The clinical utility of the auditory P300 latency subcomponent event-related potential in preclinical diagnosis of patients with mild cognitive impairment and Alzheimer's disease

Aaron S. Howe*, Ali Bani-Fatemi, Vincenzo De Luca
PharmacoEEG Study Group, Centre for Addiction and Mental Health, Toronto M5T 1R8, Canada

ARTICLE INFO

Article history:
Accepted 27 January 2014
Available online 22 February 2014

Keywords:
P300 latency
Event-related potential
Alzheimer's disease
Mild cognitive impairment
Biological marker

ABSTRACT

The present meta-analysis investigated the clinical utility of the auditory P300 latency event-related potential in differentiating patients with Alzheimer's disease (AD), patients with mild cognitive impairment (MCI), and unaffected controls. Effect size estimates were computed from mean P300 latency measurements at midline electrodes between patients and unaffected controls using the random effects restricted maximum likelihood model. The effects of clinical and ERP/EEG methodological variables were assessed in a moderator analysis. P300 latency was found to be significantly prolonged in patients with AD (and MCI) compared to unaffected controls. Shortened P300 latencies were observed when comparing patients with MCI to patients with AD. Clinically relevant differences in P300 latency effect sizes were associated with mean age, interstimulus interval, stimulus difference, target frequency, reference electrode, and sampling rate. The meta-analytic findings provide robust statistical evidence for the use of the auditory P300 latency subcomponent as a biological marker of prodromal AD.

1. Introduction

The cognitive deterioration seen in Alzheimer's disease (AD) pathology has been characterized by the P300 event-related potential for the past 35 years. The first report to suggest and provide evidence of the clinical utility of the auditory P300 latency subcomponent in dementia was Goodin, Squires, and Starr (1978). During this time, event-related potentials were considered as a specific and sensitive measure of afferent function in neurologic patients. The NINCDS–ADRDA working group also commented on the use of event-related potentials for diagnosis of Alzheimer's disease in research settings (McKhann et al., 1984). However, the current view of the P300 waveform for clinical and diagnostic use of AD in research settings has drastically changed over the past 35 years. Since initial reports (Blackwood, St Clair, Blackburn, & Tyrer, 1987; Brown, Marsh, & LaRue, 1982; Goodin, Squires, Starr, et al., 1978; Goodin, Starr, Chippendale, & Squires, 1983; Ortiz, Martin Loeches, Miguel, Abdad, & Puente, 1994; Patterson, Michalewski, & Starr, 1988; Pfefferbaum, Wenegrat, Ford, Roth, & Kopell, 1984; Sluets & Fortgens, 1984; St Clair, Blackwood, & Chris-
Numerous clinical ERP studies have reported significant P300 latency differences between patients with AD and unaffected controls. More specifically, patients with AD exhibit a prolongation of the P300 latency compared to age-matched unaffected controls. P300 latency has also been reported to be associated with several clinical variables in AD: family history of AD/genetic mutations (APOE) (Ally et al., 2006; Golob et al., 2009; Irimajiri, Golob, & Starr, 2010), cholinesterase inhibitors (Ally et al., 2006; Golob & Starr, 2000; Katada, Sato, Ojika, & Ueda, 2004; Onofrj et al., 2002; Reeves, Struve, Patrick, Booker, & Nave, 1999; Thomas, Iacono, Bonanni, D’Andreamatteo, & Onofrj, 2001; Werber, Klein, & Rabey, 2001), language and acoustic-motor ability (Blackwood et al., 1987), attention (Boller et al., 2002; Nishige et al., 1988; Picton & Hillyard, 1974), severity of cognitive deficits (Ball, Marsh, Schubarth, Brown, & Strandburg, 1989; Gurgor et al., 2005; Lai et al., 2010; Lee et al., 2013; Pfefferbaum, Wenegrat, et al., 1984; Pokryszko-Dragan, Sfoltwinski, & Podemski, 2003; Polich & Pitzer, 1999; Polich et al., 1986; Sydulko et al., 1982; Szelies, Mielke, Grond, & Heiss, 1995; Tanaka, Kachi, Yamada, & Sobue, 1998; Van Deursen et al., 2009; Wright, Scott, Richardson, Rai, & Exton-Smith, 1988), delayed motor response (Krauhiin et al., 1990; Van Deursen et al., 2009; Williams et al., 1991), CSF concentrations of monoamines (5-HIAA) (Ito et al., 1990) and neurotransmitters (Mochizuki et al., 2001), hypometabolism in parietal lobe (Marsh et al., 1990; Szelies et al., 1995) executive function (Lee et al., 2013), decreased cerebral blood flow (precuneus, frontal lobe) (Gurgor et al., 2005). However, some studies have disconfirming evidence for these associations: memory (Blackwood et al., 1987; Boller et al., 2002), attention (Pfefferbaum, Wenegrat, et al., 1984), decreased cerebral blood flow (Mochizuki et al., 2001), age (Muscolo et al., 2006), severity of early stage cognitive deficits (Muscolo et al., 2006), differentiating subcortical and cortical dementias (Tachibana et al., 1993; Tachibana et al., 1996). Some studies have also suggested that P300 latency prolongation may be linked to neurodegeneration of cortical areas in the temporoparietal lobe (Frodé et al., 2002; Gurgor et al., 2005; Jiménez-Esrig et al., 2002; Juckel et al., 2012; Muscolo et al., 2006).

