
Article

Cautionary Observations Concerning the Introduction of Psychophysiological Biomarkers into Neuropsychiatric Practice

Paul E. Rapp^{1,*}, Christopher Cellucci², David Darmon³ and David Keyser⁴

¹ Department of Military and Emergency Medicine, Uniformed Services University, Bethesda, MD, USA; paul.rapp@usuhs.edu

² Aquinas LLC, Berwyn, PA, USA ; cellucci@gmail.com

³ Aquinas LLC, Berwyn, PA, USA; david.m.darmon@gmail.com

⁴ Department of Military and Emergency Medicine, Uniformed Services University, Bethesda, MD; david.keyser@usuhs.edu

* Correspondence: paul.rapp@usuhs.edu

Abstract: The combination of statistical learning technologies with large databases of psychophysiological data has appropriately generated enthusiastic interest in future clinical applicability. It is argued here that this enthusiasm should be tempered with the understanding that significant obstacles must be overcome before the systematic introduction of psychophysiological measures into neuropsychiatric practice becomes possible. The nonspecificity of psychophysiological measures complicates their use in diagnosis. Low test-retest reliability complicates use in longitudinal assessment, and quantitative psychophysiological measures can normalize in response to placebo intervention. Ten cautionary observations are introduced and, in some instances, possible directions for remediation are suggested.

Keywords: biomarkers; electroencephalography; event related potentials; heart rate variability; diagnosis; sensitivity; specificity

Introduction

Psychophysiology is the branch of physiology dealing with the relationships between physiological processes and psychological phenomena (thoughts, emotions, and behaviors). Clinical psychophysiology is more narrowly defined here as the use of psychophysiological measures to inform assessment and treatment of neuropsychiatric disease. Three applications are critical to this program: diagnosis, longitudinal assessment of treatment response or disease progression, and identification of individuals in the subsyndromal state who are at risk of neuropsychiatric disorders. The possibility of combining statistical machine learning technologies with psychophysiological measures to address these objectives has appropriately generated a great deal of enthusiasm. It is argued here, however, that this enthusiasm should be tempered with an appreciation of the challenges that are still ahead. This contribution presents ten cautionary observations. Indications of how some of these challenges might in part be addressed are also presented.

1. The central role of patient report Observation

The limitations of patient self-report are commonly recognized [1], [2]. Patients can give inadequate report due to denial, lack of insight or willful intention to mislead. This has motivated the introduction of psychophysiological measures into clinical practice. Additionally, psychophysiological measures are by definition related to physiological processes and this physiological information may inform treatment. Nonetheless in neuropsychiatric practice, patient report remains critical to all that follows. This is arguably true in all areas of practice, but it is particularly true in neuropsychiatry where there are no dispositive clinical measures analogous to blood pressure, plasma glucose concentration or tumor volume. While physiology may be part of the story, it is not all of the story. Patient report

remains the final arbiter.

Response to Observation 1.

Recognizing that patient reports are integral to practice, an effort can be made to obtain these reports systematically with standardized questionnaires [3]. Constructing and establishing the reliability and validity of a health scale is a significant undertaking [4]. This argues strongly for using previously developed and validated questionnaires. In the case of clinical trials, the US Food and Drug Administration has provided guidance [5]. More generally, the COSMIN study produced an international consensus on measurement properties for health related, patient-reported outcomes (COSMIN, Consensus-based Standards for the Selection of a Health Measurement Instrument, [6], [7], [8]). This panel generated a checklist for assessing measurement instruments [10], [9]. It is suggested that investigations use questionnaires that satisfy checklist criteria.

2. The physiological implementation of consciousness is not understood

Observation

We do not understand the physiological basis of consciousness. Here we use the word consciousness to refer to higher cognitive processes rather than, for example, the distinction between being awake and asleep or anesthetized. Indeed, some have argued that this understanding is not possible ([11]; for an alternative view, however, see [12]). Neuropsychiatric presentations can include deformations of consciousness ranging from minimal in cases of mild cognitive impairment to profound in some presentations of psychotic disorders. This has clinical implications. As summarized by Harrington [13]: "After all, current brain science has little understanding of the biological foundations of many – indeed most - everyday mental activities. This being the case, how could current psychiatry possibly expect to have a mature understanding of how such activities become disordered – and may possibly be reordered. We have previously observed that "Twenty-first century clinicians are charged with constructing a physiologically informed response to these presentations (abnormalities of consciousness). An understanding of the physiological basis of consciousness is therefore not simply a matter of deep scientific and philosophical importance. Rather, it is of immediate clinical significance," [14]. While the ultimate accessibility of an understanding of the basis of consciousness is a matter of debate, it is certainly true that a solution is not at present available.

Response to Observation 2.

While the physiological implementation of consciousness is not understood, a great deal is now known about the neural correlates of consciousness [15], [16], [17]. Psychophysiological variables known to be correlated with specific mental processes such as perception and decision making have provided starting points of efforts to identify potentially useful clinical biomarkers. This ongoing activity is accelerating.

3. The interaction between conscious and unconscious processes is not understood but is clinically important

Observation

The mystery of consciousness deepens when we recognize that unconscious processes are a significant component of psychological functioning [18] and that the physiological basis of unconscious processes are largely unknown. As in the cases of conscious psychological processes, this has clinical implications. Emotional processes can be unconscious [19], and unconscious processes can have a significant impact on health [20]. Wiers, et al. [21] concluded that "implicit processes might be particularly important in psychopathology." An understanding of the physiological basis of the unconscious is therefore also an unmet clinical requirement.

Response to Observation 3.

The psychophysiology of the unconscious can also be investigated empirically ([22], [23], [24], [25], [26], these are representative examples drawn from a larger literature). As in the case of conscious cognitive processes, quantitative measures developed by these

investigations may prove to be of clinical value.

4. High inter-individual variation

Observation

The next cautionary observation is admirably summarized by Shackman and Fox in the title of their 2018 contribution [27] "Getting serious about variation: lessons for clinical neuroscience," in which they cite the Holmes and Patrick paper "The myth of optimality in clinical neuroscience" [28]. Psychophysiological measures are characterized by broad distributions. In this context it is appropriate to emphasize that statistically significant between-group differences do not ensure their usefulness in a classification. The commonly used independent samples t-test assesses the difference in the means of two distributions. The distributions can, however, overlap. A particularly instructive example is given in Holmes and Patrick (Figure 1) that presents a measure of frontotemporal connectivity obtained in patients with psychosis and healthy comparison participants. The group means are different, but the distributions show substantial overlap. They conclude: "Analytic approaches that focus on group differences may mask the presence of substantial overlap in phenotypic distributions providing an illusion of diagnostic specificity."

Response to Observation 4.

When considering the potential utility of a measure in a classifier, as noted by Holmes and Patrick, the independent samples t-test is not a completely satisfactory indicator. In the case of a dichotomous classification, rejection of the null hypothesis is not an adequate indication of usefulness in classification. An appropriate indicator is the probability of error in a classification. For example, Rapp, et al. [29] have published a computational example in which the p value is $p = 2.1 \times 10^{-12}$ but the classification error is 0.408, where the error rate for classifying with equal probability to the two classes is 0.5.

In the case of a dichotomous classification where it can be assumed that the distribution of the adjudicating measure in both groups is normal, or normal-enough, an analytic estimate of the probability of classification error is available ([30] see also [31]). Alternatively, an empirical determination of error can be obtained with a leave-one-out cross validation. The LOOCV has the advantage of generalizing to classification problems that include more than two groups. Determination of classification error rates should be included with the results of a t-test when reporting between-group differences observed with a psychophysiological measure.

5. Psychophysiological measures have low diagnostic specificity

Observation

The low specificities of psychophysiological measures complicate their use in diagnosis and as prodromes used in the identification of at-risk individuals. For example, Rapp, et al. ([32] Table 8) have identified results in the literature showing altered EEG synchronization patterns in AD/HD, alcohol abuse, alexithymia, autism, bipolar disorders, dementia, depression, hallucinations, HIV dementia, migraine, multiple sclerosis Parkinson's disease, post-traumatic stress disorder, schizophrenia, and other psychotic disorders. A similar literature study identified non-specific changes in functional connectivity in eleven clinical conditions ([32] Table 3).

A further complication must be recognized. Sensitivity and specificity results obtained in clinical studies with well-defined participant groups are misleading if not interpreted with care. Results obtained in a study where the comparison is between healthy persons and a "pure" clinical population that has satisfied rigorous inclusion/exclusion criteria can be unwarrantably optimistic. Specificity is expected to be far lower when a general psychiatric intake population is considered.

An additional consideration merits attention. It could be argued that the discouraging results obtained when psychophysiological measures are used in diagnosis could reflect deficiencies in diagnostic systems, notably the DSM-5. Kapur, Phillips and Insel [33] have suggested that, in part, a commitment to support DSM-5 diagnostic structures has been

an impediment to the development of clinically useful biomarkers. Newson, et al. [34] conducted a quantitative analysis of 107,349 adults using the Mental Health Quotient. Of those participants whose symptoms mapped to at least one of the ten DSM-5 diagnostic disorders considered in the study, the heterogeneity of symptom profiles was almost as high within a disorder as between two disorders, and “not separable from randomly selected groups of individuals with at least one of any of the 10 disorders.” In summary they concluded, “Overall, these results quantify the scale of misalignment between clinical symptom profiles and DSM-5 disorder labels and demonstrate that DSM-5 disorder criteria do not separate individuals from random when the complete mental health symptom profile of an individual is considered.” This conclusion could suggest that psychophysiological based diagnosis in support of the DSM-5 classifications is not possible in principle. The Newson, et al. results should, however, be compared with the results of DSM-5 interrater reliability field trials that report either “very good agreement” or “good agreement” in fourteen of twenty diagnostic categories ([35] Figure 1). Details of the field trials are given in [36], [37], [38].

The identification of individuals at risk of neuropsychiatric disease is a special case of a diagnostic process. To all of the previously described difficulties of psychophysiological assisted diagnosis we must introduce additional problems, specifically the challenges of a long term longitudinal study and the very large study populations required. In contrast with other areas of medical practice, reliable predictors of neuropsychiatric disease onset are often unavailable. Obstacles in the implementation of the much hoped for arrival of preemptive psychiatry [39] should be recognized. Consider the structure of the study. An asymptomatic or subsyndromal population is identified, and psychophysiological measures are acquired on intake. The population is followed for a specified duration, possibly months or years. Participants who remain stable (Stables) and those who present the disorder (Converters) are identified. The Stable versus Converter discrimination can itself be difficult. An attempt is then made to construct a classifier that discriminates between Stables and Converters using baseline intake data. By definition this is a longitudinal study. The challenges including expense and loss of participants to follow up are well known to investigators with experience in long term clinical studies.

