TECHNICAL COMMENTARY

NeuRA Discover. Conquer. Cure.

BIPOLAR DISORDERS LIBRARY

P300

Introduction

The P300 wave is an event-related potential, measured using EEG. It is a spike in brain activity approximately 300ms after presentation of a target stimulus, which is usually auditory but can also be visual.

A typical auditory paradigm uses a series of tones where 20% of them are infrequent targets called “oddballs”. Research subjects are asked to respond to these oddballs, and the related spike in brain activity is usually apparent in frontal regions thought to be related to contextual updating and memory storage. This is called a P3b response, while a P3a response reflects response to an infrequent *non-target* stimulus usually apparent in parietal regions and thought to be related to automatic attention processing.

P300 measures also include amplitude and latency of the P300 wave. Amplitude is proportional to the amount of attentional resource devoted to the task and the degree of information processing required, while latency is considered a measure of stimulus classification speed.

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 to high quality evidence shows medium to large effects of reduced P300 P3b (target) amplitude and longer P300 P3b (target) latency in people with bipolar disorder compared to controls in both

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

auditory and visual paradigms. There was also a small effect of reduced P300 P3a (non-target) amplitude in an auditory paradigm. These results remained in studies of people with or without a history of psychotic symptoms, in people with different bipolar types (bipolar I vs. bipolar II), and in people in different phases of the disorder (depression, euthymia, mania).

- Moderate quality evidence shows large effects of longer P300 latency and lower amplitude in people with bipolar depression compared to controls. In remitted patients, results remained significant only for longer latency.
- Compared to people with unipolar depression, moderate to high quality evidence shows a medium-sized effect of longer P300 latency in people with bipolar depression, with no differences in amplitude. Results were similar in the analyses of remitted patients. The results were larger in studies conducted in China, in studies with >100 patients, and in unmedicated patients.

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

Wada M, Kurose S, Miyazaki T, Nakajima S, Masuda F, Mimura Y, Noda Y

The P300 event-related potential in bipolar disorder: A systematic review and meta-analysis

Journal of Affective Disorders 2019; 256: 234-49

View online review abstract

Comparison	**P3a and P3b latency and amplitude in people with bipolar disorder vs. controls.**
Summary of evidence	**Moderate to high quality evidence (medium to large samples, mostly consistent and precise, direct) shows medium to large effects of reduced amplitude and longer latency in both P3b (target) auditory and visual paradigms. There was also a small effect of reduced amplitude in a P3a (non-target) auditory paradigm.**
P3b	
People with bipolar disorder showed reduced amplitude and longer latency in both auditory and visual paradigms (medium to large effects); Amplitude in an auditory paradigm: 23 studies, N = 2,917, SMD = -0.599, 95%CI -0.709 to -0.489, $p < 0.001$ Amplitude in a visual paradigm: 7 studies, N = 271, SMD = -0.509, 95%CI -0.795 to -0.224, $p < 0.001$ Latency in an auditory paradigm: 21 studies, N = 2,632, SMD = 0.590, 95%CI 0.450 to 0.730, $p < 0.001$ Latency in a visual paradigm: 6 studies, N = 231, SMD = 1.253, 95%CI 0.091 to 2.415, $p = 0.035$ Note that authors used an alpha of 0.01, adjusting for multiple comparisons, and therefore concluded latency in a visual paradigm was only significant at a trend level. There were no moderating effects of having a history of psychotic symptoms, bipolar type (bipolar I vs. bipolar II), or phase of illness (depression, euthymia, mania).	
P3a	
People with bipolar disorder showed reduced P3a amplitude in an auditory paradigm (small effect); 4 studies, N = 417, SMD = -0.335, 95%CI -0.565 to -0.106, $p = 0.004$ There were insufficient studies for meta-analyses of latency or amplitude in a visual paradigm.	
Consistency in results‡	Consistent, apart from P3b amplitude with auditory paradigm (I^2 = 58.2%).

