



Identifying Electrophysiological Prodromes of Post-traumatic Stress Disorder: Results from a Pilot Study

Chao Wang^{1,2}, Michelle E. Costanzo^{2,3}, Paul E. Rapp^{1*}, David Darmon^{1,2}, Kylee Bashirelahi^{1,2}, Dominic E. Nathan^{1,2,4}, Christopher J. Cellucci⁵, Michael J. Roy³ and David O. Keyser¹

¹ Traumatic Injury Research Program, Department of Military and Emergency Medicine, Uniformed Services University of the Health Sciences, Bethesda, MD, USA, ² The Henry M. Jackson Foundation for the Advancement of Military Medicine, Inc., Bethesda, MD, USA, ³ Department of Medicine and Center for Neuroscience and Regenerative Medicine, Uniformed Services University of the Health Sciences, Bethesda, MD, USA, ⁴ Graduate School of Nursing, Uniformed Services University of the Health Sciences, Bethesda, MD, USA, ⁵ Aquinas LLC, Berwyn, PA, USA

OPEN ACCESS

Edited by:

Kim T. Mueser,
Boston University, USA

Reviewed by:

Jonathan K. Wynn,
University of California
Los Angeles, USA
Takako Mitsudo,
Kyushu University, Japan

*Correspondence:

Paul Rapp
paul.rapp@usuhs.edu

Specialty section:

This article was submitted to
Psychopathology,
a section of the journal
Frontiers in Psychiatry

Received: 22 January 2017

Accepted: 13 April 2017

Published: 15 May 2017

Citation:

Wang C, Costanzo ME, Rapp PE, Darmon D, Bashirelahi K, Nathan DE, Cellucci CJ, Roy MJ and Keyser DO (2017) Identifying Electrophysiological Prodromes of Post-traumatic Stress Disorder: Results from a Pilot Study. *Front. Psychiatry* 8:71. doi: 10.3389/fpsy.2017.00071

The objective of this research project is the identification of a physiological prodrome of post-traumatic stress disorder (PTSD) that has a reliability that could justify preemptive treatment in the sub-syndromal state. Because abnormalities in event-related potentials (ERPs) have been observed in fully expressed PTSD, the possible utility of abnormal ERPs in predicting delayed-onset PTSD was investigated. ERPs were recorded from military service members recently returned from Iraq or Afghanistan who did not meet PTSD diagnostic criteria at the time of ERP acquisition. Participants ($n = 65$) were followed for up to 1 year, and 7.7% of the cohorts ($n = 5$) were PTSD-positive at follow-up. The initial analysis of the receiver operating characteristic (ROC) curve constructed using ERP metrics was encouraging. The average amplitude to target stimuli gave an area under the ROC curve of greater than 0.8. Classification based on the Youden index, which is determined from the ROC, gave positive results. Using average target amplitude at electrode Cz yielded Sensitivity = 0.80 and Specificity = 0.87. A more systematic statistical analysis of the ERP data indicated that the ROC results may simply represent a fortuitous consequence of small sample size. Predicted error rates based on the distribution of target ERP amplitudes approached those of random classification. A leave-one-out cross validation using a Gaussian likelihood classifier with Bayesian priors gave lower values of sensitivity and specificity. In contrast with the ROC results, the leave-one-out classification at Cz gave Sensitivity = 0.65 and Specificity = 0.60. A bootstrap calculation, again using the Gaussian likelihood classifier at Cz, gave Sensitivity = 0.59 and Specificity = 0.68. Two provisional conclusions can be offered. First, the results can only be considered preliminary due to the small sample size, and a much larger study will be required to assess definitively the utility of ERP prodromes of PTSD. Second, it may be necessary to combine ERPs with other biomarkers in a multivariate metric to produce a prodrome that can justify preemptive treatment.

Keywords: post-traumatic stress disorder, prodromes, event-related potentials, delayed onset, traumatic brain injury, P300

INTRODUCTION

Historically, psychiatric practice has been reactive rather than preemptive. It has been recognized that a transition to preemptive psychiatry requires the identification of prodromes of psychiatric disorders that have a predictive reliability that justifies intervention in the absence of a fully expressed disorder. A prodrome is not a risk factor. A prodrome is a physiological change antecedent to a full expression of the disorder. Costello and Angold (1) provide the following definition: “. . . a prodrome is a premonitory manifestation of the disease. It is not a characteristic of the individual or their environment or a causal agent of the disease. A prodromal symptom may or may not continue to be manifest once the full disease appears. Conversely, the same disease may or may not manifest prodromal symptoms in different episodes.” Emerging genetic, epigenetic, and psychophysiological technologies offer the possibility of identifying prodromes or combinations of prodromes (where a combination of metrics may improve specificity) that can warrant preemptive treatment (2, 3). Prior research has investigated prodromes of several psychiatric disorders including psychosis (4–7), depression (8), autism (9, 10), dementia (11), alcoholism and substance abuse (1, 12), and post-traumatic stress disorder [PTSD (13–15)].

The objective of this research project is the identification of a physiological prodrome of PTSD that has a reliability that could justify preemptive treatment in the sub-syndromal state. The search for statistically reliable prodromes requires two things: a sub-syndromal period where physiological changes prior to the disease onset have been initiated, and a measure that can quantify these changes. In the ideal case, a third element can facilitate the search for prodromes: the identification of an at-risk population because an enriched population (a population where incidence is higher than the general population) will increase the statistical likelihood of identifying a prodrome. In this contribution, we address a specific question: can event-related potentials identify individuals at risk of delayed-onset PTSD? As preceding questions we must ask whether an at-risk population can be identified and if there is evidence indicating that PTSD can, in some instances, present with delayed onset? It is the period between trauma exposure and the presentation of a fully expressed PTSD that provides the window of opportunity for preemptive treatment.

Can a PTSD At-Risk Group Be Identified?

Military deployment is a risk factor for PTSD. The reported incidence of PTSD in veterans varies greatly between studies. A critical review found that PTSD incidence in US Iraq veterans ranges from 4 to 17% (16). Reports of the incidence of PTSD in the general population are similarly varied, but the National Comorbidity Survey Replication Study (17, 18) estimated the lifetime prevalence of PTSD in adult Americans to be 6.8%. Current past year prevalence was estimated at 3.5%. This suggests that military service members (SMs) who have returned from deployment will provide a statistically enriched population increasing the likelihood of identifying prodromes of PTSD. When making this observation, it is recognized that it is possible that military-related PTSD and PTSD in civilian populations may have distinct pathophysiological etiologies. This would potentially limit the general utility of results obtained with a military population.

Can PTSD Present with Delayed Onset?

Meta-analysis indicates that approximately 25% of PTSD cases present with delayed onset, where delayed onset is defined as meeting diagnostic criteria after a sub-syndromal or asymptomatic period of at least 6 months after the precipitating traumatic event (19, 20). In a military population, Grieger et al. (21) found that the majority of individuals PTSD-positive 7 months after serious combat injury did not meet diagnostic threshold at 1 month post-injury. In cases of PTSD following mild traumatic brain injury (TBI), the fraction of cases presenting with delayed onset can be higher. Bryant et al. (22) found that of those who met PTSD criteria at 24 months following a TBI, 44.1% reported no PTSD at 3 months. The analysis of Smid et al. (20) and Andrews et al. (19) indicates that PTSD can present after a symptom-free period, but it has been found to be more likely after a period of sub-syndromal PTSD in which two or three of the symptom clusters are endorsed (22). The factors contributing to delayed-onset PTSD in the absence of mild TBI are incompletely understood (15). On reviewing the trajectories of full and sub-syndromal PTSD, Bryant et al. (22) reached the following conclusions: “The present study demonstrates longitudinally that there is not a linear relationship between acute trauma response and long-term PTSD and highlights that PTSD levels fluctuate markedly in the initial years after trauma exposure. This pattern can explain the modest predictive capacity of acute markers to identify subsequent PTSD status. The complexity of these trajectories is further indicated by the delayed occurrence of PTSD responses, which appears to result from a combination of the immediate stress response and cumulative stress in the aftermath of the trauma.” These clinical observations further encourage the search for reliable physiological prodromes of PTSD.

Is There a Prior Literature Reporting Alterations of Event-Related Potentials in Fully Expressed PTSD?

As noted above, an additional requirement in the search for prodromes is the identification of a measure that can quantify physiological changes antecedent to disease onset. This search can be informed by asking whether there are markers that show alteration in the fully expressed disease, since it seems possible that these alterations may have begun prior to reaching diagnostic threshold. In the specific context of this investigation, this question becomes is there a prior literature showing abnormalities in event-related potentials in PTSD patients? An examination of the prior literature summarized in **Table 1** suggests that event-related potentials can be altered in the fully expressed PTSD state.

The divergence of electrophysiological results across studies is consistent with the emerging understanding that PTSD is not a discrete clinical entity and that different pathophysiological processes may be active in different individuals. The results do, however, suggest that alterations of brain electrical behavior can be associated with the disorder. As indicated in **Table 1**, alterations in P300 are most frequently reported.

There is an emerging understanding of the neurological origin of the empirical results reported in **Table 1** that suggests why alterations of P300 may be associated with both fully expressed PTSD and the sub-syndromal state. P300 has been hypothesized

TABLE 1 | Studies reporting ERP abnormalities in PTSD-positive participants.