Additionally, the number of individuals required in a prodrome search is much larger than might be expected. A critical determinant of the required sample size is the conversion rate in the study population. Conversion rates to psychiatric disorders in the general population are typically low. For this reason, investigators will attempt to identify an enriched population with a higher conversion rate. For example, in a study of electrophysiological prodromes of delayed onset PTSD, Wang, et al. [40] followed military personnel who had recently returned from combat. For this population the conversion rate is high, about 10%. Even in an enriched population the number of required participants can be high. There is more than one procedure for estimating the required sample size to obtain a confidence interval with a prespecified precision. Wang, et al. used Hoeffding’s inequality [41]. Using this criterion, the sample size required for ± 0.1 sensitivity estimate with a 95% confidence interval requires 185 Converters. If the conversion rate is 10% a total sample size of 1850 participants may be required. By fixing the precision of the interval estimate independently of the underlying sensitivity, a sample size determination based on Hoeffding’s inequality is conservative in the sense of requiring a large sample size. Alternatively, one could determine the sample size required to give an expected width using the Clopper-Pearson interval [42], [43]. With this criterion the required number of Converters for ± 0.1 is 104, giving a total sample size of 1040. Even with this more encouraging sample size, the study sample is significantly larger than that usually observed in the prodrome literature. For small sample size studies including the report of the confidence intervals of sensitivity and specificity is therefore especially important.

Problems associated with underpowered studies are not limited to identification of prodromes of neuropsychiatric disorders. Button, et al. [44] argue that the problem is ubiquitous in neuroscience. They stress that “... it is less well appreciated that lower power

also reduces the likelihood that a statistically significant result reflects a true effect. ... The consequences of this include overestimation of effect size and low reproducibility of results."

Response to Observation 5.

The low diagnostic specificity of psychophysiological measures can to a degree be addressed by expanding the analysis to multivariate classification. A well-known example from physiology provides an example. Acid-based disorders are broadly classified as alkalotic versus acidotic with respiratory or metabolic etiologies. A discrimination cannot be made by measuring pH alone or bicarbonate alone. Both must be measured simultaneously. Similarly, it is suggested that the utility of computationally informed neuropsychiatric assessment might be advanced by incorporating multiple psychophysiological measures and, importantly, other classes of measures (patient history, family history, imaging, genetics epigenetics, etc.).

The incorporation of multiple measures introduces an additional challenge: model selection. Predictors used in a classifier must be selected from the set of candidate predictors in a statistically responsible fashion. More is not necessarily better. Watanabe, et al [45] give a classification example (eyes open versus eyes closed, no task EEGs) where the classification error rate decreases as measures are eliminated from the classifier ([45], Figure 6). Fortunately, an extensive literature exists to guide model selection [46]. The validation of multivariate classifiers is an essential activity. Unfortunately, this is not always done correctly. Attention is directed to Section 7.10.2, "The Wrong and Right Way to Do Cross-Validation" in Hastie, et al. ([46]; see also [47]).

It is impossible to measure everything that might be measured. Statistical measures can guide the selection of predictors from a predetermined set of candidate predictors, but mathematics alone cannot direct the construction of that set. Ideally, as in all areas of practice, the selection of signals to be acquired and measures to be calculated should be driven by physiological hypotheses. For example, if it is hypothesized that the interaction of cognitive processes has been compromised following injury, then an ERP assessment based on a task like the flanker arrow task [48] where the difficulty of two processes is manipulated (stimulus identification and response selection) might be disclosing. If impairment of the autonomic nervous system is suspected, measures of heart rate variability are indicated. Simply put, there is no substitute for clinical insight.

6. The test-retest reliability of psychophysiological measures is frequently untested and can be unacceptably low

Observation

While the deficiencies of psychophysiological measures in diagnosis are occasionally recognized, it is often argued that measures with low specificity can still be useful as longitudinal measures. A clear example is body temperature. A fever is not specific to a single disorder, but it is nonetheless an essential clinical measure, but body temperature is stable (reliable) in health and is responsive to disease progression or recovery. High test-retest reliability of a measure in a clinically stable population is essential to its use in longitudinal clinical assessment. The literature on the test-retest reliability of psychophysiological and neuropsychological measures is limited and discouraging. For example, Cole, et al. [49] conducted a test-retest reliability study of four computerized neurocognitive assessment programs in a healthy active duty military population. They concluded "However, overall test-retest reliabilities in four NCATs (Neurocognitive Assessment Tools) in a military population are consistent with reliabilities reported in the literature (non-military populations) and are lower than desired for clinical decision making."

Response to Observation 6.

Estimating the test-retest reliability of a psychophysiological measure in healthy clinically stable controls is the essential first step because typically this population gives the best reliability (Gibson's Law). If a measure is not reliable in that population, its clinical utility is at best marginal. Unfortunately, methodological errors can be observed in some

investigations of reliability. In the case of a continuous variable, which is typically the case for psychophysiological measures, linear product moment (Pearson) correlations do not provide an appropriate quantification of reliability. The intraclass correlation coefficient, ICC, is the appropriate measure. There are, however, several measures known collectively as intraclass correlation coefficients. Shrout and Fleiss [50] give six versions, and McGraw and Wong [51] have ten versions. The choice depends on the reliability evaluation protocol being used. Müller and Büttner [52] and Koo and Li [53] provide selection guidance. Inappropriate versions of the intraclass correlation coefficient are often used. Because the numerical value of the ICC can vary considerably with the version being used it is essential to include a specification of the ICC version in the report [54].

Interpretation of ICCs is problematic. Fleiss [55] described values in the range 0.4 to 0.75 as good to fair. Koo and Li [53] gave bands for four characterizations (poor, moderate, good, excellent). De Mast [56] has described generalizations of this kind as being “hopelessly arbitrary.” Given this uncertainty, though often unreported, confidence intervals for ICCs are critical for their interpretation. A large literature has developed procedures for confidence interval construction for ICCs [50], [57], [58], [59].

The interpretation of intraclass correlation coefficients can be further advanced by using them to calculate the standard error of measurement and the minimum detectable difference [60]. While helpful when interpreting ICCs, it should be understood that the standard error of measurement and the minimum detectable difference are statistical properties of distributions of the measure obtained in test-retest studies. They are not directly connected to the clinical response. The minimum detectable difference is not the same as the minimum clinically important difference. Estimating the clinical importance of a change in a psychophysiological variable requires additional validation. Anchor-based methods for connecting changes in psychophysiological variables to changes in clinical state are presented in Copay, et al. [61]. Sample size estimates should be specific to reliability studies and are larger than often supposed. Zou [62] has derived sample size formulas for estimating intraclass correlation coefficients with precision and assurance. If the measure is being used longitudinally in a clinical trial, the test-retest interval of the validating study should be equal to the duration of the trial.

When possible, a test-retest reliability study of a new variable should incorporate the simultaneous measurement of another variable which is already known to be reliable under the conditions of the test in the population being tested, for example, simultaneous measurement of heart rate variability when evaluating a measure calculated from event related potentials. If the new variable is found to be reliable, all is well. If the new variable is found to have low reliability, two possibilities exist: (1.) the new variable is unreliable in these conditions with this population or (2.) there were errors in the design and implementation of the reliability study. These two possibilities can be distinguished by observing the results obtained from the variable which is known to be reliable.

While establishing acceptable reliability in a healthy control population is the first step in the evaluation of a psychophysiological measure, it should be noted that reliability is population-specific. This is particularly true in neuropsychiatry. High variability is a characteristic of an injured or diseased central nervous system [63], [64]. For example, a longitudinal change in a measure that would be remarkable in a healthy twenty year old might not be remarkable in a clinically stable neuropsychiatric patient. Reliability should therefore be determined for the population of interest.

To summarize test-retest reliability requirements:

1. Test-retest reliability should be quantified with the intraclass correlation coefficient using an adequate sample size.
2. The version of the ICC used should be specified.
3. The report of the ICC should include confidence intervals and a specification of the procedure used to calculate the confidence interval.
4. The population used to determine the ICC should be appropriate for the clinical question being addressed.

5. Consideration should be given to including the simultaneous measurement of a variable of known reliability in order to evaluate the validity of the test-retest study.
6. The report should include determination of the standard error, the minimum detectable difference, and their confidence intervals.
7. If the measure is being used for pre- and post-trial evaluation in a clinical trial, the test-retest interval should be equal to the duration of the trial.
8. Consideration should be given to incorporating a determination of the minimum clinically important difference into the study.

7. Variation of psychophysiological measures with age, sex, and ethnicity

Observation

The task of establishing the reliability of psychophysiological measures is seen to be more demanding than might be supposed when it is noted that psychophysiological variables can show a dependence on age, sex, and ethnicity.

Measures of heart rate variability can show a dependence on age and gender (representative examples include: [65], [66], [67], [68]) as can event related potentials [69], [70], [71]. As will be described presently, gender differences in placebo-induced alterations of ERPs have been reported [72].

The literature describing ethnic and culture-specific variations in psychophysiological variables is smaller. Fukusaki, et al. [73] found that gender-dependent HRV measures in Western populations were not observed in Japanese populations, While Choi, et al. [74] found that age-dependent decreases in HRV measures commonly observed in Western populations were also observed in Korean populations, they did note that "The cause of the difference in HRV depending on the gender between Westerners and Asians should be included in future studies." Mu, et al. [75] found cultural differences in event-related potentials in Chinese and US populations in a social norm violation paradigm. Specifically, "the N400 at the frontal and temporal regions, however, was only observed among Chinese but not US participants, illustrating culture-specific neural substrates of the detection of norm violations."

Response to Observation 7.

Operationally, these results indicate that when used longitudinally in a clinical study, the reliability of psychophysiological measures should be determined for the age, gender, and ethnicity of the study population. The procedures outlined in response to Observation 6 are applicable. If heterogeneous participant populations are used in a study, multiple determinations of the minimal detectable difference will be required.

