HHS Public Access

Author manuscript

Cogn Behav Neurol. Author manuscript; available in PMC 2023 September 01.

Published in final edited form as:

Cogn Behav Neurol. ; 35(3): 188–197. doi:10.1097/WNN.0000000000000311.

Distinguishing Between Genuine and Feigned Dementia Using Event-related Potentials

August M. Price, PsyD[*,†], **Rocco Palumbo, PhD**[*,‡], **Anna Marin, MS**[*], **Prayerna Uppal, BA**[*], **Cheongmin Suh, MS**[*], **Andrew E. Budson, MD**[*,‡], **Katherine W. Turk, MD**[*,‡]

[*]Center for Translational Cognitive Neuroscience, Veterans Affairs Boston Healthcare System, Boston, Massachusetts

[†]Department of Clinical Psychology, William James College, Newton, Massachusetts

[‡]Boston University Alzheimer's Disease Research Center, Department of Neurology, Boston University School of Medicine, Boston, Massachusetts.

Abstract

Background: Individuals with probable Alzheimer disease (AD) may perform below cutoffs on traditional, memory-based performance validity tests (PVTs). Previous studies have found success using event-related potentials (ERPs) to detect feigned neurocognitive impairment in younger populations.

Objective: To evaluate the utility of an auditory oddball task in conjunction with the P3b peak amplitude to distinguish probable AD from simulated dementia.

Method: Twenty individuals with probable AD and 20 older healthy controls (HC) underwent an ERP auditory oddball protocol and the Test of Memory Malingering (TOMM). The HC were asked to perform honestly for one condition and to simulate dementia for the other. The individuals with probable AD were asked to perform honestly. The P3b peak amplitude and button press accuracy were collected from each participant and were analyzed to determine their effectiveness in detecting performance validity.

Results: The P3b peak amplitude remained stable regardless of behavioral condition in the HC group. When combined with the TOMM Trial 2 score, the P3b peak amplitude further improved the ability to correctly differentiate individuals with probable AD from HC simulating dementia with 100% sensitivity and 90% specificity.

Conclusion: The P3b peak amplitude was found to be an effective physiologic measure of cognitive impairment in individuals with probable AD compared with HC simulating dementia. When combined with the TOMM Trial 2 score, the P3b peak amplitude served as a

[Corresponding author: August M. Price, 150 South Huntington Avenue, Boston, MA 02130-4817; Phone number: 413-824-1715; august.price@utsouthwestern.edu].

The authors declare no conflicts of interest.

promising performance validity measure for differentiating individuals with probable AD from HC simulating dementia.

Keywords

performance validity test; event-related potential; P300; Alzheimer disease; dementia

Ensuring that a patient is responding honestly during a neuropsychological evaluation is integral to obtaining objective and reliable results. Although many patient populations are able to use forced-choice memory paradigms effectively to achieve near perfect scores on performance validity tests (PVTs; Green, 2003; Green et al, 1996; Tombaugh, 1997), dementia patients with prominent amnestic deficits may have difficulty performing above cutoffs on this type of memory-based PVT. Few dementia samples have been included in the vast majority of performance validity research (Bortnik et al, 2013), and studies that have evaluated the efficacy of forced-choice memory-based PVTs in dementia populations have produced inconsistent outcomes. Thus, the clinical applicability of memory-based PVTs among individuals with suspected dementia is currently limited by these factors.

Evaluation of the performance validity of individuals with probable Alzheimer disease (AD) or other dementias is a particular clinical challenge given the nature of these individuals' memory impairments; individuals with suspected dementia who undergo neuropsychological testing may "fail" PVTs despite providing optimal effort. Nonetheless, clinical neuropsychologists and other professionals routinely administer PVTs to individuals with dementia despite evidence that a memory-based PVT could falsely characterize an individual with dementia as providing poor effort. Thus, it remains clinically relevant to improve on detection accuracy rates for novel and previously established PVTs. A related issue is that individuals who are being evaluated for neurologic conditions may feign impairment on cognitive testing due to contextual incentives such as disability status, compensation claims, and/or litigation (Teichner and Wagner, 2004).

Unfortunately, the detection accuracy of PVTs with dementia populations varies, with, for example, specificity of the Test of Memory Malingering Trial 2 (TOMM–2; Tombaugh, 1996) ranging between 24% (Teichner and Wagner, 2004) and 82% (Greve et al, 2006). In a mixed dementia sample, the Medical Symptom Validity Test (Green, 2003) was found to produce specificity rates between 17% and 61% (Howe et al, 2007). Using a cutoff of <20 items on the recognition trial and <9 items on the free recall trial for the Rey 15-Item test (Lezak et al, 2012), Dean et al (2009) discovered specificity rates of 14% and 26% in a non-litigating dementia sample. Finally, Merten et al (2007) discovered that up to 95% of an AD sample did not perform above cutoffs on the Word Memory Test (Green et al, 1996).

The purpose of the present study was to evaluate the utility of a three-tone auditory oddball event-related potential (ERP) protocol to differentiate individuals with probable AD who have true cognitive impairment from older, cognitively intact adults who were simulating dementia. Past PVT studies have used P300 amplitude derived from forced-choice or old/new recognition paradigms to detect feigned neurocognitive impairment in adult populations. The P300 ERP component is a positive deflection in a recorded waveform that occurs anywhere between 300 ms and 800 ms after an eliciting event; it is

typically provoked by meaningful and/or rarely occurring stimuli, depending on stimulus modality (Ellwanger et al, 1999; Polich, 2007). The P3b, a subcomponent of the P300 ERP component, is thought to reflect a nonvoluntary response to recognized items and is proposed to be modulated by the amount of attentional resources that are allocated when working memory is updated (Donchin and Coles, 1988).

Studies examining ERP malingering have typically found the P3b peak amplitude to be unaffected by deceptive responding, suggesting that although individuals may be able to manipulate their performance on neuropsychological paper and pencil tests, they are less able to manipulate their neurophysiologic responses (Ellwanger et al, 1996, 1999; Tardif et al, 2000; van Hooff et al, 2009). Previous studies have also shown that individuals with mild AD dementia are able to complete a P300 auditory oddball protocol and that the P3b is able to differentiate individuals with mild AD dementia from older, healthy controls (HC). Typically, the P3b peak amplitude is reduced in individuals with AD compared with HC (Cecchi et al, 2015; Cid-Fernández et al, 2014; Polich and Corey-Bloom, 2005; van Deursen et al, 2009).

We hypothesized that HC, both when they are performing honestly and when they are simulating dementia, would have increased P3b peak amplitudes compared with individuals with probable AD, indicating underlying cognitive health. To our knowledge, this is the first study to examine the utility of an auditory oddball ERP protocol, as opposed to a task that relies primarily on episodic memory, as a potential PVT measure to differentiate individuals with genuine dementia from those with simulated dementia.

METHOD

Participants

We recruited patients from the Veterans Affairs (VA) Boston Healthcare System Neurology Memory Disorders Clinic between March 2017 and February 2018 who met the National Institute on Aging–Alzheimer's Association clinical diagnostic criteria for either mild AD dementia (McKhann et al, 2011) or mild cognitive impairment due to AD (Albert et al, 2011). We referred to this combined cohort as having *probable AD.* HC were recruited by either word of mouth or referral from the Boston University Alzheimer's Disease Center during the same time frame. Participants were compensated up to $60 for their participation in the study.

