

Language

Introduction

Language may be altered in people with bipolar disorder and may present in the form of disorganised speech. Tasks designed to assess language ability include; letter fluency tasks that assess the ability to generate words starting with a particular letter; and category fluency tasks that assess the ability to name words within a specified category. Working memory is needed for both letter and category fluency as participants must organise and retrieve relevant information.

Other tests designed to assess language include: Boston Naming task; Wechsler Adult Intelligence Scale (WAIS) comprehension (including the subtest information, similarities and vocabulary), WAIS verbal memory, verbal fluency, National Adult Reading Test (NART)/Wide Range Achievement Test (WRAT), Controlled Oral Word Association Test (COWA), Category Instance Generation Test (CIGT), Multiple Choice Vocabulary Test (MWT-B), Hopkins Verbal Learning test (HVLt), California Verbal Learning Test (CVLT), Rey Auditory Verbal Learning Test (AVLT), semantic priming tasks and Lexical Decision Task and the Peabody Individual Achievement reading comprehension (PIAT).

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-analysis¹. 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 of staff of NeuRA (Neuroscience Research Australia).

Results

We found eight systematic reviews that met our inclusion criteria³⁻¹⁰.

Language

- Moderate to high quality evidence shows a medium-sized effect of poorer language fluency in people with bipolar I disorder than controls. There was also a medium-sized effect of poorer language fluency in people with bipolar II disorder compared to controls. In direct comparison between bipolar I disorder and bipolar II disorder, those with bipolar I disorder showed a small effect of poorer language fluency.
- High quality evidence shows a small effect of poorer language fluency in people with first-episode bipolar disorder compared to controls. Moderate to high quality evidence suggests a medium-sized effect of better fluency in people with first-episode bipolar disorder compared to people with first-episode schizophrenia.
- High quality evidence suggests a medium to large effect of poorer language fluency in elderly people with bipolar disorder compared to controls matched for age and education.
- High quality evidence suggests a small to medium-sized effect of poorer language fluency in first-degree relatives of people with bipolar disorder compared to controls. There was better fluency in first-degree relatives of people with bipolar disorder than first-degree relatives of people with schizophrenia.
- High quality evidence suggests a small association between poor verbal ability/fluency and poor general functioning.
- Moderate quality evidence suggests no changes in performance on fluency tasks over time (~3-4 years) in people with bipolar disorder.

Language

Bora E, Pantelis C

Meta-analysis of Cognitive Impairment in First-Episode Bipolar Disorder: Comparison With First-Episode Schizophrenia and Healthy Controls

Schizophrenia Bulletin 2015; 41(5): 1095-1104

[View review abstract online](#)

Comparison 1	Fluency in people with first-episode bipolar disorder vs. controls.
Summary of evidence	High quality evidence (large sample, consistent, precise, direct) suggests a small effect of poorer fluency in people with first-episode bipolar disorder.
Fluency	
<p><i>A significant, small effect of poorer fluency in people with first-episode bipolar disorder;</i> 9 studies, N = 1,280, $d = 0.36$, 95%CI 0.17 to 0.55, $p < 0.001$, $I^2 = 47%$, $p = 0.06$ Authors report no evidence of publication bias.</p>	
Consistency in results	Consistent
Precision in results	Precise
Directness of results	Direct
Comparison 2	Fluency in people with first-episode bipolar disorder vs. people with first-episode schizophrenia.
Summary of evidence	Moderate to high quality evidence (large sample, consistent, mostly precise, direct) shows a medium-sized effect of poorer fluency in people with first-episode schizophrenia compared to people with first-episode bipolar disorder.
Fluency	
<p><i>A significant, medium-sized effect of poorer fluency in people with first-episode schizophrenia compared to first-episode bipolar disorder;</i></p> <p>All fluency tasks: 7 studies, N = 865, $d = 0.50$, 95%CI 0.33 to 0.66, $p < 0.001$, $I^2 = 22.0%$, $p = 0.26$ Letter: 5 studies, N = 542, $d = 0.42$, 95%CI 0.24 to 0.60, $p < 0.001$ Category: 3 studies, N = 328, $d = 0.77$, 95%CI 0.0 to 1.53, $p = 0.05$ Authors report no publication bias.</p>	

Language

No differences were found for males vs. females or younger vs. older patients.	
Consistency in results	Consistent
Precision in results	Precise, apart from category task.
Directness of results	Direct

<i>Bora E</i>	
A comparative meta-analysis of neurocognition in first-degree relatives of patients with schizophrenia and bipolar disorder	
European Psychiatry: the Journal of the Association of European Psychiatrists 2017; 45: 121-8 View review abstract online	
Comparison 1	Fluency in first-degree relatives of people with bipolar disorder vs. controls.
Summary of evidence	High quality evidence (large sample, consistent, precise, direct) suggests a small to medium-sized effect of poorer fluency in first-degree relatives of people with bipolar disorder.
Fluency	
<i>Significant, small to medium-sized effect of poorer fluency in first-degree relatives of people with bipolar disorder;</i> 7 studies, N = 509, $d = 0.33$, 95%CI 0.15 to 0.50, $p < 0.001$, $I^2 = 0\%$, $p = 0.66$ There was no evidence of publication bias.	
Consistency in results	Inconsistent
Precision in results	Precise
Directness of results	Direct
Comparison 2	Fluency in first-degree relatives of people with bipolar disorder vs. first-degree relatives of people with schizophrenia.
Summary of evidence	High quality evidence (large sample, consistent, precise, direct) suggests better fluency in first-degree relatives of people with bipolar disorder.

