*)[c]	I^2
Crystallized IQ	4	109	6.70	0.13	−0.13 to 0.38	0.98	0.33	0.50	0.00
TMT_A	5	169	4.87	0.04	−0.18 to 0.25	0.32	0.75	0.55	0.00
TMT_B	5	169	4.87	−0.19	−0.47 to 0.10	−1.30	0.19	0.18	36.78
Category fluency	3	153	3.75	−0.14	−0.50 to 0.23	−0.74	0.46	0.08	60.60
Phonemic fluency	5	197	4.17	0.07	−0.17 to 0.30	0.55	0.58	0.26	24.44
CPT	3	102	2.68	0.01	−0.27 to 0.28	0.05	0.96	0.56	0.00
List learning	5	166	4.83	0.00	−0.21 to 0.22	0.03	0.98	0.43	0.00
Delayed list recall	4	154	5.06	0.04	−0.18 to 0.27	0.37	0.71	0.53	0.00
Backward digit span	4	181	4.10	−0.08	−0.28 to 0.11	−0.83	0.41	0.92	0.00
Forward digit span	3	138	5.06	0.22	0.00 to 0.44	1.94	0.05	0.52	0.00
Stroop (interference score)	4	131	4.90	0.01	−0.26 to 0.29	0.09	0.93	0.64	0.00
Psychomotor speed	3	74	3.51	0.03	−0.29 to 0.36	0.19	0.85	0.37	0.00
WCST (perseverative errors)	4	104	6.67	0.09	−0.22 to 0.41	0.60	0.55	0.17	37.33
Global Cognitive Score	7	279	3.44	0.00	−0.16 to 0.17	0.05	0.96	0.64	0.00

CPT=Continuous performance test; TMT=trail making test; WCST=Wisconsin card sorting test.

[a] Effect size (*d*).
[b] Test of significance of effect size.
[c] Test of homogeneity, based on χ^2 with $k-1$ degrees of freedom.

Table 3
Mean weighted effect sizes for patients' and healthy controls' neurocognitive performance at two different time points.

Variable	Studies (*k*)	Subjects	Follow-up	ES[a]	95% CI	*Z*[b]	*P*	*Q* test (*P*)[c]	I^2
Phonemic fluency	3	93 HC	2.28	0.00	−0.33 to 0.32	−0.03	0.98	0.30	18.01
		136 BD		0.12	−0.12 to 0.36	0.98	0.33	0.62	0.00
Backwards digit span	3	93 HC	2.28	0.01	−0.34 to 0.35	0.04	0.97	0.25	27.87
		136 BD		−0.07	−0.31 to 0.17	−0.56	0.58	0.64	0.00
Stroop (interference score)	3	66 HC	3.20	0.27	−0.08 to 0.62	1.52	0.13	0.36	2.35
		86 BD		0.15	−0.15 to 0.45	0.98	0.33	0.75	0.00
Global cognitive score	5	150 HC	2.28	0.16	−0.06 to 0.39	1.41	0.16	0.84	0.00
		218 BD		0.01	−0.18 to 0.20	0.09	0.93	0.48	0.00

[a] Effect size (*d*).
[b] Test of significance of effect size.
[c] Test of homogeneity, based on χ^2 with $k-1$ degrees of freedom.

performed, and the effect size distributions for BD patients were highly homogeneous. No between-group differences were observed with respect to test–retest effect sizes for any of the four variables (Fig. 2).

4. Discussion

The current study sought to quantify the magnitude of the differences between bipolar patients' cognitive performance at baseline and after a follow-up period in order to explore the longitudinal trajectory of neurocognitive functioning in BDs. To the best of our knowledge, this is the first approach to studying the longitudinal course of cognition in BDs by means of meta-analytic methods. Pooled standardized test–retest differences, with a mean retest interval of 4.62 years, were calculated for 14 cognitive variables, namely crystallized IQ, TMT_A, TMT_B, continuous performance test, phonemic fluency, category fluency, serial learning, delayed list recall, forward digit span, backward digit span, Stroop (interference score), psychomotor speed, WCST (perseverative errors), and global cognitive score. No significant test–retest effect sizes were observed for any of these variables, and the hypothesis of homogeneity was not rejected in any of the analyses performed. In 10 out of the 14 analyses the effect size distributions were highly homogeneous, which indicates that the same results could be obtained using either a fixed or a random effects model, thus supporting the robustness of our results. Furthermore, BD patients' test–retest effect sizes for four