Recently, the NINCA-ADRSA working group has suggested that accumulated amyloid-β deposition begins the pathophysiological process and cognitive/behavioral deficits are further implicated as a result of this accumulation (Sperling et al., 2011). According to this criteria, the early cognitive deficits observed in patients are more likely to be indicative of the later asymptomatic AD phase (Sperling et al., 2011). Therefore, studying pathophysiological differences in patients with MCI may provide an effective model of the later stages of prodromal AD (Albert et al., 2011). There has been a few clinical auditory ERP studies investigating the P300 latency in patients with MCI compared to unaffected controls (Bennys, Rondouin, Benattar, Gabelle, & Touchon, 2011; Bennys et al., 2007; Frodl et al., 2002; Gironell et al., 2005; Golob, Irimajiri, & Starr, 2007; Golob, Johnson, & Starr, 2002; Lai et al., 2010; Medvidovic, Titlic, & Maras-Simunic, 2013; Papaliagkas, Kimiskidis, Tsolaki, & Anogianakis, 2008, 2010, 2011; Van Deursen et al., 2009) and to patients with AD (Bennys et al., 2007; Frodl et al., 2002; Gironell et al., 2005; Lai et al., 2010; Van Deursen et al., 2009). Most of these studies have reported longer prolongations of the P300 latency in patients with MCI compared to unaffected controls, and shortened P300 latencies when compared to patients with AD. However, quantifying these results for support of ERPs as useful clinical assays still remains to be a question that has not been fully answered (Bennys et al., 2011; Juckel et al., 2012).

Therefore, a quantitative meta-analysis is required to ascertain the potential of the P300 latency as an accurate assay of the cognitive dysfunction observed in the late preclinical stages of AD (Polich & Corey-Bloom, 2005). The primary goal of the present
study was to provide robust quantitative evidence of the clinical utility of the auditory P300 latency in patients with MCI and AD.

2. Methods

2.1. Search protocol

Studies included in the meta-analytic assessment were identified using a systematic computerized search of electronic journal databases: PubMed, PsycINFO, Web of Science, MEDLINE, EMBASE, and Scopus. The search in each database was performed for the years between 1970 and 2013 using the keywords: “P300”, “latency”, “ERP”, “evoked potentials”, “P3”, “P3a”, “P3b”, “Alzheimer”, “dementia”, “mild cognitive impairment”, “aging”, and “biomarker”. A thorough search of the reference list of each identified study and published review was performed to identify additional articles. All authors independently performed the systematic literature search and came to consensus on the studies that met the meta-analysis criteria. Studies were included in the meta-analysis based on the following criteria: (1) ERPs were compared in a patient group (AD and/or MCI) and an unaffected control group of age-matched individuals. (2) P300 latency differences were reported as mean (±SD) in both patients and unaffected controls, so that effect size estimates could be calculated exactly. (3) P300 latency differences were measured using an auditory oddball task. (4) Studies with patient groups with a small proportion (<15%) of dementia not related to AD were included to increase sample size. (5) Only baseline data from longitudinal reports were used in the analysis (Ball et al., 1989; Gironell et al., 2005; Lai et al., 2010; Papaliagkas, Kimiskidis, Tsolaki, & Anogianakis, 2011; Van Deursen et al., 2009). Studies were excluded from the meta-analysis based on the following criteria: (1) Studies that used the same participants over more than 1 paper (only the most comprehensive paper was selected) (Papaliagkas et al., 2008; Papaliagkas et al., 2011) (2) Passive auditory ERP procedures were excluded (Gottlieb, Wertman, & Bentin, 1991; Knott, Mohr, Hache, Mahoney, & Mendis, 1999), while active auditory ERP procedures were retained. (3) Statistical analysis and data was presented graphically or as p-values (Bennys et al., 2007; Bonanni et al., 2010; Gordon et al., 1986; Gordon et al., 1989; Krauhiin et al., 1990; Marsh et al., 1990; Polich et al., 1986; Swanwick, Rowan, Coen, Coakley, & Lawlor, 1999; Swanwick et al., 1996; Tanaka et al., 1998). (4) The study did not include a unaffected comparison group (Engel et al., 1992;Kataeda et al., 2003; Knott, Mohr, Mahoney, England, & Ilivitsky, 2002; Reeves et al., 1999; Saletu et al., 1995; Semlitsch, Anderer, & Saletu, 1992; Sloan, Fenton, Kennedy, & MacLennan, 1994; St Clair, Blackburn, Blackwood, & Tyrer, 1988; Thomas, Iacono, Bonanni, D'AndreaMatteo, & Onofri, 2001; Van Gool, Waardenburg, Meyies, Weinstein, & De Wilde, 1991; Werber et al., 2001). (5) Studies using only visual tasks and other atypical ERP recording methods were excluded. Visual tasks were excluded due to the small amount of literature comparing patients with AD (or MCI) and unaffected controls (Daffner et al., 2001; Gozke, Tomruku, & Erdal, 2013; Parra, Ascencio, Urquina, Manes, & Ibanez, 2012; Saito et al., 2001) Visual ERPs have also been shown to produce weaker P300 differences between patients with cognitive dysfunction and unaffected controls (Pokryszko-Dragan et al., 2003; Verleger, Kompf, & Neuèkäer, 1992), but some studies have found little differences between auditory and visual ERPs (Corey-Bloom & Polich, 2005; Onofri et al., 1991; Pfefferbaum, Ford, et al., 1984; Pfefferbaum, Wenergrat, et al., 1984; Polich & Pitzer, 1999). Therefore, to make meaningful clinical inferences on P300 latency as a potential biomarker of preclinical AD only auditory procedures were considered. For studies that contained only graphical analysis of P300, authors were contacted multiple times regarding the descriptive statistics of each group before the study was excluded.