Language

Fluency	
<i>Significant, small to medium-sized effect of better fluency in first-degree relatives of people with bipolar disorder;</i> 7 studies, N = 431, $d = 0.24$, 95%CI 0 to 0.47, $p = 0.05$, $I^2 = 33%$, $p = 0.18$	
Consistency in results	Consistent
Precision in results	Precise
Directness of results	Direct

<p>Cotrena C, Damiani Branco L, Ponsoni A, Samame C, Milman Shansis F, Paz Fonseca R</p> <p>Executive functions and memory in bipolar disorders I and II: new insights from meta-analytic results</p> <p>Acta Psychiatrica Scandinavica 2020; 141: 110-30 View review abstract online</p>	
Comparison 1	Language fluency in people with bipolar I disorder vs. controls.
Summary of evidence	Moderate to high quality evidence (large sample, inconsistent, precise, direct) shows a medium-sized effect of poorer language fluency in people with bipolar I disorder.
Language fluency	
<i>A medium-sized effect showed people with bipolar I disorder were more impaired on language fluency;</i> 38 studies, N = 4,672, $g = 0.54$, 95%CI 0.45 to 0.64, $p < 0.05$, $I^2 = 51%$, $p < 0.001$	
Consistency in results	Inconsistent
Precision in results	Precise
Directness of results	Direct
Comparison 2	Language fluency in people with bipolar II disorder vs. controls.
Summary of evidence	High quality evidence (large sample, consistent, precise, direct) shows a medium-sized effect of poorer language fluency in people with bipolar II disorder.

Language

Language fluency	
<i>A medium-sized effect showed people with bipolar II disorder were more impaired on language fluency;</i> 11 studies, N = 1,247, $g = 0.48$, 95%CI 0.35 to 0.61, $p < 0.05$, $I^2 = 18\%$, $p = 0.27$	
Consistency in results	Consistent
Precision in results	Precise
Directness of results	Direct
Comparison 3	Language fluency in people with bipolar I disorder vs. people with bipolar II disorder.
Summary of evidence	High quality evidence (large sample, consistent, precise, direct) shows a small effect of poorer language fluency in people with bipolar I disorder.
Language fluency	
<i>Small effect showed people with bipolar I disorder were more impaired on language fluency;</i> 7 studies, N = 744, $g = 0.18$, 95%CI 0.04 to 0.31, $I^2 = 0\%$, $p = 0.64$	
Consistency in results	Consistent
Precision in results	Precise
Directness of results	Direct

<p><i>Depp CA, Mausbach BT, Harmell AL, Savla GN, Bowie CR, Harvey PD, Patterson TL</i></p> <p>Meta-analysis of the association between cognitive abilities and everyday functioning in bipolar disorder</p> <p>Bipolar Disorders 2012; 14: 217-26</p> <p>View review abstract online</p>	
Comparison	Associations between verbal ability/fluency and functioning in people with bipolar disorder.
Summary of evidence	High quality evidence (large samples, consistent, precise, direct) suggests a small association between poor verbal

Language

	ability/fluency and poor general functioning.
General verbal ability and fluency	
<p><i>Significant, small associations between poor verbal ability/fluency and poor general functioning;</i> General verbal ability: 8 studies, N = 418, $r = 0.24$, 95%CI 0.14 to 0.33, $p < 0.0045$, $Qp = 0.762$ Verbal fluency: 10 studies, N = 634, $r = 0.22$, 95%CI 0.13 to 0.30, $p < 0.0045$, $Qp = 0.833$</p>	
Consistency in results	Consistent
Precision in results	Precise
Directness of results	Direct

