TECHNICAL COMMENTARY

NeuRA Discover. Conquer. Cure.
BIPOLAR DISORDERS LIBRARY

Executive functioning

Introduction

Executive functions are a group of cognitive processes including control, mental flexibility, planning, inhibition, decision-making, initiation, abstraction, self-monitoring and pursuit of goals. Executive functions are important in situations involving error correction and behaviour evaluation in response to environmental feedback.

Executive functioning is most commonly measured using the Wisconsin Card Sorting Task (WCST). This task requires the ability to shift cognitive sets. Study participants are told to match stimulus cards containing varying coloured shapes, based first on colour, then quantity, then design. The participant is then given additional cards and asked to match each one without being told any matching rules, so participants usually match according to the previous rule. Feedback is provided as to whether their match was correct or incorrect, based on a new and undisclosed matching rule that changes during the task. Other common tasks assessing executive functioning include the Trail Making Test (TMT), which requires participants to connect, in order, letters and/or numbers as quickly as possible. Also, the Stroop Colour Word Test (SCWT), presents colour names printed in an ink congruent to the colour name (e.g. blue), or incongruent to the colour name (e.g. blue). Participants are asked to either read the word or name the ink colour. Any impairment in executive functioning can also reflect impairments in other cognitive functions such as processing speed, attention and memory.

Method

We have included only systematic reviews (systematic literature search, detailed methodology with inclusion/exclusion criteria) published in full text, in English, from the year 2010 that report results separately for people with a diagnosis of bipolar and related disorders. Reviews were identified by searching the databases MEDLINE, EMBASE, and PsycINFO. Hand searching reference lists of identified reviews was also conducted. When multiple copies of review topics were found, only the most recent and comprehensive review was included. Reviews with pooled data are prioritised for inclusion.

Review reporting assessment was guided by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist that describes a preferred way to present a meta-analysis[1]. Reviews with less than 50% of items checked have been excluded from the library. The PRISMA flow diagram is a suggested way of providing information about studies included and excluded with reasons for exclusion. Where no flow diagram has been presented by individual reviews, but identified studies have been described in the text, reviews have been checked for this item. Note that early reviews may have been guided by less stringent reporting checklists than the PRISMA, and that some reviews may have been limited by journal guidelines.

Evidence was graded using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) Working Group approach where high quality evidence such as that gained from randomised controlled trials (RCTs) may be downgraded to moderate or low if review and study quality is limited, if there is inconsistency in results, indirect comparisons, imprecise or sparse data and high probability of reporting bias. It may also be downgraded if risks associated with the intervention or other matter under review are high. Conversely, low quality evidence such as that gained from observational studies may be upgraded if effect sizes are large or if there is a dose dependent response. We have also taken into account sample size and whether results are consistent, precise and direct with low associated risks (see end of table for an explanation of these terms)[2]. The resulting table represents an objective summary of the available evidence, although the conclusions are solely the opinion

Margarete Ainsworth Building, Barker Street, Randwick NSW 2031. Phone: 02 9399 1000. Email: info@neura.edu.au

of staff of NeuRA (Neuroscience Research Australia).

Results

We found 11 systematic reviews that met our inclusion criteria[3-13].

- Moderate to high quality evidence finds a medium-sized effect of poorer executive functioning in people with bipolar I or II disorder compared to controls. Poor executive functioning was associated with poor general functioning.
- High quality evidence finds a small effect of poorer executive functioning in people with bipolar disorder and a history of psychotic symptoms compared to people with bipolar disorder and no history of psychotic symptoms. There were also small effects of poorer executive functioning in people with bipolar I disorder compared to people with bipolar II disorder, and in overweight people with bipolar disorder compared to normal weight people with bipolar disorder.
- Moderate quality evidence found no differences in executive functioning between people with bipolar disorder and people with major depression.
- High quality evidence finds a small effect of poorer performance on the Stroop test, but not the WCST, in young relatives of people with bipolar disorder compared to controls.