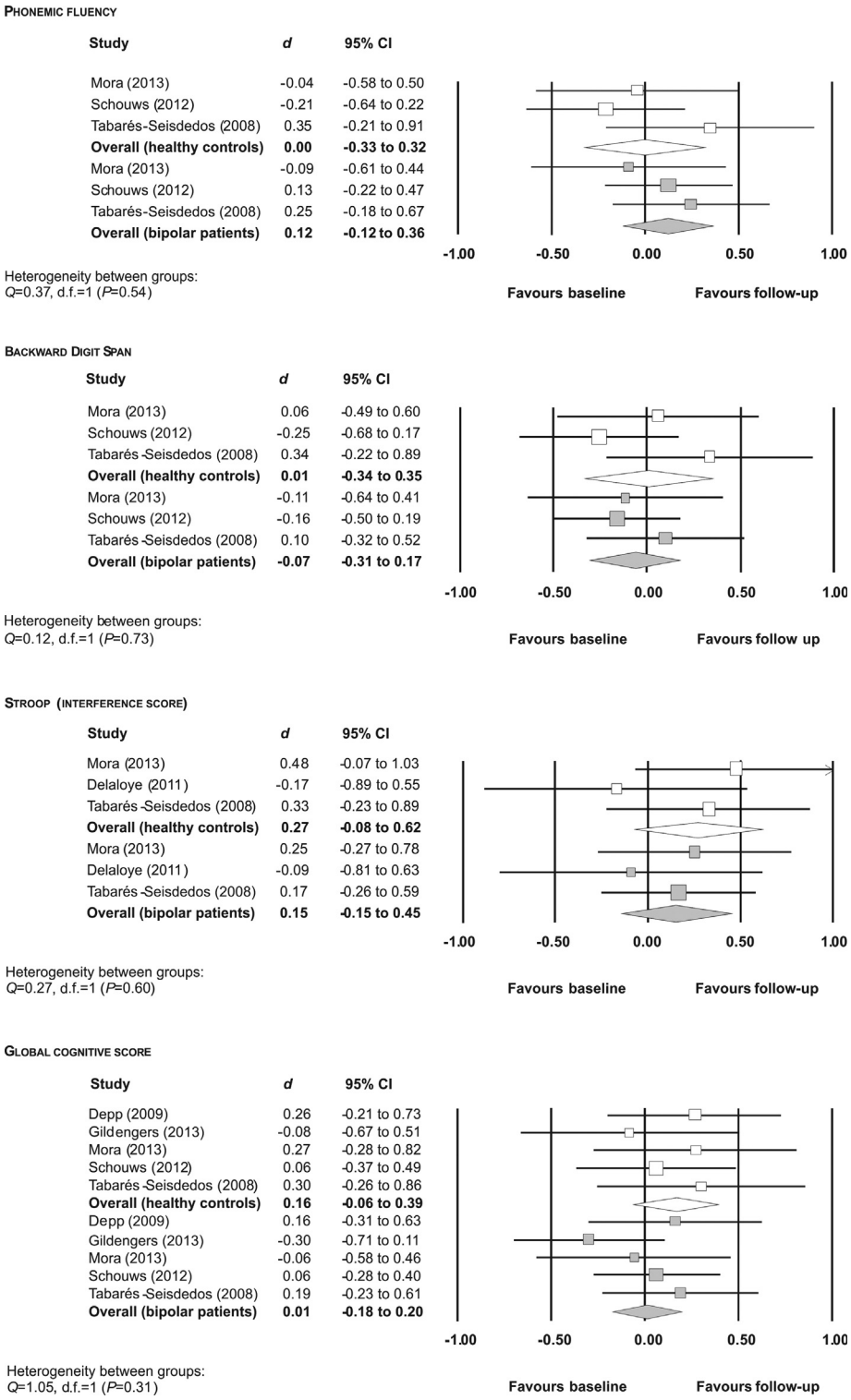

Fig. 2. Forest plots of individual and pooled patients' and controls' test–retest effect sizes for phonemic fluency, backward digit span, Stroop interference score, and global cognition. The area of each square reflects weighting from random effects analysis.

neuropsychological variables –backward digit span, phonemic fluency, Stroop interference score, and global cognition- were compared with those of healthy controls. No significant patient-control differences were found for any of these variables.

