TECHNICAL COMMENTARY

Cognition in childhood bipolar disorder

NeuRA Discover. Conquer. Cure.
BIPOLAR DISORDERS LIBRARY

Introduction

Deficits across various cognitive domains are a common feature of bipolar disorder and are strongly associated with difficulties in activities of daily living. Early age at onset of the illness is associated with more severe symptoms and poor prognosis than later age at onset, and cognitive deficits in children with the disorder may also differ from those observed in older patients. Identifying cognitive deficits in children contributes to the development of specific treatments and rehabilitation approaches.

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 or related disorders. Due to the high volume of systematic reviews we have now limited inclusion to systematic meta-analyses. Where no systematic meta-analysis exists for a topic, systematic reviews without meta-analysis are included for that topic. 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 reviews assessing the same topic were found, only the most recent and/or comprehensive review was included.

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 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, there is a dose dependent response or if results are reasonably 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 of staff of NeuRA (Neuroscience Research Australia).

Results

We found three systematic reviews that met our inclusion criteria[3-5].

- Moderate quality evidence suggests large impairments in global cognition, verbal and visual learning and memory and working memory in youth with bipolar disorder. There were no impairments found in attention/vigilance, reasoning and problem solving, and processing speed.
- High quality evidence finds a medium to large effect of poorer emotion recognition, and moderate quality evidence finds a large effect of poorer theory of mind in children with bipolar disorder compared to age-matched controls.

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

Cognition in childhood bipolar disorder

- Moderate quality evidence finds a medium to large effect of poorer accuracy on emotion recognition in children with bipolar disorder compared to age-matched controls. There was a smaller, non-significant effect of poorer response time. Unmedicated children showed longer response times than medicated children. Caucasian children showed both longer response time and poorer accuracy than non-Caucasian children.

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

Elias LR, Miskowiak KW, Vale AM, Kohler CA, Kjaerstad HL, Stubbs B, Kessing LV, Vieta E, Maes M, Goldstein BI, Carvalho AF

Cognitive Impairment in Euthymic Pediatric Bipolar Disorder: A Systematic Review and Meta-Analysis

Journal of the American Academy of Child & Adolescent Psychiatry 2017; 56: 286-96

View review abstract online

Comparison	**Cognition in euthymic youth with bipolar disorder vs. controls of similar age (mean 13 years) and IQ (mean 104).**
Summary of evidence	**Moderate quality evidence (small to medium-sized samples, mostly inconsistent, mostly precise, direct) suggests large impairments in global cognition, verbal and visual learning and memory and working memory in youth with bipolar disorder. There were no impairments in attention/vigilance, reasoning and problem solving, and processing speed.**

Cognition

Large, significant effects of more impairment in euthymic youth with bipolar disorder in;

Global cognition: 7 studies, N = 342, g = 0.78, 95%CI 0.16 to 1.41, p = 0.014, I^2 = 85%, p < 0.05

Verbal learning and memory: 7 studies, N = 401, g = 0.76, 95%CI 0.29 to 1.22, p = .001, I^2 = 76%, p < 0.05

Working memory: 8 studies, N = 433, g = 0.99, 95%CI 0.64 to 1.35, p < 0.001, I^2 = 62%, p < 0.05

Visual learning and memory: 4 studies, N = 135, g = 0.78, 95%CI 0.43 to 1.13, p < 0.001, I^2 = 0%, p > 0.05

No significant impairments were observed for attention/vigilance, reasoning and problem solving, and processing speed.

Authors report that differences in the definition of euthymia across studies explained the heterogeneity in the effect size estimate for verbal learning and memory.

Co-occurring attention-deficit/hyperactivity disorder, anxiety disorders, the use of medications, and the use of different neuropsychological tests also contributed to heterogeneity across study results on some estimates.

Consistency in results	Inconsistent, apart from visual learning and memory.
Precision in results	Precise, apart from global cognition.
Directness of results	Direct

Halac E, Ciray RO, Turan S, Tuncturk M, Agac N, Elmas FN, Rosson S, Ermis C

Impaired theory of mind and emotion recognition in pediatric bipolar disorder: A systematic review and meta-analysis