Precision in results[§]	Precise, apart from latency in a visual paradigm.
Directness of results[‖]	Direct

Zhong BL, Xu YM, Xie WX, Li Y

Can P300 aid in the differential diagnosis of unipolar disorder versus bipolar disorder depression? A meta-analysis of comparative studies

Journal of Affective Disorders 2019; 245: 219-27

View online review abstract

Comparison 1	**P300 latency and amplitude in people with bipolar depression vs. controls.**
Summary of evidence	**Moderate quality evidence (large sample, unclear consistency, imprecise, direct) shows large effects of longer latency and lower amplitude in people with bipolar depression. In remitted patients, there were no longer differences in amplitude, but latency remained significant, with a large effect.**
P300	
Large effects of longer latency and lower amplitude in people with bipolar depression; Latency: 7 studies, N = 831, SMD = 2.375, 95%CI 1.149 to 3.601, $p < 0.05$, I^2 not reported Amplitude: 7 studies, N = 831, SMD = -1.030, 95%CI -1.511 to -0.549, $p < 0.05$, I^2 not reported The results were similar after excluding the one poor-quality study. In remitted patients, there were no differences in amplitude, but latency remained significantly longer, with a large effect.	
Comparison 2	**P300 latency and amplitude in people with bipolar depression vs. people with unipolar depression.**
Summary of evidence	**Moderate to high quality evidence (large samples, inconsistent, precise, direct) shows a medium-sized effect of longer P300 latency in people with bipolar depression, with no differences in amplitude. These results remained in analyses of remitted patients, although the effect size was larger for latency. The effects were also larger in studies conducted in China vs. other countries, in studies with >100 vs. <100 patients, and in unmedicated vs. medicated patients.**
P300	

A medium-sized effect of longer latency in people with bipolar depression, with no differences in amplitude; Latency: 8 studies, N = 787, SMD = 0.580, 95%CI 0.309 to 0.850, $p < 0.05$, $I^2 = 68\%$, $p < 0.01$ Amplitude: 8 studies, N = 787, SMD = -0.202, 95%CI -0.520 to 0.115, $p > 0.05$, $I^2 = 78\%$, $p < 0.01$ Sensitivity analysis gave similar results after excluding the two poor-quality studies. In remitted patients, there were no differences in amplitude, while latency remained significantly longer, and the effect was larger. Subgroup analyses of latency results revealed larger effect sizes in; studies conducted in China vs. other countries, studies with >100 patients than <100 patients, and studies of unmedicated vs. medicated patients. There were no differences in the results according to; publication year (1991-2013 vs. 2014-2017), diagnostic method (DSM vs. other), first-episode patients vs. chronic patients, auditory vs. visual stimulus, or study quality. There were no significant differences in any moderator in the amplitude subgroup analyses. Authors report no evidence of publication bias.

Consistency in results	Inconsistent for comparison 2, unable to assess comparison 1.
Precision in results	Precise, apart from latency in comparison 1.
Directness of results	Direct

Explanation of acronyms

CI = confidence interval, DSM = diagnostic and statistical manual of mental disorders, EEG = electroencephalogram, I^2 = degree of heterogeneity across study results not explained by chance, N = number of participants, p = probability of obtaining that result ($p < 0.05$ generally regarded as significant), SMD = standardised mean difference, 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[4].

---

† 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 effect[4].

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]. 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

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

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 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 relaxed[6].

‖ 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.

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

References

1. GRADEWorkingGroup (2004): Grading quality of evidence and strength of recommendations. *British Medical Journal* 328: 1490.
2. Zhong BL, Xu YM, Xie WX, Li Y (2019): Can P300 aid in the differential diagnosis of unipolar disorder versus bipolar disorder depression? A meta-analysis of comparative studies. *Journal of Affective Disorders* 245: 219-27.
3. Wada M, Kurose S, Miyazaki T, Nakajima S, Masuda F, Mimura Y*, et al.* (2019): The P300 event-related potential in bipolar disorder: A systematic review and meta-analysis. *Journal of Affective Disorders* 256: 234-49.
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 32 for Windows*

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