Bora E, Ozerdem A

A meta-analysis of neurocognition in youth with familial high risk for bipolar disorder

European Psychiatry 2017; 44: 17-23

View review abstract online

Comparison	**Executive functioning in first-degree relatives aged 10 to 25 years of a person with bipolar disorder vs. controls.**
Summary of evidence	**High quality evidence (large sample, consistent, precise, direct) suggests a small effect of poorer performance on the Stroop test in young relatives of people with bipolar disorder.**
Executive functioning	
Significant, small effect of poorer performance in executive functioning in young relatives of bipolar patients; 10 studies, N = 1,146, d = 0.15, 95%CI 0.01 to 0.30, p = 0.04, I^2 = 33%, p = 0.14 Subgroup analysis of individual tasks revealed significant differences were found on the Stroop test but not the WCST.	
Consistency	Consistent
Precision	Precise
Directness	Direct

Bora E

Neurocognitive features in clinical subgroups of bipolar disorder: A meta-analysis

Journal of Affective Disorders 2018; 229: 125-34

View review abstract online

Comparison 1	**Executive functioning in people with bipolar I disorder vs. bipolar II disorder.**
Summary of evidence	**High quality evidence (large samples, mostly consistent, precise, direct) suggests a small effect of poorer executive**

	functioning in people with bipolar I disorder.
Executive functioning	
Small, significant effects of poorer executive functioning in people with bipolar I disorder; Speed: 14 studies, N = 1,422, *d* = 0.16, 95%CI 0.00 to 0.32, *p* = 0.04, I^2 = 48%, *p* = 0.02 Accuracy: 11 studies, N = 984, *d* = 0.17, 95%CI 0.03 to 0.33, *p* = 0.02, I^2 = 0%, *p* = 0.53 Subgroup analysis of individual speeded tasks revealed significant differences were found in semantic fluency, but not in phonetic fluency, Stroop interference or TMT-B. Accuracy tests using the WCST categories were significant; WCST perseverative was not significant.	
Comparison 2	**Executive functioning in people with bipolar disorder and a history of psychotic symptoms vs. people with bipolar disorder and no history of psychotic symptoms.**
Summary of evidence	**High quality evidence (large samples, consistent, precise, direct) suggests a small significant effect of poorer executive functioning in people with bipolar disorder and a history of psychotic symptoms.**
Executive functioning	
Small, significant effects of poorer executive functioning in people with a history of psychosis; Speed: 13 studies, N = 1,209, *d* = 0.15, 95%CI 0.03 to 0.26, *p* = 0.01, I^2 = 0%, *p* = 0.54 Accuracy: 14 studies, N = 1,109, *d* = 0.20, 95%CI 0.06 to 0.35, *p* = 0.007, I^2 = 31%, *p* = 0.13 Subgroup analysis of individual speeded tasks revealed significant differences were found in semantic fluency, Stroop, and TMT-B, but not in phonetic fluency. Accuracy tests using the WCST were all significant.	
Consistency in results	Mostly consistent
Precision in results	Precise
Directness of results	Direct

Bora E, McIntyre RS, Ozerdem A

Neurococognitive and neuroimaging correlates of obesity and components of metabolic syndrome in bipolar disorder: a systematic

review

Psychological medicine 2019; 49: 738-49

View review abstract online

Comparison	**Executive functioning in overweight people with bipolar disorder vs. normal weight people with bipolar disorder.**
Summary of evidence	**High quality evidence (large samples, consistent, precise, direct) shows a medium-sized effect of poorer executive functioning in overweight patients compared to normal weight patients.**
Executive functioning	
A medium-sized effect showed overweight/obese patients were significantly associated with more impaired executive functioning; 5 studies, N = 330, d = 0.61, 95%CI 0.29 to 0.92, $p < 0.001$, I^2 = 40%	
Consistency in results	Consistent
Precision in results	Precise
Directness of results	Direct

Cotrena C, Damiani Branco L, Ponsoni A, Samame C, Milman Shansis F, Paz Fonseca R

Executive functions and memory in bipolar disorders I and II: new insights from meta-analytic results

Acta Psychiatrica Scandinavica 2020; 141: 110-30

View review abstract online

Comparison 1	**Executive functioning in people with bipolar I disorder vs. controls.**
Summary of evidence	**Moderate to high quality evidence (large samples, some inconsistency, precise, direct) shows a medium-sized effect of poorer executive functioning in people with bipolar I disorder.**
Executive functioning	