The study protocol was approved by the VA Boston Healthcare System Institutional Review Board and was performed according to the ethical guidelines of the Declaration of Helsinki and its later amendments. All individuals provided informed written consent before enrolling in the study.

Participant Screening

To be enrolled in this study, participants had to be within the age range of 55 to 100 years. Following recruitment, a cognitive behavioral neurologist (A.E.B. or K.W.T.) evaluated the individuals with probable AD to determine if they met the study inclusion/exclusion criteria. Each evaluation included a medical history; neurologic examination; and review

of laboratory studies, neuropsychological test results, and neuroimaging (eg, brain MRI or head CT). A.E.B. and K.W.T. used neuroimaging to assist in determining the clinical status (mild cognitive impairment due to mild AD dementia vs AD) of each of the individuals with probable AD.

We administered the following neurocognitive test battery to all of the study participants in order to assist in determining their current cognitive status:

- the Consortium to Establish a Registry for Alzheimer's Disease Word List Memory Test (Morris et al, 1989) to assess memory,
- the Verbal Fluency Test (Category and Letter Fluency; Lezak et al, 2012) to assess aspects of executive functions and semantic language,
- the Boston Naming Test—Short Form (Kaplan et al, 2001; Mack et al, 1992) to assess semantic word knowledge,
- Trail-Making Tests A and B (Reitan, 1958) to assess rapid scanning and set-shifting/divided attention,
- the Mini-Mental State Examination (Folstein et al, 1975) to assess global cognitive abilities,
- the Geriatric Depression Scale—Short Form (Yesavage and Sheikh, 1986) to assess depression,
- the Geriatric Anxiety Inventory (Pachana et al, 2007) to assess anxiety, and
- the Wechsler Test of Adult Reading (Wechsler, 2001) to assess premorbid intelligence using single-word reading.

All normative data were derived from Weintraub et al (2009).

To be enrolled in the study, the individuals with mild AD dementia had to perform at least 1.5 SDs below the mean for age- and education-adjusted norms on two or more neurocognitive tests (McKhann et al, 2011), and the individuals with mild cognitive impairment due to AD had to perform at least 1.5 SDs below the mean for age- and education-adjusted norms on at least one neurocognitive test (Albert et al, 2011).

Inclusion criteria for the HC included performance within or above 0.0–1.0 SD for age- and education-adjusted norms on the neurocognitive test battery and a Mini-Mental State Examination score ≥26. HC who did not score within 1.0 SD of the mean on all of the neurocognitive tests were excluded, thereby eliminating HC with cognitive decline due to vascular and other potential comorbidities.

Additional exclusion criteria for both groups (probable AD and HC) were major psychiatric disorder, clinically significant alcohol or substance abuse, clinically significant cerebrovascular disease, traumatic brain injury, and clinically relevant neurologic diseases (other than AD for the probable AD group).

Experimental Paradigm

We used a three-tone auditory oddball ERP protocol and the TOMM to test the study participants. The HC completed two visits ~1 week apart during which they performed honestly or simulated dementia, counter-balanced for condition order. For the simulating dementia condition, we gave the HC a checklist of common dementia symptoms that are easily discoverable through internet search engines (supplementary digital content, http://links.lww.com/CBN/A118). Then, we asked the HC to "convincingly" simulate the performance of an individual with memory impairments while completing the auditory oddball ERP protocol and the TOMM. The symptom checklist was reviewed and approved by A.E.B. and K.W.T. We instructed the HC to complete the tests honestly for the performing honestly condition. We instructed the individuals with probable AD to complete the auditory oddball ERP paradigm and the TOMM honestly.

The stimuli for the auditory oddball protocol consisted of standard tones (1000 Hz), target tones (2000 Hz), and distractor tones (white noise) that occurred with probabilities of 0.75, 0.15, and 0.10. The auditory tones were presented in pseudorandom order for ~15–20 minutes depending on the number of artifacts incurred. The participants were instructed to respond to the target stimuli by pressing a button as fast as possible with their dominant hand and to refrain from pressing the button for nontarget stimuli (ie, standard tones and distractor tones).

For each test, 400 total stimuli were presented binaurally through insert earphones at a perceived 70-dB volume. There was a minimum of 21 correct target tone trials for each participant based on a minimum button press accuracy of 35%, resulting in adequate trials to obtain good quality ERP data for each group (Hayama et al, 2008; Tsivilis et al, 2001). For each stimulus, a tone duration of 100 ms was presented, with rise and fall times of 10 ms. The interstimulus gap was between 2.5 and 3.0 seconds.

Before administering the auditory oddball ERP protocol to each participant, we used COGNISION system software to administer a pure-tone audiometry test that was similar to the auditory oddball protocol, including standard tones amplified to 1000 Hz and target tones amplified to 2000 Hz. The audiometry test was included to ensure that the participants could distinguish the auditory stimuli (three tones) adequately. Auditory stimuli were amplified to compensate participants with <30 dB of hearing loss at any frequency. Because our study protocol used an auditory oddball task exclusively, we did not assess for visual disturbance.

Assessments

We obtained EEG activity from seven electrode sites (Fz, Cz, Pz, F3, P3, F4, and P4) that were connected to mastoid electrodes at M1 and M2, using a COGNISION headset worn by each participant. Fpz served as the common electrode. The COGNISION headset has been validated to obtain reliable ERP recordings when skin contact impedance is <70 kΩ (Cecchi et al, 2015). Impedance checks were automated at all electrodes after the target and distractor tones and were kept at <70 kΩ throughout the protocol. In order to ensure a sufficient sampling rate, data were collected from –240 to 1000 ms around the stimuli,

digitized at 125 Hz, and bandpass filtered from 0.3 to 35 Hz (Bougrain et al, 2012). An artifact threshold limit of ±100 μV was set for all of the trials. Trial sets of target or distractor tones and the immediately preceding standard tones with artifacts surpassing the threshold were excluded in real time and were repeated. Extraction and trial averaging of ERP measures was completed using COGNISION system software (Cecchi et al, 2015).

The TOMM is a standalone PVT that is designed to detect cognitive feigning or exaggeration. At study, participants are shown 50 items one at a time followed by 50 recognition panels at test, with two options per panel: one previously presented target item and one foil item. At test, participants are asked to select the previously presented stimuli from each panel. The same 50 items are presented in a different order for Trial 2. An optional retention trial may be administered 15 minutes later without re-displaying the 50 original items. Forty-five or more correct responses on Trial 2 out of a possible 50 at test suggests a valid cognitive performance (Tombaugh, 1997).

Statistical Analysis

P3b peak amplitude was analyzed for HC performing honestly, HC simulating dementia, and the probable AD group and was measured as the difference between the mean pre-stimulus baseline and the maximum peak amplitude. P3b peak amplitude was measured from the target tone response to correct trials only and was defined as the maximum positivity between 325 and 580 ms. The time window was determined by reviewing individual and group grand averages, and P3b peak amplitudes were obtained by averaging all of the electrodes (Cecchi et al, 2015).