8. Adaptation not Repair: Psychophysiological measures do not invariably normalize during recovery

Observation

Consider the following scenario. A patient is diagnosed with a specific psychiatric disorder. Consistent with the prior literature, a psychophysiological measure obtained from this patient is found to be abnormal. Treatment is initiated, and according to standardized patient reported outcomes, clinical recovery is observed. In an ideal world the psychophysiological measure would also normalize and, even better, track progress during the course of treatment. This is not an ideal world. As Steven Weinberg observed, "the universe was not designed to make physicists happy." Evidently this is also true for clinical psychophysiologicalists. Psychophysiological measures do not invariably normalize during recovery. This limits their utility in the longitudinal assessment of treatment. This is a potentially significant setback for clinical psychophysiology because the hope of utility in longitudinal assessment was deemed to be important when it was recognized that the nonspecificity of psychophysiological measures often precluded their use in diagnosis.

An inconsistent pattern is observed when the literature describing changes in psychophysiological measures in response to treatment is examined. Kemp, et al. [76] found that HRV measures are lower in depressed patients, and the magnitude of the decrease as

compared to healthy controls was correlated with the clinically perceived severity of the depression. This decrease was most apparent with nonlinear measures of HRV. Kemp, et al. report that tricyclic medication decreased HRV measures, and that the SSRIs mirtazapine and nefazodone had no significant effect on HRV even though patients responded positively clinically to treatment. Similarly, Brunoni, et al [77] found that depressed patients responded to sertraline or direct current electrical current therapy for depression, but lower HRV measures did not normalize. Bozkurt, et al. [78] found no relationship between treatment response and change in HRV measures. The review by Alvares, et al. [79] found reduced HRV in the patient groups considered (mood disorders, anxiety, psychosis, dependent disorders), and found that HRV did not normalize in response to successful medication. Psychotropic medication further reduced HRV specifically associated with tricyclic antidepressants and clozapine. In contrast, Udupa, et al. [80] found that the form of treatment was an important consideration. In their patient population, HRV measures increased in response to rTMS, decreased in response to tricyclic antidepressants and were effectively unchanged in response to SSRIs. The measures used to quantify heart rate variability may also be important. Nahshoni, et al [81] treated elderly patients presenting major depressive disorder with ECT. The pointwise dimension of HRV increased in responders and showed a tendency toward a correlation with symptom improvement. In this study, spectral measures of HRV, however, did not show a significant difference after ECT. This observation emphasizes the possible importance of developing comprehensive analysis protocols incorporating a large number of psychophysiological measures. In the case of heart rate variability, the Kubios analysis suite is a significant contribution [82].

While there is a substantial literature describing alterations of event related potentials in clinical populations, for example schizophrenia [83] and depression [84], the literature describing treatment dependent change or the absence of change in ERPs is much smaller. As in the case of measures of heart rate variability, the observations of longitudinal change in EEGs and event related potentials during the course of treatment present a complex pattern of positive results (the measure normalizes in response to successful treatment) and negative results (the measure does not normalize). In a study of depression and of schizophrenia, Buchheim, et al. [85] in a clinical study of depressed patients found that psychodynamic psychotherapy results in a normalization of the late positive potential, LPP, and gamma abnormalities that had previously been identified by Siegle, et al. [86]. Decreased amplitude of the P3 observed in medication-free depressed patients normalized after four weeks of antidepressant treatment [87]. Similarly, P3 amplitude increased in response to electroconvulsive therapy [88].

A meta-analysis by Umbricht and Kriljes [89] concluded that defects in the mismatch negativity (MMN) ERP are “a robust feature in chronic schizophrenia,” and it had been hypothesized that glutathione dysregulation and subsequent N-methyl-D-aspartate hypofunction may be an element in the pathophysiology of schizophrenia. Responding to this hypothesis, Lavoie, et al. [90] administered N-methyl-cysteine, a glutathione precursor, to schizophrenia patients and observed that treatment significantly improved MMN generation, without measurable effects on the P300. MMN improvement was “observed in the absence of robust changes in assessments of clinical severity, though the later was observed in a larger and more prolonged clinical study.” Similarly, Zhou, et al [91] found that treatment of schizophrenic patients with aripiprazole improved amplitude of the MMN and reduced Positive and Negative Syndrome Scale scores.

The results with ERPs just cited contrast with results obtained in the treatment of anxiety. The error related negativity (ERN) is an ERP component observed at frontal and central electrodes after the participant makes an incorrect response (reviewed in Gehring, et al. [92]). Anxiety disorders are associated with an increased amplitude of the ERN [93], [94]. Successful pharmacological treatment of anxiety does not, however, normalize the ERN [95], [96], [97], [98], [99]. Valt, et al. [100] investigated a related internalizing disorder, panic disorder, and consistent with prior literature found that compared to controls untreated participants had a greater ERN and also a greater vertex positive potential. In a subsequent

treatment study, Valt, et al. [101] found that as before the ERN did not normalize in response to psychological treatment, but treatment related normalization of the vertex positive potential was observed.

Response to Observation 8.

In the specific context of anxiety disorders, Hajcak, et al. [102] raise the following interesting possibility: “The evidence suggests that typical treatments for anxiety do not normalize an increased ERN. One possibility is that the ERN is related to the risk for anxiety but not the expression of an anxious phenotype. In this case, treatment-related effects on the ERN would not be expected unless treatments alter underlying risk processes that are reflected in the ERN.” Considered more generally, in those presentations where psychophysiological measures are indicative of risk and not the presence of disease, the utility of these measures as longitudinal markers of treatment progress will be limited.

In the case of traumatic brain injury, observations of recovery of function in the absence of normalization of psychophysiological variables are frequently observed. The work of Kurt Goldstein [103] is of particular interest. Goldstein treated patients who had sustained significant brain injuries during World War I and concluded that in many cases when restoration of function is observed the injured brain is not repaired, it adapts. Goldstein focused on traumatic brain injury. It seems possible that the principle of adaptation not repair is more generally applicable in neuropsychiatry. Restoration or partial restoration of function in the absence of normalization of clinical markers may be common.

In those instances where either of the two possibilities just considered, (1.) psychophysiological measures as indicators of risk or (2.) adaptation in the absence of repair, are correct a simple response to Observation 8 that psychophysiological measures do not invariably normalize in response to a clinically successful treatment will not be available.

An alternative treatment strategy utilizing psychophysiological measures of considerable clinical value may be available. In typical clinical practice treatments are directed to the resolution of symptoms. It is possible to consider treatment protocols explicitly directed to the normalization of aberrant psychophysiological measures. This has been addressed by Hajcak, et al. [102]. In their summary, the focus is on error related negativity, but the concept is more generally applicable: “Traditional CBT and SSRIs do not seem to impact the ERN, and the ERN does not appear to robustly predict treatment response to traditional CBT or SSRIs. Yet the ERN is a robust predictor of subsequent changes in symptoms and psychopathology, and a range of strategies appear to modulate the ERN, at least in the short term. To us, these data collectively suggest the need to test and develop novel interventions that are more focused on altering the ERN. To our knowledge, no intervention has been designed to directly target the ERN, and pharmacological studies have not routinely considered ERPs as potential targets. Moreover, brain stimulation and neurofeedback might provide additional and more direct methods of altering ERPs.”

The treat-the-biomarker strategy is the organizing principle of many neurofeedback protocols [104] such as theta/beta ratio training in ADHD [105] and frontal alpha asymmetry in depression [106], [107]. See, however, Papo [108] for a review of unresolved issues in neurofeedback. Similarly, HRV biofeedback seeks change in a physiological variable, not a change in a behavior. The role of psychophysiology in clinical practice may increase substantially as more disorder-specific psychophysiological signatures are identified and incorporated into neurofeedback protocols.

9. Psychophysiological biomarkers can change and, in some cases, normalize in response to placebo interventions

Observation

Placebos can alter measures of heart rate variability, eye tracking behavior and event related potentials. Vaschillo, et al. [109] measured the impact of alcohol and placebo on measures of heart rate variability obtained during an emotional cue challenge. Participants were presented with pictures (negative, positive, neutral) in a 5 seconds on/5 seconds off protocol (0.1 Hz). Three commonly used measures of heart rate variability were recorded

(SDNN, pNN50 and high frequency heart rate variability). Additionally, they reported a novel measure, the 0.1-Hz Index, which is the maximum amplitude of the RR interval spectrum in the 0.075 to 0.108 Hz range. The .01-Hz Index was diminished by both alcohol and placebo, and there was no statistically significant difference between the alcohol and placebo group as determined by this measure. Vaschillo, et al. concluded that this suggests a dependence of this measure of heart rate variability on the participant's cognitive expectancy. Darragh, et al. [110] studied placebo modification of heart rate variability during the recovery from cognitive stress. An acute reduction in heart rate variability was induced by an arithmetic test conducted in the presence of a research assistant. The experimental group was administered a placebo nasal spray and was told that it contained serotonin and that it was expected that this would accelerate recovery from stress. Two metrics of heart rate variability were employed, high frequency spectral power and the root mean square of successive differences. An increase in vagally-mediated heart rate variability was observed in the placebo treated group contra the untreated control group.

Daniali and Flaten [111] conducted a systematic review of the effects of placebo analgesia and nocebo hyperalgesia on cardiac activity. Specifically, they reviewed papers that reported blood pressure, heart rate and heart rate variability. They identified six papers that reported effects on heart rate variability and provide the following summary: "The results indicate that the placebo analgesic effect is associated with a decrease in LF-HRV (low frequency heart rate variability), that the nocebo hyperalgesic effect is associated with an increase in LF-HRV, and that HRV is a predictor for placebo effects. However, there is no reliable effect of placebo on the LF/HF (low frequency/high frequency) ratio and HF-HRV."

Schienze, et al. [112] found that placebos can affect eye tracking behavior. In this experiment participants viewed neutral and disgust-inducing images with and without a "disgust placebo," an inert pill presented with the verbal suggestion of disgust relief. In an experiment in which participants look at side-by-side images of neutral-disgust and neutral-neutral images, it was found that the placebo resulted in a marked decrease in reported distress and an increased number of fixations on disgusting images. In similarly designed fMRI experiments Schienze, et al. [113], [114] found that the disgust placebo reduced the activation of the insula and the visual cortex and reduced experienced disgust. In an eye tracking experiment investigating placebo effects on phobias, specifically spider phobia [115], participants were shown spider pictures paired with neutral pictures with and without a placebo labeled as propranolol. Fixation count and dwell time increased in the placebo condition, and there was a slight decrease in reported symptom severity.