Study	Reported observation(s)
Araki et al. (23)	Lower amplitude ERPs at Pz in an auditory oddball task
Blomhoff et al. (24)	Amplitudes to emotionally related words were significantly related to CAPS scores
Charles et al. (25)	P300 amplitude lower in PTSD-positive participants
Felmingham et al. (26)	Auditory oddball, PTSD positive participants show the following Target stimuli: reduced P200 amplitude, reduced P300 amplitude, increased N200 amplitude, increased N200 latency, increased P300 latency Standard stimuli: reduced P200 amplitude
Ghisolfi et al. (27)	PTSD positive participants showed auditory P50 sensory gating deficits
Hansenne (28)	Literature review includes PTSD
Javanbakht et al. (29)	Literature review of 36 studies. Increased P300 response to trauma-related stimuli. P50 studies suggest impaired gating
Johnson et al. (30)	P300a, P300b amplitudes larger with trauma related stimuli P300b small with neutral stimuli P300 working memory amplitudes smaller
Karl et al. (31)	Reduced P50 suppression Increase P300 amplitude to trauma-related stimuli
Kimble et al. (32)	Significant P300 amplitude enhancements to distracting stimuli
Kimble et al. (33)	Larger frontal, smaller central, and parietal CNVs
McFarlane et al. (34)	Delayed N200 P300 elicited by target and distracter tones indistinguishable
Metzger et al. (35)	Parietal P300 amplitude to target tones were smaller in unmedicated PTSD positive participants
Metzger et al. (36)	Modified Stroop task for personal traumatic, personal positive, and neutral words. PTSD-positive participants have reduced and delayed P300 across word type
Metzger et al. (37)	Contrary to previous results, the PTSD group had larger P300b amplitude and increased P200 amplitude/intensity slopes
Neylan et al. (38)	Impaired P50 gating to non-startle trauma-neutral auditory stimuli
Neylan et al. (39)	Nine of 24 P300 measures were significantly less predictable over time in the PTSD-positive group
Shu et al. (40)	mTBI only compared against mTBI + PTSD, larger emotional face processing ERPs in mTBI + PTSD
Shu et al. (41)	mTBI only compared against mTBI + PTSD, larger inhibitory processing ERPs in mTBI + PTSD
Shucard et al. (42)	PTSD group has longer P300 latency to NoGo stimuli and greater P300 amplitude to irrelevant non-target stimuli

CAPS, Clinician-Administered PTSD Scale; CNV, contingent negative variation; ERP, event-related potential.

to reflect neural activity associated with attention and subsequent memory processing (43), with larger P300 amplitude associated with greater attentional resources employed in the task (44, 45). The prior studies with PTSD positive participants reporting

reduced P300 amplitude to target stimuli in the PTSD group compared to the control group, suggest impairment of attentional processes which is consistent with clinical observation. In addition, a meta-analysis examining ERP components and PTSD revealed that the P300 amplitude may also be sensitive to contextual cues such that information processing is modulated based on the situation and environment (31). These dynamics are consistent with functional changes of two reported neural generators of the P300 (46, 47): the anterior cingulate cortex (ACC) and the hippocampus, which are also altered in individuals with PTSD (48). The ACC is critical to attentional processing and fear inhibition (49, 50) and the hippocampus is involved in memory and contextual representations (51). Araki et al. (23) revealed that lower P300 amplitude in patients with PTSD was associated with smaller ACC volume, which linked the P300 abnormality to underlying brain morphological abnormality.

It should be recognized that the results in **Table 1** were obtained from participants who were diagnostically PTSD-positive at the time of recording. The question of the utility of ERPs as a predictor of a transition to PTSD is not addressed by these studies, but these studies do suggest that altered ERPs may be present in the sub-syndromal state. This possibility is investigated in this study. The study was sponsored by the Department of Defense to investigate the utility of using a reduced montage that could be implemented in a military field hospital environment. Event-related potentials can be elicited by visual, auditory, somatosensory, and olfactory stimuli, with visual and auditory stimuli being the most commonly used. Hearing and vision can be compromised after blast exposure, but visual disturbances typically resolve faster. We therefore used visual stimuli in this study. As indicated in **Table 1**, several ERP components [P50, P200, N200, and contingent negative variation (CNV)] can be altered in PTSD-positive participants. Typically, however, the P300 is the most robust component. Since the object of this research program is the development of a robust technology that can be implemented in an austere medical environment, we focused on the P300.

METHODS

Subjects

We recruited 85 military SMs within 2 months of their return from an Operation Enduring Freedom (OEF)/Operation Iraqi Freedom (OIF) deployment of at least 3 months' duration in either Iraq or Afghanistan. The Clinician-Administered PTSD Scale (CAPS) (52) and the PTSD Checklist-Military Version (PCL-M) (53) were administered to assess PTSD. Patient Health Questionnaire-9 (PHQ-9) (54) and the International Classification of Diseases, 10th Clinical Modification (ICD-10) criteria for postconcussional syndrome (PCS) were administered to determine the presence of depression and PCS, respectively. Exclusion criteria included a history of head injury resulting in loss of consciousness for 60 min or more; a current Glasgow Coma Scale less than 13; visual acuity lower than 20/100 after correction; psychosis; active suicidal, or homicidal ideation; pregnancy; a diagnosis of PCS according ICD-10, PHQ-9 score greater than or equal to 10; and a PCL-M score greater than or equal to 50, or a diagnosis of PTSD made by an experienced psychologist using the CAPS based on the

DSM-IV criteria. All subjects provided written informed consent in accordance with the protocol approved by institutional review boards at Uniformed Services University, Walter Reed National Military Medical Center, and the National Institutes of Health.

Out of the 85 participants, 8 were excluded after baseline assessment: 2 for PCL-M ≥ 50 , 2 for PHQ-9 scores ≥ 10 , and 4 for problems with electroencephalogram (EEG) recording. Among the remaining 77 participants, 65 completed at least one follow-up psychological evaluation (52 at 3 months, 33 at 6 months, and 53 at 12 months). On serial follow-up evaluations, 5 of the 65 participants developed PTSD as determined by PCL-M scores (4 PTSD, 1 PTSD with depression). We therefore separated the 65 participants into 5 cases (referred to as Converters, mean age 35.6 ± 6.2 years, 4 men and 1 woman) and 60 controls (referred to as Stables, mean age 30.5 ± 8.0 years, 54 men and 6 women). The 5 Converters and 60 Stables are the final set of subjects in this study. In this paper, we focus on electrophysiological data from baseline assessment as we are trying to identify neural markers that predict the development of PTSD.

All participants in the group of 65 were exposed to relatively severe traumatic experiences. The types of index trauma reported by those who developed PTSD included experiencing a base attack (e.g., mortar or rocket fire, $n = 1$), engaging in combat-related violence (e.g., firefights, hit by improvised explosive device, IED, killing enemy, $n = 2$), witnessing combat-related violence (e.g., watching truck in convoy hit by an IED, witnessing death $n = 1$), and deployment bullying and abuse ($n = 1$). Those who did not develop PTSD also reported experiencing base attacks ($n = 24$), engaging in combat-related violence ($n = 23$), and witnessing combat-related violence ($n = 13$). Two factors, however, preclude a meaningful search for correlations between ERP abnormalities and cause of trauma. The first is the small size of the study population. The second would be applicable even in a larger study. Many, if not most of these participants have received multiple traumas from many causes.

Electrophysiological Recording

A visual oddball task was performed by subjects in an acoustically and electrically shielded room. Visual stimuli were presented by a digital tachistoscope of our own design and construction. The tachistoscope is a 5×5 square array of yellow, light-emitting diodes. Each diode is 1 cm in diameter. Given spacing between LEDs, the array is 6 cm \times 6 cm. The standard visual stimulus was a vertical stimulus which consists of the five vertical center line LEDs illuminated simultaneously for 40 ms. The target visual stimulus was a horizontal stimulus which is composed of the five horizontal center line LEDs illuminated simultaneously for 40 ms. Each subject received 125 stimuli in total, of which about 21% (26 ± 1 trials) were target and 79% (99 ± 1 trials) were standard stimuli. The subjects were instructed to maintain a silent count of the number of target stimulus presentations and to report their count at the end. The inter-stimulus onset time was varied randomly between 1.4 and 1.8 s. The number of trials in the current study is sufficient to elicit a valid P300 response. For example, a classic P300 study by Pollich et al. (55) used 25 target trials. Cohen and Polich (56) found that the P300 stabilized with approximately 20 trials.

The scalp EEG was recorded using the EPA6 amplifier (Sensorium Inc.) and the Grass electrodes (Natus Neurology Inc.) at Fz, Cz, Pz, Oz, C3, and C4 according to the standard 10-20 electrode system, with linked earlobes as reference and a forehead ground. Electrode impedances were maintained under 5 k Ω . EOG was recorded from two electrodes placed below and above the right eye. The sampling rate was 2,048 Hz, and the analog filter band-pass was 0.02–500 Hz.

Data Processing of Electrophysiological Data

Data processing was performed offline using custom scripts written in MATLAB (www.mathworks.com). Channels contaminated by artifacts were removed from analysis. This resulted in one Fz channel (from the Stable group) and four Oz channels (one from the Converter group and three from the Stable group) being removed. EOG artifacts were corrected by using a regression approach (57). The data after EOG correction were high-pass filtered at 0.5 Hz, low-pass filtered at 50 Hz, and down sampled to 256 Hz. The analysis period was -200 to 1,000 ms where time zero denotes stimulus onset. Trials with peak potentials exceeding 75 μ V or exhibiting abnormal trends were excluded from ERP averaging. The overall trial rejection rate was 4.84%. Target trials and standard trials were averaged separately. P300 amplitude was measured as the voltage of the largest positive peak of target ERP within 250–500 ms. P300 latency was measured as the time from stimulus onset to the maximum positive amplitude within 250–500 ms.

Statistical Analyses

Differences between groups in demographics, psychological measures, and task performance (accuracy of target count) were examined by Student's *t*-tests if data are numerical or Fisher's exact tests if data are categorical. Because the Oz channel was lost in some recordings (including one in the Converter group), the statistical analysis is limited to Fz, Cz, Pz, C3, and C4 electrode sites. Group differences in P300 amplitude and latency at each electrode site were tested by Student's *t*-tests. Correlations between P300 amplitude and the psychological measures were examined by Pearson's correlation coefficient. *p*-Values less than 0.05 were considered statistically significant.

To examine the efficacy of using P300 amplitude as the predictor for PTSD, we performed several statistical analyses including approximate classification error rate, receiver operating characteristic (ROC) curve, leave-one-out cross validation, and bootstrapping. The detailed mathematical methods and equations can be found in the Mathematical Appendices.

RESULTS

Subject Characteristics and Baseline Psychological Measures

The subject characteristics and baseline psychological measures were summarized in Table 2. Age, gender, handedness, and history of mild TBI (mTBI) were not significantly different between the Converter and Stable groups. At the baseline assessment, the

TABLE 2 | Subject characteristics and baseline psychological measures.

Variable	Converter (<i>n</i> = 5)		Stable (<i>n</i> = 60)		Group comparison ^a		
	Mean	SD	Mean	SD	df	<i>t</i> -Value	<i>p</i> -Value
Age	35.6	6.2	30.5	8.0	63	1.37	0.18
Gender, male/female	4/1		54/6				0.36
Handedness, R/L	5/0		55/5				0.66
History of mTBI < 10 years, yes/no	2/3		18/42				0.33
Clinician-Administered PTSD Scale total	30.6	15.4	18.7	12.5	63	2.02	0.047
Patient Health Questionnaire-9 score	5.2	2.3	2.5	2.3	62	2.51	0.015
PTSD Checklist-Military Version (PCL-M) score	33.4	11.0	25.9	7.4	63	2.10	0.040

^aFisher's exact tests were used for gender, handedness, and history of mTBI. Student's *t*-tests were used for other variables.

