Introduction

The P200 wave is an event-related brain potential (ERP) measured using electroencephalography (EEG). P200 refers to a spike in activity approximately 150 to 250ms following presentation of a target stimulus that is most commonly auditory, although response is also obtained following somatosensory and visual events. The P200’s latency and amplitude vary with aspects of selective attention or stimulus encoding. Latency is considered a measure of stimulus classification speed, and amplitude is proportional to the amount of attentional resources devoted to the task and the degree of information processing required [1]. Amplitude and latency may be measured using tasks using ‘standard’ and ‘oddball’ stimuli, where the subject is asked to react only to ‘oddball’ target stimuli that are hidden as rare occurrences amongst a series of more common, ‘standard’ stimuli.

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 2000 that report results separately for people with a diagnosis of schizophrenia, schizoaffective disorder, schizophreniform disorder or first episode schizophrenia. As part of a wider search for all topics included in the library, reviews on P200 ERPs in schizophrenia were identified by searching the databases MEDLINE, EMBASE, CINAHL, Current Contents, PsycINFO and the Cochrane library. Hand searching reference lists of identified reviews was also conducted. When multiple copies of reviews were found, only the most recent version was included. The decision to include or exclude reviews was conducted in duplicate by two independent reviewers with any disagreements settled by discussion. All quality assessments and data extraction have been completed in duplicate by two reviewers who were not masked to review authors.

Review reporting assessment was guided by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist (formerly the QUOROM statement) which describes a preferred way to present a meta-analysis [2]. Reviews were assigned a low, medium or high possibility of reporting bias* depending on how many items were checked. For instance, a low possibility of bias would be assigned to reviews checking over 66% of items, a medium possibility between 33 and 66% and a high possibility would be given to reviews checking less than 33%. Reviews rated as having a high possibility of reporting bias 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 randomized controlled trials (RCT) may be downgraded to moderate, low or very 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)[3]. The resulting table represents an objective summary of the available evidence, although the conclusions are solely the opinion of staff of the Schizophrenia Research Institute.

Results
See table below for a detailed summary of the available evidence for P200 ERPs. We found one systematic reviews that met our inclusion criteria [1].
See PRISMA checklist for review quality assessments.

Conclusions
• Moderate quality evidence suggests a small reduction in P200 amplitude and latency at the Cz midline electrode during standard stimuli conditions, and a small to medium increase in amplitude and latency at the Fz, Cz and Pz electrodes during oddball stimuli conditions in patients compared to controls.
Ferreira-Santos, F., Silveira, C., Almeida, P.R., Palha, A., Barbosa, F., Marques-Teixeira, J.

The auditory P200 is both increased and reduced in schizophrenia? A meta-analytic dissociation of the effect for standard and target stimuli in the oddball task

Clinical Neurophysiology 2012; 123:1300-1308
View review abstract online

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Comparison of P200 ERP amplitude and latency in schizophrenia vs. healthy controls</th>
</tr>
</thead>
</table>

| Summary of evidence | Moderate quality evidence (mostly inconsistent, mostly precise, direct) suggests a small reduction in P200 amplitude and latency at the Cz midline electrode during standard stimuli conditions, and a small to medium increase in amplitude and latency at the Fz, Cz and Pz electrodes during oddball stimuli conditions |

P200 activity

†Meta-analysis compared effect sizes in patients with schizophrenia and controls for P200 amplitude and latency in standard and oddball stimuli conditions at midline electrodes Fz, Pz and Cz

**Standard stimuli**

**P200 amplitude**

Small effect size suggests significantly reduced amplitude at Cz in patients
Fz: N = 502, 5 studies, d = -0.09, 95%CI -0.26 to 0.08, p > 0.05, (Q = 0.03, p = 0.008)
Cz: N = 1238, 15 studies, d = -0.36, 95%CI -0.52 to -0.20, p < 0.05, (Q = 0.09, p < 0.001)
Pz: N = 502, 5 studies, d = -0.22, 95%CI -0.43 to 0.00, p = 0.05, (Q = 0.05, p < 0.001)

**P200 latency**

Small effect size suggests significantly reduced latency at Cz in patients
Fz: N = 282, 3 studies, d = 0.05, 95%CI -0.43 to 0.54, p > 0.05, (Q = 0.17 p < 0.001)
Cz: N = 998, 13 studies, d = -0.32, 95%CI -0.54 to -0.10, p < 0.05, (Q = 0.14, p < 0.001)
Pz: N = 282, 3 studies, d = -0.16, 95%CI -0.80 to 0.47, p > 0.05, (Q = 0.30, p < 0.001)