Journal of Psychiatric Research 2021; 138: 246-55

View review abstract online

Comparison	**Theory of mind and emotion recognition in children with bipolar disorder vs. age-matched controls.** **Mean age = 13.6 years. 85% of the sample had bipolar disorder I.**
Summary of evidence	**High quality evidence (large sample, consistent, precise, direct) finds a medium to large effect of poorer emotion recognition in children with bipolar disorder. Moderate quality evidence (small sample, consistent, imprecise, direct) finds a large effect of poorer theory of mind.**
Theory of mind	
A significant, large effect of poorer theory of mind in children with bipolar disorder; 3 studies, N = 156, g = -0.98, 95%CI -1.41 to -0.55, $p < 0.001$, $I^2 = 39\%$	
Emotion recognition	
A significant, medium to large effect of poorer emotion recognition in children with bipolar disorder; 8 studies, N = 541, g = -0.74, 95%CI -0.91 to -0.57, $p < 0.001$, $I^2 = 0\%$ Meta-regressions showed no moderating effects of age, gender, sample size, the severity of mood symptoms, estimated IQ, the frequencies of bipolar-I disorder, attention-deficit hyperactivity disorder, medications, study quality and euthymia.	
Consistency in results	Consistent
Precision in results	Precise for emotion recognition
Directness of results	Direct

Khafif TC, Rotenberg LDS, Nascimento C, Beraldi GH, Lafer B

Emotion regulation in pediatric bipolar disorder: A meta-analysis of published studies

Journal of Affective Disorders 2021; 285: 86-96 View review abstract online	
Comparison	**Emotion recognition in children with bipolar disorder vs. age-matched controls.** **Mean age = 13.6 years. 85% of the sample had bipolar disorder I.**
Summary of evidence	**Moderate quality evidence (medium-sized sample, inconsistent, imprecise, direct) finds a medium to large effect of poorer accuracy on emotion recognition in children with bipolar disorder. There was a smaller, trend effect of poorer response time. Unmedicated patients showed longer response times than medicated patients. Caucasian children showed both longer response time and poorer accuracy than non-Caucasian children.**
Emotion recognition	
A significant, medium to large effect of poorer accuracy on emotion recognition tasks in children with bipolar disorder; Accuracy: 7 studies, N = 331, *g* = -0.75, 95%CI -1.18 to -0.33, *p* < 0.001, I^2 = 71% *A trend effect of poorer response time on emotion recognition tasks in children with bipolar disorder;* Response time: 8 studies, N = 351, *g* = 0.38, 95%CI -0.01 to 0.77, *p* = 0.057, I^2 = 67% Unmedicated patients showed longer response time than medicated patients (both compared to controls). There was no moderating effect of medication on accuracy measures. Caucasian children showed longer response time and poorer accuracy than non-Caucasians. There were no moderating effects of age, gender, IQ, or study quality.	
Consistency in results	Inconsistent
Precision in results	Imprecise
Directness of results	Direct

Explanation of acronyms

CI = Confidence interval, *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, 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 tath 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[6].

---

† Different effect measures are reported by different reviews.

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. 0.2 represents a small effect, 0.5 a medium effect, and 0.8 and over represents a large treatment effect[6].

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, an 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. An 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[7]. 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.

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).

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

between variables. They are an 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 dependent 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 treatment 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 substantial heterogeneity and 75% to 100%: considerable heterogeneity. I^2 can be calculated from Q (chi-square) for the test of heterogeneity with the following formula[6];

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

§ 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, this criteria should be relaxed[8].

‖ 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 so is 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.

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

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. Halac E, Ciray RO, Turan S, Tuncturk M, Agac N, Elmas FN*, et al.* (2021): Impaired theory of mind and emotion recognition in pediatric bipolar disorder: A systematic review and meta-analysis. *Journal of Psychiatric Research* 138: 246-55.
4. Khafif TC, Rotenberg LDS, Nascimento C, Beraldi GH, Lafer B (2021): Emotion regulation in pediatric bipolar disorder: A meta-analysis of published studies. *Journal of Affective Disorders* 285: 86-96.
5. Elias LR, Miskowiak KW, Vale AM, Kohler CA, Kjaerstad HL, Stubbs B*, et al.* (2017): Cognitive Impairment in Euthymic Pediatric Bipolar Disorder: A Systematic Review and Meta-Analysis. *Journal of the American Academy of Child & Adolescent Psychiatry* 56: 286-96.
6. CochraneCollaboration (2008): Cochrane Handbook for Systematic Reviews of Interventions. Accessed 24/06/2011.
7. Rosenthal JA (1996): Qualitative Descriptors of Strength of Association and Effect Size. *Journal of Social Service Research* 21: 37-59.
8. GRADEpro (2008): [Computer program]. Jan Brozek, Andrew Oxman, Holger Schünemann. *Version 32 for Windows*