Placebo interventions can also alter event related potentials. Placebo effects on ERPs have been observed in studies of analgesia, anxiolysis, emotional processing and cognitive enhancement. Building on a substantial prior literature investigating the effects of placebo analgesia on ERPs [116], [117] Aslaksen, et al. [72] investigated the effects of a placebo intervention in a pain study where ERPs were evoked by heat pulses. The amplitude of the N2/P2 complex was reduced by the placebo. Interestingly, a reduction in pain report and the P2 amplitude was observed at a group level in the male participant group but not in the female participant group.

In a study reported by Meyer, et al. [118] inactive treatment was accompanied by verbal suggestion that the placebo intervention would have an anxiolytic effect of experimentally induced phobic fear or sustained anxiety. A placebo-dependent sustained increase of frontal midline EEG theta power and an increase in frontoposterior theta coupling consistent with activation of cognitive control mechanisms was observed. Downregulation of unspecific cue reactivity was observed in fear ratings, skin conductance response P300 amplitude (280-400 ms) and the late positive potential (400-700 ms).

A complex mix of results have been obtained in ERP studies of processing of emotionally valenced images with and without a placebo intervention. Übel, et al. [119] studied the processing of emotionally valenced images by children. Children viewed disgusting fear-eliciting images and neutral images with and without placebo (syrup presented with

the verbal suggestion that it would ease disgust symptoms). In this experiment the placebo increased the late positive potential (defined here as 400-1000 ms) in response to disgusting and fear-eliciting photographs. Übel et al. propose the following interpretation: "These findings suggest that the placebo had the function of a safety signal which helped the children to direct their automatic attention to the aversive stimuli and to overcome visual avoidance." In Schienle, et al [120] an unpleasant context, in this study a bitter after-taste elicited by wormwood tea administered prior to ERP recording, can reduce the late positive potential elicited by affectively valenced pictures. In this study there were three experimental groups: water instead of tea, tea, tea and a placebo treatment, light therapy on the tongue to "reduce sensitivity of taste buds." Two classes of pictures were shown: neutral and disgusting. For both classes of pictures, the early late positive potential was smaller in the tea/no placebo case than in the case water case with the tea/placebo amplitudes being intermediate to them. The authors state, "This is the first EEG study to demonstrate effects of a context-targeting placebo," the context being the wormwood tea pretreatment.

The possibility of using transcranial direct current stimulation, tCDS, to enhance cognitive performance has received excited public attention. van Elk, et al. [121] conducted an ERP study where the placebo was an inert sham tCDS. Participants were advised that their performance would be enhanced in tCDS trials. Participants reported improved subjective performance during placebo enhancement, but objective performance was unchanged. During induction phase, placebo-induced expectation of enhancement increased frontal theta power "potentially reflecting a process of increased cognitive central allocation." The placebo manipulation did not, however, change the ERN associated with incorrect responses.

One of the most intriguing studies of placebo impact on ERPs is the Guevarra, et al. [122] study utilizing a "non-deceptive" placebo. Non-deceptive placebos are not original to this study [123], [124], additionally, Guevarra, et al. cite studies describing the beneficial effects of non-deceptive placebos for several disorders), but insofar as we know, the Guevarra, et al. study is the first study showing that non-deceptive placebos can change ERPs. They give the following description of the non-deceptive placebo protocol used in their study: "Participants in the non-deceptive placebo group read about placebo effects and were then asked to inhale a nasal spray consisting of saline solution. They were told that the nasal spray was a placebo that contained no active ingredients, but would help reduce their negative emotional reactions to viewing distressing images if they believed it would. Participants in the control group read about the neural processes underlying the experience of pain and were also asked to inhale the same saline solution spray; however, they were told that the purpose of the nasal spray was to improve the clarity of the physiological readings we were recording in the study. The articles were matched for narrative structure, emotional content and length."

The Guevarra, et al. experiment was a negative picture viewing task. In the first experiment, participants viewed a picture (neutral or negative image) and were asked "Rate how does this picture makes you feel?" on a one to nine scale. The non-deceptive placebo reduced self-report measures of emotional distress. In the second experiment, the participants viewed neutral and negative images as before but were not asked to rate them. ERPs were recorded in the second experiment. The ERP analysis of their study focused on the late positive potential (LPP). Two distinct components have been identified in the LPP, and they have different cognitive associations. The early component with a latency of 400-1000 ms corresponds to attention allocation [125]. The late component, with a latency of 1000-1600 ms is associated with conscious appraisal and emotional processing [125], [126]. The late positive potential is down-regulated by cognitive emotion-regulation strategies [127]. As reported above, some studies suggest that deceptive placebos amplify attention to negative stimuli and other studies suggest the opposite where attention is quantified by the amplitude of the early component of the LPP. Guevarra, et al. found no reliable non-deceptive placebo effect on the early LPP. In contrast, however, they did observe that the non-deceptive placebo reduced activity during the later component of the LPP which

quantifies “meaning-making stages of emotional reactivity.”

Response to Observation 9.

The studies showing an effect of placebos on psychophysiological measures summarized here are laboratory investigations conducted on a time scale of hours. The degree to which they are applicable to clinical trials conducted over a period of weeks or months is unclear. It is commonly suggested that placebo effects are short lived while clinically induced physiological change is longer lasting. If this is true, and this view is subject to challenge, it might follow that psychophysiological measures will be useful in long term follow-up. Investigations of the long term impact of placebos on psychophysiological measures are warranted. In any case, we must be alert to the possibility that, as with PROs, physiological measures will be an imperfect measure of treatment response. The results summarized by Daniali and Flaten [111] indicate that more than one measure of heart rate variability should be examined. While they report placebo and nocebo changes in LF-HRV, there was no reliable effect on the LF/HF ratio. Physiological measures should be used in combination with PROs. Instances where the effect size calculated with PROs is small while the effect size calculated with a measure of HRV is large might be taken to indicate a physiological treatment response that was not captured in a PRO report.

The magnitude of a placebo response can be estimated using a placebo control group. As an additional check it is possible to ask if active treatment responders and placebo responders are different at intake. If pre-treatment measures that characterize active treatment responders are indistinguishable from measures that characterize placebo responders and if the frequency of active versus placebo response is nearly equal, then perhaps the active treatment response is a placebo response.

Determination of the magnitude of either a placebo response or a treatment response can be confounded by spontaneous recovery. It is commonly recognized that spontaneous recovery can be a significant complication in depression treatment trials. Indeed Kraepelin (cited by Posternak et al. [128]) concluded that untreated depressive episodes would last six to eight months. The frequency of spontaneous recovery can be estimated from a waitlist control group. Posternak and Miller [129] collected and analyzed results obtained in antidepressant trials that included a waitlist control group. They provided the following summary: “Our analysis of 19 studies involving 221 depressed subject randomized to a waiting list for 2 – 20 week found a mean decrease in symptomatology of 10-15%. A sub-analysis of 11 studies that obtained depression rating scores between weeks 4 and 8 – the time frame used in most antidepressant trials – yielded similar results. We therefore would postulate that subjects enrolled in short-term antidepressant trials probably improve on their own by this amount.” In a subsequent study Posternak, et al. [128] found that depressed patients who went without somatic therapy throughout the course of a depressive episode presented a median episode duration of 13 weeks.

The possibility of both a placebo intervention altering psychophysiological measures and the possibility of spontaneous recovery argue for the incorporation of both waitlist and placebo arms in studies investigating the longitudinal response of psychophysiological measures to treatment.

10. The mathematical procedures of statistical learning are not robust to misapplication and to data artifacts

Observation

Two related final concerns merit attention, the impact of data artifacts and the misapplication of analysis algorithms. Increasingly sophisticated forms of analysis are increasingly sensitive to artifacts in data. Muthukumaraswamy [130] noted that “As analytical techniques in MEG/EEG analysis ... become increasingly complex, it is important that artifact-free data are being fed into these algorithms.” Experience suggest that this is particularly true of methods that quantify nonlinear structure in experimental data. It has been shown, for example, that filtered noise can mimic low dimensional chaotic attractors [131]. These observations led to a re-examination of evidence of low-dimensional structure in

the EEG. [132]. For the specific case of gamma band EEG activity, Hipp and Siegel [133] and Muthukumaraswamy [130] provide practical guidelines for investigations of gamma band activity. Luck [134] offers a comprehensive guide to artifact detection in ERP studies. He also offers cautions concerning artifact correction. We concur and in particular note that artifact correction procedures can distort results obtained with sensitive measures of information movement.

In addition to concerns about data quality, it should be understood that analytical methods are not robust to misapplication. Three failure patterns are particularly prominent: (1.) p-hacking, (2.) errors in the application of signal processing algorithms, and (3.) errors in the application of statistical learning algorithms. Errors in statistical analysis, volitional or inadvertent, collectively known as p-hacking (which includes multiple comparisons in hypothesis testing, over-hacking (the practice of continuing to manipulate the data to get a p value lower than .05), selection bias in the choice of analysis pathways and selective debugging of analysis software) are well documented [135], [136] but nonetheless occur in the peer reviewed literature.

Analysis errors in signal analysis includes errors in the procedures used to construct random phase surrogates that can result in the false-positive indication of deterministic structure in random data [137]. Similarly, inappropriate procedures for calculating the complexity of a time series can also give false positive indications of deterministic structure [138]. As noted in the introduction to this contribution, the three primary objectives of clinical psychophysiology are diagnosis, the longitudinal assessment of treatment response and the identification of individuals at risk of disease onset. All of these objectives are classification problems. Cross-validation is an essential step in confirming classification results. It is, however, easily misused. An incorrectly constructed cross-validation calculation can seemingly validate a classification procedure constructed with random numbers. Indeed, the error is so commonly encountered that the standard textbook on the subject of statistical learning [46] includes a section with the title "The Wrong and Right Way to Do Cross-Validation." With the increasing availability of powerful freely accessible statistical learning software tools, errors of this kind will almost certainly become more common.

Response to Observation 10.

In 2019 the Society for Psychophysiological Research conducted a one day workshop on open science to addresses some of the issues raised here. Their results were published in Garrett-Rufin, et al. [139]. The panel's recommendations included data sharing with data format harmonization, analysis pipeline sharing, pre-registration of an analysis plan, multisite studies, and encouraging replication studies. The resources of the Open Science Framework (Center for Open Science, [140]) can facilitate this effort. Saunders and Inzlicht [141] have outlined procedures that can increase the transparency of meta-analyses.