2.2. Meta-analysis methodology

The selected studies were reviewed further to extract modera-
tor variables, and mean P300 latency differences between experi-
mental groups. The data in each study reviewed was refined for the meta-analysis based on the following implemented steps: (1) P300 latency studies that reported grand averages at all three mid-
line electrodes were pooled into a single mean P300 latency for each group. P300 latency differences were also compared only at the Pz site (where P300 is largest) in a secondary analysis. (2) P300 latency reported at P3a and P3b was pooled into a single mean P300 latency for each group (Juckel et al., 2008). (3) P300 latencies reported in subgroups of AD (mild, moderate) were ana-
lyzed collectively (pooled groups) and then as separate studies in two meta-analysis comparisons (comparison 1: AD and unaffected controls, comparison 2: mild AD and unaffected controls) (Frodl et al., 2002; Gungor et al., 2005; Pokryszko-Dragan et al., 2003) (4) Studies analyzing patients with mild AD were identified when the author specifically stated mild AD patients constituted the sample and/or an MMSE total score ≥18 (if reported). (5) P300 latency measurements were only compared for target stimuli between groups in the meta-analysis. (6) P300 latency differences were compared in four comparison meta-analyses: (1) patients with AD and unaffected controls (k = 40; Pz only: k = 27), (2) pa-
tients with mild AD and unaffected controls (k = 9; Pz only: k = 5), (3) patients with MCI and unaffected controls (k = 8; Pz only: k = 4), (4) patients with AD and patients with MCI (k = 5; Pz only: k = 4).

All statistical meta-analysis procedures were performed using the R-project core version 3.0.1., and the “metafor” package (version 1.9-2, released 2013-10-07) (Viechtbauer, 2010). The effect sizes were computed as standardized mean differences, which is expressed as the product of the difference in means of both groups divided their pooled standard deviation. An effect size is a robust measure of study variability as each study is given a standardized weight corresponding to their sample size, which may be used for future power analyses (Viechtbauer, 2005; Viechtbauer, 2010). The direction of the effect sizes indicates whether the P300 latency dif-
ference between patients and unaffected controls is prolonged (positive) or shortened (negative). A 95% confidence interval was also computed based on the pooled standard deviation. The com-
puted effect sizes were then analyzed using the “rma” function in the “metafor” package using the restricted maximum likelihood (REML) random effects model to account for study heterogeneity (tau²). Tau² is the estimated standard deviation across all of the studies analyzed in the sample. This model was used to provide an inference about the average effects of the sample from a popula-
tion of normally distributed studies that are randomly selected (Viechtbauer, 2010). As a result, wider confidence intervals are re-
ported compared to that of fixed effects models.

Forest plots were constructed to provide a qualitative assess-
ment of the effect sizes presented in the study. In the forest plots, the studies have been placed in the order of largest weight (or ef-
fect) to smallest weight on the meta-analysis. The large diamond at the bottom of the forest plot indicates the magnitude and direction of the effect size computed by the REML model (Viechtbauer, 2005; Viechtbauer, 2010). Studies with a positive effect size indicate a prolonged P300 peak latency in the patient group compared to the unaffected control group. Conversely, a negative effect size indicates a shorter P300 peak latency in the experimental group compared to the control group. The 95% confidence interval is rep-
resented by the whiskers extending from the box (the box corre-
sponding to study weight) (Viechtbauer, 2005; Viechtbauer, 2010).
Publication bias tests and a funnel plots were assessed in the meta-analysis to determine potential selection bias of positive studies (Rothstein, Sutton, & Borenstein, 2005). The qualitative funnel plot provides a measure of consistency in the outcome measurement among studies with varying methodology. Funnel plots that show a homogenous distribution of studies on both sides of the pooled effect size computed indicates statistically non-significant bias in the sample of studies (Rothstein et al., 2005). A non-parametric trim and fill analysis was used to estimate the number of studies missing on either side of the funnel plot. The numbers of studies missing were then simulated into the REML model to determine a more accurate estimate of the effect sizes of all studies considered in the analysis.