Medium-sized effects showed people with bipolar I disorder were more impaired on; Flexibility composite: 76 studies, N = 8,804, g = 0.52, 95%CI 0.38 to 0.66, $p < 0.05$, I^2 = 83%, $p < 0.001$ Planning composite: 14 studies, N = 2,600, g = 0.61, 95%CI 0.44 to 0.78, $p < 0.05$, I^2 = 62%, p = 0.0013 Inhibition composite: 55 studies, N = 5,294, g = 0.56, 95%CI 0.50 to 0.62, $p < 0.05$, I^2 = 25%, p = 0.05	
Consistency in results	Consistent for inhibition only.
Precision in results	Precise
Directness of results	Direct
Comparison 2	**Executive functioning in people with bipolar II disorder vs. controls.**
Summary of evidence	**Moderate to high quality evidence (large samples, mostly inconsistent, precise, direct) shows a small effect of poorer executive functioning in people with bipolar II disorder.**
Executive functioning	
Medium-sized effects showed people with bipolar II disorder were more impaired on; Flexibility composite: 15 studies, N = 1,419, g = 0.59, 95%CI 0.50 to 0.69, $p < 0.05$, I^2 = 0%, p = 0.47 Inhibition composite: 9 studies, N =826, g = 0.60, 95%CI 0.45 to 0.74, $p < 0.05$, I^2 = 3%, p = 0.41	
Consistency in results	Inconsistent, apart from inhibition.
Precision in results	Precise
Directness of results	Direct
Comparison 3	**Executive functioning in people with bipolar I disorder vs. people with bipolar II disorder.**
Summary of evidence	**Moderate to high quality evidence (large samples, mostly inconsistent, precise, direct) shows a small effect of poorer executive functioning in people with bipolar I disorder.**
Executive functioning	
Small effects showed people with bipolar I disorder were more impaired than people with bipolar II disorder on;	

Flexibility composite: 10 studies, N = 956, *g* = 0.22, 95%CI 0.03 to 0.41, *p* < 0.05, I^2 = 52%, *p* = 0.03 Trail making test B: 7 studies, N = 784, *g* = 0.28, 95%CI 0.03 to 0.52, *p* < 0.05, I^2 = 58%, *p* = 0.03 *There were no significant differences on;* Inhibition composite: 5 studies, N = 442, *g* = 0.09, 95%CI -0.09 to 0.27, I^2 = 0%, *p* = 0.45

Consistency in results	Inconsistent, apart from inhibition.
Precision in results	Precise
Directness of results	Direct

Depp CA, Mausbach BT, Harmell AL, Savla GN, Bowie CR, Harvey PD, Patterson TL

Meta-analysis of the association between cognitive abilities and everyday functioning in bipolar disorder

Bipolar Disorders 2012; 14: 217-26

View review abstract online

Comparison	**Associations between executive functioning and general daily functioning in people with bipolar disorder.**
Summary of evidence	**High quality evidence (large sample, consistent, precise, direct) suggests a small association between poorer executive functioning and poorer general functioning.**

Executive functioning
Significant, small association between poorer executive functioning and poorer general functioning; 11 studies, N = 759, *r* = 0.26, 95%CI 0.19 to 0.33, *p* < 0.0045, Q*p* = 0.545

Consistency in results	Consistent
Precision in results	Precise
Directness of results	Direct

Dickinson T, Becerra R, Coombes J

Executive functioning deficits among adults with Bipolar Disorder (types I and II): A systematic review and meta-analysis

Journal of Affective Disorders 2017; 218: 407-27

View review abstract online

Comparison	**Executive functioning in people with bipolar I or bipolar II disorder vs. controls.**
Summary of evidence	**Moderate quality evidence (large samples, inconsistent, some imprecision, direct) suggests poorer performance on executive functioning tasks in people with bipolar I disorder or bipolar II disorder compared to controls, with no differences between bipolar I and bipolar II disorder.**

Executive functioning

Small to medium-sized effects show poorer performance on overall executive functioning in people with bipolar I disorder or bipolar II disorder compared to controls;

Planning (bipolar I disorder): 20 studies, N < 3,538, *d* = 0.40, 95%CI 0.20 to 0.61, *p* not reported, I^2 = 81%

Planning (bipolar II disorder): 3 studies, N < 2,301, *d* = 0.75, 95%CI 0.08 to 1.39, *p* not reported, I^2 = 94%