In spite of the former considerations, there are a number of limitations that should be considered when interpreting the results of this study. First, follow-up periods analyzed (2.18–4.62 years) were relatively short to assess the longitudinal trajectory of cognitive deficits. For instance, the only study yielding significant test–retest differences favoring performance at baseline for executive domains (Torrent et al., 2012) included patients with a much longer follow-up period in comparison to other studies. Such difference in neuropsychological test scores could be explained by a possible worsening of cognition as a result of illness progression or could else be attributed to the normal decrease of fluid abilities with age (Craik and Bialystok, 2006). Unfortunately, this

primary study did not include a healthy control group. Another shortcoming consists in the fact that differences in medication variables between assessment moments could not be controlled and may have influenced the results of this review. Maintenance of pharmacologic status over prolonged periods of time is difficult to accomplish in BD patients and may impact on cognition. Indeed, it has been proposed that prolonged exposure to lithium may have a protective effect on the risk of cognitive deterioration (Chuang and Manji, 2007; Nunes et al., 2007; Forlenza et al., 2011). Though this agent has been associated with neurotrophic properties in BD and in other neuropsychiatric disorders (Machado-Vieira et al., 2009), long-term effects on cognitive function have yet to be confirmed in a large controlled study of bipolar patients. By contrast, other psychotropic drugs, such as antipsychotics or benzodiazepines, might have deleterious effects on cognition (Donaldson et al., 2003; Frangou et al., 2005).

Another important limitation of the current review involves the fact that some of the reports reviewed did not provide mean scores for mood rating scales both at baseline and after the follow-up period. Therefore, we were not able to explore differences in mood state between time points that could have influenced our results. In addition, it is unclear whether the stability observed for tests scores was due to a genuine stability of neuropsychological functions in bipolar patients or better explained by the effect of repeated testing. It could be hypothesized that there was a true decline of cognitive functioning but it was masked by learning effects. However, with the exception of the WCST, which relies on a 'novelty effect', and the CVLT, practice effects are marginal after a test–retest interval of one year. The phenomenon known as 'Flynn effect' (Flynn, 1987; Ronnlund and Nilsson, 2009) could also be accounting for the stablity observed for test scores, particularly crystallized IQ and verbal memory, in the presence of genuine cognitive deterioration. Hence, data from adequately matched healthy controls are needed to rule out the influence of these artifacts in the interpretation of test–retest effect sizes for cognitive functions in BDs. In our study, we were only able to estimate controls' test–retest effect sizes for four out of the 14 variables explored and found no significant patient-control differences. However, it is worth noting the presence of between-group differences, both at single study and meta-analysis levels. For instance, significant differences for years of education were found favoring controls in three of the analyses performed. Meanwhile, in the largest study meta-analyzed (Schouws et al., 2012), the control group was significantly older than the patients group

Finally, the potential influence of attrition on the results of this study should be considered. As evident, the number of missing observations increases over time, with attrition rates ranging between 5 and 45% in the studies reviewed, thus raising concerns of drop-out bias. Nevertheless, some studies comparing patients who completed the study protocol and those who dropped out with respect to their baseline clinical, demographic, and cognitive characteristics did not find any differences (Torrent et al., 2012; Braw et al., 2013; Gildengers et al., 2013), except for one study that, indeed, found that completers were more likely to be unemployed (Braw et al., 2013). Hence, there is no evidence so far suggesting that patients who dropped out were those who would display worse cognitive outcome, though we cannot rule out this hypothesis.