2.3. Moderator analysis

Table 1 provides a summary of the binary categories and definitions of moderator variables used in the meta-analysis comparing P300 latencies in patients with AD and unaffected controls. The dichotomous categories and definitions of moderator variables were constructed based on a logical approach described elsewhere (Polich, 1996; Polich, Pollock, & Bloom, 1994). Statistically, the moderator heterogeneity tests ($Q_b$ and $Q_w$) were analyzed using a fixed-effects moderator model. A fixed effects model was used based on the assumption that the moderator variables are relatively similar across all studies. The hypothesis of the model is that the moderator variable categories are equally homogenous (impact P300 effect sizes equally). If a significant outcome is observed then there is significant heterogeneity between moderator categories that likely represents a difference in the magnitude of the P300 effect size. The moderator variables analyzed in the meta-analysis were considered key factors that would likely contribute to the between study variability (related to the inconsistency seen in the literature) in P300 latency measurements based on previous meta-analyses of the P300 wave in other patients groups (Jeon & Polich, 2003).

The three-midline electrodes (Fz, Cz, Pz) were analyzed for mean differences in P300 peak latency measurement and to determine, which electrode elicits the strongest mean difference between patients and unaffected controls. A two-way ANOVA was used to compare electrode effects between groups, and subsequent one-way ANOVAs were used for pairwise comparisons of electrodes. For all statistical procedures, statistical significance was established at $\alpha = 0.05$.

3. Results

3.1. Search result

There were 48 studies that were identified in the search that met all inclusion criteria. There was 40 studies that compared P300 latencies between patients with AD and unaffected controls ($k = 9$ considering mild AD patients only). There were 8 studies comparing P300 latencies between patients with MCI and unaffected controls ($k = 5$ comparing patients with MCI and patients with AD). Collectively, there were 846 (137 mild AD) patients with AD and 956 unaffected controls in the 40 studies comparing patients with AD and unaffected controls. In the MCI P300 studies, there were 250 patients with MCI and 201 unaffected controls.

3.2. P300 latency in Alzheimer’s disease

In the first comparison analysis, the REML model reported severe prolongation of the P300 latency in patients with AD compared to unaffected controls at pooled midline electrodes (ES = 0.99, 95% CI: 0.82–1.19, $z = 10.71$, $p < 0.001$), and at Pz (ES = 1.07, 95% CI: 0.87–1.28, $z = 10.44$, $p < 0.001$). There was statistically significant heterogeneity present in the meta-analysis at both pooled midline electrodes ($tau^2 = 0.21$, $Q_b$ (39) = 111.58, $p < 0.001$), and at Pz ($tau^2 = 0.16$, $Q_b$ (26) = 68.39, $p < 0.001$). The adjusted rank correlation test (Kendall’s $tau = 0.10$, $p = 0.385$) and regression test ($z = 0.63$, $p = 0.527$) reported statistically non-significant funnel plot asymmetry of the pooled midline electrode meta-analysis. A qualitative assessment of funnel plot asymmetry validated the quantitative publication bias statistics (Fig. 5). There was one study (Boller et al., 2002) that deviated from the funnel plot, but this study was not considered a significant outlier in the meta-analysis (sensitivity test: $p > 0.05$). A forest plot of the studies included in the pooled midline electrode meta-analysis and their corresponding effect sizes are presented in Fig. 1. A summary of all meta-analysis results are shown in Table 3.

A trim-and-fill analysis reported that $k = 10$ studies were missing from the left side (shortened P300 latencies in AD compared to unaffected controls) of the funnel plot. The $k = 10$ studies were added to the $k = 40$ originally analyzed studies, and the REML model was re-fitted to observe the effect of the missing studies on the meta-analysis. There was a moderate reduction in the test statistic and effect size after adding the 10 missing studies to the pooled electrode meta-analysis. (ES: 0.79, 95% CI: 0.60–0.98, $z = 8.13$, $p < 0.001$). The heterogeneity tests were increased as a result of introducing 10 more studies to the analysis ($tau^2 = 0.37$, $Q_b$ (49) = 191.12, $p < 0.001$).