The Society for Psychophysiology call for standardization was not the first to stress the need for standardization in order to increase the clinical utility of psychophysiological biomarkers [142], [143], [33], [144], [145], [146]. The most comprehensive effort to date has been the development of ERP CORE by Luck and his colleagues [147]. They have made freely available "a set of optimized paradigms, experiment control scripts, data processing pipeline and sample data (N=40 neurotypical young adults) for seven widely used ERP components." These authors and others have noted that rather than use a generic two stimulus P3 oddball protocol, the ERP paradigm used should be consistent with the clinical presentation under consideration. Representative examples could include reduced contingent negative variation in ADHD [148], reduced mismatch negativity in schizophrenia [89], absence of P50 suppression in schizophrenia [149], decreased P3b in depression [150], enhanced error related negativity in anxiety disorders [98], pattern separation/pattern completion in dementia [151] and increased error related negativity in obsessive-compulsive disorder [95]. Further, increased clinical utility of ERPs may be obtained with the expansion of ERP analysis beyond an examination of amplitude and latency of averaged ERP waveforms with the incorporation of measures of synchronization

[152], event related oscillations [153], microstates [154], information dynamics [155] and network analysis [156].

As is generally recognized, statistical tests should be incorporated routinely in psychophysiological studies, though as noted above important tests like surrogate data calculations and cross-validation can be misused. Running all analysis software against publicly available standard data sets provides an important mechanism for validating the software used in a specific study. Bootstrapped confidence intervals [157], [158] for reported psychophysiological measures can be an important safeguard. Wang, et al. [40] provide an illustrative example. In a search for electrophysiological prodromes of delayed onset PTSD, initial calculations indicated sensitivity and specificity of approximately 0.8. The corresponding confidence intervals, however, were found to be [0,1], thus indicating that the initial encouraging result was the fortuitous artifact of a small sample size. The previously cited analysis of Button, et al. [44] concerning the dangers of underpowered studies is relevant.

In the case of the classification calculations which are essential to implementing the three previously stated objectives of clinical psychophysiology, Fernández-Delgado, et al. [159] offer guidance on the choice of a classification algorithm. The OpenML website [160] provides access to experimental data including data from the University of California, Irvine machine learning database [161] that can be used in pipeline validation studies. For confounds specific to EEG classification see Li, et al. [162] and Ahmed, et al. [163].

Discussion

Within the psychophysiological community, and indeed in this paper, the term biomarker is used casually. The US Food and Drug Administration has issued technical definitions of biomarkers [164] and has issued draft criteria for biomarker qualification [165], [166]. That guidance states: "Qualification of a biomarker is a determination that within the stated context of use, the biomarker can be relied on to have a specific interpretation and application in drug development and regulatory review." Many proposed biomarkers fail to meet the standards set by these more rigorous definitions. The report by Prata, et al [167] specifically focusing on schizophrenia is instructive. They provide the following summary: "We categorized all PubMed-indexed articles investigating psychosis-related biomarkers to date (over 3200). Fewer than 200 studies investigated biomarkers longitudinally for prediction of illness course and treatment response. These biomarkers were then evaluated in terms of their statistical reliability and clinical effect size. Only one passed our a priori threshold for clinical applicability." The single biomarker that met the Prata et al. criteria is a single nucleotide polymorphism predicting risk for clozapine-induced agranulocytosis.

Subsequent research suggests, however, that a more optimistic assessment may be appropriate. Of many promising studies we make note of two based on serum biomarkers. Chan, et al. [168] presented results obtained from a serum biomarker constructed to discriminate between depressed patients and healthy controls. A 10-fold cross validation and application of a least absolute shrinkage and selection operator (LASSO) resulted in an optimal panel of 33 immune-neuroendocrine biomarkers and gender. Moderate to good discrimination was observed (AUC Area Under the Curve), $0.62 < \text{AUC} < 0.86$, these are 95

Chan, et al. [169] compared 127 first onset drug naïve schizophrenia patients with 204 controls. Using LASSO, they identified 26 biomarkers that best discriminated between patients and controls. This was validated against two independent cohorts. The schizophrenia detection study gave $0.95 < \text{AUC} < 1.00$ (95% confidence intervals for the AUC). The predictive performance was tested in a pre-onset population where $0.86 < \text{AUC} < 0.95$. In a prodromal population the biomarker panel gave $0.71 < \text{AUC} < 0.93$ which was improved to $0.82 < \text{AUC} < 0.98$ by incorporating the positive symptom subscale of the Comprehensive Assessment of At-Risk Mental State.

It should be noted, however, that these encouraging results [168], [169] were obtained from carefully constructed clinical studies comparing either depressed patients versus controls or schizophrenics versus controls. The performance in general psychiatric practice

which would include a mix of these patients populations along with generalized anxiety disorder, PTSD, bipolar disorder, and others is unclear.

The role of psychophysiology advancing our understanding of central nervous system physiology is clear. We have, however, been addressing a different question: what are the prospect of the near-term utility of psychophysiological measures in clinical practice? Overall, our assessment for near term utility is guarded optimism. The effort to introduce quantitative measures into psychiatric practice will be a journey well worth traveling but may be a longer journey than might be supposed.

Author Contributions: PER: Conceptualization, literature review, first draft, CC: Revision of the final draft, DD: Review of statistical commentary, DK: Conceptualization and revision of the final draft. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding

Institutional Review Board Statement: Not applicable: This contribution is a review of publicly accessible previously published literature

Informed Consent Statement: Not applicable

Data Availability Statement: Not applicable

Acknowledgments: PER would like to acknowledge Dr. Adele Gilpin and the students, faculty and staff of the Center for the Neurobiology of Learning and Memory at the University of California, Irvine and its Director Professor Michael Yassa for discussion critical to the development of this contribution.

Conflicts of Interest: The authors declare no conflict of interest.

Disclaimer: The opinions and assertions contained herein are those of the authors and do not necessarily reflect the official policy or position of the Uniformed Services University, the Department of Defense, or the Henry M Jackson Foundation for the Advancement of Military Medicine.

References

1. Takeuchi, H., Fervaha, G. and Remington, G. Reliability of patient-reported outcome measures in schizophrenia: results from back-to-back self-ratings. *Psychiatry Research* **2016**, *244*, 415-419.
2. Greenhalgh, J., Gooding, K., Gibbons, E., Dalkin, S., et al. How do patient-reported outcome measures (PROMs) support clinician-patient communication and patient care? A realist synthesis. *Journal of Patient Reported Outcomes* **2018**, *2*: 42.
3. Rush, A.J., First, M.B. and Blacker, D. *Handbook of Psychiatric Measures. Second Edition*. Washington, DC: American Psychiatric Publishers, Washington, DC, USA; 2008.
4. Streiner, D.L. and Norman, G.R. *Health Measurement Scales: A Practical Guide to Their Development and Use. Fourth Edition*, Oxford University Press., Oxford, UK; 2008.
5. Food and Drug Administration (2009). Guidance for Industry. Patient-reported outcome measures: Use in medical product development to support labeling claims. Silver Spring, MD, USA: Food and Drug Administration.
6. Mokkink, L.B., Terwee, C.B., Patrick, D.L., Alonso, J., et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient reported outcomes. *Journal of Clinical Epidemiology* **2010**, *63*(7), 737-745.
7. Mokkink, L.B., Terwee, C.B., Patrick, D.L., Alonso, J., et al. The COSMIN checklist for assessing methodological studies on measurement properties of health status measurement instruments: an international Delphi study. *Quality of Life Research*. **2010**, *19*(4), 539-549.
8. Mokkink, L.B., Terwee, C.B., Knol, D.L., Stratford, P.W., et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. *BMC Medical Research Methodology*. **2010**, *10*: 22.
9. Mokkink, L.B., Terwee, C.B., Patrick, D.L., Alonso, J., et al. *COSMIN Checklist Manual*. VU University Medical Center, Amsterdam, The Netherlands; 2012
10. Terwee, C.B., Mokkink, L.B., Knol, D.L., Ostelo, R.W.J.G., et al. Rating the methodological quality in systemic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Quality of Life Research* **2012**, *21*(4), 651-657.

