

Ethical Considerations in the Use of Psychophysiological Methods to Identify Biological Markers for Internalizing Disorders

Huiting Liu MA, Department of Psychology
Lynne Lieberman MA, Department of Psychology
Stewart A. Shankman PhD, Department of Psychology, Department of Psychiatry
University of Illinois at Chicago, USA

Abstract

The National Institutes of Mental Health's Research Domain Criteria (RDoC) is a promising research initiative that emphasizes the use of biomarkers to advance assessment of internalizing psychopathology. However, despite exciting research results, important ethical issues need to be taken into account for future research and policy endeavors involving biomarkers. The goal of this paper is to briefly draw awareness to several of these issues. Specifically, we discuss (1) potential implications of changing mental health classification in regards to the ongoing debate about diagnostic *labels* in mental health; (2) the general lack of standardization, normative data, and psychometrics (reliability and validity) of biomarker data; (3) the importance of demographic factors – in terms of moderating factors of the relationship between brain and behavior, and representation in biomarker research; and (4) the importance of increasing understanding of the temporal and contextual factors that affect biomarker assessment. At the present time, RDoC (and other efforts to identify biomarkers) are only research enterprises. However, these ethical considerations are essential to consider if (or perhaps when) they are introduced into clinical and policy endeavors.

Key words: biomarker, internalizing disorders, depression, anxiety

Internalizing disorders such as depression and anxiety lead to significant disease burden including subjective distress, significant dysfunction across different domains, and poor economic and health outcomes (Mathers, Fat, & Boerma, 2008; Murray et al., 2012). Accurate assessment

and detection of these disorders is a crucial first step toward identifying appropriate treatment of these disorders and subsequently reducing their negative functional and emotional impact. In addition to identifying individuals currently experiencing an internalizing disorder, it is also important to predict treatment response, course of illness, and individuals who are at risk for developing these disorders. Over the last several decades, mental health assessment has largely relied on the Diagnostic and Statistical Manual of Mental Disorders (DSM) for diagnostic criteria of specific disorders. Although the DSM provides a somewhat reliable diagnostic system for many disorders, there remains an absence of objective markers of illnesses.

More recently, the National Institute of Mental Health (Insel et al., 2010) proposed the Research Domain Criteria (RDoC) initiative as an alternative research framework¹ of classifying mental disorders more dimensionally, with an emphasis on identifying biomarkers of disorders in order to improve prediction of risk and onset of mental disorders. The goal of RDoC is to facilitate future personalized medicine in which diagnosis and prediction will be informed by objective laboratory measures of biomarkers. The term “biomarker” has been defined in many different ways, but one widely used definition is that it is “a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention” (Strimbu & Tavel, 2010) and can be measured through many different methodologies (Biomarkers Definitions Working Group, 2001). Of note, although some biomarkers are heritable and represent a causal mechanism of an illness, other biomarkers may simply be a correlate of the illness (Miller & Rockstroh, 2013).

Consistent with the RDoC Initiative, research on psychopathology has increasingly focused on identifying biomarkers of psychological disorders, including internalizing psychopathologies. Electrophysiological measures may be particularly promising for this endeavor as they can provide measures of risk for psychopathology independent of one’s subjective awareness of their tendencies, are not impacted by reporting biases, and are relatively inexpensive and easy to administer. For example, there is burgeoning evidence across multiple laboratories that aberrant electrophysiological reactivity to threat and reward (two RDoC constructs) help distinguish different internalizing psychopathologies (Gorka et al., in press; Grillon et al., 2008; Shankman et al., 2013)

¹ Although RDoC is only designed to be a research framework at the present time, the ultimate goal of the initiative is for it to be translated into clinical practice (Cuthbert & Insel, 2013; Zalta & Shankman, 2016).

as well as vulnerability for these psychopathologies (Weinberg et al., 2015; Nelson et al., 2013). There may, however, be unforeseen ethical consequences (or at least ethical considerations) that need to be taken into account with the RDoC initiative and biomarker research in psychopathology in general. Using examples from electrophysiology, the goal of this paper is to briefly highlight several of these issues with the purpose of increasing awareness of these considerations for future studies and policy makers.