The TOMM Trial 2 scores were analyzed; Trial 1 scores were not included in the statistical analysis because Trial 1 is not considered a stand-alone PVT (Tombaugh, 1996). Button press accuracy (hit rates and false alarms) for the target and nontarget (standard and distractor) auditory tones during the auditory oddball protocol were also analyzed. Hits were calculated as the percentage of correct responses to the target tones. False alarms were defined as button presses to nontarget tones.

In order to account for the HC completing the ERP auditory oddball protocol across two visits (versus one visit for the probable AD group), once performing honestly and once simulating dementia, group comparisons were analyzed using a repeated measures ANOVA. Individual univariate analyses of covariance were (a) performed between the probable AD group and the HC simulating dementia condition, (b) performed separately for the probable AD group and the HC performing honestly condition, and (c) corrected for multiple comparisons using Bonferroni correction. All data analyses were performed while correcting for age and education.

To better understand how ERPs and button press accuracy can determine whether a participant had probable AD or was an HC simulating dementia, a series of logistic regressions was performed. We then analyzed the binary logistic regression scores through receiver operating characteristic (ROC) curves to determine the specificity and sensitivity of these measures to differentiate individuals with probable AD from HC simulating dementia.

RESULTS

Participant Demographics and Clinical Data

Twenty individuals with probable AD met the study criteria and participated in the study—15 with mild AD dementia and five with mild cognitive impairment due to AD. None of the individuals dropped out of the study, and all of the data were usable. Of the 23 HC who were recruited, two did not return for the second study session, and one performed below 1.0 SD of the mean on the neurocognitive test battery, yielding a total of 20 HC who participated in the study. The participants were between 58 and 90 years of age.

As shown in Table 1, all of the participants were male, and there were no significant differences in age or self-reported depression symptoms (Geriatric Depression Scale—Short Form) between the two groups (all *P*s > 0.05). A self-report checklist of anxiety (Geriatric Anxiety Inventory) showed significantly higher levels of anxiety-type symptoms for the probable AD group compared with the HC. Educational attainment and estimated verbal intelligence (Wechsler Test of Adult Reading) were significantly higher for the HC compared with the probable AD group. As expected, a comparison of neurocognitive test battery data between the groups indicated significantly lower scores (*P*s < 0.001) for the probable AD group compared with the HC on all of the neurocognitive tests, including the Consortium to Establish a Registry for Alzheimer's Disease, Verbal Fluency Test, Boston Naming Test—Short Form, Trail-Making Tests A and B, and Mini-Mental State Examination (Table 1).

P3b Peak Amplitude

The grand average waveforms for the correct P3b target tone selections are presented for each of the seven electrodes in Figure 1. A comparison of the P3b peak amplitude for all of the electrodes averaged (Figure 2) revealed no significant differences between the HC in the performing honestly (M = 7.41 μV; SD = 3.23) and simulating dementia (M = 6.45 μV; SD = 3.54) conditions ($F_{1, 19} = 0.99$; $P = 0.33$, $\eta_p^2 = 0.05$). When corrected for age and education, the probable AD group (M = 3.00 μV; SD = 3.36) showed significantly decreased P3b peak amplitudes compared with the HC in the performing honestly ($F_{1, 36} = 7.49$; $P = 0.01$, $\eta_p^2 = 0.17$) and simulating dementia ($F_{1, 36} = 5.18$; $P < 0.05$, $\eta_p^2 = 0.13$) conditions, which remained significant after Bonferroni correction for multiple comparisons.

TOMM Trial 2

We observed significant differences in TOMM–Trial 2 scores between the HC in the simulating dementia condition and the probable AD group ($F_{1, 39} = 44.21$; $P < 0.001$, $\eta_p^2 = 0.02$) and between the HC in the performing honestly and simulating dementia conditions ($F_{1, 19} = 128.06$; $P < 0.001$, $\eta_p^2 = 0.87$). Differences between the HC in the performing honestly condition and the probable AD group approached significance ($F_{1, 39} = 3.85$; $P = 0.057$, $\eta_p^2 = 0.10$). As expected, the HC in the performing honestly condition performed best (M = 50.00; SD = 0.00), followed by the probable AD group ($M = 45.50$; SD = 8.19), and finally the HC in the simulating dementia condition ($M = 29.50$; $SD = 8.10$). Figure 3 shows individual TOMM Trial 2 performances for the probable AD group and the HC in the performing honestly and simulating dementia conditions, revealing that in several instances,

the probable AD group scored in the same range as the HC simulating dementia, whereas the HC performing honestly performed uniformly at ceiling.

Auditory Oddball Accuracy

Inspection of target tone accuracy revealed significant differences between the HC in the performing honestly and simulating dementia conditions ($F_{1, 19} = 65.11$; $P < 0.001$, $\eta_p^2 = 0.77$), the HC in the performing honestly condition and the probable AD group ($F_{1, 39} = 4.79$; $P < 0.05$, $\eta_p^2 = 0.12$), and the HC in the simulating dementia condition and the probable AD group ($F_{1, 39} = 25.93$; $P < 0.001$, $\eta_p^2 = 0.42$). The HC in the performing honestly condition performed almost without error (M = 97.05; SD = 4.51), followed by the probable AD group (M = 82.84; SD = 19.80), and finally the HC in the simulating dementia condition (M = 46.75; SD = 18.76).

Regression and ROC Analyses for Probable AD Group Versus HC Simulating Dementia

The probable AD versus HC simulating dementia comparison was most meaningful for the group and condition comparisons because it allowed us to address the initial hypothesis that HC simulating dementia would have increased P3b peak amplitudes compared with the probable AD group. We performed a series of binary logistic regression models including age, education, TOMM Trial 2 scores, and P3b peak amplitudes as the variables (Table 2).

The regression scores for Models 1 and 3 listed in Table 2 were submitted to a ROC curve analysis (Figure 4) in order to evaluate the effect of adding the P3b peak amplitude on the sensitivity and specificity for accurately differentiating between probable AD and HC simulating dementia. The model including both the TOMM Trial 2 score and the P3b peak amplitude (Model 3, Figure 4B) led to a sensitivity of 100% and a specificity of 90%, compared with 95% and 80%, respectively, for the model using the TOMM Trial 2 score alone (Model 1, Figure 4A). The area under the curve increased when the P3b peak amplitude was added to the ROC curve analysis, though not significantly, when comparing Model 1 to Model 3 ($z = 0.54$, $P = 0.59$).

Using cutoff scores determined by the ROC curve for the TOMM Trial 2 score and P3b peak amplitude combined, and the validated cutoff score for the TOMM Trial 2 (Tombaugh, 1996, 1997), we determined the false positive (FP) and false negative (FN) rates of each PVT measure in classifying participants performing honestly versus those simulating dementia. The TOMM Trial 2 score and P3b peak amplitude in combination produced FP rates equal to 2.5% and FN rates equal to 2.5%, whereas the TOMM Trial 2 score alone produced FP rates equal to 7.5% and FN rates equal to 5%.