11. McGinn, C. Can we solve the mind-body problem? *Mind (New Series)*. **1989**, 98(391), 349-366.
12. Dennett, D.C. (1991). The brain and its boundaries. Review of McGinn: *The Problem of Consciousness*. The Times Literary Supplement. 10 May 1991.
13. Harrington, A. *Mind Fixers: Psychiatry's Troubled Search for the Biology of Mental Illness*. W.W. Norton and Company, New York, NY, USA; 2019.
14. Rapp, P.E., Darmon, D., Cellucci, C.J. and Keyser, D.O. The physiological basis of consciousness: a clinical ambition and the insufficiency of current philosophical proposals. *Journal of Consciousness Studies* **2018**, 25(1-2), 191-205
15. Koch, C., Massimini, M., Boly, M. and Tononi, G. Neural correlates of consciousness: progress and problems. *Nature Reviews Neuroscience* **2016**, 17, 307-321.
16. Tononi, G., Boly, M., Massimini, M. and Koch, C. Integrated Information Theory: from consciousness to its physical substrate. *Nature Reviews Neuroscience* **2016**, 17(7), 450-461.
17. Boly, M., Massimini, M., Tsuchiya, N., Postle, B.R., et al. Are the neural correlates of consciousness in the front or in the back of the cerebral cortex? Clinical and Neuroengineering evidence. *Journal of Neuroscience* **2017**, 37(40), 9603-9613.
18. Weinberger, J. and Stoycheva, V. *The Unconscious: Theory, Research and Clinical Implications*. Guilford Press, New York, NY, USA; 2020.
19. Berridge, K.C. and Winkielman, P. What is an unconscious emotion? The case for unconscious 'liking.' *Cognition and Emotion* **2003**, 17, 181-211.
20. Brosschot, J.F. Markers of chronic stress: prolonged physiological activation and (un)conscious perseverative cognition. *Neuroscience and Behavior Reviews* **2010**, 35, 46-50.
21. Wiers, R.W., Teachman, B.A. and De Houwer, J. Implicit cognitive processes in psychopathology: an introduction. *Journal of Behavior Therapy Experimental Psychiat.* **2007**, 38(2), 95-104.
22. Sperdin, H.F., Spierer, L., Becker, R., Michel, C.M., et al. Submillisecond unmasked subliminal visual stimuli evoke electrical brain responses. *Human Brain Mapping* **2015**, 36, 1470-1483.
23. Elgendi, M., Kumar, P., Barbic, S., Howard, N., et al. Subliminal priming – state of the art and future perspectives. *Behavioral Science* **2018**, 8: 54.
24. Herzog, M.H., Drissi-Daoudi, L. and Doerig, A. All in good time: long-lasting postdictive effects reveal discrete perception. *Trends in Cognitive Science* **2020**, 24(10), 826-837.
25. Jiang, J., Bailey, K. and Xiao, X. Midfrontal theta and posterior parietal alpha band oscillations support conflict resolution in a masked affective priming task. *Frontiers in Human Neuroscience* **2018**, 12: 175.
26. Siegel, P., Cohen, B. and Warren, R. Nothing to fear but fear itself: a mechanistic test of unconscious exposure. *Biological Psychiatry* **2022**, 91, 294-302.
27. Shackman, A.J. and Fox, A.S. Getting serious about variation: lessons for clinical neuroscience. (A commentary on the myth of optimality in clinical neuroscience). *Trends in Cognitive Science* **2018**, 22(5), 368-369.
28. Holmes, A.J. and Patrick, L.M. The myth of optimality in clinical neuroscience. *Trends in Cognitive Science* **2018**, 22(3), 241-257.
29. Rapp, P.E., Cellucci, C.J., Keyser, D.O., Gilpin, A.M.K. et al. Statistical issues in TBI clinical studies. *Frontiers in Neurology* **2013**, 4, 177.
30. Mahalanobis, P.C. On the generalized distance in statistics. *Proceedings of the National Institute of Science India*. **1936**, 2, 49-55.
31. Lachenbruch, P.A. *Discriminant Analysis*. Hafner Press, New York, NY, USA; 1975.
32. Rapp, P.E., Keyser, D.O., Albano, A.M., Hernandez, R., et al. Traumatic brain injury detection using electrophysiological methods. *Frontiers in Human Neuroscience* **2015**, 9: 11.
33. Kapur, S., Phillips, A.G. and Insel, T.R. Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it? *Molecular Psychiatry* **2012** 17(12), 1174-1179.
34. Newson, J.J., Pastukh, V. and Thiagarajan, T.C. Poor separation of symptom profiles by DSM-5 disorder criteria. *Frontiers in Psychiatry* **2021**, 12: 775762.
35. Freedman, R., Lewis, D.A., Michels, R., Pine, D.S., et al. (2013). Initial field trials of DSM-5: new blooms and old thorns. *American Journal of Psychiatry* **2013**, 170(1), 1-5.
36. Clarke, D.E., Narrow, W.E., Regier, D.A., Kuramoto, S.J., et al. DSM-5 field trials in the United States and Canada, Part I: Study design, sampling strategy, implementation, and analytic approaches. *American Journal of Psychiatry* 170(1), **2013**, 43-58.
37. Regier, D.A., Narrow, W.E., Clarke, D.E., Kraemer, H.C. et al. DSM-5 field trials in the United States and Canada. Part II: test-retest reliability of selected categorical diagnoses. *American Journal of Psychiatry* **2013**, 170(1), 59-70.
38. Narrow, W.E., Clarke, D.E., Kuramoto, S.J., Kraemer, H.C., et al. DSM-5 trials in the United States and Canada. Part III: Development and reliability of a cross-cutting symptom assessment for DSM-5 *American Journal of Psychiatry* **2013**, 170(1), 71-82.
39. Insel, T.R. The arrival of preemptive psychiatry. *Early Intervention Psychiatry* **2007**, 1(1), 5-6.
40. Wang, C., Costanzo, M.E., Rapp, P.E., Darmon, D., et al. Identifying electrophysiological prodromes of post-traumatic stress disorder: results from a pilot study. *Frontiers in Psychiatry* **2017**, 8: 71.
41. Hoeffding, W. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association* **1963**, 58(301), 13-30.

42. Clopper, C.J. and Pearson, E.S. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika* **1934**, 26(4), 404-413.
43. Thulin, M. The cost of using exact confidence intervals for a binomial proportion. *Electronic Journal of Statistics* **2014**, 8, 817-840.
44. Button, K.S., Ioannidis, J.P.A., Mokrysz, C., Nosek, B.A., et al. Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience* **2013**, 14(5), 365-376.
45. Watanabe, T.A.A., Cellucci, C.J., Kohegyi, E., Bashore, T.R., et al. The algorithmic complexity of multichannel EEGs is sensitive to changes in behavior. *Psychophysiology* **2003**, 40, 77-97.
46. Hastie, T., Tibshirani, R. and Friedman, J. *Elements of Statistical Learning. Second Edition*. Springer, New York, NY, USA; 2009.
47. Ambrose, C. and McLachlan, G. Selection bias in gene extraction on the basis of microarray gene-expression data. *Proceedings of the National Academy of Science USA*. **2002**, 99, 6562-6566.
48. Bashore, T.R. and Osman, A. On the temporal relation between perceptual analysis and response selection: A psychophysiological investigation of stimulus congruency and S-R compatibility effects on human information processing. In *Proceedings Fourth International Conference on Cognitive Neurosciences*. Dourdan-Paris, France (1987)
49. Cole, W.R., Arrioux, J.P., Schwab, K., Ivins, B.J., et al. Test-retest reliability of four computerized neurocognitive assessment tools in an active duty military population. *Archives of Clinical Neuropsychology* **2013**, 28, 732-742.
50. Shrout, P.E. and Fleiss, J.L. Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin* **1979**, 86(2), 420-428.
51. McGraw, K.O. and Wong, S.P. Forming inferences about some intraclass correlation coefficients. *Psychological Methods* **1996**, 1(1), 30-46.
52. Müller, R. and Büttner, P. A critical discussion of intraclass correlation coefficients. **1994**, *Statistics in Medicine* 13(23-24), 2465-2476.
53. Koo, T.K. and Li, M.Y. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Research* **2016**, 15(2), 155-163.
54. Krebs, D.E. Declare your ICC type. *Physical Therapy* **1986** 66(9), 1431.
55. Fleiss, J.L. *The Design and Analysis of Clinical Experiments*. John Wiley and Sons, New York, NY, USA 1986.
56. de Mast, J. Agreement and kappa-type indices. *American Statistician* **2007**, 61, 148-153.
57. Donner, A. and Wells, G. A comparison of confidence interval methods for the intraclass correlation coefficient. *Biometrics*. **1986**, 42(2), 401-412.
58. Doros, G. and Lew, R. Design based on intraclass correlation coefficients. *American Journal of Biostatistics* **2010**, 1(1), 1-8.
59. Ionan, A.C., Polley, M.-Y., McShane, L.M. and Dobbin, K.K. Comparison of confidence interval methods for an intraclass correlation coefficient (ICC). *BMC Medical Research Methodology* **2014**, 14: 121.
60. Portney, L.G. and Watkins, M.P. *Foundations of Clinical Research. Applications to Practice. Third Edition*. Prentice Hall Health, Upper Saddle River, NJ, USA; 2009.
61. Copay, A.G., Subach, B.R., Glassman, S.D., Polly, D.W. et al. Understanding the minimum clinically important difference: a review of concepts and methods. *The Spine Journal* **2007**, 7, 541-546.
62. Zou, G.Y. Sample size formulas for estimating intraclass correlation coefficients with precision and assurance. *Statistics in Medicine* **2012**, 31, 3972-3981.
63. Head, H. *Aphasia and Kindred Disorders of Speech*. Cambridge University Press., Cambridge, UK; 1926
64. Bleiberg, J., Garmoe, W.S., Halpern, E.L., Reeves, D.L., et al. Consistency of within-day and across-day performance after mild brain injury. *Neuropsychiatry, Neuropsychology and Behavioral Neurology* **1997**, 10(4), 247-253.
65. Garavaglia, L., Gulich, D., Defeo, M.M., Mailland, J.T. et al. The effect of age on heart rate variability of healthy subjects. *PLoS ONE*. **2021**, 16(10): e0255894.
66. Voss, A., Schroeder, R., Heitmann, A., Peters, A., et al. Short-term heart rate variability – influence of gender and age in healthy subjects. *PLoS ONE* **2015**, 10(3), e0118308.
67. Dietrich, D.F., Schindler, C., Schwartz, J., Barthélémy, J.C., et al. Heart rate variability in an ageing population and its association with lifestyle and cardiovascular risk factors: results of the SAPALDIA study. *Europace* **2006**, 8, 521-529.
68. Yukishita, T., Lee, K., Kim, S., Yumoto, Y., et al. Age and sex-dependent alterations in heart rate variability profiling the characteristics of men and women in their 30s. *Anti-Aging Med.* **2010**, 7, 94-99.
69. Gutchess, A.H., Ieui, Y. and Federmeier, K.D. Event-related potentials reveal age differences in the encoding and recognition of scenes. *Journal of Cognitive Neuroscience* **2007**, 19(7), 1089-1103.
70. Guillem, E. and Mograss, M. Gender differences in memory processing: evidence from event related potentials to faces. *Brain and Cognition* **2005**, 57, 84-92
71. Campanella, S., Rossignol, M., Mejjas, S., Joassin, F., et al. Human gender differences in an emotional visual oddball task: an event-related potentials study. *Neuroscience Letters* **2004**, 367, 14-18.
72. Aslaksen, P.M., Bystad, M., Vambheim, S.M. and Flaten, M.A. Gender differences in placebo analgesia: event-related potentials and emotional modulation. *Psychosomatic Medicine* **2011**, 73, 193-199.
73. Fukusaki, C., Kawakubo, K. and Yamamoto, Y. Assessment of the primary effect of aging on heart rate variability in humans. *Clinical Autonomic Research* **2000**, 10, 123-130.
74. Choi, J., Cha, W. and Park, M.-G. Declining trends of heart rate variability according to aging in healthy Asian adults. *Front. Aging Neuroscience* **2020**, 12: 610626.