Impact of Mental Health Labels

The dimensional classification system proposed by the RDoC initiative necessitates changes in how the field currently uses diagnostic labels. Disorders previously identified categorically by the DSM, such as Major Depressive Disorder, will instead fall under transdiagnostic dimensional constructs, such as reduced reward sensitivity. It is unclear how this change will impact patients' and public perception of mental health. Indeed, the impact of current categorical diagnostic labels remains a debate among mental health professionals. Although some argue that diagnosis exacerbates symptoms and increases stigma associated with labeling (Robinson, 2009), others believe that diagnosis does the opposite and reduces feelings of isolation, self-blame, and stigma, by providing a common language between patients, mental health professionals, and the public (Craddock & Mynors-Wallis, 2014). Diagnoses may provide reassurance to patients that their condition is not unique and can be treated by evidence-based practices. Additionally, efforts to reduce stigma in mental disorders have led to the creation of education and advocacy groups (e.g. Anxiety and Depression Association of America) dedicated to improving the lives of individuals with specific mental health disorders.

Given that DSM categories have been the standard nomenclature for several decades, changes in classification and labeling may lead to unexpected consequences, including a return to stigma, or perhaps an undermining of efforts designed to increase public awareness of mental health issues. For example, over the last several decades, there has been increasing concern in academic and popular media about "disease mongering" – a pejorative term to describe the widening of diagnostic boundaries in order to expand the market (most often pharmaceutical) for those who need treatment (Moynihan, Doran, & Henry, 2008; Wolinsky, 2005). It is therefore

possible that changing a patient's "label" from, for example, the DSM label *major depressive disorder* to the RDoC-ian label *reduced reward functioning* may be perceived by some as a rapacious attempt to increase a market - even if this new label was given because it has better reliability, prognostic power and validity. When revising the DSM5, the organizers were aware of this issue and created a "Clinical and Public Health Committee" whose job, in part, was to anticipate the potential impact that any changes in the DSM would have on public opinion and public health in general (Yager & McIntyre, 2014). As mentioned above, RDoC is presently only a research framework. However, if RDoC were to be implemented into clinical practice, a similar type of committee would need to be created – especially given how dramatically the labels for psychiatric illness would change.

Lack of Psychometric Data for Biomarkers

Despite growing promising research on biomarkers, the psychometrics (e.g., reliability and validity) of biomarkers in mental health are often assumed, but not assessed. Although psychometrics of tests might not, at first pass, be thought of as an 'ethical issue,' the proper development and use of assessment tools is part of the ethical code of conduct for numerous disciplines. For example, the American Psychological Association's Ethics Code states "Psychologists...use appropriate psychometric procedures and current scientific or professional knowledge for test design, standardization, validation, reduction or elimination of bias and recommendations for use" (code 9.05; American Psychological Association, 2010). In other words, psychometrics is an important ethical issue as it impacts a clinician's scope of practice.

Several methodological issues contribute to the questionable validity of many psychophysiological measures. First, although conventions of psychophysiological measurement are generally agreed upon, psychophysiological tasks tend to lack standardization. Individual research labs typically develop unique tasks to measure the constructs of interest and even when tasks are shared between research groups, labs often modify the task to their needs, contributing to a lack of standardization of measures. For example, numerous studies have used affective pictures to examine physiological [e.g. EMG startle, ERP (event-related potential or measured brain response)] indicators of emotional states. Studies vary a great deal in the type of pictures that they examine

(even if drawn from the same databank of pictures such as the International Affective Picture Set; Lang, Bradley, & Cuthbert, 2008) as well as number of stimuli, duration of presentation, and time between stimuli. To address this issue, the RDoC unit at NIMH recently convened a workgroup to develop a list of standardized paradigms for many RDoC constructs (NIMH, 2016). However, this only identified a small number of tasks and was only meant to be a guideline from which researchers can choose and not a definitive list.