Set-shifting (bipolar II disorder): 20 studies, N not reported, *d* = 0.38, 95%CI 0.15 to 0.60, *p* not reported, I^2 = 80%

No differences in overall executive functioning between bipolar I disorder and bipolar II disorder;

Planning: 10 studies, N not reported, *d* = -0.01, 95%CI −0.13 to 0.11, *p* not reported, I^2 not reported

Set-shifting: 10 studies, N not reported, *d* = -0.09, 95%CI −0.15 to 0.33, *p* not reported, I^2 = 51%, *p* < 0.10

Consistency in results	Inconsistent
Precision in results	Precise, apart from planning in bipolar II disorder.
Directness of results	Direct

Hajek T, Alda M, Hajek E, Ivanoff J

Functional neuroanatomy of response inhibition in bipolar disorders - Combined voxel based and cognitive performance meta-analysis

Journal of Psychiatric Research 2013; 47: 1955-66 View online review abstract	
Comparison	**Response inhibition in people with bipolar disorder vs. controls.**
Summary of evidence	**Moderate to high quality evidence (large sample, unable to assess consistency, precise, direct) suggests people with bipolar disorder showed poorer response inhibition than controls, particular patients in a manic phase.**
Response inhibition	
A significant, small effect of poorer response inhibition in people with bipolar disorder; 28 studies, N = 1,203, *d* = 0.32, 95%CI 0.14 to 0.49, *p* = 0.0004, I^2 not reported *The effect was similar in the subgroup analysis of mania patients, but was not significant in euthymic patients;* Mania: 10 studies, N = 298, *d* = 0.40, 95%CI 0.15 to 0.65, *p* = 0.002 Euthymia: 12 studies, N = 604, *d* = 0.09, 95%CI -0.18 to 0.35, *p* = 0.51 There were not enough studies of patients with depression for a meta-analysis.	
Consistency in results	Unable to assess; no measure of consistency is reported.
Precision in results	Precise
Directness of results	Direct

Samame C, Martino DJ, Strejilevich SA

A quantitative review of neurocognition in euthymic late-life bipolar disorder

Bipolar Disorders 2013; 15: 633-44

View review abstract online

Comparison	**Executive functioning in older people with bipolar disorder vs. controls matched for age and years of education.**
Summary of evidence	**Moderate to high quality evidence (medium-sized samples, consistent, precise, direct) suggests a large effect of poorer executive functioning in elderly people with bipolar disorder.**
Executive functioning	

Large, significant effects of poorer executive functioning in elderly people with bipolar disorder; Cognitive flexibility (TMT-B): 3 studies, N = 301, g = 0.88, 95%CI 0.64 to 1.12, p < 0.001, I^2 = 0%, p = 0.86 Digit span backwards: 3 studies, N = 301, g = 0.77, 95%CI 0.53 to 1.01, p < 0.001, I^2 = 0%, p = 0.88 Subgroup analyses showed no changes in the effect sizes according to age or years of education.

Consistency in results	Consistent
Precision in results	Precise
Directness of results	Direct

Samame C, Szmulewicz AG, Valerio MP, Martino DJ, Strejilevich SA

Are major depression and bipolar disorder neuropsychologically distinct? A meta-analysis of comparative studies

European Psychiatry 2017; 39: 17-26

View review abstract online

Comparison	**Executive functioning in people with bipolar disorder vs. people with major depression.**
Summary of evidence	**Moderate quality evidence (medium to large samples, some inconsistencies and imprecision, direct) suggests no differences on tasks of executive functioning.**

Executive functioning

There were no significant differences between groups;

Euthymia

Trail-making test B: 5 studies, N = 383, g = -0.06, 95%CI -0.55 to 0.43, p = 0.82, I^2 = 78%, p < 0.001

Response inhibition: 4 studies, N = 197, g = 0.08, 95%CI -0.30 to 0.45, p = 0.69, I^2 = 38%, p = 0.18

Cognitive flexibility: 3 studies, N = 274, g = 0.10, 95%CI -0.13 to 0.34, p = 0.39, I^2 = 0%, p = 0.76

Depression

Trail-making test B: 3 studies, N = 540, g = 0.65, 95%CI -0.59 to 1.90, p = 0.30, I^2 = 94%, p < 0.001

Planning: 4 studies, N = 598, g = 0.06, 95%CI -0.10 to 0.22, p = 0.46, I^2 = 0%, p = 0.61