In the comparison of patients with only mild AD and unaffected controls, the P300 prolongation in mild AD patients was similar to that of the collective AD group at midline electrodes (ES: 0.78, 95% CI: 0.48–1.08, $z = 5.10$, $p < 0.001$), and at Pz (ES: 0.78, 95% CI: 0.33–1.23, $z = 3.43$, $p < 0.001$). The heterogeneity tests reported in the analysis were not statistically significant due to the small sample size at midline electrodes ($tau^2 = 0.29$, $Q_b$ (8) = 12.82, $p = 0.118$).

Table 1
Summary of the definitions and binary categories of the moderator variables.

<table>
<thead>
<tr>
<th>Moderator</th>
<th>Category</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample characteristics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender (years)</td>
<td>1 = Less males</td>
<td>$1 = n &lt; 60%$</td>
</tr>
<tr>
<td></td>
<td>2 = More males</td>
<td>$2 = n &gt; 60%$</td>
</tr>
<tr>
<td>Age</td>
<td>1 = Young</td>
<td>$1 = M &lt; 70$ years</td>
</tr>
<tr>
<td></td>
<td>2 = Older</td>
<td>$2 = M &gt; 70$ years</td>
</tr>
<tr>
<td>MMSE (severity)</td>
<td>1 = Moderate</td>
<td>$1 = M &lt; 21$</td>
</tr>
<tr>
<td></td>
<td>2 = Mild</td>
<td>$2 = M &gt; 21$</td>
</tr>
<tr>
<td>Medication</td>
<td>1 = “No”</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2 = “Yes”</td>
<td></td>
</tr>
<tr>
<td>Sample size (n)</td>
<td>1 = Small</td>
<td>$1 = n &lt; 20$</td>
</tr>
<tr>
<td></td>
<td>2 = Large</td>
<td>$2 = n &gt; 20$</td>
</tr>
<tr>
<td>EEG and oddball task parameters</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Target probability</td>
<td>1 = Low probability</td>
<td>$1 = p &lt; 0.20$</td>
</tr>
<tr>
<td></td>
<td>2 = High probability</td>
<td>$2 = p &gt; 0.20$</td>
</tr>
<tr>
<td>Target intensity</td>
<td>1 = Low intensity</td>
<td>$1 = \leq 80$ dB</td>
</tr>
<tr>
<td></td>
<td>2 = High intensity</td>
<td>$2 = &gt;80$ dB</td>
</tr>
<tr>
<td>Target frequency</td>
<td>1 = Low frequency</td>
<td>$1 = \leq 1500$ Hz</td>
</tr>
<tr>
<td></td>
<td>2 = High frequency</td>
<td>$2 = &gt;1500$ Hz</td>
</tr>
<tr>
<td>Target response</td>
<td>1 = “Button press”</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2 = “Mental count”</td>
<td></td>
</tr>
<tr>
<td>Tone duration</td>
<td>1 = Short</td>
<td>$1 = \leq 50$ ms</td>
</tr>
<tr>
<td></td>
<td>2 = Long</td>
<td>$2 = &gt;50$ ms</td>
</tr>
<tr>
<td>Stimulus difference</td>
<td>1 = Small</td>
<td>$1 = \leq 500$ Hz</td>
</tr>
<tr>
<td></td>
<td>2 = Large</td>
<td>$2 = &gt;500$ Hz</td>
</tr>
<tr>
<td>Sampling rate</td>
<td>1 = Low</td>
<td>$1 = \leq 500$ Hz</td>
</tr>
<tr>
<td></td>
<td>2 = High</td>
<td>$2 = &gt;500$ Hz</td>
</tr>
<tr>
<td>Interstimulus interval</td>
<td>1 = Short</td>
<td>$1 = \leq 1.5$ s</td>
</tr>
<tr>
<td></td>
<td>2 = Long</td>
<td>$2 = &gt;1.5$ s</td>
</tr>
<tr>
<td>High-pass filter</td>
<td>1 = High</td>
<td>$1 = \leq 0.5$ Hz</td>
</tr>
<tr>
<td></td>
<td>2 = Low</td>
<td>$2 = &gt;0.5$ Hz</td>
</tr>
</tbody>
</table>

Note: Stimulus difference refers to the difference (in Hz) between target and standard stimuli.
and at Pz (tau² = 0.14, $Q_k(4) = 9.22, p = 0.056$). A forest plot of the studies included in the pooled midline electrode meta-analysis and their corresponding effect sizes are presented in Fig. 2.

### 3.3. Moderator analysis: sample characteristics

Table 2 illustrates the effect of the moderator variables on the P300 latency effect sizes computed for the REML model comparing P300 latencies in patients with AD and unaffected controls. Similar effect sizes were observed for severity of disease (mild and moderate) as measured by the MMSE total score ($p = 0.866$). There were no significant medication effects observed between the studies. Studies with medication-free AD patients produced the strongest P300 latency effect size, but this finding was not statistically significant ($p = 0.113$). There was no significant difference in effect size when comparing proportion of male patients in the patient group ($p = 0.399$). Interestingly, P300 effect sizes were found to be strongest in patients younger than 70 years old ($p = 0.015$). Similar effect sizes were reported for sample size of the patients in the study, but did not reach statistical significance ($p = 0.796$).