We then compared the FPs and FNs using a χ^2 analysis. There were no significant differences between the FPs for the combination of TOMM Trial 2 score and P3b peak amplitude compared with the TOMM Trial 2 score alone ($\chi^2 = 0.51$, $P = 0.47$), nor were there any significant differences between FNs across the measures ($\chi^2 = 0.17$, $P = 0.68$).

Within the probable AD group, TOMM Trial 2 score alone showed 90% accuracy in correctly identifying participants as having probable AD compared with 95% accuracy for the combination of TOMM Trial 2 score and P3b peak amplitude ($\chi^2 = 0.35$; $P = 0.55$).

DISCUSSION

The overall aim of this study was to explore the utility of using an auditory oddball ERP protocol for distinguishing probable AD from simulated dementia in a geriatric population. We found that decreased P3b peak amplitude among individuals with probable AD compared with HC simulating dementia was a significant predictor of AD diagnosis. When used in combination with the TOMM Trial 2 score, the P3b peak amplitude was able to improve the sensitivity and specificity of detecting simulated dementia. In keeping with prior literature showing that the P3b component remains stable when HC feign cognitive impairment (Ellwanger et al, 1996, 1999; Tardif et al, 2000; van Hooff et al, 2009), our findings suggest that neurophysiologic electrical activity is relatively unaffected by deceptive responding and perhaps less prone than performance-based PVTs to being intentionally manipulated by an individual simulating dementia. Thus, the P3b peak amplitude appears to be a potentially viable measure for differentiating individuals with probable AD from those simulating dementia when it is used in combination with the TOMM Trial 2 score.

As expected, the TOMM Trial 2 score was able to distinguish well between HC in the performing honestly (M = 50.00, SD = 0.00) and simulating dementia conditions (M = 29.50, SD = 8.10); however, the TOMM Trial 2 score was not as useful in distinguishing individuals with probable AD (M = 45.50, SD = 8.19) from HC simulating dementia. While it is true that it may not be common for healthy individuals to feign dementia, the inverse relationship is more common, wherein individuals with AD are incorrectly labeled as providing suboptimal effort due to below-cutoff TOMM scores since the TOMM relies primarily on memory, which tends to be impaired for individuals with AD. Additionally, there are relatively rare instances (<2% of those seeking compensation) where older individuals may malinger or feign dementia for secondary gain related either to service connection benefits in the VA system or to disability decisions related to cognitive impairment in other national health care systems outside the United States (Mittenberg et al, 2002).

Regarding behavioral performances during the auditory oddball task, the HC performing honestly produced the fewest number of errors (M = 97.05, SD = 4.51). The HC simulating dementia (M = 46.75, SD = 18.76) and the probable AD group (M = 82.84, SD = 19.80) performed less accurately on the task compared with the HC performing honestly. These results, in addition to group differences on the TOMM Trial 2 score (ie, HC performing honestly perfectly, followed by the probable AD group who, in some instances, did not meet cutoffs, and finally by HC simulating dementia performing at the lowest levels), ensure that the HC simulating dementia group was truly performing as expected following instructions to simulate dementia. However, the TOMM Trial 2 score may, in fact, have overperformed in our study as a predictor of dementia status when it was used to predict HC simulating dementia versus individuals with probable AD.

A typical malingerer in a real-world clinical setting—possibly researching how to “fool” the test using online articles or being coached by unscrupulous individuals—would likely score more similarly to the probable AD group, potentially missing only a few items and

therefore performing closer to cutoffs. Thus, in a real-world clinical setting, the predictive ability of the TOMM Trial 2 score may be less robust compared with the ROC curve that was generated from our analysis for individuals with probable AD versus HC simulating dementia with a relatively high sensitivity and specificity.

To compare the utility of the P3b peak amplitude to differentiate individuals with probable AD from HC simulating dementia, we combined the P3b peak amplitude with the TOMM Trial 2 score in an effort to improve the latter's reliability among dementia populations. We found 100% sensitivity and 90% specificity in differentiating individuals with probable AD from HC simulating dementia using the P3b peak amplitude and TOMM Trial 2 score in combination compared with 95% sensitivity and 80% specificity using the TOMM Trial 2 score alone. According to Seshan et al (2013), when comparing two predictors pulled from nested logistic regression models, the area under the curve test may be biased, greatly reducing its discriminatory ability and reducing sensitivity. Thus, while the ROC curve analyses between Models 1 and 3 did not produce significant differences, we present the sensitivity and specificity of each model in order for the reader to gauge the potential real-world contribution of combining the measures. Future studies should consider using this potential biomarker in more diverse samples to better generalize the findings.

Five of the 20 individuals with probable AD performed below the cutoffs on the TOMM Trial 2. In contrast, only one of the 20 individuals with probable AD was incorrectly categorized by the P3b peak amplitude in combination with the TOMM Trial 2. To our knowledge, this is the first study to demonstrate the potential utility of employing ERPs as a performance validity measure for differentiating individuals with probable AD from HC simulating dementia.

Study Limitations

This study had a number of limitations including that it was conducted in a predominately male veteran population. One of the challenges with any PVT study is generalizing the performance of the participants who were instructed to feign neurocognitive impairment without identifiable incentives. Compensation-seeking individuals have been shown to produce more subtle deficits than a simulating cohort (Rogers and Vitacco, 2002).

In addition, there may be differences in behavioral performance between simulated malingering and true malingering; for example, a true malingerer might exaggerate symptoms to avoid detection rather than flagrantly feign symptoms. Our protocol encouraged the participants to "convincingly" simulate dementia based on prior knowledge and the symptom checklist we provided. However, no additional incentives were offered other than compensation for participating. Thus, the commitment to accurately feign dementia was left entirely up to each participant.

Additionally, our study did not account for suboptimal effort as a result of general fatigue, lack of interest, or failure to appreciate the implications of a poor performance, as is common in geriatric and other populations (Bauer and McCaffrey, 2006). Future studies may include a study group where individuals complete the auditory oddball protocol while

dividing their attention toward a distractor task in order to simulate genuine inattention on testing in order to determine its influence on the P3b response.

Our HC group had a statistically higher level of education than our probable AD group. Years of education or better cognitive performance has been shown to be associated with an elevated P3b peak amplitude (Begum et al, 2014; Daffner et al, 2011; Riis et al, 2008), potentially skewing results toward higher P3b peak amplitudes for the HC. Although we corrected for education during analysis, future studies should attempt to match participants based on education.

Finally, our P3b values were obtained solely from correct trials (as opposed to target tone misses). Future research may consider incorporating a P3a response, which is similar to the P3b response but is elicited specifically from distractor (false alarm) tones, as one study found reliable but significantly reduced P300 peak amplitudes for responses to incorrect stimuli when participants feigned neurocognitive impairment on a multiple-choice recognition task (Shelley-Tremblay et al, 2019). This phenomenon has been called a *conscious suppression mechanism* and is tentatively proposed to indicate an individual's attempt to resist the conflict between their plan for performing a feigned response and their true knowledge of the correct response (van Hooff et al, 2009).