Response inhibition: 3 studies, N = 105, g = 0.69, 95%CI -0.59 to 1.89, p = 0.26, I^2 = 88%, p < 0.001

Cognitive flexibility: 6 studies, N = 816, *g* = 0.39, 95%CI -0.15 to 0.92, *p* = 0.16, I^2 = 89%, *p* < 0.001 Phonological fluency: 3 studies, N = 107, *g* = 0.73, 95%CI -0.18 to 1.64, *p* = 0.12, I^2 = 80%, *p* = 0.007 Backward digit span: 3 studies, N = 680, *g* = -0.05, 95%CI -0.24 to 0.13, *p* = 0.57, I^2 = 15%, *p* = 0.31 Spatial span: 3 studies, N = 278, *g* = 0.11, 95%CI -0.15 to 0.37, *p* = 0.40, I^2 = 0%, *p* = 0.45

Consistency in results	Consistent, apart fromTMT-B (euthymia and depression), response inhibition (depression), and cognitive flexibility (depression).
Precision in results	Precise, apart from TMT-B (depression), and response inhibition (depression).
Directness of results	Direct

Samame C, Martino DJ, Strejilevich SA

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

Journal of Affective Disorders 2014; 164: 130-8

View review abstract online

Comparison	**Changes in executive functioning over time in people with bipolar disorder.**
Summary of evidence	**Moderate quality evidence (small samples, consistent, precise, direct) suggests no changes in measures of executive functioning over time (~4-7 years).**

Executive functioning
There were no significant changes over time; TMT-B: 5 studies, N = 169, follow up = 4.87 years, *d* = -0.19, 95%CI -0.47 to 0.10, *p* = 0.19, I^2 = 37%, *p* = 0.18 Stroop (interference): 4 studies, N = 131, follow up = 4.90 years, *d* = 0.01, 95%CI -0.26 to 0.29, *p* = 0.93, I^2 = 0%, *p* = 0.64 Backward digit span: 4 studies, N = 181, follow up = 4.10 years, *d* = -0.08, 95%CI -0.28 to 0.11, *p* = 0.41, I^2 = 0%, *p* = 0.92 WCST: 4 studies, N = 104, follow up = 6.67 years, *d* = 0.09, 95%CI -0.22 to 0.41, *p* = 0.55, I^2 = 37%, *p* = 0.17

Consistency in results	Consistent
Precision in results	Precise
Directness of results	Direct

Wang K, Song LL, Cheung EFC, Lui SSY, Shum DHK, Chan RCK

Bipolar disorder and schizophrenia share a similar deficit in semantic inhibition: A meta-analysis based on hayling sentence completion test performance

Progress in Neuro-Psychopharmacology and Biological Psychiatry 2013; 46: 153-60

View review abstract online

Comparison	**Semantic inhibition in people with bipolar disorder vs. controls.**
Summary of evidence	**Moderate to low quality evidence (small to medium-sized samples, inconsistent, some imprecision, direct) suggests medium to large effects of poorer semantic inhibition in people with bipolar disorder compared to controls.**

Semantic inhibition

Significant, medium to large effects of poor semantic inhibition in bipolar disorder;

Total latency of task A: 6 studies, N = 341, d = 0.719, 95%CI 0.231 to 1.207, $p < 0.05$, Q$p < 0.01$

Total latency of task B: 5 studies, N = 253, d = 0.930, 95%CI 0.403 to 1.457, $p < 0.05$, Q$p < 0.05$

Total error of task B: 5 studies, N = 253, d = 0.866, 95%CI 0.402 to 1.330, $p < 0.05$, Q$p < 0.05$

Type A error of task B: 2 studies, N = 146, d = 0.678, 95%CI 0.336 to 1.021, $p < 0.05$, Q$p < 0.05$

No significant differences on;

Type B error of task B: 2 studies, N = 146, d = 0.869, 95%CI -0.472 to 2.211, $p > 0.05$, Q$p < 0.05$

Suppression time: 4 studies, N = 218, d = 0.156, 95%CI 0.240 to -0.313, $p > 0.05$, Q$p < 0.05$

Consistency in results	Inconsistent
Precision in results	Precise, apart from Type B Error of Task B.
Directness of results	Direct