### 3.4. Moderator analysis: EEG and oddball task parameters

Studies with a low probability target stimulus had larger effect sizes, but this trend was not significant ($p = 0.389$). In addition, studies with a high target frequency (>1500 Hz) contribute to sta-
The findings of this meta-analysis provide strong quantitative evidence that P300 latency is a statistically accurate measure of differences in brain activity between patients with AD (and MCI) and unaffected controls. According to Cohen’s general guidelines (Cohen, 1998), the effect sizes in meta-analyses comparing patients with AD (mild AD) and unaffected controls was strong (i.e. 0.80). Compared to unaffected controls, patients with MCI statistically significantly greater effect sizes (p = 0.030). Target stimuli delivered at low intensities contributes to greater effect sizes, but this finding was not statistically significant (p = 0.101). Longer interstimulus intervals (>1.5 s) contributed to statistically significant greater effect sizes in P300 studies (p = 0.008). Greater differences between target and standard stimuli produced larger effect sizes (p = 0.040). There was a small differential effect of long tone duration on effect size, but it was not statistically significant (p = 0.327). Similar to tone duration there was a small non-significant differential effect of studies with low frequency low pass filters (p = 0.229). Interestingly, higher EEG sampling rates were associated with significantly stronger P300 effect sizes (p = 0.006). Previous meta-analysis of the P300 failed to consider sampling rate as a significant influence on P300 latency measurements.

The final two task parameters tested in the moderator analysis were electrode reference (ear or other) and response (mental count or button press). Larger P300 effect sizes were observed in studies that had patient’s respond to target stimuli with a button press compared to a mental count of the stimuli. However, this finding was statistically non-significant (p = 0.146). Studies that used a linked earlobe reference in their EEG apparatus contributed to a significantly stronger P300 effect sizes compared to other reference sites (mastoid, mandible, etc.) (p = 0.002).

3.5. Electrode effects

The two-way ANOVA performed reported no statistically significant interaction observed between participant status (patient × unaffected control) and electrode (Fz × Cz × Pz) (F (2, 108) = 0.06, p = 0.947). The findings of the univariate ANOVA electrode comparisons were also statistically non-significant: Fz × Cz (F (1, 28) = 0.08, p = 0.785), Cz × Pz (F (1, 41) = 0.49, p = 0.490), Fz × Pz (F (1, 39) = 0.12, p = 0.732), Fz × Cz × Pz (F (2, 54) = 0.24, p = 0.787). When considering all meta-analysis studies, the P300 latency difference between patients and controls was largest at Fz (k = 14; AD: 420.98 ms, CTRL: 364.96 ms, Δdiff = 56.02 ms) compared to Cz (k = 16; AD: 417.78 ms, CTRL: 369.96 ms, Δdiff = 47.82 ms), and Pz (k = 27; AD: 416.18 ms, CTRL: 368.66 ms Δdiff = 47.52 ms).

3.6. P300 latency in mild cognitive impairment

Compared to unaffected controls, patients with MCI had moderately prolonged P300 latencies at midline electrodes (ES: 0.46, 95% CI: 0.18–0.75, z = 3.18, p = 0.002) and at Pz (ES: 0.53, 95% CI: 0.04–1.07, z = 2.12, p = 0.034). The heterogeneity tests were statistically significant in midline electrode analysis (tau² = 0.08, Qb (7) = 14.21, p = 0.048), and at Pz (tau² = 0.16, Qb (3) = 9.12, p = 0.034). Compared to patients with AD, patients with MCI had shorter P300 latencies, but the comparison was statistically non-significant (ES: −0.70, 95% CI: −1.45 to 0.05, z = −1.83, p = 0.067), and at Pz (ES: −0.87, 95% CI: −1.78 to 0.04, z = −1.87, p = 0.062). The heterogeneity test were statistically significant in the pooled electrode analysis (tau² = 0.63, Q(4) = 27.44, p = 0.001) and at Pz only (tau² = 0.75, Q(3) = 22.99, p = 0.006). The forest plot of the studies included in these pooled meta-analyses (MCI vs. unaffected controls, and MCI vs. AD) and their corresponding effect sizes are presented in Figs. 3 and 4, respectively.

4. Discussion

The findings of this meta-analysis provide strong quantitative evidence that P300 latency is a statistically accurate measure of differences in brain activity between patients with AD (and MCI) and unaffected controls. According to Cohen’s general guidelines (Cohen, 1998), the effect sizes in meta-analyses comparing patients with AD (mild AD) and unaffected controls was strong (i.e. 0.80). Compared to unaffected controls, patients with MCI...
Table 3

Summary of meta-analysis results.