Second, the psychophysiological literature also lacks information on normative data (e.g., frequency, means, standard deviations, etc.) for most biomarkers. Psychophysiological studies almost always examine relative differences on levels of a biomarker (e.g., comparing those with depression to controls on levels of ERP responses to reward) rather than on whether the biomarker is 'present' vs. 'absent.' This type of approach necessitates large scale normative data from which means, percentiles, etc. can be looked up for individual participants. Without normative data, electrophysiological biomarkers (or any biomarker that is dimensional) are unlikely to inform clinical diagnosis and treatment.

Lastly, unlike measures of self-report, the general reliability of many psychophysiological measures are largely unknown. Few studies include reliability analyses for psychophysiological measures before publishing them. In contrast, it is standard procedure that self-report measures of mental health symptoms (e.g. Beck Depression Inventory – BDI; Beck, 1988) undergo rigorous testing before they are disseminated to ensure good psychometric properties. Given that validity is, in part, a function of reliability (Cronbach & Meehl, 1955), without adequate data on reliability, it is unclear how to interpret the associations between many psychophysiological biomarkers and various validators (e.g., prospective course of illness, treatment response, etc.).

Biased Attitudes in Favor of Biological Measures

It is unclear why a discrepancy appears to exist in the psychometric standards for self-report versus "biological" methodologies such as psychophysiology. One possibility is that researchers (and consumers of research) hold beliefs that biological measures are inherently more objective than psychological measures. This bias may decrease the likelihood that they will scrutinize the psychometric properties of the biological measures relative to what they would do for self-report

measures. Indeed, research evidence suggests that people may hold positive biases for biological research methods, particularly methods involving neuroscience. For example, McCabe and Castel (2008) demonstrated that the presence of brain images in neuroscience journal articles increased people's ratings of the scientific merit of the articles when compared to ratings for the same articles presented without the brain image. Similarly, Fernandez-Duque, Evans, Christian, & Hodges (2015) found that including irrelevant neuroscience information in a description of psychological phenomena yielded higher ratings of argument quality compared to inclusion of social science information. Weisberg et al. (2008) also found that inclusion of *logically irrelevant* neuroscience information led to more positive judgments of scientific explanations, suggesting that individuals were unable to critically analyze the neuroscience information included. Taken together, these results suggest that observed bias in favor of neuroscience or biological explanations extend beyond brain imaging and also apply to language related to neuroscience, even when it is superfluous to the point of the study. The presence of a conceptual bias favoring biological explanations over psychological may contribute to less scrutiny for psychophysiological methods, and are further in line with prior research demonstrating that information consistent with a preferred conclusion tends to be examined less critically (Ditto & Lopez, 1992).

Given people's inherent bias in favor of neuroscience methods, it is particularly important to critically examine the psychometric properties and validity of proposed biomarkers before using them for target outcomes for treatments or preventative interventions. Additionally, it is important to use specific and deliberate language when disseminating research on biomarkers for internalizing psychopathology to the public. Given the interest in neuroscience in the popular press (Beck, 2010), researchers have the responsibility of accurately and concisely conveying findings (including known limitations), and not overstating the implications for current research on biomarkers. Overreaching the implications of biomarker research may lead to unforeseen consequences in public perception and policy.

Diversity Considerations in Biomarkers

A lack of attention to diversity persists in psychophysiological research, raising both ethical and methodological concerns. A recent special issue of *Psychophysiology*, a flagship journal of

psychophysiological research, discussed the moderating role of demographic factors on the association between psychophysiological biomarkers and behavior (Gatzke-Kopp, 2016). Current research demonstrating ethnic group differences in psychophysiology (e.g., Liu, Lieberman, Stevens, Auerbach, & Shankman, 2016) challenges the notion that observed biomarkers are universal, as demographic factors may have a dynamic influence on development of brain structure and function. Thus, examining biomarkers solely in largely homogenous demographic groups severely limits the generalizability of biomarkers in mental health, as both biological (e.g. sex, race) and environmental factors (e.g. socioeconomic status, culture) may contribute to individual differences in expressions of biomarkers. On the other hand, inclusion of demographically diverse samples to ensure adequate representation may lead to masking of significant findings due to unexamined group differences (or group differences for which the investigator does not have the statistical power to test). For example, a recent meta-analysis showed that the association between anxiety and the error-related negativity (ERN), an ERP component reflecting error monitoring, may be unique to women (Moser et al., 2016). Thus, inclusion of male participants may reduce the overall group magnitude of ERN, reducing the likelihood to detecting significant effects in the overall sample. It is therefore important to specifically examine demographic group differences in biomarker research, rather than relying on diverse sampling techniques to derive average group effects.