Taking together these constraints, it is not possible to conclude from our findings alone that cognitive deficits in BD are stable across the course of the illness. Despite these limitations, which are essentially those of the primary studies, this work contributes to our knowledge on the longitudinal trajectory of cognitive deficits as a synthesis of the evidence available to date. Moreover, the results of this meta-analysis are in keeping with recent studies of elderly bipolar subjects in which the absence of differences with respect to young adult patients regarding neurocognitive performance did not support the hypothesis of a progressive decline of cognitive deficits (Samamé et al., 2013; Strejilevich and Martino, 2013). Furthermore, our results could contribute to future studies, for example, by providing a clue about the follow-up time needed to evaluate a potential cognitive decline. Future investigations should assess cognitive functions in larger samples of euthymic subjects for longer test–retest periods by means of a sensitive test battery and include healthy control groups in order to overcome the practice effect artifact and normal effects of aging. Likewise, other potential confounders such as medication status, subclinical mood state, or attrition should ideally be controlled in longitudinal studies. Beyond these methodological concerns, forthcoming studies should be able to take into account a number of theoretical issues. First, even in the presence of test–retest stability in controlled studies of bipolar patients, neuropsychological outcomes do not necessarily indicate that cognitive functioning is stable in BDs, given that a decline in cognitive functioning may have happened at the time of mood symptoms onset or at a pre-clinical stage and remain stable over the course of the illness. In fact, at present it is not clear in which moment the cognitive deficits reported in patients with BD develop. Therefore, some longitudinal studies should be able to focus on high-risk populations or patients with first episodes. Second, there is preliminary evidence suggesting that both the clinical course and cognitive deficits may be heterogeneous among patients with BD (Martino et al., 2008; Iverson et al., 2011; Baldessarini et al., 2012). If this were the case, the longitudinal course of cognitive impairments might be different in subgroups of patients (for example comparing patients with and without increasing rate of cicling or clinically significant cognitive impairment at baseline). Similarly, there is also preliminary evidence suggesting that patients with late-onset BD have more neurological comorbidity and cognitive deficits than those with early-onset of the disorder (Depp and Jeste, 2004; Schouws et al., 2009; Martino et al., 2013a). Then, future longitudinal studies should assess whether there are differences in the longitudinal course of cognitive deficits in these subgroups according to age at onset. Finally, It is worth noting that this review included only traditional neurocognitive tests, which may probably not capture subtle behavioral variations associated with cognitive change. Social cognitive domains, which have also been found to be impaired in euthymic bipolar subjects, such as emotion processing and theory of mind (Samamé et al., 2012; Samamé, 2013), have not been studied by means of longitudinal study designs either.

In summary, the results of this review indicate that there is not enough evidence so far for cognitive deficits being progressive in bipolar disorder. These findings, however, should be interpreted cautiously, considering the fact that there is still dearth of longitudinal studies of cognition in BDs and the methodological shortcomings that are common to research on this group of disorders. Future research in this field is warranted.

Role of funding source

Drs. Samamé and Martino report grants from the National Scientific and Technical Research Council (CONICET), Buenos Aires, Argentina, during the conduct of the study. Dr. Strejilevich reports personal fees and non-financial support from Abbott, GlaxoSmithKline, and AstraZeneca outside the submitted study.

Conflict of interest

The authors do not have any conflict of interest in relation to the contents and opinions expressed in this paper.

Acknowledgments

No acknowledgments.

References

Balanzá-Martínez, V., Tabarés-Seisdedos, R., Selva-Vera, G., Martínez Arán, A., Torrent, C., Salazar-Fraile, J., Leal-Cercós, C., Vieta, E., Gómez-Beneyto, M., 2005. Persistent cognitive dysfunctions in bipolar I disorder and schizophrenic patients: a 3-years follow-up study. Psychother. Psychosom. 74, 113–119.

Baldessarini, R.J., Salvatore, P., Khalsa, H.M.K., Imaz-Etxeberria, H., Gonzalez-Pinto, A., Tohen, M., 2012. Episode cycles with increasing recurrences in first-episode bipolar-I disorder patients. J. Affect. Disord. 136, 149–152.

Benton, A., Hamsher, K., Sivan, A., 1983. Multilingual Aphasia Examination, 3rd ed. AJA Associates, Iowa.

Berk, M., 2009. Neuroprogression: pathways to progressive brain changes in bipolar disorder. Int. J. Neuropsychopharmacol. 12, 441–445.

Bonnín, C.M., Martínez-Arán, A., Torrent, A., Pacchiarotti, I., Rosa, A.R., Franco, C., Murru, A., Sanchez-Moreno, J., Vieta, E., 2010. Clinical and neurocognitive predictors of functional outcome in bipolar euthymic patients: a long-term, follow-up study. J. Affect. Disord. 121, 156–160.

Borenstein, M., Hedges, L., Higgins, J., Rothstein, H., 2005. Comprehensive Meta Analysis, Version 2. Biostat, Englewood.

Braw, Y., Erez, G., Tal, S., Gvirts, Z.H., Harel, E.V., Bloch, Y., 2013. A longitudinal study of cognition in asymptomatic and mildly symptomatic bipolar disorder patients. Psychiatry Res. 210, 842–849.

