

Physical anomalies

Introduction

First-degree relatives of people with bipolar disorder may show attenuated signs of the illness, such as physical features commonly identified with the disorder. These may include structural and/or functional anomalies as well as sleep disturbances.

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. 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 version 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. Reviews were assigned a low, medium or high possibility of reporting bias* depending on how many items were checked. Reviews rated as having less than 50% of items checked have now 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)¹. 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 two systematic reviews that met our inclusion criteria^{2,3}.

- Moderate quality evidence suggests a medium-sized effect of lower relative amplitude of the sleep-wake cycle in people with bipolar disorder than in people at risk.
- Low quality evidence is unclear of brain anomalies in relatives of people with bipolar disorder compared to controls.

Physical anomalies

Miskowiak KW, Kjaerstad HL, Meluken I, Petersen JZ, Maciel BR, Kohler CA, Vinberg M, Kessing LV, Carvalho AF

The search for neuroimaging and cognitive endophenotypes: A critical systematic review of studies involving unaffected first-degree relatives of individuals with bipolar disorder

Neuroscience and Biobehavioral Reviews 2017; 73: 1-22

[View online review abstract](#)

Comparison	Brain structure and function in relatives of people with bipolar disorder vs. controls.
Summary of evidence	Low quality evidence (small samples, inconsistent, unable to assess precision, direct) is unclear of brain anomalies in relatives.
Grey matter volume	
<p>2 studies (N = 61 and 54) found increased grey matter volume in the inferior frontal gyrus.</p> <p>2 studies (N = 58 and 30) found decreased grey matter volume in the orbitofrontal cortex.</p> <p>1 study (N = 58) reported increased grey matter volume in the dorsolateral prefrontal cortex, however another study (N = 60) found no differences in the dorsolateral prefrontal cortex.</p> <p>1 study (N = 63) found no volume differences in the orbitomedial prefrontal cortex.</p> <p>1 study (N = 42) found no volume differences in the ventromedial prefrontal cortex.</p> <p>2 studies (N = 60 and 30) found reduced grey matter volume in the insular.</p> <p>2 studies (N = 50 and 54) found reduced grey matter volume in the cerebellum.</p> <p>1 study (N = 119) found increased grey matter volume in the insula and cerebellum</p> <p>1 study (N = 63) found increased volume in the amygdala, however another study (N = 42) found no significant differences in amygdala volume.</p> <p>1 study (N = 42) reported increased volume of the parahippocampal gyrus extending into hippocampus, however another study (N = 63) found no differences in hippocampal volume.</p>	
White matter integrity	
<p>1 study (N = 60) found decreased white matter volumes of the right (but not left) medial frontal gyrus.</p> <p>1 study (N = 90) found anomalies in white matter tracks adjacent to the ventral prefrontal cortex cingulate gyrus, superior middle temporal gyrus, precuneus, and posterior regions of the occipital lobe.</p> <p>1 study (N = 15) found bilaterally compromised white matter integrity in the superior longitudinal</p>	

Physical anomalies

fasciculus.	
1 study (N = 45) found bilaterally compromised white matter integrity in the inferior longitudinal fasciculus, and the corpus callosum.	
1 study (N = 44) found bilaterally compromised white matter integrity in the internal capsule and unicate fasciculus.	
3 studies (N = 93, 38 and 39) found no white matter anomalies.	
Connectivity	
2 studies (N = 75 and 49) found differences in functional connectivity within regions of the frontostriatal circuitry during resting state.	
1 study (N = 170) found reduced connectivity between frontooccipital regions and the anterior default mode prefrontal network, and between the mesoparalimbic network and frontotemporal-paralimbic regions.	
1 study (N = 40) found increased connectivity between the dorsolateral prefrontal cortex and amygdala bilaterally.	
1 study (N = 38) found longer latency and reduced amplitude of N100 (an ERP measure of attention processes) and P300 (a measure of post-perceptual activity).	
1 study (N = 38) found abnormal threat processing as indicated by reduced pre-pulse inhibition of the startle reflex.	
2 studies (N = 38 and 144) found no differences in simple visual stimuli processing as reflected by N200 and P200.	
1 study (N = 49) found no activity difference within the default mode network.	
Consistency in results	Inconsistent for grey matter
Precision in results	Unable to assess; no CIs reported
Directness of results	Direct

Ng TH, Chung KF, Ho FY, Yeung WF, Yung KP, Lam TH

Sleep-wake disturbance in interepisode bipolar disorder and high-risk individuals: a systematic review and meta-analysis

Sleep Medicine Reviews 2015; 20: 46-58

[View online review abstract](#)

Comparison	Sleep efficacy in people at high risk of bipolar disorder (family history and questionnaire scores) vs. people with bipolar disorder
-------------------	---

Physical anomalies

	in the euthymic phase.
Summary of evidence	Moderate quality evidence (small to medium sample size, consistent, precise, direct) suggests a medium-sized effect of lower relative amplitude of the sleep-wake cycle in people with bipolar disorder than in people at risk.
Sleep variables	
<p><i>Significant, medium-sized effect of lower amplitude of the sleep-wake cycle in people with bipolar disorder compared to relatives;</i></p> <p>3 studies, N = 174, SMD = -0.43, 95%CI -0.83 to -0.02, $p = 0.04$, $I^2 = 43%$, $p = 0.17$</p> <p>There were no significant differences in sleep time, efficacy, latency, or wakefulness.</p>	
Consistency in results	Consistent
Precision in results	Precise
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, p = statistical probability of obtaining that result ($p < 0.05$ generally regarded as significant), SMD = standardised mean difference, vs. = versus

Physical anomalies

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 small⁴.

† 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.

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 effect⁴.

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).

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 ⁵. 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

Physical anomalies

between variables. They are an 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 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;

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

§ 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 relaxed⁶.

|| 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.

Physical anomalies

References

1. GRADEWorkingGroup (2004): Grading quality of evidence and strength of recommendations. *British Medical Journal* 328: 1490.
2. Miskowiak KW, Kjaerstad HL, Meluken I, Petersen JZ, Maciel BR, Kohler CA, *et al.* (2017): The search for neuroimaging and cognitive endophenotypes: A critical systematic review of studies involving unaffected first-degree relatives of individuals with bipolar disorder. *Neuroscience and Biobehavioral Reviews* 73: 1-22.
3. Ng TH, Chung KF, Ho FY, Yeung WF, Yung KP, Lam TH (2015): Sleep-wake disturbance in interepisode bipolar disorder and high-risk individuals: a systematic review and meta-analysis. *Sleep Medicine Reviews* 20: 46-58.
4. CochraneCollaboration (2008): Cochrane Handbook for Systematic Reviews of Interventions. Accessed 24/06/2011.
5. Rosenthal JA (1996): Qualitative Descriptors of Strength of Association and Effect Size. *Journal of Social Service Research* 21: 37-59.
6. GRADEpro (2008): [Computer program]. Jan Brozek, Andrew Oxman, Holger Schünemann. *Version 3.2 for Windows*.