**Oddball stimuli**

**P200 amplitude**

Small effect size suggests significantly increased amplitude at Cz in patients during oddball stimuli conditions
Fz: N = 282, 3 studies, d = -0.36, 95%CI -0.53 to -0.19, p < 0.05, (Q = 0.28 p < 0.001)
Cz: N = 998, 13 studies, d = -0.16, 95%CI -0.38 to 0.06, p > 0.05, (Q = 0.42, p < 0.001)
Pz: N = 282, 3 studies, d = -0.16, 95%CI -0.80 to 0.47, p > 0.05, (Q = 0.30, p < 0.001)
Electrophysiology – P200

Small to medium effect size suggests significantly increased amplitude at Fz, Cz and PZ in patients
- Fz: N = 465, 5 studies, d = 0.34, 95%CI 0.06 to 0.62, p < 0.05, (Q = 0.09, p < 0.001)
- Cz: N = 653, 9 studies, d = 0.48, 95%CI 0.16 to 0.81, p < 0.05, (Q = 0.23, p < 0.001)
- Pz: N = 465, 5 studies, d = 0.46, 95%CI 0.13 to 0.79, p < 0.05, (Q = 0.13, p < 0.001)

P200 latency
Small to medium effect size suggests significantly increased latency at Fz, Cz and PZ in patients
- Fz: N = 385, 4 studies, d = 0.27, 95%CI 0.11 to 0.42, p < 0.05, (Q = 0.02, p = 0.070)
- Cz: N = 516, 7 studies, d = 0.42, 95%CI 0.23 to 0.62, p < 0.05, (Q = 0.05, p < 0.001)
- Pz: N = 385, 4 studies, d = 0.45, 95%CI 0.35 to 0.55, p < 0.05, (Q = 0.00, p = 0.394)

Meta-regression identified that the greater the male percentage in studies, the larger the effect size for target amplitude (r = 0.674, p = 0.047). Increased chlorpromazine equivalent dosage was related to a decrease in effect size for standard latency (r =−0.765, p = 0.027).

Authors also indicate evidence of publication bias for Fz and Pz latency comparisons
Egger test p < 0.01.

<table>
<thead>
<tr>
<th>Consistency in results</th>
<th>Inconsistent for all except latency at Fz and PZ with oddball stimuli</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precision in results</td>
<td>Precise for all except latency at Pz with standard stimuli</td>
</tr>
<tr>
<td>Directness of results</td>
<td>Direct</td>
</tr>
</tbody>
</table>

Explanation of acronyms

CI = Confidence Interval, CZ = central electrode, d = Cohen’s d and g = Hedges’ g = standardized mean differences (see below for interpretation of effect sizes), EEG = electroencephalogram, ERP = event-related potential, FZ = frontal lobe electrode, N = number of participants, p = statistical probability of obtaining that result (p < 0.05 generally regarded as significant), PZ = parietal lobe electrode, Q = Q statistic (chi-square) for the test of heterogeneity in results across studies, Qb = between group heterogeneity, Qw = within group heterogeneity, SE = standard error, T3/TCP1= Left temporal lobe electrodes, T4/TCP2 = Right temporal lobe electrodes, vs = versus
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 which 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.

Weighted mean difference scores refer to mean differences between treatment and comparison groups after treatment (or occasionally pre to post treatment) and in a randomized trial there is an assumption that both groups are comparable on this measure prior to treatment. Standardized mean differences are divided by the pooled standard deviation (or the standard deviation of one group when groups are homogenous) which 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 moderate 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 which are correctly identified (100% sensitivity = correct identification of all actual positives) and specificity is the proportion of negatives which 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 large effect is considered if RR > 2 or < 0.5 and a very large effect if RR > 5 or < 0.2[5]. InOR 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 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. Unstandardized (b) regression coefficients indicate the average change in the dependent variable associated with a 1 unit change in the dependent variable,
Electrophysiology – P200

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.

‡ Inconsistency refers to differing estimates of treatment effect across studies (i.e. heterogeneity or variability in results) which 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, this criteria should be relaxed[5].

|| 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 which 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.

This topic is yet to be reviewed by a content expert
Electrophysiology – P200

References