<p><i>Dickinson T, Becerra R, Coombes J</i></p> <p>Executive functioning deficits among adults with Bipolar Disorder (types I and II): A systematic review and meta-analysis</p> <p>Journal of Affective Disorders 2017; 218: 407-27</p> <p>View review abstract online</p>	
Comparison	Fluency in people with bipolar I disorder and bipolar II disorder vs. controls.
Summary of evidence	Moderate quality evidence (unclear sample size, inconsistent, precise, direct) suggests a medium-sized effect of poorer fluency in people with bipolar disorder compared to controls, with no differences between bipolar I disorder and bipolar II disorder.
Fluency	
<p><i>Medium-sized effects show poorer fluency in people with bipolar I disorder or bipolar II disorder than controls;</i></p> <p>Bipolar I disorder: 30 studies, N not reported, $d = -0.77$, 95%CI -1.04 to -0.50, p not reported, $I^2 = 9075\%$</p> <p>Bipolar II disorder: 12 studies, N not reported, $d = -0.56$, 95%CI -0.74 to -0.37, p not reported, $I^2 = 44\%$</p> <p><i>No differences in attention between bipolar I disorder and bipolar II disorder;</i> 10 studies, N not reported, $d = -0.04$, 95%CI -0.21 to 0.13, p not reported</p>	

Language

Consistency in results	Inconsistent
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	Fluency in older people with bipolar disorder vs. controls matched for age and years of education.
Summary of evidence	High quality evidence (large samples, consistent, precise, direct) suggests a medium to large effect of poorer fluency in elderly people with bipolar disorder.
Fluency	
<p><i>Medium to large, significant effects of poorer fluency in elderly people with bipolar disorder;</i> Semantic fluency: 5 studies, N = 427, $g = 0.75$, 95%CI 0.55 to 0.95, $p < 0.001$, $I^2 = 0\%$, $p = 0.74$ Phonemic fluency: 4 studies, N = 362, $g = 0.80$, 95%CI 0.43 to 1.16, $p < 0.001$, $I^2 = 55\%$, $p = 0.08$ The analysis remained significant for semantic fluency when the analysis included only studies of patients in a depression phase ($g = 0.32$). Subgroup analyses showed no changes in the effect sizes according to age or years of education.</p>	
Consistency in results	Consistent
Precision in results	Precise
Directness of results	Direct

Samame C, Martino DJ, Strejilevich SA

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

Language

analytic study	
Journal of Affective Disorders 2014; 164: 130-8 View review abstract online	
Comparison	Changes in fluency over time in people with bipolar disorder.
Summary of evidence	Moderate quality evidence (small samples consistent, precise, direct) suggests no changes in measures of fluency over time (~3-4 years).
Fluency	
<i>There were no significant changes over time;</i>	
Category fluency: 3 studies, N = 153, follow up = 3.73 years, $d = -0.14$, 95%CI 0.50 to -0.23, $p = 0.46$, $I^2 = 61\%$, $p = 0.08$	
Phonemic fluency: 5 studies, N = 197, follow up = 4.17 years, $d = 0.07$, 95%CI -0.17 to 0.30, $p = 0.58$, $I^2 = 24\%$, $p = 0.26$	
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	Fluency in people with bipolar disorder vs. people with major depression.
Summary of evidence	Low quality evidence (small sample, inconsistent, imprecise, direct) is unable to determine whether there are any differences between groups.
Phonological fluency	
<i>There were no significant differences between groups;</i>	

Language

3 studies, $N = 107$, $g = 0.73$, 95%CI -0.18 to 1.64, $p = 0.12$, $I^2 = 80\%$, $p = 0.007$

Consistency in results	Inconsistent
Precision in results	Imprecise
Directness of results	Direct

Explanation of acronyms

CI = confidence interval, 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

Language

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 small¹¹.

† 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 effect¹¹.

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 ¹². InOR stands for logarithmic OR where a InOR 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

Language

between variables. They can provide an indirect indication of prediction, but do not confirm causality due to possible and often unforeseen 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 formula¹¹;

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

§ 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 relaxed¹³.

|| 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.

Language

References

1. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group (2009): Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *British Medical Journal* 151: 264-9.
2. GRADE Working Group (2004): Grading quality of evidence and strength of recommendations. *British Medical Journal* 328: 1490.
3. Bora E, Pantelis C (2015): Meta-analysis of Cognitive Impairment in First-Episode Bipolar Disorder: Comparison With First-Episode Schizophrenia and Healthy Controls. *Schizophrenia Bulletin* 41: 1095-104.
4. 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.
5. Samame C, Martino DJ, Strejilevich SA (2013): A quantitative review of neurocognition in euthymic late-life bipolar disorder. *Bipolar Disorders* 15: 633-44.
6. 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.
7. 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.
8. Bora E (2017): A comparative meta-analysis of neurocognition in first-degree relatives of patients with schizophrenia and bipolar disorder. *European Psychiatry: the Journal of the Association of European Psychiatrists* 45: 121-8.
9. 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.
10. 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.
11. Cochrane Collaboration (2008): Cochrane Handbook for Systematic Reviews of Interventions. Accessed 24/06/2011.
12. Rosenthal JA (1996): Qualitative Descriptors of Strength of Association and Effect Size. *Journal of Social Service Research* 21: 37-59.
13. GRADEpro (2008): [Computer program]. Jan Brozek, Andrew Oxman, Holger Schünemann. *Version 3.2 for Windows*