Burdick, K.E., Goldberg, J.F., Harrow, M., Faull, R.N., Malhotra, A.K., 2006. Neurocognition as a stable endophenotype in bipolar disorder and schizophrenia. J. Nerv. Ment. Dis. 194, 255–260.

Chuang, D.M., Manji, H.K., 2007. In search of the Holy Grail for the treatment of neurodegenerative disorders: has a simple cation been overlooked? Biol. Psychiatry 62, 4–6.

Conners, C.K, Staff, MHS, 2000. Conners' Continuous Performance Test II. MHS, Toronto.

Craik, F.I.M., Bialystok, E., 2006. Cognition through the lifespan: mechanisms of change. Trends Cogn. Sci. 10, 131–138.

Delaloye, C., Moy, G., de Bilbao, F., Weber, K., Baudois, S., Haller, S., Xekardaki, A., Canuto, A., Giardini, U., Lovblad, K.O., Gold, G., Giannakopoulos, P., 2011. Longitudinal analysis of cognitive performances and structural brain changes in late-life bipolar disorder. Int. J. Geriatr. Psychiatry 26, 1309–1318.

Delis, D.C., Kramer, J.H., Kaplan, E., Ober, B.A., 1987. California Verbal Learning Test. The Psychological Corporation, San Antonio.

Depp, C.A., Jeste, D.V., 2004. Bipolar disorder in older adults: a critical review. Bipolar Disord. 6, 343–367.

Depp, C.A., Savla, G.N., Moore, D.J., Palmer, B.W., Stricker, J.L., Lebowitz, B.D., Jeste, D.V., 2008. Short-term course of neuropsychological abilities in middle-aged and older adults with bipolar disorder. Bipolar Disord. 10, 684–690.

Dickerson, F., Boronow, J.J., Stallings, G., Origoni, A.E., Cole, S.K., Yolken, R.H., 2004. Cognitive functioning in schizophrenia and bipolar disorder: comparison of performance on the Repeatable Battery for the Assessment of Neuropsychological Status. Psychiatry Res. 129, 45–53.

Donaldson, S., Goldstein, L., Landau, S., Raymont, V., Frangou, S., 2003. 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 64, 86–93.

Dunlap, W.P., Cortina, J.M., Vaslow, J.B., Burke, M.J., 1996. Meta-analysis of experiments with matched groups or repeated measure designs. Psychol. Methods 1, 170–177.

Flynn, J.R., 1987. Massive IQ gains in 14 nations: what IQ tests really measure. Psychol. Bull. 101, 171–191.

Forlenza, O.V., Diniz, B.S., Radanovic, M., Santos, F.S., Talib, L.L., Gattaz, W.F., 2011. Disease-modifying properties of long-term lithium treatment for amnestic mild cognitive impairment: randomised controlled trial. Br. J. Psychiatry 198, 351–356.

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

Fulford, D., Peckham, A.D., Johnson, K., Johnson, S.L., 2014. Emotion perception and quality of life in bipolar I disorder. J. Affect. Disord. 152, 491–497.

Gildengers, A.G., Chisholm, D., Butters, M.A., Anderson, S.J., Begley, A., Holm, M., Rogers, J.C., Reynolds 3rd, C.F., Mulsant, B.H., 2013. Two-year course of cognitive function and instrumental activities of daily living in older adults with bipolar disorder: evidence for neuroprogression? Psychol. Med. 30, 1–11.

Gildengers, A.G., Mulsant, B.H., Begley, A., Mazumdar, S., Hyams, A.V., Reynolds Iii, C.F., Kupfer, D.J., Butters, M.A., 2009. The longitudinal course of cognition in older adults with bipolar disorder. Bipolar Disord. 11, 744–752.

Gitlin, M.J., Swendsen, J., Heller, T.L., Hammen, C., 1995. Relapse and impairment in bipolar disorder. Am. J. Psychiatry 152, 1635–1640.

Golden, C.J., 1978. Stroop Color Word Test. Stroeting, Chicago.

Heaton, R., 1981. The Wisconsin Card Sorting Test Manual. Psychological Assessment Resources, Odessa, TX.

Hellvin, T., Sundet, K., Simonsen, C., Aminoff, S.R., Lagerberg, T.V., Andreassen, O.A., Melle, I., 2012. Neurocognitive functioning in patients recently diagnosed with bipolar disorder. Bipolar Disord. 14, 227–238.

Huxley, N., Baldessarini, R.J., 2007. Disability and its treatment in bipolar disorder patients. Bipolar Disord. 9, 183–196.