Explanation of acronyms

CI = confidence interval, CPT = continuous performance test, *d* = Cohen's *d* and *g* = Hedges' *g* standardised mean difference, I^2 = the percentage of the variability in effect estimates that is due to heterogeneity rather than sampling error (chance), N = number of participants, *p* = probability of rejecting a null hypothesis of no differences between groups, Q = test for heterogeneity, *r* = correlation coefficient, TMT = trail-making test, WCST = Wisconsin Card Sorting Test, vs. = versus

Margarete Ainsworth Building, Barker Street, Randwick NSW 2031. Phone: 02 9399 1000. Email: info@neura.edu.au

Explanation of technical terms

* Bias has the potential to affect reviews of both RCT and observational studies. Forms of bias include; reporting bias – selective reporting of results; publication bias - trials that are not formally published tend to show less effect than published trials, further if there are statistically significant differences between groups in a trial, these trial results tend to get published before those of trials without significant differences; language bias – only including English language reports; funding bias - source of funding for the primary research with selective reporting of results within primary studies; outcome variable selection bias; database bias - including reports from some databases and not others; citation bias - preferential citation of authors. Trials can also be subject to bias when evaluators are not blind to treatment condition and selection bias of participants if trial samples are small[14].

† Different effect measures are reported by different reviews.

Prevalence refers to how many existing cases there are at a particular point in time. Incidence refers to how many new cases there are per population in a specified time period. Incidence is usually reported as the number of new cases per 100,000 people per year. Alternatively some studies present the number of new cases that have accumulated over several years against a person-years denominator. This denominator is the sum of individual units of time that the persons in the population are at risk of becoming a case. It takes into account the size of the underlying population sample and its age structure over the duration of observation.

Reliability and validity refers to how accurate the instrument is. Sensitivity is the proportion of actual positives that are correctly identified (100% sensitivity = correct identification of all actual positives) and specificity is the proportion of negatives that are correctly identified (100% specificity = not identifying anyone as positive if they are truly not).

Weighted mean difference scores refer to mean differences between treatment and comparison groups after treatment (or occasionally pre to post treatment) and in a randomised trial there is an assumption that both groups are comparable on this measure prior to treatment. Standardised mean differences are divided by the pooled standard deviation (or the standard deviation of one group when groups are homogenous) that allows results from different scales to be combined and compared. Each study's mean difference is then given a weighting depending on the size of the sample and the variability in the data. Less than 0.4 represents a small effect, around 0.5 a medium effect, and over 0.8 represents a large effect[14].

Odds ratio (OR) or relative risk (RR) refers to the probability of a reduction (< 1) or an increase (> 1) in a particular outcome in a treatment group, or a group exposed to a risk factor, relative to the comparison group. For example, a RR of 0.75 translates to a reduction in risk of an outcome of 25% relative to those not receiving the treatment or not exposed to the risk factor. Conversely, a RR of 1.25 translates to an increased risk of 25% relative to those not receiving treatment or not having been exposed to a risk factor. A RR or OR of 1.00 means there is no difference between groups. A medium effect is considered if RR > 2 or < 0.5 and a large effect if RR > 5 or < 0.2[15]. lnOR stands for logarithmic OR where a lnOR of 0 shows no difference between groups. Hazard ratios measure the effect of an explanatory variable on the hazard or risk of an event.

Correlation coefficients (eg, r) indicate the strength of association or relationship

between variables. They can provide an indirect indication of prediction, but do not confirm causality due to possible and often unforseen confounding variables. An r of 0.10 represents a weak association, 0.25 a medium association and 0.40 and over represents a strong association. Unstandardised (*b*) regression coefficients indicate the average change in the dependent variable associated with a 1 unit change in the independent variable, statistically controlling for the other independent variables. Standardised regression coefficients represent the change being in units of standard deviations to allow comparison across different scales.