Converter group reported significantly higher CAPS, PHQ-9, and PCL-M scores than the Stable group.

Behavioral Data

The accuracy of target count at baseline assessment was not significantly different between Converters and Stables. For Converters, the mean accuracy of target count was 93.1% (SD 5.0%) and for Stables the mean accuracy was 97.4% (SD 5.5%) The difference was not statistically significant ($t = 1.70$, $df = 63$, $p = 0.095$).

P300 Data: Amplitude and Latencies of Averaged Responses

We computed the approximate signal-to-noise ratios (SNRs) for both target and standard trials within the P300 time window for each subject. The SNR was calculated from the power of the ERP during the P300 window (300–400 ms) minus the power of the ERP during baseline (–200 to 0 ms) and then divided by the power of the ERP during baseline window. The mean SNR for single subject ERP for target trials at Pz is 145 (21.6 dB). The mean SNR for single subject ERP for standard trials at Pz is 87 (19.4 dB).

The P300 waveforms of average responses to standard stimuli do not have a well-defined single peak that can provide a unique amplitude and latency measure that can be incorporated into statistical analysis. Statistical analysis is therefore limited to the average responses to target stimuli where well-defined P300 waveforms make precise measurements possible. **Figure 1** displays the grand average ERPs in response to target and standard stimuli at the six electrodes in Converters and Stables. Because the Oz channel was lost in some recordings, the statistical analysis is further limited to Fz, Cz, Pz, C3, and C4 electrode sites. We found that for all these electrode sites, the P300 amplitude was significantly smaller ($p < 0.05$) for the Converter group compared to the Stable group. The P300 latency was not significantly different ($p > 0.05$) between the two groups. The statistical results for each electrode were summarized in **Table 3**. We also explored the correlation between the P300 amplitudes and the psychological measures (CAPS, PHQ-9, and PCL-M) across subjects. No significant correlations were found ($p > 0.05$).

DIAGNOSTIC VALIDITY

Approximate Classification Error Rate

As summarized in **Table 3**, there was a statistically significant difference in the target amplitude between the participants who remained PTSD-negative throughout the study and those who became PTSD-positive. A statistically significant between-group separation does not, however, establish the efficacy of these measures as predictors. The most commonly applied quantitative measure of between-group separation is the *t*-test. As shown in **Table 3**, a naive calculation (a two-tailed *t*-test that assumes unequal variances) suggests a significant separation between the two participant groups. Two essential observations should be made. First, the asymptotic assumptions of the *t*-test cannot be meaningfully satisfied when $N_C = 5$. Second, a separation of means, which is what the *t*-test assesses, does not of itself ensure a successful classification even in those instances where the assumptions of the test are satisfied. An estimate of classification error rates can be made by again assuming normality of the two populations. The equations used are given in the Mathematical Appendices. This estimate often results in a substantial under estimate of the true error rate. This is particularly true when population numbers are small (58). The results shown in **Table 3** show that application of this admittedly optimistic error rate estimate predicts that using target amplitude results in unacceptable classification error rates of $P_{\text{ERROR}} = 0.29$ to $P_{\text{ERROR}} = 0.32$, where it should be remembered that random assignment results in a 0.50 error if we assume that the two populations occur in equal proportions. This negative conclusion will be supported by the more reliable empirical determinations of classification error. It should be noted, however, that the error rates are different between the amplitudes and latencies, namely approximately 30% for the amplitudes and 50% for the latencies.

ROC Curve

Prediction using prodromes can be treated as a diagnostic problem in which the disease-positive state corresponds to being a member of the group that becomes PTSD positive. Calculation of the ROC curve is a commonly employed method for characterizing a diagnostic classification. The first row of **Table 4** shows the area under the curve (AUC), for the electrophysiological measures. The mathematical methods used to determine the AUC and its confidence intervals are given in the Mathematical Appendices. A value of $AUC > 0.5$ indicates better than random assignment. The P300 amplitude at Cz showed the highest predictive power, with an AUC of 0.85 (confidence interval of [0.67, 0.94]). The ROC curve of the P300 amplitude at Cz is shown in **Figure 2**. While the values of the AUC are encouraging, the very large confidence intervals diminish confidence in the result.

Diagnostic Efficacy and Determination of the Diagnostic Cut Score

The results of a diagnostic calculation (and by implication for the present context the identification of a prodrome) can be expressed in the canonical four element diagnostic matrix: true positive, false positive, false negative, and true negative. There is

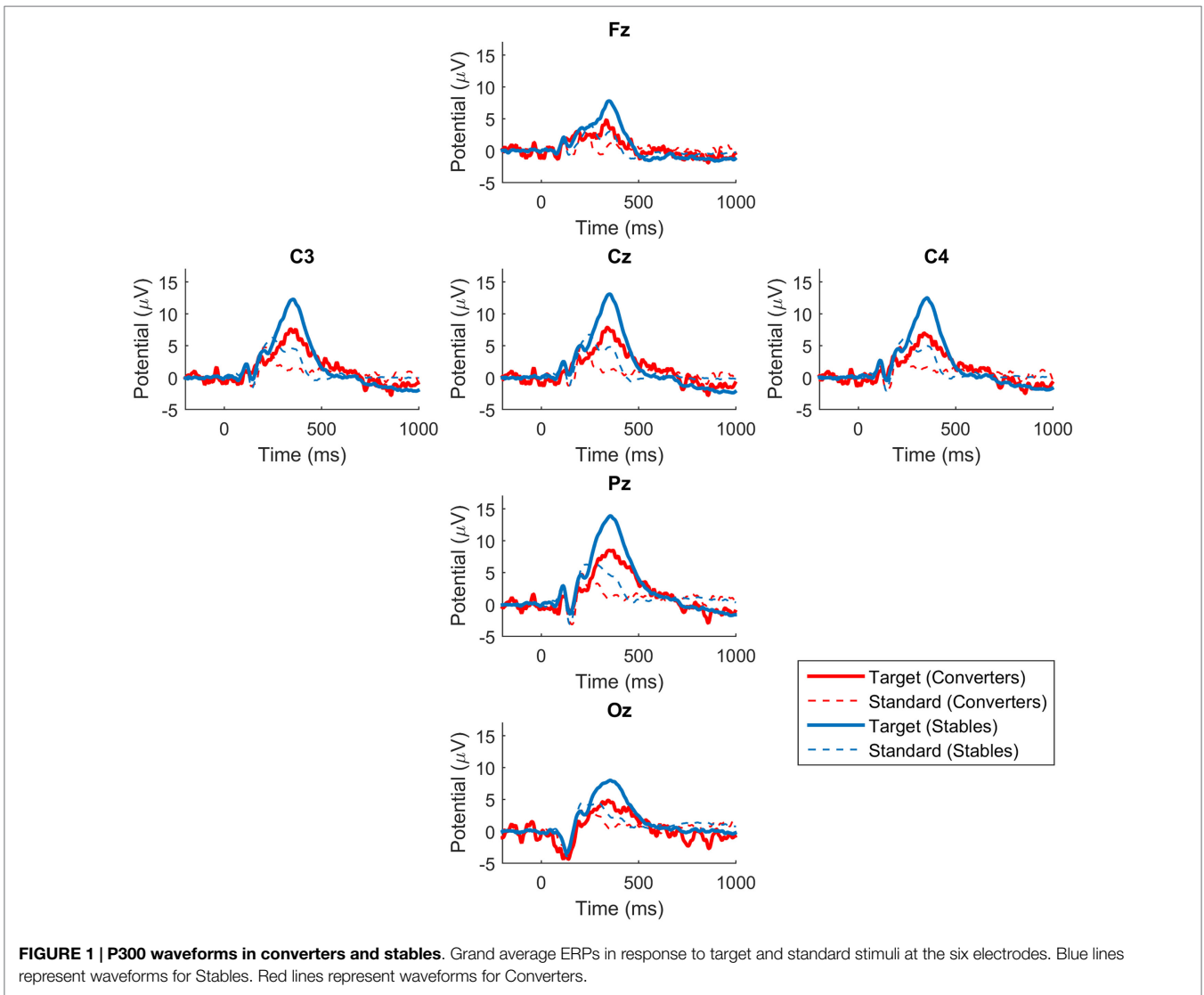


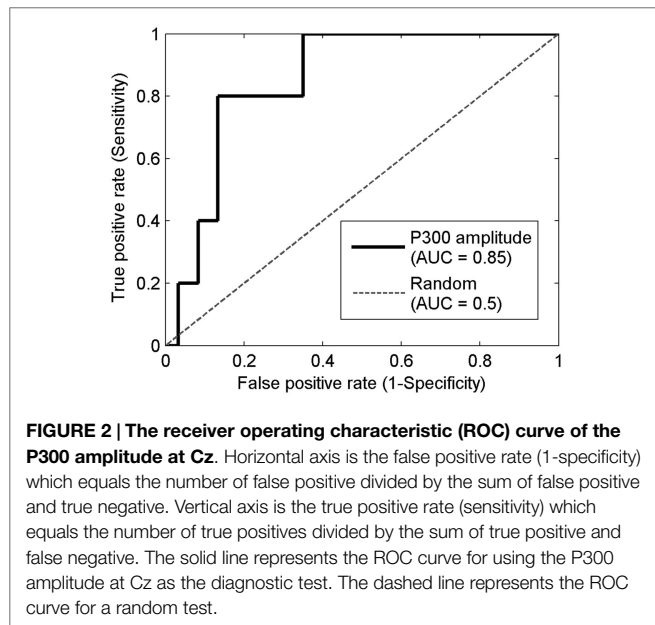
FIGURE 1 | P300 waveforms in converters and stables. Grand average ERPs in response to target and standard stimuli at the six electrodes. Blue lines represent waveforms for Stables. Red lines represent waveforms for Converters.

TABLE 3 | Baseline results from participants who remained PTSD-negative for one year after enrollment (*N* = 60) and those who converted to PTSD-positive (*N* = 5).