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Site</th>
<th>k</th>
<th>ES</th>
<th>95% CI</th>
<th>z</th>
<th>p</th>
<th>Qb</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>AD vs. CTRL</td>
<td>Pz</td>
<td>27</td>
<td>1.07</td>
<td>0.87–1.28</td>
<td>10.44</td>
<td>&lt;0.001</td>
<td>Q(26) = 68.39, τ² = 0.16</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>Pooled</td>
<td>40</td>
<td>0.99</td>
<td>0.82–1.19</td>
<td>10.71</td>
<td>&lt;0.001</td>
<td>Q(39) = 111.58, τ² = 0.21</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>mAD vs. CTRL</td>
<td>Pz</td>
<td>5</td>
<td>0.78</td>
<td>0.33–1.23</td>
<td>3.43</td>
<td>&lt;0.001</td>
<td>Q(4) = 9.22, τ² = 0.14</td>
<td>0.056</td>
</tr>
<tr>
<td></td>
<td>Pooled</td>
<td>9</td>
<td>0.78</td>
<td>0.48–1.08</td>
<td>5.10</td>
<td>&lt;0.001</td>
<td>Q(8) = 12.82, τ² = 0.29</td>
<td>0.118</td>
</tr>
<tr>
<td>MCI vs. CTRL</td>
<td>Pz</td>
<td>4</td>
<td>0.53</td>
<td>0.04–1.07</td>
<td>2.12</td>
<td>0.034</td>
<td>Q(3) = 9.12, τ² = 0.16</td>
<td>0.034</td>
</tr>
<tr>
<td></td>
<td>Pooled</td>
<td>8</td>
<td>0.46</td>
<td>0.18–0.75</td>
<td>3.18</td>
<td>0.002</td>
<td>Q(7) = 14.12, τ² = 0.08</td>
<td>0.048</td>
</tr>
<tr>
<td>MCI vs. AD</td>
<td>Pz</td>
<td>4</td>
<td>–0.87</td>
<td>–1.78 to 0.04</td>
<td>–1.87</td>
<td>0.062</td>
<td>Q(3) = 22.99, τ² = 0.75</td>
<td>0.006</td>
</tr>
<tr>
<td></td>
<td>Pooled</td>
<td>5</td>
<td>–0.70</td>
<td>–1.45 to 0.05</td>
<td>–1.83</td>
<td>0.067</td>
<td>Q(4) = 27.44, τ² = 0.63</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Note: AD = Alzheimer's disease, CTRL = unaffected controls, mAD = patients with mild Alzheimer's disease, MCI = mild cognitive impairment, pooled = electrode sites Fz, Cz, and Pz were pooled into one mean P300 latency for patients and unaffected controls, k = number of studies, ES = effect size, Qb = heterogeneity between studies.

Fig. 3. A forest plot illustrating the effect sizes of the studies included in the meta-analysis comparing patients with MCI and unaffected controls (k = 8).

Fig. 4. A forest plot illustrating the effect sizes of the studies included in the meta-analysis comparing patients with MCI and patients with AD (k = 5).
exhibited a low-to-moderate effect size (i.e. 0.50). However, the meta-analysis was not able to provide strong statistical evidence for measuring brain activity differences between patients with MCI and patients with AD despite the large effect size computed. P300 latency effect sizes were also statistically sensitive to several moderator variables (mean age, ISI, stimulus difference, target frequency, reference electrode, and sampling rate). Clinical ERP research studies should focus on implementing optimal parameters to reduce the inter-lab variability in P300 latency measurements seen in patients with cognitive deficits.

The subtle onset of AD poses a difficult challenge for clinicians to accurately diagnose and implement treatment to slow the progression of the disease (Lee et al., 2013). However, the moderator variables provided some novel results in eliciting P300 latency in patients that are in the prodromal phase of AD. The sensitivity of moderator variable on P300 latency effect sizes also indicates that clinical ERP investigators must carefully select the appropriate sample characteristics, EEG parameters, and task settings to elicit the strongest differences between patients and unaffected controls.

In terms of sample characteristics, clinical ERP studies should employ the auditory oddball task in patients that are less than 70 years of age. We hypothesize that individuals greater than 70 years of age are likely to have more neural degeneration due to aging that may make oddball tasks too demanding on cognitive resources to produce a reliable difference between patients and unaffected controls (Goodin, 1986; Juckel et al., 2008; Muscoso et al., 2006; Pfefferbaum, Wenegrat, et al., 1984; Schiff et al., 2008). This finding also supports the notion that the P300 latency may be more accurate in the prodromal phase where patients are typically younger than 70 years of age (Bennys et al., 2007; Juckel et al., 2008; Sperling et al., 2011). The other sample characteristics that did not reach statistical significance are likely to have an effect on the P300 latency measurement, but not large enough to constitute an observable difference in effect size between patients and unaffected controls.