Iverson, G.L., Brooks, B.L., Langenecker, S.A., Young, A.H., 2011. Identifying a cognitive impairment subgroup in adults with mood disorders. J. Affect. Disord. 132, 360–367.

Jansen, K., Magalhães, P.V., Tavares Pinheiro, R., Kapczinski, F., Silva, R.A., 2012. Early functional impairment in bipolar youth: a nested population-based case-control study. J. Affect. Disord. 142, 208–212.

Kapczinski, F., Dias, V.V., Kauer-Sant'Anna, M., Brietzke, E., Vázquez, G.H., Vieta, E., Berk, M., 2009. The potential use of biomarkers as an adjunctive tool for staging bipolar disorder. Prog. Neuropsychopharmacol. Biol. Psychiatry 33, 1366–1371.

López-Jaramillo, C., Lopera-Vásquez, J., Gallo, A., Ospina-Duque, J., Bell, V., Torrent, C., Martínez-Arán, A., Vieta, E., 2010. Effects of recurrence on the cognitive performance of patients with bipolar I disorder: implications for relapse prevention and treatment adherence. Bipolar Disord. 12, 557–567.

Machado-Vieira, R., Manji, H.K., Zarate Jr., C.A., 2009. The role of lithium in the treatment of bipolar disorder: convergent evidence for neurotrophic effects as a unifying hypothesis. Bipolar Disord. 11, 92–109.

Mann-Wrobel, M.C., Carreno, J.T., Dickinson, D., 2011. Meta-analysis of neuropsychological functioning in euthymic bipolar disorder: an update and investigation of moderator variables. Bipolar Disord. 13, 334–342.

Martino, D.J., Marengo, E., Igoa, A., Scapola, M., Ais, E.D., Perinot, L., Strejilevich, S.A., 2009. Neurocognitive and symptomatic predictors of functional outcome in bipolar disorder: a prospective 1 year follow-up study. J. Affect. Disord. 116, 37–42.

Martino, D., Strejilevich, S., Manes, F., 2013a. Neurocognitive functioning in early-onset and late-onset older patients with euthymic bipolar disorder. Int. J. Geriatr. Psychiatry 28, 142–148.

Martino, D.J., Strejilevich, S.A., Marengo, E., Igoa, A., Fassi, G., Teitelbaum, J., Caravotta, P., 2013b. Relationship between neurocognitive functioning and episode recurrences in bipolar disorder. J. Affect. Disord. 147, 345–351.

Martino, D.J., Strejilevich, S.A., Scapola, M., Igoa, A., Marengo, E., Ais, E.D., Perinot, L., 2008. Heterogeneity in cognitive functioning among patients with bipolar disorder. J. Affect. Disord. 109, 149–156.

Moorhead, T.W., McKirdy, J., Sussmann, J.E., Hall, J., Lawrie, S.M., Johnstone, E.C., McIntosh, A.M., 2007. Progressive gray matter loss in patients with bipolar disorder. Biol. Psychiatry 62, 894–900.

Mora, E., Portella, M.J., Forcada, I., Vieta, E., Mur, M., 2013. 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. 43, 1187–1196.

Mur, M., Portella, M.J., Martínez-Arán, A., Pifarre, J., Vieta, E., 2008a. Long-term stability of cognitive impairment in bipolar disorder: a 2-year follow-up study of lithium-treated euthymic bipolar patients. J. Clin. Psychiatry 69, 712–719.

Mur, M., Portella, M.J., Martínez-Arán, A., Pifarré, J., Vieta, E., 2008b. Neuropsychological profile in bipolar disorder: preliminary study of monotherapy lithium-treated euthymic bipolar patients evaluated at a two-year interval. Acta Psychiatr. Scand. 118, 373–381.

Nelson, H.E., 1982. National Adult Reading Test (NART): Test Manual. NFER-Nelson, Windsor, United Kingdom.

Nunes, P.V., Forlenza, O.V., Gattaz, W.F., 2007. Lithium may reduce the risk for Alzheimer's disease in elderly patients with bipolar disorder. Br. J. Psychiatry 390, 359–360.

Post, R.M., Fleming, J., Kapczinski, F., 2012. Neurobiological correlates of illness progression in the recurrent affective disorders. J. Psychiatr. Res. 46, 561–573.