	Baseline scores individuals	Baseline scores individuals	Between-group separation	<i>P</i> _{ERROR}
	PTSD-negative at 1 year stables	PTSD-positive at 1 year converters	<i>T</i> -test, two-tailed, unequal variance	Equal priors
	<i>N</i> = 60 (<i>N</i> = 59 for Fz)	<i>N</i> = 5	<i>p</i>	
Average Fz amplitude response to target stimulus (µV)	9.87 ± 4.25	5.98 ± 2.38	0.0157	0.3193
Average Cz amplitude response to target stimulus (µV)	15.48 ± 5.45	9.71 ± 2.65	0.0038	0.2937
Average Pz amplitude response to target stimulus (µV)	16.18 ± 5.27	11.13 ± 3.71	0.0338	0.3130
Average C3 amplitude response to target stimulus (µV)	14.52 ± 5.16	9.34 ± 2.54	0.0054	0.3032
Average C4 amplitude response to target stimulus (µV)	14.72 ± 5.23	9.07 ± 2.67	0.0046	0.2898
Average Fz latency response to target stimulus (ms)	356.2 ± 43.8	357.0 ± 57.2	0.9760	0.4963
Average Cz latency response to target stimulus (ms)	359.7 ± 39.0	357.3 ± 57.6	0.9235	0.4868
Average Pz latency response to target stimulus (ms)	360.5 ± 42.4	374.2 ± 58.6	0.6345	0.4377
Average C3 latency response to target stimulus (ms)	359.4 ± 37.3	352.3 ± 68.2	0.8291	0.4646
Average C4 latency response to target stimulus (ms)	355.2 ± 36.9	355.5 ± 60.3	0.9928	0.4987

TABLE 4 | Area under the receiver operating curve and measures of diagnostic efficacy computed using the smallest value of threshold giving the maximum value of the Youden index.

Measure	Average Fz amplitude response to target stimulus	Average Cz amplitude response to target stimulus	Average Pz amplitude response to target stimulus	Average C3 amplitude response to target stimulus	Average C4 amplitude response to target stimulus
Area under the curve	0.7864 [0.5616, 0.8960]	0.8533 [0.6708, 0.9347]	0.7833 [0.4737, 0.9108]	0.8233 [0.6170, 0.9185]	0.8433 [0.5980, 0.9390]
Max Youden Index	0.5763	0.6667	0.5500	0.6000	0.7000
T_{MAX} (μ V)	9.0140	10.4186	12.1866	12.4016	9.1688
Diagnostic accuracy	0.6094	0.8615	0.7538	0.6308	0.8923
Sensitivity	1.0000	0.8000	0.8000	1.0000	0.8000
Specificity	0.5763	0.8667	0.7500	0.6000	0.9000
Positive likelihood ratio	2.3600	6.0000	3.2000	2.5000	8.0000
Negative likelihood ratio	0.0000	0.2308	0.2667	0.0000	0.2222
Diagnostic odds ratio	Undefined	26.0000	12.0000	Undefined	36.0000



no single fully satisfactory summary measure for characterizing the diagnostic matrix. Each has advantages and limitations. The limitations are particularly evident in studies like this one where disease prevalence is low. We will therefore examine six common measures of diagnostic efficacy: diagnostic accuracy, sensitivity, specificity, the positive likelihood ratio, the negative likelihood ratio, and the diagnostic odds ratio. Their definitions are given in the Mathematical Appendices.

The values of elements in the diagnostic matrix, and therefore measures of diagnostic efficacy like sensitivity and specificity, are critically dependent on the cut score used to assign individuals to the disease-positive and disease-negative groups. The choice of the cut value is therefore a central problem in the implementation of a diagnostic procedure. As outlined in the Mathematical Appendices, more than one candidate procedure has been proposed. In the calculations summarized in **Table 4**, the diagnostic threshold was determined by the value of threshold that gave the maximum value of J , the Youden index (59). The value of sensitivity, specificity, and other measures of diagnostic efficacy reported in **Table 4** are the values obtained when the threshold was set to the smallest value of threshold giving the maximum J . Because the

results of **Table 3** indicate that target latencies cannot discriminate between-group means, the analysis is limited to target amplitudes.

Leave-One-Out Cross Validation

The results presented in **Table 4** are encouraging particularly in the cases of average Cz amplitude and average C4 amplitude which give sensitivity and specificity values in excess of 0.8. Measures of diagnostic efficacy obtained by examination of the ROC can be misleadingly optimistic if sample sizes are small. A fast, albeit imperfect, reality check can be implemented by a leave-one-out cross validation. In this calculation, one of the values is removed from the sample. A between-group classifier is constructed from the remaining data, and the omitted value is classified. It is then replaced. Another value is removed and classified. This procedure continues to exhaustion and the classification results are used to populate the diagnostic matrix (true positive, false positive, false negative, true negative). The measures of diagnostic efficacy introduced in the previous section are then calculated.

In order to implement a leave-one-out cross validation the choice of classifier must be addressed. In these calculations, a classifier based on Gaussian populations with prior probabilities was used. The mathematical structure of the classifier is given in the Mathematical Appendices. Two sets of prior probabilities were considered. In the first set of calculations, equal priors were used. In the second, it was supposed that the prior probability of delayed-onset PTSD was 0.25 which is the value derived from a review of the clinical literature (19, 20).

With both sets of prior probabilities, the sensitivity and specificity values are considerably less encouraging (**Table 5**). In the previous calculations, the sensitivity and specificity obtained at Cz are 0.80 and 0.87, respectively. In the leave-one-out calculation using equal priors, the corresponding values are 0.60 and 0.65. Similarly, the previous sensitivity and specificity results obtained at C4 were 0.80 and 0.90, respectively. The leave-one-out values with equal priors are 0.80 and 0.62. This divergence counsels interpretive caution when evaluating the results summarized in **Table 3**.

Populating the Diagnostic Matrix by Bootstrapping

A deficiency of the results presented in the previous section is immediately apparent on examining **Table 5**. The sensitivities

TABLE 5 | Classification based on average target amplitudes determined by a leave-one-out calculation.

Measure	Average Fz amplitude response to target stimulus	Average Cz amplitude response to target stimulus	Average Pz amplitude response to target stimulus	Average C3 amplitude response to target stimulus	Average C4 amplitude response to target stimulus
Prior probabilities, $P_s = 0.5, P_c = 0.5$					
Number of true positives	4	3	4	3	4
Number of false positives	25	21	25	24	23
Number of false negatives	1	2	1	2	1
Number of true negatives	34	39	35	36	37
Diagnostic accuracy	0.5938	0.6462	0.6000	0.6000	0.6308
Sensitivity	0.8000	0.6000	0.8000	0.6000	0.8000
Specificity	0.5763	0.6500	0.5833	0.6000	0.6167
Prior probabilities, $P_s = 0.75, P_c = 0.25$					
Number of true positives	0	1	0	1	3
Number of false positives	6	7	3	7	8
Number of false negatives	5	4	5	4	2
Number of true negatives	53	53	57	53	52
Diagnostic accuracy	0.8281	0.8308	0.8769	0.8308	0.8462
Sensitivity	0.0000	0.2000	0.0000	0.2000	0.6000
Specificity	0.8983	0.8833	0.9500	0.8833	0.8667

TABLE 6 | Classification based on average target amplitudes determined by a bootstrap calculation.

Measure	Average Fz amplitude response to target stimulus	Average Cz amplitude response to target stimulus	Average Pz amplitude response to target stimulus	Average C3 amplitude response to target stimulus	Average C4 amplitude response to target stimulus
Prior probabilities $P_s = 0.5, P_c = 0.5$					
Diagnostic accuracy	0.6288 [0.4348, 0.8261]	0.6719 [0.4500, 0.8621]	0.6087 [0.3333, 0.8462]	0.6431 [0.4500, 0.8261]	0.6954 [0.4286, 0.9091]
Sensitivity	0.6236 [0.0000, 1.0000]	0.5916 [0.0000, 1.0000]	0.6835 [0.0000, 1.0000]	0.5790 [0.0000, 1.0000]	0.6746 [0.0000, 1.0000]
Specificity	0.6325 [0.4118, 0.9048]	0.6802 [0.4211, 0.9444]	0.6068 [0.3158, 0.9444]	0.6513 [0.4211, 0.8846]	0.6996 [0.3913, 1.0000]
Positive likelihood ratio	1.6108 [0.4423, 3.2051]	1.8999 [0.5882, 4.6667]	1.6975 [0.3333, 3.8182]	1.5957 [0.4058, 3.3409]	2.6276 [0.7667, 7.0000]
Negative likelihood ratio	0.6720 [0.2174, 1.3889]	0.6592 [0.2121, 1.2857]	0.6695 [0.2069, 1.8254]	0.6992 [0.2114, 1.4457]	0.5572 [0.2100, 1.1613]
Diagnostic odds ratio	3.7930 [0.3176, 11.9231]	4.3037 [0.4667, 15.6154]	4.0970 [0.2000, 13.8889]	3.9384 [0.2870, 13.8889]	5.9901 [0.6863, 19.8000]
Prior probabilities $P_s = 0.75, P_c = 0.25$					
Diagnostic accuracy	0.8282 [0.6364, 0.9583]	0.8159 [0.6400, 0.9545]	0.8647 [0.6818, 0.9615]	0.8023 [0.6190, 0.9565]	0.8562 [0.7143, 0.9583]
Sensitivity	0.1533 [0.0000, 1.0000]	0.3104 [0.0000, 1.0000]	0.1481 [0.0000, 1.0000]	0.2175 [0.0000, 1.0000]	0.3513 [0.0000, 1.0000]
Specificity	0.8887 [0.6667, 1.0000]	0.8622 [0.6667, 1.0000]	0.9269 [0.7200, 1.0000]	0.8542 [0.6522, 1.0000]	0.9030 [0.7273, 1.0000]
Positive likelihood ratio	4.0682 [0.5778, 12.5000]	3.3854 [0.6786, 12.0000]	5.5863 [0.7407, 15.7500]	3.0264 [0.4902, 12.0000]	4.9856 [0.9286, 16.5000]
Negative likelihood ratio	0.8286 [0.2949, 1.1667]	0.7378 [0.2660, 1.1111]	0.7963 [0.2805, 1.0811]	0.8170 [0.2838, 1.2069]	0.6722 [0.2000, 1.0135]
Diagnostic odds ratio	6.1093 [0.4921, 19.8000]	6.6509 [0.6104, 29.4000]	8.6072 [0.6863, 39.0000]	5.1757 [0.4026, 23.4000]	10.2212 [0.9184, 43.0000]

and specificities are reported without confidence intervals. This deficiency can be addressed with a bootstrap calculation. The procedure is outlined in the Mathematical Appendices. Two thousand bootstrap samples were used to estimate the bootstrapped distribution. The results are shown in **Table 6**. The confidence intervals provide an essential clarification to the preceding results. The sample size precludes a dispositive response to the hypothesis that the amplitudes of average ERPs can serve as a predictor of delayed-onset PTSD.