Developmental factors such as age and developmental stage may also contribute to individual differences in psychophysiology. Much of human physiology varies across the lifespan, including cortical volume, heart rate, blood pressure, and atrophy of skin and muscle cells (Boss & Seegmiller, 1981). Given the known age-related impact on physiology, it is critical to consider developmental differences in psychophysiological biomarkers for psychopathology. For example, as EEG power is positively correlated with brain maturity, the range of EEG alpha frequency also changes depending on age (Klimesch, 1999). The traditionally used alpha frequency range of 7.5 to 12.5 Hz may be appropriate for adults, but less so for children, as they typically exhibit a lower range of alpha frequency. Childhood and adolescence covers a wide range of development and specific norms of psychophysiological measures need to be developed for each developmental group. However, establishing baselines and norms for this population may be particularly challenging, as children experience rapid physiological, cognitive, and emotional development. Two individuals of the same age may mature at different rates and belong to different developmental stages. Thus, age may be an inappropriate normative group for psychophysiological measures. Instead, incorporating

other methods of establishing developmental stage, such as the use of Tanner stages (Marshall & Tanner, 1969), may help establish more developmentally appropriate norms for certain psychophysiological biomarkers.

More broadly, underrepresentation of certain groups in psychophysiological research is an important ethical issue for the generalizability of results to the public at large. Surprisingly, research participant representation is not well-documented in psychophysiological research, as more than 80% of empirical articles do not report racial/ethnic information for participants (Gatzke-Kopp, 2016). However, racial minorities, residents in rural areas, as well as individuals of low SES are likely to be underrepresented in research due to lack of access and recruitment for research, geographical segregation, and other psychological and physical barriers to participation. Thus, biomarker research may not accurately reflect true correlates of psychopathology for underrepresented individuals. Given the complex interplay between biology and environment on individual differences in brain structure and function, assessments and interventions for psychopathology informed by biomarkers may be ineffective or perhaps harmful for underrepresented individuals.

Biomarkers: Stability and Context Dependency

Lastly, biomarkers may vary in their stability across contexts. A genetic mutation in DNA in an individual, for example, is present irrespective of the context or time point at which the DNA is measured. That is, an assessment of DNA will yield the same result whether the sample was obtained at home school, or under laboratory-induced stress. Genomic variation is therefore trait-like in that it displays temporal and cross-situational stability. A psychophysiological biomarker, on the other hand, may be entirely state-like, and therefore only be present on a particular day and in a specific context. For example, in a sample of adolescents, van den Bulk et al. (2013) showed poor retest reliability ($ICC < 0.4$) for amygdala activity during an emotional face task with fMRI (but better reliability for prefrontal cortex). On the other hand, psychophysiological biomarkers may be trait-like in their temporal stability (e.g., the rank order stability may be significant from one time to the next), but dependent on the context in which it is measured. As mentioned above, heightened startle responding is characteristic of multiple fear-based anxiety psychopathologies (Gorka et al., in press), is specific to unpredictably threatening contextual situations and (generally) not predictably

threatening situations (Gorka et al., in press). Heightened startle potentiation to unpredictable threat has been found to display strong retest reliability (Shankman et al., 2013). There is also growing evidence suggesting that this biomarker represents a vulnerability factor for fear-based anxiety psychopathologies (Gorka et al., in press; Nelson et al., 2013), and may therefore precede disorder onset. Thus, although startle responsivity is context dependent, the change in startle from baseline to a threatening context is trait-like.