‡ Inconsistency refers to differing estimates of effect across studies (i.e. heterogeneity or variability in results) that is not explained by subgroup analyses and therefore reduces confidence in the effect estimate. I^2 is the percentage of the variability in effect estimates that is due to heterogeneity rather than sampling error (chance) - 0% to 40%: heterogeneity might not be important, 30% to 60%: may represent moderate heterogeneity, 50% to 90%: may represent considerable heterogeneity and over this is considerable heterogeneity. I^2 can be calculated from Q (chi-square) for the test of heterogeneity with the following formula[14];

$$I^2 = \left(\frac{Q - df}{Q}\right) \times 100\%$$

§ Imprecision refers to wide confidence intervals indicating a lack of confidence in the effect estimate. Based on GRADE recommendations, a result for continuous data (standardised mean differences, not weighted mean differences) is considered imprecise if the upper or lower confidence limit crosses an effect size of 0.5 in either direction, and for binary and correlation data, an effect size of 0.25. GRADE also recommends downgrading the evidence when sample size is smaller than 300 (for binary data) and 400 (for continuous data), although for some topics, these criteria should be relaxed[16].

‖ Indirectness of comparison occurs when a comparison of intervention A versus B is not available but A was compared with C and B was compared with C that allows indirect comparisons of the magnitude of effect of A versus B. Indirectness of population, comparator and/or outcome can also occur when the available evidence regarding a particular population, intervention, comparator, or outcome is not available and is therefore inferred from available evidence. These inferred treatment effect sizes are of lower quality than those gained from head-to-head comparisons of A and B.

References

1. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMAGroup (2009): Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *British Medical Journal* 151: 264-9.
2. GRADEWorkingGroup (2004): Grading quality of evidence and strength of recommendations. *British Medical Journal* 328: 1490.
3. Bora E, Ozerdem A (2017): A meta-analysis of neurocognition in youth with familial high risk for bipolar disorder. *European Psychiatry* 44: 17-23.
4. Bora E (2018): Neurocognitive features in clinical subgroups of bipolar disorder: A meta-analysis. *Journal of Affective Disorders* 229: 125-34.
5. Depp CA, Mausbach BT, Harmell AL, Savla GN, Bowie CR, Harvey PD*, et al.* (2012): Meta-analysis of the association between cognitive abilities and everyday functioning in bipolar disorder. *Bipolar Disorders* 14: 217-26.
6. Samame C, Szmulewicz AG, Valerio MP, Martino DJ, Strejilevich SA (2017): Are major depression and bipolar disorder neuropsychologically distinct? A meta-analysis of comparative studies. *European Psychiatry* 39: 17-26.
7. Samame C, Martino DJ, Strejilevich SA (2013): A quantitative review of neurocognition in euthymic late-life bipolar disorder. *Bipolar Disorders* 15: 633-44.
8. Samame C, Martino DJ, Strejilevich SA (2014): Longitudinal course of cognitive deficits in bipolar disorder: a meta-analytic study. *Journal of Affective Disorders* 164: 130-8.
9. Dickinson T, Becerra R, Coombes J (2017): Executive functioning deficits among adults with Bipolar Disorder (types I and II): A systematic review and meta-analysis. *Journal of Affective Disorders* 218: 407-27.
10. Wang K, Song LL, Cheung EFC, Lui SSY, Shum DHK, Chan RCK (2013): Bipolar disorder and schizophrenia share a similar deficit in semantic inhibition: A meta-analysis based on hayling sentence completion test performance. *Progress in Neuro-Psychopharmacology and Biological Psychiatry* 46: 153-60.
11. Hajek T, Alda M, Hajek E, Ivanoff J (2013): Functional neuroanatomy of response inhibition in bipolar disorders - Combined voxel based and cognitive performance meta-analysis. *Journal of Psychiatric Research* 47: 1955-66.
12. Bora E, McIntyre RS, Ozerdem A (2019): Neurococognitive and neuroimaging correlates of obesity and components of metabolic syndrome in bipolar disorder: a systematic review. *Psychological medicine* 49: 738-49.
13. Cotrena C, Damiani Branco L, Ponsoni A, Samame C, Milman Shansis F, Paz Fonseca R (2020): Executive functions and memory in bipolar disorders I and II: new insights from meta-analytic results. *Acta Psychiatrica Scandinavica* 141: 110-30.
14. CochraneCollaboration (2008): Cochrane Handbook for Systematic Reviews of Interventions. Accessed 24/06/2011.
15. Rosenthal JA (1996): Qualitative Descriptors of Strength of Association and Effect Size. *Journal of Social Service Research* 21: 37-59.
16. GRADEpro (2008): [Computer program]. Jan Brozek, Andrew Oxman, Holger Schünemann. *Version 32 for Windows*