The confidence intervals reported for sensitivity values, [0,1] in all cases, are particularly telling. The definition of sensitivity is

$$\text{Sensitivity} = \text{True Positive Ratio} = \frac{N_{TP}}{N_{TP} + N_{FN}}$$

where N_{TP} is the number of true positives and N_{FN} is the number of false negatives. There are only five elements in the Converter set, and two of these elements are used to build the classifier. Therefore, N_{TP} is frequently zero, giving Sensitivity = 0. Similarly,

if in other cases $N_{TP} \neq 0$ and $N_{FN} = 0$ giving Sensitivity = 1 as another frequent value. This results in a bootstrapped confidence interval of [0,1].

DISCUSSION

In this analysis, the identification of individuals who will present delayed-onset PTSD is treated as a diagnostic process where the diagnostic groups are Converters (those who present delayed-onset PTSD) and Stables (those who do not). Sensitivity values based on average target stimulus amplitude range from 0.58 to 0.68. Specificity values range from 0.61 to 0.70, suggesting that event-related potentials may be helpful in identifying at-risk individuals.

The results in this study can only be considered preliminary due to the small sample size of Converters. The limitations of the sample size are indicated by the calculations presented in **Table 6**. Suppose the objective is to know sensitivity to an accuracy of ± 0.1

with 95% confidence. A calculation given in the Mathematical Appendices indicates that $N \geq 185$ is required, where it must be emphasized that this N is the number of Converters. If Converters are 10% of the population, then the projected requirement is for 1,850 participants in the study. The implications of this simple calculation extend beyond the study of PTSD and generalize to all of neuropsychiatry where conversion rates even in enriched populations are low. Large participant numbers will be required. Additionally, by definition, the search for prodromes requires a longitudinal study extended, perhaps, over a period of years. The challenges of supporting and implementing very large longitudinal studies are formidable.

Further limitations should be acknowledged. Electrophysiological abnormalities associated with neuropsychiatric disorders are non-specific. For example, in addition to PTSD, alterations in EEG synchronization have been observed in AD/HD, alcohol abuse, alexithymia, autism, bipolar disorder, dementia, depression, migraine, multiple sclerosis, Parkinson's disease, TBI, schizophrenia, and other psychotic disorders (60). The potential loss of electrophysiological specificity is particularly likely in a military population where PTSD is often associated with TBI and is comorbid with depression and substance abuse. Additionally, medications can alter event-related potentials and will complicate diagnosis based on ERPs.

Statistical identification of individuals who will present with PTSD might, however, be improved by two extensions to the present analysis. First, the analysis of ERPs reported here was limited to calculation of average ERPs. More recently, developed methods of analysis, for example, information dynamics (61) and network analysis of brain electrical activity (62) might improve results. Second, specificity and sensitivity may be improved by combining electrophysiological measures with other biomarkers and clinical information. Incorporating scores from psychological questionnaires with electrophysiological results in a multivariate discrimination would be an obvious possibility. The psychological measures including CAPS, PHQ-9, and PCL-M scores showed significant difference between Stables and Converters at the baseline assessment, but none of the scores significantly correlated with the P300 amplitude. The discordance between neural responses and self-reported symptoms may be partially a consequence of psychological defensive denial (63, 64). Some SMs recruited in this study may deny the presence of their PTSD symptoms due to military training or concerns that this may jeopardize their job, promotion, and self-image. This defensive denial may be softened after a prolonged period. Consistent with this possibility, a review by Andrews et al. (19) reported that most delayed-onset PTSD cases occurred in military samples rather than in civilian samples. If this is the case, objective biomarkers would be fundamentally more favorable

than self-report psychological measures in identifying SMs at risk of PTSD.

While additional forms of electrophysiological analysis in combination with other classes of data may improve the likelihood of success, this will not eliminate the previously documented requirement for large sample sizes in a longitudinal study. Such detection would be critical to the military because early intervention to prevent PTSD has revealed a critical window for fear activation and extinction of conditioned responses related to traumatic memories (65).

ETHICS STATEMENT

This research protocol was approved by the Institutional Review Board of the Uniformed Services University and by the Institutional Review Board of the Walter Reed National Military Medical Center. All participants gave written informed consent in accordance with the Declaration of Helsinki.

AUTHOR CONTRIBUTIONS

CW performed the analysis of the event-related potentials and the preliminary statistical analysis. MC screened participants for eligibility and conducted the psychological assessments. PR performed the literature search, statistical analysis, and wrote the final drafts of the paper. DD participated in developing and implementing the statistical analysis plan. KB and DN obtained the electrophysiological data. CC designed and built the ERP acquisition system. MR participated in the design of the investigation. DK lead the research effort and participated in acquisition of the electrophysiological data.

ACKNOWLEDGMENTS

We would like to acknowledge support from the Uniformed Services University, the Defense Medical Research and Development Program and the Center for Neuroscience and Regenerative Medicine. The Opinions and assertions contained herein are the private opinions of the authors and are not to be construed as official or reflecting the views of the United States Department of Defense.

FUNDING

Funding was provided by the Center for Neuroscience and Regenerative Medicine Project 351030 and by the Defense Medical Research and Development Program Project D10_1_AR_J5_605.

REFERENCES

- Costello EJ, Angold A. Developmental transitions to psychopathology: are there prodromes of substance use disorder? *J Child Psychol Psychiatry* (2010) 51(4):526–32. doi:10.1111/j.1469-7610.2010.02221.x
- Insel TR. The arrival of preemptive psychiatry. *Early Interv Psychiatry* (2007) 1(1):5–6. doi:10.1111/j.1751-7893.2007.00017.x
- Insel TR. From prevention to preemption: a paradigm shift in psychiatry. *Psychiatric Times* (2008) 25(9):13.
- Addington J, Cadenhead KS, Cornblatt BA, Mithalon DH, McGlashan TH, Perkins DO, et al. North American prodrome longitudinal study (NADLS2): overview and recruitment. *Schizophr Res* (2012) 142(1–3):77–82. doi:10.1016/j.schres.2012.09.012
- Murphy JR, Rawdon C, Kelleher I, Twomey D, Markey PS, Cannon M, et al. Reduced duration mismatch negativity in adolescents with psychotic symptoms: further evidence for mismatch negativity as a possible biomarker for vulnerability to psychosis. *BMC Psychiatry* (2013) 13(1):1. doi:10.1186/1471-244X-13-45