There are specific ethical issues to be considered depending on the stability of a particular biomarker. For example, a biomarker may be used to identify children at risk for the development of a particular psychopathology in adulthood. If that biomarker is state-like, then it is critical to consider the time point and context in which that biomarker is assessed. For example, a biomarker may only be evident when it is measured at home, but not at school or in a laboratory (perhaps due to the child's increased comfort in their home environment). If this contextual factor is not considered, then there may be false negatives for the presence of a particular abnormal biomarker. This would be problematic if a biomarker is being used to identify children who would benefit from a preventative intervention as children would be falsely classified as 'not having the biomarker' and would thus miss out on being able to receive the preventative intervention. False positives are also problematic in this example as it could result in pulling children out of classes to receive an unnecessarily costly intervention.

Summary

In summary, the present work highlights several ethical issues that need to be addressed by future research and policy makers to improve the diagnostic and clinical utility of biomarkers in psychopathology. Although biomarkers are currently only used in research settings, implementing them into clinical practice would necessitate careful thought and planning on not only how changing diagnostic procedures would impact public opinion and public health in general, but also on whether the psychometrics and demographic representation of previous studies warrants inclusion of particular biomarkers in clinical practice. Nevertheless, despite these challenges, we are optimistic that biomarkers can someday improve the validity and prognostic power of psychiatric assessment.

References

- American Psychological Association. (2010). Ethical principles of psychologists and code of conduct. Retrieved from <http://www.apa.org/ethics/code/>
- Beck, D. M. (2010). The appeal of the brain in the popular press. *Perspective on Psychological Science*, 5, 762-766.
- Biomarkers Definitions Working Group (2001). Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. *Clinical Pharmacology & Therapeutics*, 69, 89-95.
- Boss, G. R., & Seegmiller, J. E. (1981). Age-related physiological changes and their clinical significance. *Geriatric Medicine*, 135, 434-440.
- Craddock, N., & Mynors-Wallis, L. (2014). Psychiatric diagnosis: Impersonal, imperfect and important. *The British Journal of Psychiatry*, 204, 93-95.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281-302.
- Cuthbert, B. N., & Insel, T. R. (2013). Toward the future of psychiatric diagnosis: the seven pillars of RDoC. *BMC Medicine*, 11.
- Ditto, P. H. & Lopez, D. F. (1992). Motivated skepticism: Use of differential decision criteria for preferred and nonpreferred conclusions. *Journal of Personality and Social Psychology*, 63, 568-584.
- Fernandez-Duque, D., Evans, J., Christian, C., & Hodges, S. D. (2015). Superfluous neuroscience information makes explanations of psychological phenomena more appealing. *Journal of Cognitive Neuroscience*, 27, 926-944.
- Gatzke-Kopp, L. M. (2016). Diversity and representation: Key issues for psychophysiological science. *Psychophysiology*, 53, 3-13.
- Gorka, S. M., Lieberman, L., Shankman, S. A., & Phan, K. L. (In press). Startle potentiation to uncertain threat as a psychophysiological indicator of fear-based psychopathology: An examination across multiple internalizing disorders. *Journal of Abnormal Psychology*, 126(1), 8-18.
- Grillon, C., Lissek, S., Rabin, S., McDowell, D., Dvir, S., & Pine, D. S. (2008). Increased anxiety during anticipation of unpredictable but not predictable aversive stimuli as a psychophysiological marker of panic disorder. *American Journal of Psychiatry*, 165(7), 898-904.
- Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., . . . Wang, P. (2010). Research domain criteria (RDoC): Toward a new classification framework for research on mental disorders. *American Journal of Psychiatry*, 167, 748-751.
- Klimesch, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain Research Reviews*, 29, 169-195.
- Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2008). International affective picture system (IAPS): Affective ratings of pictures and instruction manual. Technical Report A-8. University of Florida, Gainesville, FL.
- Lieberman, L., Gorka, S. M., Shankman, S. A., & Phan, K. L. (2016). Impact of panic on psychophysiological and neural reactivity to unpredictable threat in depression and anxiety. *Clinical Psychological Science*. Ahead of print. doi: 10.1177/2167702616666507