75. Mu, Y., Kitayama, S., Han, S. and Gelfand, M.J. How culture gets enbrained: cultural differences in event-related potentials of social norm violations. *Proceedings of the National Academy of Science USA*. **2015**, 112(50), 15348-15353.
76. Kemp, A.H., Quintana, D.S., Gray, M.A., Felmingham, K.L., et al. Impact of depression and antidepressant treatment on heart rate variability: a review and meta-analysis. *Biological Psychiatry* **2010**, 67(11), 1067-1074.
77. Brunoni, A.R., Kemp, A.H., and Dantas, E.M. (2013). Heart rate variability is a trait marker of major depressive disorder: evidence from the sertraline vs. electric current therapy to treat depression: clinical study. *International Journal of Neuropsychopharmacology*. **2013** 16(9), 1937-1949.
78. Bozkurt., A., Barcin, C., Isintas, M., Ak, M. et al. Changes in heart rate variability before and after ECT in the treatment of resistant major depressive disorder. *Israeli Journal of Psychiatry and Related Sciences* **2013**, 50(1), 40-46.
79. Alvares, G.A., Quintana, D.S., Hickie, I.B. and Guastella, A.J. Autonomic nervous system dysfunction in psychiatric disorders and the impact of psychotropic medications: a systematic review and meta-analysis. *Journal of Psychiatry and Neuroscience* **2016**, 41(2), 89-104.
80. Udupa,K. Thirthalli, J., Sathyaprabha, T.N., Kishore, K.R., et al. Differential actions of antidepressant treatments on cardiac autonomic alterations in depression: a prospective comparison. *Asian Journal of Psychiatry* **2011**, 4, 100-106.
81. Nahshoni, E., Aizenberg, D., Sigler, M., Strasberg, B., et al. Heart rate variability increases in elderly patients who respond to electroconvulsive therapy. *Journal of Psychosomatic Medicine* **2004**, 56(1), 89-94.
82. Tarvainen,M.P., Nishanen, J.-P., Lipponen, J.A., Ranta-aho, P.O., et al. Kubios HRV heart rate variability analysis software. *Computer Methods and Programs in Biomedicine* **2014**, 113, 210-220.
83. O'Donnell, B.F., Salisbury, D.F., Niznikiewicz, M.A., Brenner, C.A. et al. (2012). Abnormalities of event related potential components in schizophrenia. In: *The Oxford Handbook of Event-Related Potential Components*. S. J. Luck and E. S. Kappenman, eds. Oxford University Press, New York, NY, USA, 2012; pp. 537-562.
84. Bruder, G.E., Kayser, J. and Tenke, C.E. Event-related brain potentials in depression: clinical, cognitive, and neurophysiological implications. In: *The Oxford Handbook of Event-Related Potential Components*. S. J. Luck and E. S. Kappenman, eds. Oxford University Press. New York, NY, USA 2012; pp. 563-692
85. Buchheim, A., Labek, K., Taubner, S., Kessler, H., et al. Modulation of gamma band activity and late positive potential in patients with chronic depression after psychodynamic psychotherapy. *Psychotherapy and Psychosomatics* **2018**, 87(4), 252-254.
86. Siegle, G.J., Condray, R., Thase, M.E., Keshavan, M. et al. Sustained gamma-band EEG following negative words in depression and schizophrenia. *International J. Psychophysiology* **2010**, 75, 107-118.
87. Blackwood, D.H., Whalley, L.J., Christie, J.E., Blackburn, I.M., et al. Changes in auditory P3 event-related potential in schizophrenia and depression. *British Journal of Psychiatry* **1987**, 150, 154-160.
88. Gangadhar, B.N., Ancy, J., Janakiramaiah, N. and Umapathy, C. P300 amplitude in non-bipolar melancholic depression. *Journal of Affective Disorders* **1993**, 28, 57-60.
89. Umbricht, D. and Krljes, S. Mismatch negativity in schizophrenia: a meta-analysis. *Schizophrenia Research* **2005**, 76, 1-23.
90. Lavoie, M.E., Murray, M.M., Deppen, P., Knyazera, et al. Glutathione precursor N-acetyl-cysteine improves mismatch negativity in schizophrenia patients *Neuropsychopharmacology* **2008**, 33, 2187-2199.
91. Zhou, Z., Zhu, H. and Chen, L. Effects of aripiprazole on mismatch negativity (MMN). *PLoS One*. **2013**, 8: e52186.
92. Gehring, W.J., Liu, Y., Orr, J.M. and Carp, J. The error related negativity (ERN/Ne). In: *The Oxford Handbook of Event-Related Potential Components*. S. J. Luck and E. S. Kappenman, eds. Oxford University Press, New York, NY: USA, 2012; pp. 231-292
93. Olvet, D.M. and Hajcak, G. (2008). The error related negativity (ERN) and pathophysiology: toward an endophenotype. *Clinical Psychophysiology Reviews* **2008**, 28, 1343-1354.
94. Moser, J.S., Moran, T.P., Schroder, H.S., Donnellan, M.B. et al. On the relationship between anxiety and error monitoring: a meta-analysis and conceptual framework. *Frontiers in Human Neuroscience* **2013**, 7: 466.
95. Hajcak, G., Franklin, M.E., Foa, E.B. and Simons, R.F. Increased error-related brain activity in pediatric obsessive-compulsive disorder before and after treatment. *American Journal of Psychiatry* **2008**, 165(1), 116-123
96. Riesel, A., Endrass, T., Auerbach, L.A. and Kathmann, N. Overactive performance monitoring as an endophenotype for obsessive-compulsive disorder: evidence from a treatment study. *American Journal of Psychiatry* **2015**, 172(7), 665-673.
97. Kujawa, A., Weinberg, A., Bunford, N., Fitzgerald, K.D., et al. Error-related brain activity in youth and young adults before and after treatment for generalized or social anxiety disorder. *Progress in Neuro-psychopharmacology and Biological Psychiatry* **2016**, 71, 162-168
98. Gorka, S.M., Burkhouse, K.L., Klumpp, H., Kennedy, A.E., et al. Error-related brain activity as a treatment moderator and index of symptom change during cognitive-behavioral therapy or selective serotonin reuptake inhibitors. *Neuropsychopharmacology* **2018**, 43(6), 1355-1363.
99. Ladouceur, C.B., Tan, P.Z., Shama,V., Bylsma, L.M., et al. Error-related brain activity in pediatric anxiety disorders remains elevated following individual therapy: a randomized clinical trial. *Journal of Child Psychology and Psychiatry* **2018**, 59(11), 1152-1161.
100. Valt, C., Huber, D., Erhardt, K. and Stürmer, B. Internal and external signal processing in patients with panic disorder: an event-related potential (ERP) study. *PLoS One*. **2018**, 13(11): e0208257.
101. Valt, C., Huber, D. and Stürmer, B. (2020). Treatment-related changes towards normalization of the abnormal external signal processing in panic disorder. *PLoS One*. **2020**, 15(1), e0227673.

102. Hajcak, G., Klawohn, J. and Meyer, A. The utility of event-related potentials in clinical psychology. *Annual Reviews of Clinical Psychology* **2019**, 15, 71-95.
103. Goldstein, K. (1934, 1995). *Der Aufbau des Organismus. Einführung in die Biologie unter besonderer Berücksichtigung der Erfahrungen am kranken Menschen*. The Hague: Nijhof. Republished in English as *The Organism*, Forward by Oliver Sachs. Zone Books, Brooklyn, NY, USA.
104. Niv, S. (2013). Clinical efficacy and potential mechanisms of neurofeedback. *Personality Individual Differences* **2013**, 54(6), 676-686.
105. Micoulaud-Franchi, J.-A., Geoffroy, P.A., Fond, G., Lopez, R., et al. EEG neurofeedback treatments in children with ADHD: an updated meta-analysis of randomized controlled trials. *Frontiers in Human Neuroscience* **2014**, 8: 906.
106. Rosenfeld, J.P. EEG biofeedback of frontal alpha asymmetry in affective disorders. *Biofeedback*. **1997**, 25(1), 8-25.
107. Wiedemann, G., Pauli, P., Dengler, W., Lutzenberger, W., et al. Frontal brain asymmetry as a biological substrate of emotions in patients with panic disorders. *Archives of General Psychiatry* **1999**, 56, 78-84.
108. Papo, D. Neurofeedback: principles, appraisals, and outstanding issues. *European Journal of Neuroscience* **2019**, 49(11), 1454-1469.
109. Vaschillo, E.G., Bates, M.E., Vaschillo, B., Lehrer, P., et al. Heart rate variability response to alcohol, placebo, and emotional picture cue challenges: effects of 0.1-Hz stimulation. *Psychophysiology* **2008**, 45(5), 847-858.
110. Darragh, M., Vanderboor, T., Booth, R.J., Sollers, J.J. and Consedine, N.S. Placebo 'serotonin' increases heart rate variability in recovery from psychosocial stress. *Physiology and Behavior* **2015**, 145, 45-49.
111. Daniali, H. and Flaten, M.A. Placebo analgesia, nocebo hyperalgesia and the cardiovascular system: a qualitative systematic review. *Frontiers in Physiology* **2020**, 11: 549807.
112. Schienle, A., Gremsl, A., Übel, S. and Körner, C. Testing the effect of disgust placebo with eye tracking. *International Journal of Psychophysiology* **2016**, 101, 69-75.
113. Schienle, A., Übel, S., Schongassner, F., Ille, R., et al. Disgust regulation via placebo: an fMRI study. *Social, Cognitive and Affective Neuroscience* **2014**, 9, 985-990.
Schienle, A., Übel, S., and Scharmüller, W. Placebo treatment can alter primary visual cortex activity and connectivity. *Neuroscience* **2014**, 263, 125-129.
114. Schienle, A., Übel, S., and Scharmüller, W. Placebo treatment can alter primary visual cortex activity and connectivity. *Neuroscience* **2014**, 263, 125-129.
115. Gremsl, A., Schwab, D., Höfler, C. and Schienle, A. (2018). Placebo effects in spider phobia: an eye-tracking experiment. *Cognition and Emotion*. 3(8), 1571-1577.
116. Wager, T.D., Dagfinn, M.B. and Casey, K.L. Placebo effects in laser-evoked pain potentials. *Brain Behavior and Immunology* **2006**, 20, 219-230.
117. Watson, A., El-Deredy, W., Vogt, B.A. and Jones, A.K.P. Placebo analgesia is not due to compliance or habituation: EEG and behavioural evidence. *Neuroreport* **2007**, 18, 771-775.
118. Meyer, B., Yuen, K.S.L., Ertl, M., Polomac, N. et al. Neural mechanisms of placebo anxiolysis. *Journal of Neuroscience* **2015**, 35(19). 7365-7373.
119. Übel, S., Leutgeb, V. and Schienle, A. Electrocutaneous effects of a disgust placebo in children. *Biological Psychology* **2015**, 108, 78-84.
120. Schienle, A., Gremsl, A. and Schwab, D. Placebos can change affective contexts: an event related potential study. *Biological Psychology* **2020**, 150, 107843.
121. van Elk, M., Groenendijk, E. and Hoogeveen, S. Placebo brain stimulation affects subjective by not neurocognitive measures of error processing. *Journal of Cognitive Enhancement* **2020**, 4, 389-400.
122. Guevarra, D.A., Moser, J.S., Wager, T.D. and Kross, D. Placebos without deception reduce self-report and neural measures of emotional distress. *Nature Communications* **2020**, 11: e3785.
123. Colloca, L. and Howick, J. Placebos without deception: a review of their outcomes, mechanisms, and ethics. *International Reviews of Neurobiology* **2018**, 138, 219-240.
124. Kaptchuk, T.J., Friedlander, E., Kelley, J.M., Sanchez, M.N., et al. Placebos without deception: a randomized controlled trial in irritable bowel syndrome. *PLoS ONE*. **2010**, 5, e15591
125. Hajcak, G., MacNamara, A. and Olvet, D.M. Event related potentials, emotion, and emotion regulation: an integrative review. *Developmental Neuropsychology* **2010**, 35, 129-155.
126. Liu, Y., Huang, H., McGinnis-Deweese, M., Keil, A. et al. Neural substrate of the late positive potential in emotional processing. *Journal of Neuroscience* **2012**, 32, 14563-14572.
127. Lin, Y., Fisher, M.E., Roberts, S.M.M. and Moser, J.S. Deconstructing the emotion regulatory properties of mindfulness: an electrophysiological investigation. *Frontiers in Human Neuroscience* **2016**, 10: 451
128. Posternak, M.A., Solomon, D.A., Leon, A.C., Mueller, T.I., The naturalistic course of unipolar major depression in the absence of somatic therapy. *Journal of Nervous and Mental Diseases* **2006**, 194(5), 324-329.
129. Posternak, M.A. and Miller, I. Untreated short-term course of major depression: a meta-analysis of outcomes from studies using wait-list control groups. *Journal of Affective Disorders* **2001**, 66, 139-146.