CONCLUSION

The current study used a three-stimulus, auditory oddball ERP protocol as part of a novel PVT paradigm that integrated P3b peak amplitudes and standard PVT cutoffs for the detection of simulated dementia. To our knowledge, this was the first performance validity study in a probable AD and HC population to examine the utility of an auditory oddball ERP protocol versus a task that relies primarily on episodic memory. ERPs were also an important component due to being potentially less manipulatable than traditional performance-based measures. We found that the P3b peak amplitude remains relatively stable in individuals regardless of behavioral test performance. We demonstrated that the combination of using the P3b peak amplitude and the TOMM Trial 2 score was highly sensitive and specific in differentiating individuals with probable AD from HC feigning neurocognitive impairment. Future research could explore whether this novel PVT can be applied to other clinical settings and populations.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

The authors thank Marco Cecchi, PhD (Neuronetrix, Louisville, Kentucky), for providing statistical guidance, software development, and proofreading of the manuscript and Jason Osher, PhD (William James College, Newton, Massachusetts), for assistance in developing the initial study proposal.

Supported in part by a grant (AACF-16-443347) from The Alzheimer's Association and a Veterans Affairs Career Development Award (CX002065) to K.W.T; a Veterans Affairs Merit Award (CX001698) to A.E.B.; and a grant (P30-AG072978) from the National Institutes of Health/National Institute on Aging.

Glossary

AD	Alzheimer disease
ERP	event-related potential
FN	false negative
FP	false positive
HC	healthy controls
PVT	performance validity test
ROC	receiver operating characteristic
TOMM	Test of Memory Malingering
VA	Veterans Affairs

REFERENCES

Albert MS, DeKosky ST, Dickson D, et al. 2011. The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging–Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement. 7:270–279. doi:10.1016/j.jalz.2011.03.008 [PubMed: 21514249]

Bauer L, McCaffrey RJ. 2006. Coverage of the Test of Memory Malingering, Victoria Symptom Validity Test, and Word Memory Test on the internet: is test security threatened? Arch Clin Neuropsychol. 21:121–126. doi:10.1016/j.acn.2005.06.010 [PubMed: 16122901]

Begum T, Reza F, Ahmed I, et al. 2014. Influence of education level on design-induced N170 and P300 components of event related potentials in the human brain. J Integr Neurosci. 13:71–88. doi:10.1142/S0219635214500058 [PubMed: 24738540]

Bortnik KE, Horner MD, Bachman DL. 2013. Performance on standard indexes of effort among patients with dementia. Appl Neuropsychol Adult. 20:233–242. doi:10.1080/09084282.2012.695757 [PubMed: 23537314]

Bougrain L, Saavedra C, Ranta R. 2012. Finally, what is the best filter for P300 detection? Paper presented at: TOBI Workshop III—Tools for Brain-Computer Interaction - 2012; March 2012; Würzburg, Germany. https://hal.inria.fr/hal-00756669/document

Cecchi M, Moore DK, Sadowsky CH, et al. 2015. A clinical trial to validate event-related potential markers of Alzheimer's disease in outpatient settings. Alzheimers Dement (Amst). 1:387–394. doi:10.1016/j.dadm.2015.08.004 [PubMed: 27239520]

Cid-Fernández S, Lindín M, Díaz F. 2014. Effects of amnestic mild cognitive impairment on N2 and P3 Go/NoGo ERP components. J Alzheimers Dis. 38:295–306. doi:10.3233/JAD-130677 [PubMed: 23963292]

Daffner KR, Chong H, Sun X, et al. 2011. Mechanisms underlying age- and performance-related differences in working memory. J Cogn Neurosci. 23:1298–1314. doi:10.1162/jocn.2010.21540 [PubMed: 20617886]

Dean AC, Victor TL, Boone KB, et al. 2009. Dementia and effort test performance. Clin Neuropsychol. 23:133–152. doi:10.1080/13854040701819050 [PubMed: 18609332]

Donchin E, Coles MG. 1988. Is the P300 component a manifestation of context updating? Behav Brain Sci. 11:357–374. doi:10.1017/S0140525X00058027

Ellwanger J, Rosenfeld JP, Sweet JJ, et al. 1996. Detecting simulated amnesia for autobiographical and recently learned information using the P300 event-related potential. Int J Psychophysiol. 23:9–23. doi:10.1016/0167-8760(96)00035-9 [PubMed: 8880362]

Ellwanger J, Tenhula WN, Rosenfeld JP, et al. 1999. Identifying simulators of cognitive deficit through combined use of neuropsychological test performance and event-related potentials. J Clin Exp Neuropsychol. 21:866–879. doi:10.1076/jcen.21.6.866.850 [PubMed: 10649540]

Folstein MF, Folstein SE, McHugh PR. 1975. “Mini-mental state”. a practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res. 12:189–198. doi:10.1016/0022-3956(75)90026-6 [PubMed: 1202204]

Green P 2003. Medical Symptom Validity Test for Windows: User’s Manual and Program. Edmonton, Canada: Green’s.

Green P, Allen LM, Astner K. 1996. The Word Memory Test: A User’s Guide to the Oral and Computer-administered Forms. US Version 1.1. Durham, NC: CogniSyst.

Greve KW, Bianchini KJ, Black FW, et al. 2006. Classification accuracy of the Test of Memory Malingering in persons reporting exposure to environmental and industrial toxins: results of a known-groups analysis. Arch Clin Neuropsychol. 21:439–448. doi:10.1016/j.acn.2006.06.004 [PubMed: 16870391]

Hayama HR, Johnson JD, Rugg MD. 2008. The relationship between the right frontal old/new ERP effect and post-retrieval monitoring: specific or non-specific? Neuropsychologia. 46:1211–1223. doi:10.1016/j.neuropsychologia.2007.11.021 [PubMed: 18234241]

Howe LLS, Anderson AM, Kaufman DAS, et al. 2007. Characterization of the Medical Symptom Validity Test in evaluation of clinically referred memory disorders clinic patients. Arch Clin Neuropsychol. 22:753–761. doi:10.1016/j.acn.2007.06.003 [PubMed: 17651938]

Kaplan E, Goodglass H, Weintraub S. 2001. The Boston Naming Test. 2nd ed. Philadelphia, Pennsylvania: Lippincott, Williams & Wilkins.

Lezak MD, Howieson DB, Bigler ED, et al. 2012. Neuropsychological Assessment. 5th ed. New York, New York: Oxford University Press.