- Liu, H., Lieberman, L., Stevens, E., Auerbach, R. P., & Shankman, S. A. (2016). Using a cultural and RDoC framework to conceptualize anxiety in Asian Americans. *Journal of Anxiety Disorders*. Ahead of print. <http://dx.doi.org/10.1016/j.janxdis.2016.09.006>
- Marshall, W. A., & Tanner, J. M. (1969). Variations in pattern of pubertal changes in girls. *Archives of Disease in Childhood*, *44*, 291-303.
- Mathers, C. D., Fat, D. M., & Boerma, J. (2008). The global burden of disease: 2004 update. Geneva, Switzerland: World Health Organization.
- McCabe, D. P., & Castel, A. D. (2008). Seeing is believing: The effect of brain images on judgments of scientific reasoning. *Cognition*, *107*, 343-352.
- Miller, G. A., & Rockstroh, B. (2013). Endophenotypes in psychopathology research: Where do we stand? *Annual Reviews of Clinical Psychology*, *9*, 177-213.
- Moser, J. S., Moran, T. P., Kneip, C., Schroder, H. S., & Larson, M. J. (2016). Sex moderates the association between symptoms of anxiety, but not obsessive compulsive disorder, and error-monitoring brain activity: A meta-analytic review. *Psychophysiology*, *53*, 21-29.
- Moynihan, R., Doran, E., & Henry, D. (2008). Disease Mongering Is Now Part of the Global Health Debate. *PLoS Med* *5*, e106.
- Murray, C. J., Vos, T., Lozano, R., Naghavi, M., Flaxman, A. D., Michaud, C., . . . Memish, Z. A. (2012). Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: A systematic analysis for the Global Burden of Disease Study 2010. *The Lancet*, *380*, 2197-2223.
- National Institute of Mental Health (2016). Behavioral assessment methods for RDoC constructs. Retrieved from https://www.nimh.nih.gov/about/advisory-boards-and-groups/namhc/reports/rdoc_council_workgroup_report_153440.pdf
- Nelson, B. D., Bishop, J. R., Sarapas, C., Kittles, R. A., & Shankman, S. A. (2014). Asians demonstrate reduced sensitivity to unpredictable threat: A preliminary startle investigation using genetic ancestry in a multi-ethnic samples. *Emotion*, *14*, 615-623.
- Nelson, B. D., McGowan, S. K., Sarapas, C., Robison-Andrew, E. J., Altman, S. E., ... & Shankman, S. A. (2013). Biomarkers of threat and reward sensitivity demonstrate unique associations with risk for psychopathology. *Journal of Abnormal Psychology*, *122*, 662-671.
- Robinson, A. (2009). Dangers of diagnostic labels in patients with mental health issues. *Progress in Neurology and Psychiatry*, *13*, 6-7.
- Shankman, S. A., Nelson, B. D., Sarapas, C., Robison-Andrew, E. J., Campbell, M. L., Altman, S. E., ... Gorka, S. M. (2013). A psychophysiological investigation of threat and reward sensitivity in individuals with panic disorder and/or major depressive disorder. *Journal of Abnormal Psychology*, *112*, 322-338.
- Strimbu, K., & Tavel, J. A. (2010). What are biomarkers? *Current Opinions in HIV and AIDS*, *5*, 463-466.
- van den Bulk, B. G., Koolschijn, P. C., Meens, P. H., van Lang, N. D., van der Wee, N. J., ... & Crone, E. A. (2013). How stable is activation in the amygdala and prefrontal cortex in adolescence? A study of emotional face processing across three measurements. *Developmental Cognitive Neuroscience*, *4*, 65-76.

- Weisberg, D. S., Keil, F. C., Goodstein, J., Rawson, E., & Gray, J. R. (2008). The seductive allure of neuroscience explanations. *Journal of Cognitive Neuroscience*, 20, 470-477.
- Wolinsky, H. (2005). Disease mongering and drug marketing. *EMBO Reports*, 6, 612-614.
- Yager, J., & McIntyre, J. S. (2014). DSM-5 clinical and public health committee: Challenges and considerations. *American Journal of Psychiatry*, 171, 142-144.
- Zalta, A. K., & Shankman, S. A. (2016). Conducting psychopathology prevention research in the RDoC era. *Clinical Psychology: Science and Practice*, 23, 94-104.

Acknowledgements: none

Competing Interests: none

Address for Correspondence: e-mail: stewarts@uic.edu.

Date of Publication: Dec 1, 2017