6. Clark SR, Schubert KL, Baune BT. Towards indicated prevention of psychosis: using probabilistic assessments of transition risk in psychosis prodrome. *J Neural Transm* (2015) 122(1):155–69. doi:10.1007/s00702-014-1325-9
7. Joa I, Gisselgård J, Brønnick K, McGlashan T, Olav J. Primary prevention of psychosis through interventions in the symptomatic prodromal stage, a pragmatic Norwegian ultra high risk study. *BMC Psychiatry* (2015) 15:89. doi:10.1186/s12888-015-0470-5
8. Kovacs M, Lopez-Duran N. Prodromal symptoms and atypical affectivity as predictors of major depression in juveniles: implications for prevention. *J Child Psychol Psychiatry* (2010) 51(4):472–96. doi:10.1111/j.1469-7610.2010.02230.x
9. Osterling J, Dawson G. Early recognition of children with autism: a study of first birthday home videotapes. *J Autism Dev Disord* (1994) 24(3):247–57. doi:10.1007/BF02172225
10. Bosl W, Tierney A, Tager-Flusberg H, Nelson C. EEG complexity as a biomarker for autism spectrum disorder risk. *BMC Med* (2011) 9:18. doi:10.1186/1741-7015-9-18
11. Riley KP, Snowdon DA, Desrosiers MF, Markesbery WR. Early life linguistic ability, late life cognitive function and neuropathology. Findings from the Nun Study. *Neurobiol Aging* (2005) 26(3):341–3. doi:10.1016/j.neurobiolaging.2004.06.019
12. Petit G, Cimochovska A, Kornreich C, Hanak C, Verbanck P, Campanella S. Neurophysiological correlates of response inhibition predict relapse in detoxified alcoholic patients: some preliminary evidence from event-related potentials. *Neuropsychiatr Dis Treat* (2014) 10:1025–37. doi:10.2147/NDT.S61475
13. Karstoft K-I, Galatzer-Levy IR, Statnikov A, Li Z, Shalev AY; Members of Jerusalem Trauma Outreach and Prevention Study (J-TOPS) Group. Bridging a translational gap: using machine learning to improve the prediction of PTSD. *BMC Psychiatry* (2015) 15:30. doi:10.1186/s12888-015-0399-8
14. Smid GE. *Deconstructing Delayed Onset Posttraumatic Stress Disorder [Dissertation]*. Utrecht: University of Utrecht (2011).
15. Smid GE, van der Velden PG, Gerson BPR, Kleber RJ. Late-onset posttraumatic stress disorder following a disaster: a longitudinal study. *Psychol Trauma* (2012) 4(3):312–22. doi:10.1037/a0023868
16. Richardson LK, Frueh BC, Acierno R. Prevalence estimates of combat-related PTSD: a critical review. *Aust N Z J Psychiatry* (2011) 44(1):4–19. doi:10.3109/00048670903393597
17. Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry* (2005) 62(6):593–602. doi:10.1001/archpsyc.62.6.593
18. Kessler RC, Chiu WT, Demler O, Merikangas KR, Walters EE. Prevalence, severity and comorbidity of 12-month DSM-V disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry* (2005) 62(6):617–26. doi:10.1001/archpsyc.62.6.593
19. Andrews B, Brewin CR, Philpott R, Stewart L. Delayed-onset posttraumatic stress disorder: a systematic review of the evidence. *Am J Psychiatry* (2007) 164:1319–26. doi:10.1176/appi.ajp.2007.06091491
20. Smid GE, Mooven TT, van der Mast RC, Gersons BP, Kleber RJ. Delayed post-traumatic stress disorder: systematic review, meta-analysis and meta-regression analysis of prospective studies. *J Clin Psychiatry* (2009) 70(11):1572–82. doi:10.4088/JCP.08r04484
21. Grieger TA, Cozza SJ, Ursano RJ, Hoge C, Martinez PE, Engel CC, et al. Posttraumatic stress disorder and depression in battle-injured soldiers. *Am J Psychiatry* (2006) 163:1777–83. doi:10.1176/ajp.2006.163.10.1777
22. Bryant RA, O'Donnell ML, Creamer M, McFarlane AC, Silove D. A multisite analysis of the fluctuating course of posttraumatic stress disorder. *JAMA Psychiatry* (2013) 70(8):839–46. doi:10.1001/jamapsychiatry.2013.1137
23. Araki T, Kasai K, Yamasue H, Kato N, Kudo N, Ohtani T, et al. Association between lower P300 amplitude and smaller anterior cingulate cortex volume in patients with posttraumatic stress disorder: a study of victims of Tokyo subway sarin attack. *Neuroimage* (2005) 25:43–50. doi:10.1016/j.neuroimage.2004.11.039
24. Blomhoff S, Reinvang I, Malt UF. Event-related potentials to stimuli with emotional impact in posttraumatic stress patients. *Biol Psychiatry* (1998) 44(10):1045–53. doi:10.1016/S0006-3223(98)00058-4
25. Charles G, Hansenne M, Anseau M, Pichot W, Machowski R, Schittecatte M, et al. P300 in posttraumatic stress disorder. *Neuropsychobiology* (1995) 32(2):72–4. doi:10.1159/000119216
26. Felmingham KL, Bryant RA, Kendall C, Gordon E. Event-related potential dysfunction in posttraumatic stress disorder: the role of numbing. *Psychiatry Res* (2002) 109:171–9. doi:10.1016/S0165-1781(02)00003-3
27. Ghisolfi ES, Margis R, Becker J, Zanardo AP, Strimitzer IM, Lara DR. Impaired P50 sensory gating in post-traumatic stress disorder secondary to urban violence. *Int J Psychophysiol* (2004) 51(3):209–14. doi:10.1016/j.ijpsycho.2003.09.002
28. Hansenne M. Event-related brain potentials in psychopathology: clinical and cognitive perspectives. *Psychol Belg* (2006) 46(1–2):5–36. doi:10.5334/pb-46-1-2-5
29. Javanbakht A, Liberzon I, Amirsadri A, Gjini K, Boutros NN. Event-related potential studies of post-traumatic stress disorder: a critical review and synthesis. *Biol Mood Anxiety Disord* (2011) 1:5. doi:10.1186/2045-5380-1-5
30. Johnson JD, Allana TN, Medlin MD, Harris EW, Karl A. Meta-analytic review of P3 components in posttraumatic stress disorder and their clinical utility. *Clin EEG Neurosci* (2013) 44(2):112–34. doi:10.1177/1550059412469742
31. Karl A, Malta LS, Maercker A. Meta-analytic review of event-related potential studies in post-traumatic stress disorder. *Biol Psychol* (2006) 71:123–47. doi:10.1016/j.biopsycho.2005.03.004
32. Kimble M, Kaloupek D, Kaufman M, Deldin P. Stimulus novelty differentially affects attentional allocation in PTSD. *Biol Psychiatry* (2000) 47:880–90. doi:10.1016/S0006-3223(99)00258-9
33. Kimble M, Ruddy K, Deldin P, Kaufman M. A CNV-distraction paradigm in combat veterans with posttraumatic stress disorder. *J Neuropsychiatry Clin Neurosci* (2004) 16:102–8. doi:10.1176/jnp.16.1.102
34. McFarlane AC, Weber DL, Clark CR. Abnormal stimulus processing in post-traumatic stress disorder. *Biol Psychiatry* (1993) 34:311–20. doi:10.1016/0006-3223(93)90088-U
35. Metzger LJ, Orr SP, Lasko NB, Pitman RK. Auditory event related potentials to tone stimuli in combat-related posttraumatic stress disorder. *Biol Psychiatry* (1997) 42:1006–15. doi:10.1016/S0006-3223(97)00138-8
36. Metzger LT, Orr SP, Lasko NB, McNally RJ, Pitman RK. Seeking the source of emotional Stroop interference effects in PTSD: a study of P3a to traumatic words. *Integr Physiol Behav Sci* (1997) 32(1):43–51. doi:10.1007/BF02688612
37. Metzger LJ, Carson MA, Paulus LA, Lasko NB, Paige SR, Pitman RK, et al. Event related potentials to auditory stimuli in female Vietnam nurse veterans with posttraumatic stress disorder. *Psychophysiology* (2002) 39(1):49–63. doi:10.1111/1469-8986.3910049
38. Neylan TC, Fletcher DJ, Lenoci M, McCallin K, Weiss DS, Schoenfeld FB, et al. Sensory gating in chronic posttraumatic stress disorder: reduced auditory p50 suppression in combat veterans. *Biological Psychiatry* (1999) 46(12):1656–64. doi:10.1016/S0006-3223(99)00047-5
39. Neylan TC, Jasiukaitis PA, Lenoci M, Scott JC, Metzler TJ, Weiss DS, et al. Temporal instability of auditory and visual event-related potentials in posttraumatic stress disorder. *Biol Psychiatry* (2003) 53:216–25. doi:10.1016/S0006-3223(02)01450-6
40. Shu I-W, Onton JA, Prabhakar N, O'Connell RM, Simmons AN, Matthews SC. Combat veterans with PTSD after mild TBI exhibit greater ERPs from posterior-medial cortical areas while appraising facial features. *J Affect Disord* (2014) 155:234–40. doi:10.1016/j.jad.2013.06.057
41. Shu I-W, Onton JA, O'Connell RM, Simmons AN, Matthews SC. Combat veterans with comorbid PTSD and mild TBI exhibit a greater inhibitory processing ERP from the dorsal anterior cingulate cortex. *Psychiatr Res Neuroimaging* (2014) 224(1):58–66. doi:10.1016/j.psychres.2014.07.010
42. Shucard JL, McCabe DC, Szymanski H. An event-related potential study of attention deficits in posttraumatic stress disorder during auditory and visual Go/NoGo continuous performance tasks. *Biol Psychol* (2008) 79(2):223–33. doi:10.1016/j.biopsycho.2008.05.005
43. Polich J. Updating P300: an integrative theory of P3a and P3b. *Neurophysiol Clin* (2007) 118(10):2128–48. doi:10.1016/j.clinph.2007.04.019
44. Isreal JB, Chesney GL, Wickens CD, Donchin E. P300 and tracking difficulty: evidence for multiple resources in dual-task performance. *Psychophysiology* (1980) 17:259–73. doi:10.1111/j.1469-8986.1980.tb00146.x
45. Kramer AF, Wickens CD, Donchin E. Processing of stimulus properties: evidence for dual-task integrality. *J Exp Psychol Hum Percept Perform* (1985) 11:393–408.
46. Mulert C, Pogarell O, Juckel G, Rujescu D, Giegling I, Rupp D, et al. The neural basis of the P300 potential. Focus on the time-course of the underlying