Mack WJ, Freed DM, Williams BW, et al. 1992. Boston Naming Test: shortened versions for use in Alzheimer’s disease. J Gerontol. 47:154–158. doi:10.1093/geronj/47.3.p154

McKhann GM, Knopman DS, Chertkow H, et al. 2011. The diagnosis of dementia due to Alzheimer’s disease: recommendations from the National Institute on Aging–Alzheimer’s Association workgroups on diagnostic guidelines for Alzheimer’s disease. Alzheimers Dement. 7:263–269. doi:10.1016/j.jalz.2011.03.005 [PubMed: 21514250]

Merten T, Bossink L, Schmand B. 2007. On the limits of effort testing: symptom validity tests and severity of neurocognitive symptoms in nonlitigant patients. J Clin Exp Neuropsychol. 29:308–318. doi:10.1080/13803390600693607 [PubMed: 17454351]

Mittenberg W, Patton C, Canyock EM, et al. 2002. Base rates of malingering and symptom exaggeration. J Clin Exp Neuropsychol. 24:1094–1102. doi:10.1076/jcen.24.8.1094.8379 [PubMed: 12650234]

Morris JC, Heyman A, Mohs RC, et al. 1989. The Consortium to Establish a Registry for Alzheimer’s Disease (CERAD). Part I. Clinical and neuropsychological assesment of Alzheimer’s disease. Neurology. 39:1159–1165. doi:10.1212/WNL.39.9.1159 [PubMed: 2771064]

Pachana NA, Byrne GJ, Siddle H, et al. 2007. Development and validation of the Geriatric Anxiety Inventory. Int Psychogeriatr. 19:103–114. doi:10.1017/S1041610206003504 [PubMed: 16805925]

Polich J 2007. Updating P300: an integrative theory of P3a and P3b. Clin Neurophysiol. 118:2128–2148. doi:10.1016/j.clinph.2007.04.019 [PubMed: 17573239]

Polich J, Corey-Bloom J. 2005. Alzheimer’s disease and P300: review and evaluation of task and modality. Curr Alzheimer Res. 2:515–525. doi:10.2174/156720505774932214 [PubMed: 16375655]

Reitan RM. 1958. Validity of the Trail Making Test as an indicator of organic brain damage. Percept Mot Skills. 8:271–276. doi:10.2466/pms.1958.8.3.271

Riis JL, Chong H, Ryan KK, et al. 2008. Compensatory neural activity distinguishes different patterns of normal cognitive aging. Neuroimage. 39:441–454. doi:10.1016/j.neuroimage.2007.08.034 [PubMed: 17931892]

Rogers R, Vitacco MJ. 2002. Forensic assessment of malingering and related response styles. In: Van Dorsten B, ed. Forensic Psychology: From Classroom to Courtroom. New York, New York: Kluwer Academic; 83–104.

Seshan VE, Gönen M, Begg CB. 2013. Comparing ROC curves derived from regression models. Stat Med. 32:1483–1493. doi:10.1002/sim.5648 [PubMed: 23034816]

Shelley-Tremblay JF, Eyer JC, Hill BD. 2019. A laboratory word memory test analogue differentiates intentional feigning from true responding using the P300 event-related potential. Brain Sci. 9:109. Published online May 14. doi:10.3390/brainsci9050109

Tardif HP, Barry RJ, Fox AM, et al. 2000. Detection of feigned recognition memory impairment using the old/new effect of the event-related potential. Int J Psychophysiol. 36:1–9. doi:10.1016/s0167-8760(00)00083-0 [PubMed: 10700618]

Teichner G, Wagner MT. 2004. The Test of Memory Malingering (TOMM): normative data from cognitively intact, cognitively impaired, and elderly patients with dementia. Arch Clin Neuropsychol. 19:455–464. doi:10.1016/S0887-6177(03)00078-7 [PubMed: 15033228]

Tombaugh TN. 1996. Test of Memory Malingering (TOMM). New York, New York: Multi-Health Systems.

Tombaugh TN. 1997. The Test of Memory Malingering (TOMM): normative data from cognitively intact and cognitively impaired individuals. Psychol Assess. 9:260–268. doi:10.1037/1040-3590.9.3.260

Tsivilis D, Otten LJ, Rugg MD. 2001. Context effects on the neural correlates of recognition memory: an electrophysiological study. Neuron. 31:497–505. doi:10.1016/s0896-6273(01)00376-2 [PubMed: 11516405]

van Deursen JA, Vuurman EF, Smits LL, et al. 2009. Response speed, contingent negative variation and P300 in Alzheimer's disease and MCI. Brain Cogn. 69:592–599. doi:10.1016/j.bandc.2008.12.007 [PubMed: 19185407]

van Hooff JC, Sargeant E, Foster JK, et al. 2009. Identifying deliberate attempts to fake memory impairment through the combined use of reaction time and event-related potential measures. Int J Psychophysiol. 73:246–256. doi:10.1016/j.ijpsycho.2009.04.002 [PubMed: 19374924]

Wechsler D 2001. WTAR Wechsler Test of Adult Reading: WTAR. San Antonio, Texas: Psychological.

Weintraub S, Salmon D, Mercaldo N, et al. 2009. The Alzheimer's Disease Centers' Uniform Data Set (UDS): the Neuropsychological Test Battery. Alzheimer Dis Assoc Disord. 23:91–101. doi:10.1097/WAD.0b013e318191c7dd [PubMed: 19474567]

Yesavage JA, Sheikh JI. 1986. Geriatric Depression Scale (GDS): recent evidence and development of a shorter version. Clin Gerontol. 5:165–173. doi:10.1300/J018v05n01_09