130. Muthukumaraswamy, S.D. High-frequency brain activity and muscle artifacts in MEG/EEG: a review and recommendations. *Frontiers in Human Neuroscience* **2013**, 7: 138.
131. Rapp, P.E., Albano, A.M., Schmah, T.I. and Farwell, L.A. Filtered noise can mimic low dimensional chaotic attractors. *Physical Review E* **1993**, 47, 2289-2297.
132. Theiler, J. and Rapp, P.E. Re-examination of evidence for low-dimensional nonlinear structure in the human electroencephalogram. *Electroencephalography and Clinical Neurophysiology* **1996**, 98, 213-222
133. Hipp, J.F. and Siegel, M. Dissociating neuronal gamma-band activity from cranial and ocular muscle activity in EEG. *Frontiers in Human Neuroscience* **2013**, 7: 338.
134. Luck, S.J. *An Introduction to the Event-Related Potential Technique. Second Edition*. MIT Press, Cambridge, MA, USA; 2014.
135. Head, M.L., Holman, L., Lanfear, R., Kahn, A.T. et al. The extent and consequences of P-hacking in science. *PLoS Biology* **2015**, 13(3): e1002106.
136. Adda, J., Decker, C. and Ottaviani, M. P-hacking in clinical trials and how incentives shape the distribution of results across phases. *Proceedings of the National Academy of Science USA*. **2020**, 117(24), 13386-13392.
137. Rapp, P.E., Cellucci, C.J., Watanabe, T.A.A., Albano, A.M. et al. Surrogate data pathologies and the false-positive rejection of the null hypothesis. *International Journal of Bifurcation and Chaos* **2001**, 11, 983-997.
138. Rapp, P.E., Albano, A.M., Zimmerman, I.D. and Jiménez-Montaño, M.A. Phase-randomized surrogates can produce spurious identifications of non-random structure. *Physics Letters* **1994**, 192A, 27-33
139. Garrett-Ruffin, S., Cowden Hindash, A., Kaczurkin, A.N., Mears, R.P., et al. Open science in psychophysiology: an overview of challenges and emerging solutions. *International Journal Psychophysiology* **2021**, 162, 69-78.
140. Foster, E.D. and Deardorff, A. Open Science Framework (OSF) *Journal of the Medical Libraries Association* **2017**, 105(2), 203-206.
141. Saunders, B. and Inzlicht, M. Pooling resources to enhance rigour in psychophysiological research; insights from open science approaches to meta-analysis. *International Journal of Psychophysiology* **2021**, 162, 112-120.
142. Picton, T.W., Bentin, S., Berg, P., Donchin, E., et al. Guidelines for using human event-related potentials to study cognition: recording standards and publication criteria. *Psychophysiology* **2000**, 37, 127-152.
143. Duncan, C.C., Barry, R.J., Connolly, J.F., Fischer, C., et al. Event-related potentials in clinical research: Guidelines for eliciting recording and quantifying mismatch negativity, P300, and N400. *Clinical Neurophysiology* **2009**, 120(11), 1883-1908.
144. Campanella, S. and Colin, C. Event-related potentials and biomarkers of psychiatric diseases: the necessity to adopt and develop multi-site guidelines. *Frontiers in Behavioral Neuroscience* **2014**, 8: 428.
145. Kappenman, E.S. and Luck, S.J. Best practices for event-related potential research in clinical populations. *Biological Psychiatry, Cognitive Neuroscience and Neuroimaging* **2016**, 1, 101-115.
146. Campanella, S. Use of cognitive event related potentials in the management of psychiatric disorders: towards an individual follow-up and multicomponent clinical approach. *World Journal of Psychiatry* **2021**, 11(5), 153-168.
147. Kappenman, E.S., Farrens, J.L., Zhang, W., Stewart, A.X. and Luck S.J. ERP CORE: An open resource for human event-related potential research. *Neuroimage* **2021**, 225: 117465.
148. Mayer, K., Wyckoff, S.N. and Strehl, U. Underarousal in adult ADHD: How are peripheral and cortical arousal related? *Clinical EEG and Neuroscience* **2015**, 47(3), 171179.
149. Bramon, F., Rabe-Hesketh, S., Sham, P., Murray, R.M. et al. Meta-analysis of the P300 and P50 waveforms in schizophrenia. *Schizophrenia Research* **2004**, 70, 315-329.
150. Karaaslan, F., Gonul, A.S., Oguz, A., Erdinc, E. et al. P300 changes in major depressive disorders with and without psychotic features. *Journal of Affective Disorders* **2003**, 73,283-287.
151. Anderson, M.L., James, J.R. and Kirwan, C.B. An event-related potential investigation of pattern separation and pattern completion processes. *Cognitive Neuroscience* **2017**, 8(1), 9-23.
152. Ehlers, C.L, Wills, D.N., Desikan, A., Phillips, E. et al. Decreases in energy and increase in phase locking of event related oscillations to auditory stimuli occur during adolescence in human and rodent brain. *Developmental Neuroscience* **2014**, 36, 175-195.
153. Ehlers, C.L., Wills, D.N., Karriker-Jaffe, K.J., Gilder, D.A., et al. Delta event-related oscillations are related to a history of extreme binge drinking in adolescence and suicide risk. *Behavioral Science* **2020**, 10, 154
154. Murphy, M., Whitton, A.E., Deccy, S., Ironside, M.L., et al. Abnormalities in electroencephalographic microstates are state and trait markers of major depressive disorder. *Neuropsychopharmacology* **2020**, 45, 2030-2037.
155. Darmon, D. Specific differential entropy rate estimation for continuous-valued time series. *Entropy*. **2016**, 18, 190.
156. Stam, C.J. Modern network science of neurological disorders. *Nature Reviews Neuroscience* **2014**, 15(10), 683-695.
157. Efron, B. Second thoughts on the bootstrap. *Statistical Science* **2003**, 18(2), 135-140.
158. Efron, B. and Tibshirani, R. Bootstrap methods for statistical errors: confidence intervals and other measures of statistical accuracy. *Statistical Science* **1986**, 1(1), 54-75.
159. Fernández-Delgado, M., Cernadas, E. and Barro, S. Do we need hundreds of classifiers to solve real world classification problems? *J. Machine Learning Res.* **2014**, 15, 3133-3181.
160. Vanschoren, J., Blockeel, H., Pfahringer, B. and Holmes, G. Experiment databases. A new way to share, organize and learn from experiments. *Machine Learning*. **2012**, 87(2), 127-158.
161. Bache, K. and Lichman, M. UCI machine learning repository. <https://archive.ics.edu> Accessed 2022-03-03

-
162. Li, R., Johansen, J.S., Ahmed, H., Ilyevsky, T.V., et al. The perils and pitfalls of block design for EEG classification experiments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. **2021**, 43(1), 316-333.
 163. Ahmed, H., Wilbur, R.B., Bharadwaj, H.M. and Siskind, J.M. Confounds in the data - comments on "Decoding brain representations by multimodal learning of neural activity and visual features." *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2021**, doi 10.1099/TPAMI.2021.3121268.
 164. FDA-NIH Biomarker Working Group. (2021). BEST (Biomarkers, EndpointS and Other Tools) resource. Silver Spring, MD, USA: Food and Drug Administration.
 165. Food and Drug Administration. (2018). Biomarker qualification: evidentiary framework. Guidance for Industry and FDA staff. Silver Spring MD, USA Draft Guidance Food and Drug Administration.
 166. Leptak, K.C., Menetski, J., Wagner, J.A., Aubrecht, J., et al. What evidence do we need for biomarker qualification. *Science Translational Medicine* **2017**, 9: eaal4599.
 167. Prata,D., Mochelli,A. and Kapur,S. Clinically meaningful biomarkers for psychosis: a systematic and quantitative review. *Neuroscience and Biobehavioral Reviews* **2014**, 45, 134-141.
 168. Chan, M.K., Cooper, J.D., Bot, M., Steiner, J., et al. Identification of an immune-neuroendocrine biomarker panel for detection of depression: a joint effects statistical approach. *Neuroendocrinology* **2016**, 103, 693-710.
 169. Chan, M.K., Krebs, M.-O., Cox, D., Guest, P.C., et al. Development of a blood-based molecular biomarker test for identification of schizophrenia before disease onset. *Translational Psychiatry* **2015**, 5, c601.