- cortical generators. *Eur Arch Psychiatry Clin Neurosci* (2004) 254(3):190–8. doi:10.1007/s00406-004-0469-2
47. Picton T. The P300 wave of the human event-related potential. *J Clin Neurophysiol* (1992) 9:456–79. doi:10.1097/00004691-199210000-00002
 48. Shin LM, Rauch SL, Pitman RK. Amygdala, medial prefrontal cortex and hippocampal function in PTSD. *Ann N Y Acad Sci* (2006) 1079:67–79. doi:10.1196/annals.1364.007
 49. Bush G, Luu P, Posner MI. Cognitive and emotional influences in the anterior cingulate cortex. *Trends Cogn Sci* (2000) 4(6):215–22. doi:10.1016/S1364-6613(00)01483-2
 50. Jovanovic T, Norrholm SD. Neural mechanisms of impaired fear inhibition in posttraumatic stress disorder. *Front Behav Neurosci* (2011) 5:44. doi:10.3389/fnbeh.2011.00044
 51. Maren S, Phan KL, Liberzon I. The contextual brain: implications for fear conditioning, extinction and psychopathology. *Nat Rev Neurosci* (2013) 14:417–28. doi:10.1038/nrn3492
 52. Weathers FW, Keane TM, Davidson RJ. Clinician-administered PTSD scale: a review of the first ten years of research. *Depress Anxiety* (2001) 13:132–56. doi:10.1002/da.1029
 53. Forbes D, Creamer M, Biddle D. The validity of the PTSD checklist as a measure of symptomatic change in combat-related PTSD. *Behav Res Ther* (2001) 39:977–86. doi:10.1016/S0005-7967(00)00084-X
 54. Spitzer RL, Kroenke D, Williams JBW; The Patient Health Questionnaire Primary Care Study Group. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. *J Am Med Assoc* (1999) 282(18):1737–44. doi:10.1001/jama.282.18.1737
 55. Pollich J, Ellerson PC, Cohen J. P300, stimulus intensity, modality and probability. *Int J Psychophysiol* (1996) 23(1–2):55–62. doi:10.1016/0167-8760(96)00028-1
 56. Cohen J, Polich J. On the number of trials needed for P300. *Int J Psychophysiol* (1997) 25(3):249–55. doi:10.1016/S0167-8760(96)00743-X
 57. Croft RJ, Barry RJ. EOG correction: a new perspective. *Electroencephalogr Clin Neurophysiol* (1998) 107(6):387–94. doi:10.1016/S0013-4694(98)00086-8
 58. Rapp PE, Cellucci CJ, Keyser DO, Gilpin AMK, Darmon DM. Statistical issues in TBI clinical studies. *Front Neurol* (2013) 4:177. doi:10.3389/fneur.2013.00177
 59. Youden D. Index for rating diagnostic tests. *Cancer* (1950) 3:32–5. doi:10.1002/1097-0142(1950)3:1<32::AID-CNCR2820030106>3.0.CO;2-3
 60. Rapp PE, Keyser DO, Albano AM, Hernandez R, Gibson D, Zambon R, et al. Traumatic brain injury detection using electrophysiological methods. *Front Hum Neurosci* (2015) 9:11. doi:10.3389/fnhum.2015.00011
 61. Darmon D. Specific differential entropy rate estimation for continuous-valued time series. *Entropy* (2016) 18:190. doi:10.3390/e18050190
 62. Smith Bassett D, Bullmore E. Small-world brain networks. *Neuroscientist* (2006) 12(6):512–23. doi:10.1177/1073858406293182
 63. Shedler J, Mayman M, Manis M. The illusion of mental health. *Am Psychol* (1993) 48(11):1117–31. doi:10.1037/0003-066X.48.11.1117
 64. Orr SP, Roth WT. Psychophysiological assessment: clinical applications for PTSD. *J Affect Disord* (2000) 61(3):225–40. doi:10.1016/S0165-0327(00)00340-2
 65. Rothbaum BO, Kearns MC, Price M, Malcoun E, Davis M, Ressler KJ, et al. Early intervention may prevent the development of posttraumatic stress disorder: a randomized pilot civilian study with modified prolonged exposure. *Biol Psychiatry* (2012) 72(11):957–63. doi:10.1016/j.biopsych.2012.06.002
 66. Böhning D, Holling H, Patilea V. A limitation of the diagnostic odds ratio in determining an optimal cut-off value for a continuous diagnostic test. *Stat Methods Med Res* (2011) 20(5):541–50. doi:10.1177/0962280210374532
 67. Efron B. Better bootstrap confidence intervals. *J Am Stat Assoc* (1987) 82:171–200. doi:10.1080/01621459.1987.10478410
 68. Efron B, Tibshirani R. Bootstrap methods for statistical errors: confidence intervals and other measures of statistical accuracy. *Stat Sci* (1986) 1(1):54–75. doi:10.1214/ss/1177013815
 69. Efron B, Tibshirani R. *An Introduction to the Bootstrap*. London: Chapman and Hall (1993).
 70. Glas AS, Lijmer JG, Prins MH, Bonse GJ, Bussuyt PM. The diagnostic odds ratio: a single indicator of test performance. *J Clin Epidemiol* (2003) 56(11):1129–35. doi:10.1016/S0895-4356(03)00177-X
 71. Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology* (1983) 148:839–42. doi:10.1148/radiology.148.3.6878708
 72. Hoeffding W. Probability inequalities for sums of bounded random variables. *J Am Stat Assoc* (1963) 58(301):13–30. doi:10.1080/01621459.1963.10500830
 73. Jiang W, Varma S, Simon R. Calculating confidence intervals for prediction error microarray classification using resampling. *Stat Appl Genet Mol Biol* (2008) 7(1):8. doi:10.2202/1544-6115.1322
 74. Krazanowski W, Hand DJ. *ROC Curves for Continuous Data*. Boca Raton, FL: Chapman Hall/CRC (2009).
 75. Lachenbruch PA. *Discriminant Analysis*. New York, NY: Hafner Press (1975).
 76. Pepe MS. *Statistical Evaluation of Medical Tests or Classification and Prediction*. Oxford: Oxford University Press (2003).
 77. Pepe MS, Janes H, Longton G, Leisenring W, Newcomb P. Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic or screening marker. *Am J Epidemiol* (2004) 159(9):882–90. doi:10.1093/aje/kwh101
 78. Portney LG, Watkins MP. *Foundations of Clinical Research. Applications to Practice*. 3rd ed. Upper Saddle River, NJ: Prentice Hall Health (2009).
 79. Wasserman L. *All of Statistics: A Concise Course in Statistical Inference*. New York, NY: Springer (2010).
 80. Zou KH, Hall WJ, Shapiro DE. Smooth non-parametric receiver operating characteristic (ROC) curves for continuous diagnostic tests. *Stat Med* (1997) 16:2143–56.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Wang, Costanzo, Rapp, Darmon, Bashirelahi, Nathan, Cellucci, Roy and Keyser. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

MATHEMATICAL APPENDICES

Estimating Classification Error (Contents of Table 3)

For the case of a single discriminating variable, the Group A–Group B between-group Mahalanobis distance is

$$D_{A,B}^2 = \frac{(\hat{\mu}_A - \hat{\mu}_B)^2}{\hat{\sigma}_{AB}^2},$$

$$\hat{\sigma}_{AB}^2 = \frac{(N_A - 1)\hat{\sigma}_A^2 + (N_B - 1)\hat{\sigma}_B^2}{N_A + N_B - 2},$$

$\hat{\mu}_A$ is the Group A sample mean, and $\hat{\sigma}_A$ is the Group A sample SD. $\hat{\mu}_B$ and $\hat{\sigma}_B$ are defined analogously. $P_{\text{ERROR}}(G_A, G_B)$ is the error rate for the optimal classifier under the assumption of normality for the two populations and provides an estimate classification error when only means and SDs are known. It can give a serious underestimate of true classification error. This is especially true if group population numbers are low or the assumption of normality is violated. When full data sets are available, an empirical calculation of error rate is preferred *via* either cross-validation or bootstrapping. Let ρ_A and ρ_B be prior probabilities of Group A and Group B membership. $P_{\text{ERROR}}(G_A, G_B)$ is given by

$$P_{\text{ERROR}}(G_A, G_B) = \rho_A \Phi \left\{ -\frac{1}{2} \sqrt{D_{A,B}^2} + \frac{1}{\sqrt{D_{A,B}^2}} \log_e \left(\frac{\rho_B}{\rho_A} \right) \right\}$$

$$+ \rho_B \Phi \left\{ -\frac{1}{2} \sqrt{D_{A,B}^2} - \frac{1}{\sqrt{D_{A,B}^2}} \log_e \left(\frac{\rho_B}{\rho_A} \right) \right\}$$

where $\Phi(x)$ is the cumulative distribution function for a standard normal random variable (75). For the case of equal priors, the expression reduces to

$$P_{\text{ERROR}}(G_A, G_B) = 1 - \Phi \left(\frac{\sqrt{D_{A,B}^2}}{2} \right) = \Phi \left(\frac{-\sqrt{D_{A,B}^2}}{2} \right).$$

Receiver Operating Characteristic Curve (Contents of Table 4)

The area under an empirical receiver operating characteristic curve is equal to the Mann–Whitney U statistic [Ref. (74), p. 65 following from a proof on p. 27] and thus the Mann–Whitney U statistic provides an estimator for the population level AUC. Random assignment results in $\text{AUC} = 0.5$. The following notation is introduced:

- N_S number of longitudinally stable participants
- N_C number of converter participants
- S_i observed value for the i -th stable participant
- C_j observed value for the j -th converter participant.

$$\widehat{\text{AUC}} = \frac{1}{N_C N_S} \sum_{i=1}^{N_S} \sum_{j=1}^{N_C} \left\{ I(C_j < S_i) + \frac{1}{2} I(C_j = S_i) \right\}$$

where $I(Z) = 1$ if argument Z is true. It is important to note that “less than” used in this application, *contra* textbooks where

“greater than” appears, because in this analysis a participant is classed as positive if the observed value is less than the threshold value.

There are several estimates of the variance of the AUC [listed on p. 67 of Ref. (74)]. We use here the expression in Hanley and McNeil (71).

$$s^2(\widehat{\text{AUC}}) = \frac{1}{N_S N_C} \left\{ \widehat{\text{AUC}}(1 - \widehat{\text{AUC}}) + (N_C - 1)(Q_1 - \widehat{\text{AUC}}^2) + (N_S - 1)(Q_2 - \widehat{\text{AUC}}^2) \right\}$$

As in the equation for AUC, the definition of Q_1 and Q_2 uses “less than” rather than “greater than” because a participant is classed as a positive if the measure value is below threshold rather than greater than threshold. Q_1 is the proportion of all possible triples composed of two sampled members from the Converter group and one from the Stable group where the two Converter scores are less than the Stable score

$$Q_1 = \frac{1}{N_S N_C N_C} \sum_{i=1}^{N_S} \sum_{j=1}^{N_C} \sum_{k=1}^{N_C} I(C_j < S_i) I(C_k < S_i)$$

Q_2 is the proportion of all possible triples composed on one member from the Converter group and two members from the Stable group where the Converter score is less than both scores from the Stable group.

$$Q_2 = \frac{1}{N_C N_S N_S} \sum_{i=1}^{N_C} \sum_{j=1}^{N_S} \sum_{k=1}^{N_S} I(C_i < S_j) I(C_i < S_k)$$

An expression for confidence intervals has been constructed by (80), where with confidence $1 - \alpha$, the true AUC lies in the interval given by

$$1 - (1 - \widehat{\text{AUC}}) \exp \left\{ \pm z_{1-\alpha/2} \sqrt{s^2(\widehat{\text{AUC}}) / (1 - \widehat{\text{AUC}})} \right\},$$

where $z_{1-\alpha/2}$ is the $1 - \alpha/2$ quantile of a standard normal random variable. Under this transformation/inverse transformation, the upper and lower confidence intervals are always in the interval [0,1].

An analysis of the ROC can be used to determine the optimal cutoff value for a continuous, dichotomous diagnostic test. Glas et al. (70) have endorsed the diagnostic odds ratio as a single indicator of test performance and proposed using its maximum to determine the cutoff value. Pepe et al. (76) have argued against this practice and have provided examples that identify limitations of the odds ratio. A fundamental limitation is immediately apparent on examining the equation below for the ratio. It is undefined if the number of false positive or the number of false negatives is 0. Böhning et al. (66) have continued the analysis and recommend using the maximum value of the Youden index (59) as an alternative indicator of the best cutoff value. The Youden index, also called the Youden J statistic is

$$J = \text{Sensitivity} + \text{Specificity} - 1.$$

It is reported as a function of threshold, and the recommended value of threshold is the lowest threshold value giving the maximum of J .

Measures of Diagnostic Efficacy (Contents of Table 4)

Dichotomous diagnosis (two possible outcomes, disease positive and disease negative), using a single continuous variable is considered here. The diagnostic utility of the measure and classifier combination is investigated by first populating the diagnostic matrix and then computing standard measures of diagnostic efficacy. Six measures are considered here diagnostic accuracy, sensitivity, specificity, positive likelihood ratio, negative likelihood ratio, and the diagnostic odds ratio, where it should be recognized that no single measure of diagnostic effectiveness provides a complete assessment of a measure’s ability to classify participants (76). Additional measures are presented in Pepe (76) and in Portney and Watkins (78).