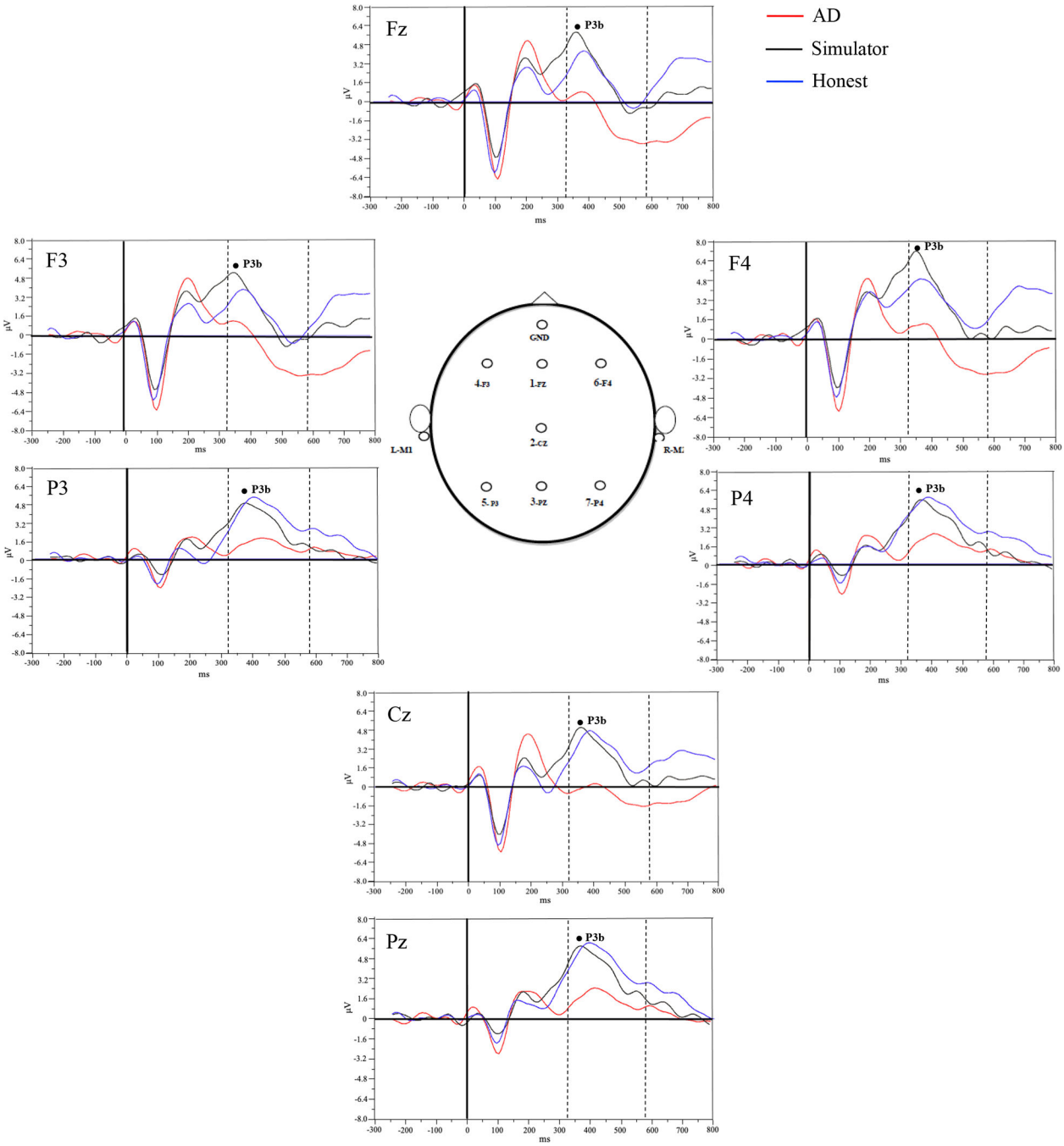

FIGURE 1.

Grand average waveforms for target tones derived from each electrode. **AD** = individuals with probable Alzheimer disease. **GND** = ground. **Honest** = healthy controls performing honestly. **Simulator** = healthy controls simulating dementia.

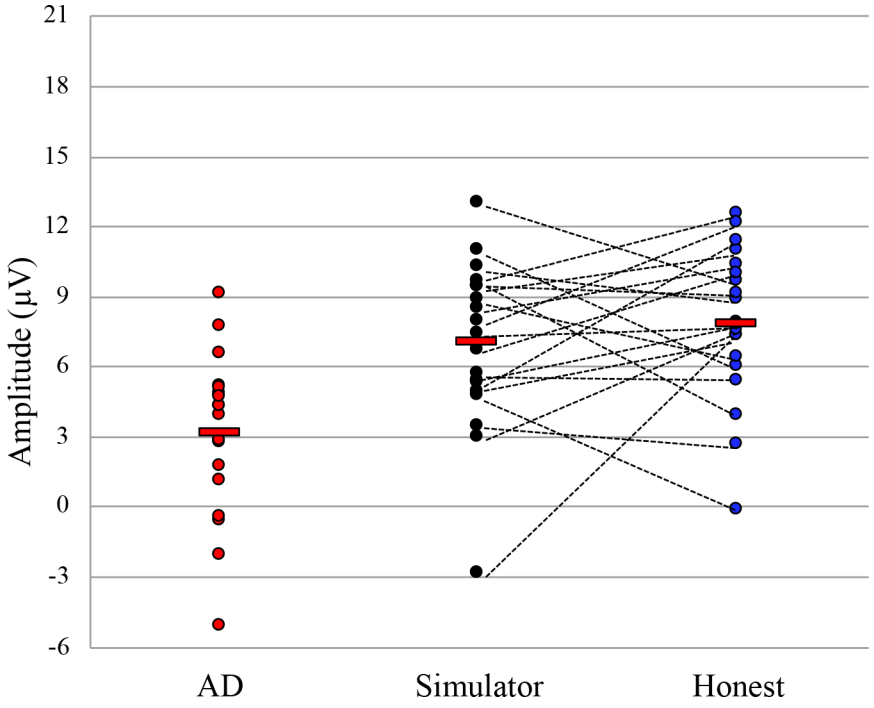

FIGURE 2.

Group differences for P3b peak amplitudes between the HC in the performing honestly and simulating dementia conditions and the probable AD group performing honestly. **AD** = individuals with probable Alzheimer disease. **HC** = healthy controls. **Honest** = healthy controls performing honestly. **Simulator** = healthy controls simulating dementia.

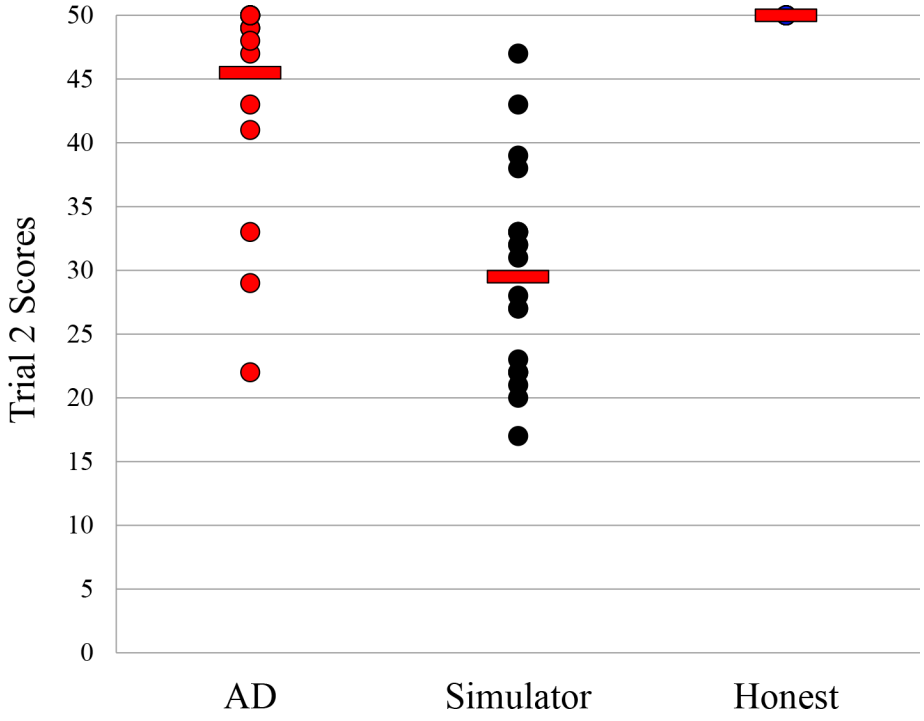

FIGURE 3.

Group differences for TOMM Trial 2 scores between the HC in the performing honestly and simulating dementia conditions and the probable AD group performing honestly. **AD** = individuals with probable Alzheimer disease. **HC** = healthy controls. **Honest** = healthy controls performing honestly. **Simulator** = healthy controls simulating dementia. **TOMM** = Test of Memory Malingering.