Statistically significant stimulus parameters (ISI, stimulus difference, and target frequency) that contributed to stronger effect sizes indicate the importance of stimulus presentation on eliciting P300 latency (Polich, 2007; Polich & Corey-Bloom, 2005; Polich & Pitzer, 1999). It appears that the oddball task must not be too cognitively challenging for the patient. Stronger effect sizes were associated with longer ISI, larger stimulus difference, and louder target stimulus frequencies. We hypothesize that this may be associated with the attentional and cognitive resources available to the patient (Goodin, 1986; Pfefferbaum, Ford, et al., 1984; Pfefferbaum, Wenegrat, et al., 1984; Picton et al., 1984; Polich & Herbst, 2000). Patients with cognitive dysfunction may only be able to direct few resources to detecting the target stimulus, and therefore require the target stimulus to be easily detectable from the standard stimulus (Picton & Hillyard, 1974, Picton et al., 1984). The longer ISI is indicative of deficits in task switching and executive function, and it is possible that patients require more time to process the stimulus before proceeding to the next stimulus (Donchin, 1987; Johnson et al., 1985; Lee et al., 2013).

EEG recording methodology also appears to have a statistically significant influence on the P300 latency measurement (Jeon & Polich, 2003). In this meta-analysis, reference electrodes attach to earlobes appear to produce greater neural differences between patients and unaffected controls. However, the comparison category was an aggregation of other reference sites and therefore it cannot be ascertained whether the earlobe is the most effective reference site (Jeon & Polich, 2003). Sampling rate was also a significant finding in the moderator analysis. Sampling rate is a reflection of how many peaks are averaged per second during the EEG recording. However, we suspect that the larger effect size associated with faster sample rates is likely due to the fact that this parameter correlates to other inter-lab differences rather than a direct effect of the sample rate on the study precision. Some other factors that may have influenced the large effect size observed for sample rate include: low pass filtering (and other EEG parameters), time of study (earlier vs. later reports), and different diagnostic criteria.

The results of the ANOVAs were unable to produce statistically significant differences between electrode sites. However, the larger mean difference at the Fz electrode site contrasts previous published literature (Ford et al., 1994; Jeon & Polich, 2003; Polich, 1996, Polich & Pitzer, 1999). Although, it is important to consider that fewer studies measured P300 latency at the Fz electrode and the finding may be biased. When considering the unequal sample sizes compared, it is likely that there are no differences in P300 latency measurements at the midline electrode sites. However, a larger effect size was estimated in the secondary Pz electrode analysis when comparing P300 latency in patients with mild cognitive impairment and Alzheimer’s disease. Future studies comparing patients with MCI and AD may want to focus on only the Pz electrode for greater differences in P300 latency between patient groups.

Based on the inferences made in the literature and previous meta-analyses, there appears to be three common factors that influence the measurement of the P300 waveform in patients with cognitive dysfunction. (1) Structural neurodegeneration in temporal lobes and other areas may be affecting P300 resolution due to the association with stimulus perception and discrimination (Frodl et al., 2002; Gungor et al., 2005; Jiménez-Escrig et al., 2002; Juckel et al., 2008; Muscoso et al., 2006). (2) The heterogeneity of cognitive dysfunction and the specific domains affected by AD and MCI often increase the variability in P300 latency measurement (Ball et al., 1989; Gungor et al., 2005; Lai et al., 2010; Lee et al., 2013; Pfefferbaum, Wenegrat, et al., 1984; Pokrzywko-Dragan et al., 2003; Polich, 1989; Polich et al., 1986; Sundukolo et al., 1982; Szellies et al., 1995; Tanaka et al., 1998; Van Deursen et al., 2009; Wright et al., 1988). (3) The difficulty of the oddball task, in terms of identifying the target stimulus and allocating enough resources to elicit a measurable P300 latency. These three factors must be considered, to effectively create the most sensitive P300 latency paradigm specifically for patients with AD and MCI (Donchin, 1987; Johnson et al., 1985; Lee et al., 2013; Picton et al., 1984).
The strongest limitation of the present study is the small sample size used in the meta-analyses. Therefore, it is important to review the present findings with caution when applying them in subsequent clinical ERP studies. The meta-analysis is not able to overcome the limitations presented in the individual research studies. Lastly, the findings are representative of the effects of moderators on P300 latency effect sizes and may not be directly generalizable to individual P300 measurements of patients.

5. Conclusion

This meta-analysis supports the use of the auditory P300 latency subcomponent as a biological marker of prodromal AD. It has been shown to consistently discriminate between patients with AD (and even mild AD) (Bennys et al., 2007; Juckel et al., 2008) and controls, and unaffected controls. It has also been shown to improve the sensitivity and specificity of amyloid assays (Papaliagkas et al., 2008; Papaliagkas et al., 2011), and unaffected controls. The findings in this meta-analysis support the use of the P300 latency subcomponent in clinical ERP research for patients with AD. More studies are required to fully elucidate the usefulness of the auditory P300 component in patients with MCI.

Acknowledgment

None.

References