	Disease present	Disease absent
Test positive	True positive N_{TP}	False positive N_{FP}
Test negative	False negative N_{FN}	True negative N_{TN}

$$\text{Diagnostic Accuracy} = (N_{TP} + N_{TN}) / (N_{TP} + N_{FP} + N_{FN} + N_{TN})$$

$$\text{Sensitivity} = \text{True Positive Fraction} = \frac{N_{TP}}{N_{TP} + N_{FN}}$$

$$\text{Specificity} = \text{True Negative Fraction} = \frac{N_{TN}}{N_{TN} + N_{FP}}$$

$$\begin{aligned} \text{Positive Likelihood Ratio} = \text{LR+} &= \frac{\text{True Positive Fraction}}{\text{False Positive Fraction}} \\ &= \left(\frac{N_{TP}}{N_{TP} + N_{FN}} \right) / \left(\frac{N_{FP}}{N_{FP} + N_{TN}} \right) \\ &= \frac{\text{Sensitivity}}{1 - \text{Specificity}} \end{aligned}$$

$$\begin{aligned} \text{Negative Likelihood Ratio} = \text{LR-} &= \frac{\text{False Negative Fraction}}{\text{True Negative Fraction}} \\ &= \left(\frac{N_{FN}}{N_{FN} + N_{TP}} \right) / \left(\frac{N_{TN}}{N_{TN} + N_{FP}} \right) \\ &= \frac{1 - \text{Sensitivity}}{\text{Specificity}} \end{aligned}$$

$$\begin{aligned} \text{Diagnostic Odds Ratio} = \text{DOR} &= \frac{\text{LR+}}{\text{LR-}} = \left(\frac{N_{TP} \cdot N_{TN}}{N_{FP} \cdot N_{FN}} \right) \\ &= \left(\frac{\text{Sensitivity}}{1 - \text{Sensitivity}} \right) \left(\frac{\text{Specificity}}{1 - \text{Specificity}} \right) \end{aligned}$$

Classification Based on Gaussian Likelihood and Bayesian Priors (Contents of Table 5)

The classifier is constructed from a single continuous variable, in this case the amplitude of the average response to the target stimulus. $\{S_j\}_{j=1}^{N_S}$ is the set of values obtained from clinically stable participants. $\hat{\mu}_S$ is the sample mean and $\hat{\sigma}_S$ is the corresponding SD. $\{C_j\}_{j=1}^{N_C}$ is the set of values obtained from participants who became PTSD-positive and has mean $\hat{\mu}_C$ and SD $\hat{\sigma}_C$. Let x denote the value of the measure obtained from the individual who is to be classified. The group specific density function for clinically stable participants is

$$f_S(x) = \frac{1}{(2\pi\hat{\sigma}_S^2)^{1/2}} \exp\left\{-\frac{1}{2}(x - \hat{\mu}_S)^2 / \hat{\sigma}_S^2\right\}$$

$f_C(x)$ is defined analogously. The posterior probabilities of group membership are

$$P(G = S|x) = \frac{\rho_S f_S(x)}{\rho_S f_S(x) + \rho_C f_C(x)}$$

and

$$P(G = C|x) = \frac{\rho_C f_C(x)}{\rho_S f_S(x) + \rho_C f_C(x)}$$

where ρ_S and ρ_C are the prior probabilities of membership in the healthy or disease-positive groups. The participant presenting measure equal to x is classified into the group with the higher posterior probability.

Populating the Diagnostic Matrix with a Bootstrap Estimator (Contents of Table 6)

A bootstrap (67) can be used to determine the value of the diagnostic metrics, and the corresponding confidence intervals. The procedure used here is similar to the bootstrap cross validation scheme for small sample sizes implemented by Jiang et al. (73). A procedure for finding the best estimate of Sensitivity from the available data is described here. The procedure immediately generalizes to other measures of diagnostic efficacy.

As before, this presentation describes a dichotomous classification using a single continuous variable between two groups, clinically stable participants and participants presenting delayed onset PTSD. $\{S_j\}_{j=1}^{N_S}$ is the set of values of this measure obtained from clinically stable participants. There are N_S elements. $\{C_j\}_{j=1}^{N_C}$ is the set of values obtained from participants who convert to the PTSD-positive state. There are N_C values. A single iteration of the bootstrap proceeds as follows:

1. $N_C + N_S$ elements are drawn randomly with replacement from the combined set $\{S_j\}_{j=1}^{N_S} \cup \{C_j\}_{j=1}^{N_C}$. This set of randomly drawn elements is denoted by $\{B_j\}_{j=1}^{N_S+N_C}$, the bootstrap sample. Typically, the bootstrap sample will contain repeated values. It is possible that the bootstrap sample does not contain an element from either set $\{S_j\}_{j=1}^{N_S}$ or set $\{C_j\}_{j=1}^{N_C}$. If this

occurs, this iteration of the bootstrap is ignored. Additionally, depending on the classifier used, a minimum number of distinct values from $\{S_j\}_{j=1}^{N_s}$ and $\{C_j\}_{j=1}^{N_c}$ will be required to construct the classifier. For example, the classifier based on Gaussian population densities will require at least two distinct elements from each set. If this minimum requirement is not satisfied this iteration of the bootstrap is ignored and the process returns to the beginning of Step 1. Also, if there is not at least one element of $\{S_j\}_{j=1}^{N_s}$ and one element of $\{C_j\}_{j=1}^{N_c}$ in the set of elements that will be classified, the randomization is rejected and the process returns to the beginning of Step 1.

2. The class membership of each element of $\{B_j\}_{j=1}^{N_s+N_c}$ is known. Use $\{B_j\}_{j=1}^{N_s+N_c}$ to construct a classifier.
3. Use this classifier to classify all members of the combined set $\{S_j\}_{j=1}^{N_s} \cup \{C_j\}_{j=1}^{N_c}$ not in $\{B_j\}_{j=1}^{N_s+N_c}$, namely $\{\{S_j\}_{j=1}^{N_s} \cup \{C_j\}_{j=1}^{N_c}\} \setminus \{B_j\}_{j=1}^{N_s+N_c}$. The results of this classification are used to calculate N_{TP} , N_{FP} , N_{FN} , N_{TN} specific to this bootstrap sample. Though in the general case, it is possible, but unlikely, that $\{\{S_j\}_{j=1}^{N_s} \cup \{C_j\}_{j=1}^{N_c}\} \setminus \{B_j\}_{j=1}^{N_s+N_c}$ is the null set, this will not occur in the present application because of the constraints on the randomization put in place in Step 1.
4. Sensitivity and other measures of diagnostic efficacy for this iteration of the bootstrap are then calculated using standard formulas.

This process is repeated until N_B values of Sensitivity are obtained. This may require more than N_B iterations of the bootstrap if the requirements of the random sample outlined in Step 1 are not met.

The average value of sensitivity, computed from the N_B successful iterations of the bootstrap is the best available estimate from $\{S_j\}_{j=1}^{N_s}$ and $\{C_j\}_{j=1}^{N_c}$. The confidence interval of sensitivity can be determined from the distribution of the N_B values of sensitivity. For example, suppose that sensitivities are calculated from 2,000 bootstrap samples and suppose that the 95% confidence interval is to be determined. Rank order the values of sensitivity. The lower bound of the confidence interval is the 50th element, and the upper bound is element 1950th.

This leaves the specification of N_B as an open question. This is not a question that has a single answer (68, 69). The required number of iterations will depend on what is being estimated and the properties of the underlying distribution. A convention in the community regards $N_B = 1,000$ as a lower bound. As an operational suggestion the estimate of sensitivity, for $N_B = 1,000$ and $N_B = 2,000$ can be compared. N_B should be large enough to give a stable value of sensitivity. $N_B = 2,000$ was used in these calculations.

This is a constrained randomization. At least two distinct elements of each class (Stables and Converters) must be in the set

used to construct the classifier (the Training Set, $\{B_j\}_{j=1}^{N_s+N_c}$). At least one element of each class must be in the set that is classified (the Testing Set = $\{S_j\}_{j=1}^{N_s} \cup \{C_j\}_{j=1}^{N_c} \setminus \{B_j\}_{j=1}^{N_s+N_c}$). Because at least one element of $\{C_j\}_{j=1}^{N_c}$ is classified, there will be at least one true positive (a converter assigned into the converter group) or one false negative (a converter classified into the stable group). Sensitivity may be 0 ($N_{TP} = 0$), but it will not be singular because $N_{TP} + N_{FN} \neq 0$. Because there are only five elements of $\{C_j\}_{j=1}^{N_c}$ and two are used to build the classifier, N_{TP} is, however, frequently 0, and Sensitivity = 0 is therefore a frequent result from an iteration of the bootstrap. Additionally in many other cases, $N_{TP} \neq 0$, but $N_{FN} = 0$ giving Sensitivity = 1. This explains the confidence interval of [0,1].

Because at least one element of $\{S_j\}_{j=1}^{N_s}$ will be classified, there is at least one true negative or one false positive. Therefore, since $N_{TN} + N_{FP} \neq 0$, specificity will be defined at each iteration of the bootstrap. In contrast with Sensitivity, because $\{S_j\}_{j=1}^{N_s}$ is a much larger set, Specificity typically shows values different from 0 to 1.

The positive likelihood ratio is undefined if Specificity is equal to 1. As noted in the preceding paragraph this is unlikely, but it is possible. The negative likelihood ratio is undefined if Specificity is equal to 1. This frequently occurs with these data. The diagnostic odds ratio is undefined if either Specificity or Sensitivity is equal to 1. Glas et al. [(70), p. 1131] suggests adding 0.5 to all four elements of the diagnostic matrix in those applications where undefined values of the diagnostic ratios are likely to occur. This was done in these calculations.

Sample Size Requirements for Measures of Diagnostic Efficacy

Sample size requirements for sensitivity and specificity assessments can be computed using an argument based on Hoeffding's inequality [(72, 79), p. 65]. If α is the significance level for a confidence interval of length 2Δ , we require

$$\alpha \leq 2e^{-2N\Delta^2}$$

giving

$$N \geq \frac{-\ln(\alpha/2)}{2\Delta^2}$$

The sample size required for a ± 0.1 sensitivity estimate with 95% ($\alpha = 0.05$) confidence is seen to be $N \geq 185$. It should be stressed that this is an estimate of sensitivity. N in this equation is the number of individuals in the sample who are disease positive. If the prevalence of the disorder in the enrollment population is 10%, then an enrollment $\geq 1,850$ is required.