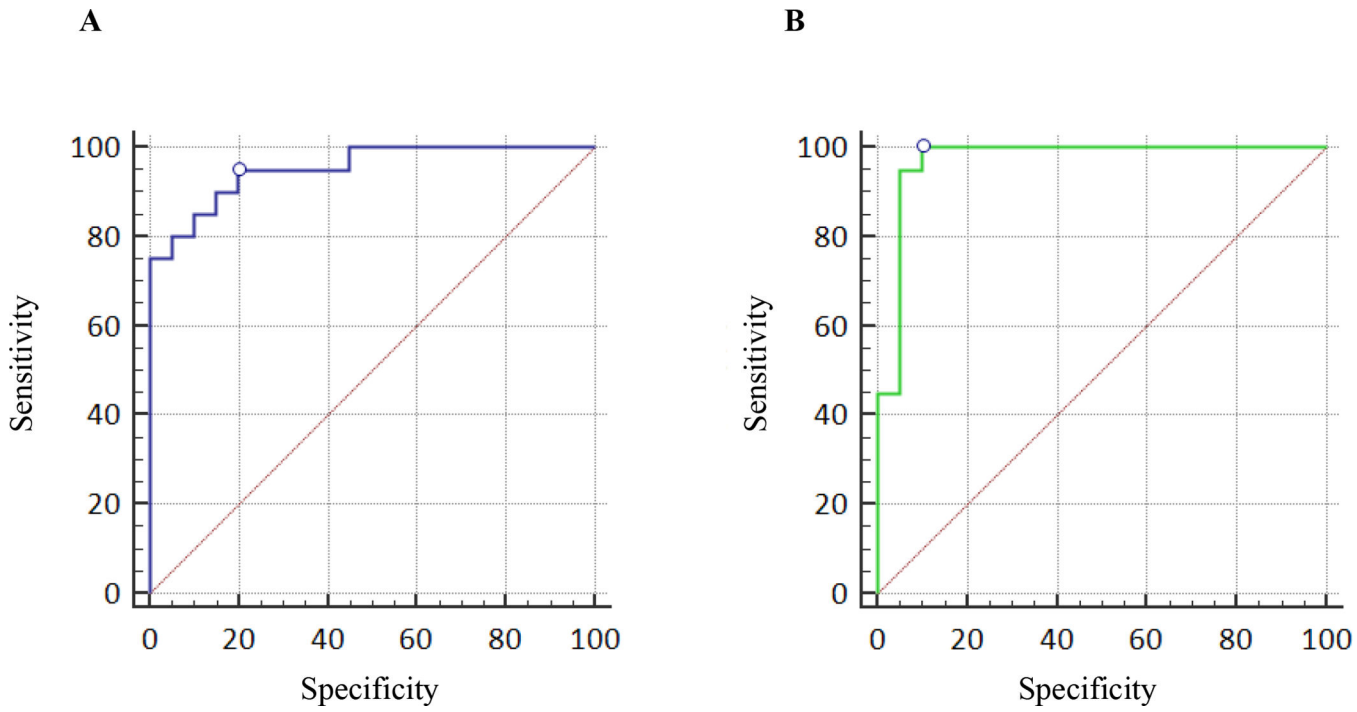

FIGURE 4.

ROC curves. **A.** A model for accurate classification of dementia diagnosis between the probable AD group and the HC simulating dementia using the TOMM Trial 2 scores. AUC: 0.95 (95% CI = 0.83–0.99), SE: 0.03, $P < 0.001$; sensitivity 95% and specificity 80% for cutoff of 0.34. **B.** A model for accurate classification of dementia diagnosis between the probable AD group and the HC simulating dementia using the TOMM Trial 2 scores and the P3b peak amplitudes. AUC: 0.97 (95% CI = 0.86–0.99), SE: 0.03, $P < 0.001$; sensitivity 100% and specificity 90% for cutoff of 0.34. **AD** = Alzheimer disease. **AUC** = area under the curve. **HC** = healthy controls. **ROC** = receiver operating characteristic. **TOMM** = Test of Memory Malingering.

TABLE 1.

Study Participants' Demographics and Neurocognitive Test Battery Data

	HC (n = 20)		Probable AD (n = 20)				
Characteristic/Test	**M**	**SD**	**M**	**SD**	$F_{1, 39}$	*P*	η_p^2
Age	76.80	8.28	77.90	9.04	0.17	0.686	0.00
Male (%)	100		100				
Education (years)	16.85	2.87	13.20	2.49	18.36	<0.001***	0.33
GDS-SF	2.30	3.08	3.95	2.96	2.83	0.101	0.07
GAI	1.80	3.16	5.65	6.18	5.89	<0.05*	0.13
EVIQ (WTAR)	113.96	4.89	100.76	11.66	38.33	<0.001***	0.50
CERAD							
Encoding	22.45	3.72	10.25	3.24	121.81	<0.001***	0.76
Recall	7.75	1.71	1.65	1.39	168.48	<0.001***	0.82
Recognition	9.60	0.99	6.30	2.79	26.36	<0.001***	0.41
Letter Fluency	51.70	11.63	26.80	11.83	51.65	<0.001***	0.58
Category Fluency	43.85	10.15	22.90	10.20	49.11	<0.001***	0.56
BNT–SF							
Hits	14.60	0.68	11.55	2.69	22.75	<0.001***	0.37
Misses	0.05	0.22	1.90	2.50	10.77	<0.005**	0.22
Phonemic cues	0.30	0.57	1.65	1.14	13.76	<0.001***	0.27
TMT, Part A (seconds)	31.90	7.92	96.05	44.21	27.32	<0.001***	0.42
TMT, Part B (seconds)	80.25	26.82	225.65	64.31	163.60	<0.001***	0.81
TMT, Part B errors	0.60	0.82	1.70	1.23	10.65	<0.005**	0.22
MMSE	28.60	1.00	21.25	3.78	108.38	<0.001***	0.74

AD = Alzheimer disease. **BNT** = Boston Naming Test. **CERAD** = Consortium to Establish a Registry for Alzheimer's Disease. **EVIQ** = estimated verbal intelligent quotient. **GAI** = Geriatric Anxiety Inventory. **GDS** = Geriatric Depression Scale. **HC** = healthy controls. **MMSE** = Mini-Mental State Examination. **SF** = Short Form. **TMT** = Trail-Making Test. **WTAR** = Wechsler Test of Adult Reading.

TABLE 2.

Binary Logistic Regression Models for the Probable AD Group Versus the HC Simulating Dementia Including Age, Education, TOMM Trial 2 Scores, and P3b Peak Amplitudes

Variable	B (SE)	*P*	OR
Model 1			
Age	−0.13 (0.09)	0.13	0.88
Education	−0.39 (0.20)	0.05*	0.68
TOMM Trial 2 score	0.23 (0.08)	0.005**	1.26
Nagelkerke (pseudo R^2) = 0.73 N = 40			
Model 2			
Age	0.24 (0.58)	0.68	1.02
Education	−0.47 (0.17)	0.01**	0.62
P3b peak amplitude	−0.35 (0.27)	0.04*	0.71
Nagelkerke (pseudo R^2) = 0.55 N = 40			
Model 3			
Age	−0.14 (0.11)	0.18	0.87
Education	−0.37 (0.26)	0.17	0.69
TOMM Trial 2 score	0.37 (0.12)	0.01**	1.36
P3b peak amplitude	−0.47 (0.22)	0.03*	0.63
Nagelkerke (pseudo R^2) = 0.84 N = 40			

AD = Alzheimer disease. **HC** = healthy controls. **TOMM** = Test of Memory Malingering.

* Significant at $P < 0.05$.

** Significant at $P < 0.01$.