Reitan, R.M., 1958. Validity of the Trail Making test as an indication of organic brain damage. Percept. Mot. Skills 8, 271–276.

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

Robinson, L.J., Ferrier, I.N., 2006. Evolution of cognitive impairment in bipolar disorder: a systematic review of cross-sectional evidence. Bipolar Disord. 8, 103–116.

Robinson, L.J., Thompson, J.M., Gallagher, P., Goswami, U., Young, A.H., Ferrier, I.N., Moore, P.B., 2006. A meta-analysis of cognitive deficits in euthymic patients with bipolar disorder. J. Affect. Disord. 93, 105–115.

Ronnlund, M., Nilsson, L.G., 2009. Flynn effects on sub-factors of episodic and semantic memory: parallel gains over time and the same set of determining factors. Neuropsychologia 47, 2174–2180.

Samamé, C., 2013. Social cognition throughout the three phases of bipolar disorder: a state-of-the-art overview. Psychiatry Res. 210, 1275–1286.

Samamé, C., Martino, D.J., Strejilevich, S.A., 2012. Social cognition in euthymic bipolar disorder: systematic review and meta-analytic approach. Acta Psychiatr. Scand. 125, 266–280.

Samamé, C., Martino, D.J., Strejilevich, S.A., 2013. A quantitative review of neurocognition in late-life bipolar disorder. Bipolar Disord. 15, 633–644.

Schouws, S.N., Comijs, H.C., Stek, M.L., Dekker, J., Oostervink, F., Naarding, P., van der Velde, I., Beekman, A.T., 2009. Cognitive impairment in early and late bipolar disorder. Am. J. Geriatr. Psychiatry 17, 508–515.

Schouws, S.N., Stek, M.L., Comijs, H.C., Dols, A., Beekman, A.T., 2012. Cognitive decline in elderly bipolar disorder patients: a follow-up study. Bipolar Disord. 14, 749–755.

Strejilevich, S.A., Martino, D.J., 2013a. Cognitive function in adulthood and elderly euthymic bipolar patients: a comparison to test models of cognitive evolution. J. Affect. Disord. 150, 1188–1191.

Strejilevich, S.A., Martino, D.J., Murru, A., Teitelbaum, J., Fassi, G., Marengo, E., Igoa, A., Colom, F., 2013b. Mood instability and functional recovery in bipolar disorders. Acta Psychiatr. Scand. 128, 194–202.

Tabarés-Seisdedos, R., Balanzá-Martínez, V., Sánchez-Moreno, J., Martínez-Arán, A., Salazar-Fraile, J., Selva-Vera, G., Rubio, C., Mata, I., Gómez-Beneyto, M., Vieta, E., 2008. Neurocognitive and clinical predictors of functional outcome in patients

with schizophrenia and bipolar I disorder at one-year follow-up. J. Affect. Disord. 109, 286–299.
Tohen, M., Hennen, J., Zarate Jr, C.M., Baldessarini, R.J., Strakowski, S.M., Stoll, A.L., Faedda, G.L., Suppes, T., Gebre-Medhin, P., Cohen, B.M., 2000. Two-year syndromal and functional recovery in 219 cases of first-episode major affective disorder with psychotic features. Am. J. Psychiatry 157, 220–228.
Torrent, C., Martinez-Arán, A., Bonnin, C.M., Reinares, M., Daban, C., Solé, B., Rosa, A.R., Tabarés-Seisdedos, R., Popovic, D., Salamero, M., Vieta, E., 2012. Long-term outcome of cognitive impairment in bipolar disorder. J. Clin. Psychiatry 73, e899–905.
Torres, I.J., Boudreau, V.G., Yatham, L.N., 2007. Neuropsychological functioning in euthymic bipolar disorder: a meta-analysis. Acta Psychiatr. Scand. Suppl. 116, 17–26.
Wechsler, D.A., 1955. Wechsler Adult Intelligence Scale-Revised. Psychological Corporation, Cleveland.
Wechsler, D.A., 1997. WAIS-III: Wechsler Adult Intelligence Scale, Administration and Scoring Manual, 3rd ed. Psychological Corporation, San Antonio.
Yucel, K., McKinnon, M.C., Taylor, V.H., Macdonald, K., Alda, M., Young, L.T., MacQueen, G.M., 2007. Bilateral hippocampal volume increase in patients with bipolar disorder after long-term lithium treatment: a longitudinal MRI study. Psychopharmacology (Berl.) 195, 357–367.