Published in final edited form as:

Psychophysiology. 2016 March ; 53(3): 406–409. doi:10.1111/psyp.12587.

Psychobiological Operationalization of RDoC Constructs: Methodological and Conceptual Opportunities and Challenges

Annmarie MacNamara and K. Luan Phan

Department of Psychiatry (AM, KLP), University of Illinois at Chicago, Chicago, IL; Departments of Psychology and Anatomy and Cell Biology, and the Graduate Program in Neuroscience (KLP), University of Illinois at Chicago, Chicago, IL; Mental Health Service Line (KLP), Jesse Brown VA Medical Center, Chicago, IL

Abstract

NIMH's Research Domain Criteria (RDoC) project seeks to advance the diagnosis, prevention and treatment of mental disorders by promoting psychobiological research on dimensional constructs that might cut across traditional diagnostic boundaries (Kozak and Cuthbert, this issue). At the core of this approach is the notion that these dimensional constructs can be assessed across different units of analysis (e.g., genes, physiology, behavior), enriching the constructs and providing more complete explanations of clinical problems. While the conceptual aspects of RDoC have been discussed in several prior papers, its methodological aspects have received comparatively less attention. For example, *how* to integrate data from different units of analysis has been relatively unclear. Here, we discuss one means of psychobiologically operationalizing RDoC constructs across different units of analysis (the psychoneurometric approach; Yancey et al., this issue), highlighting ways in which this approach might be refined in future iterations. We conclude that there is much to be learned from this technique, however greater attention to scale-development methods and to psychometrics will likely benefit this and other methodological approaches to combining measurements across multiple units of analysis.

The Research Domain Criteria (RDoC) initiative aims to move beyond the subjectivist tradition in which phenomenological (i.e., subjective) experience is deemed the primary measure of an emotion such as fear (Kozak & Cuthbert, this issue). That is, RDoC aspires to a heterophenomenological (Dennett, 2003) approach, which combines multiple indicators/units of analysis (e.g., from genes to fear physiology to fear behavior). This approach is central to RDoC's interest in uncovering biopsychological explanations of clinical problems, organized according to "process" constructs that connect to basic brain circuits (Kozak & Cuthbert, this issue). Although RDoC compels collection of multiple data units (Cuthbert, 2013, 2014a, 2014b; Insel & Quirion, 2005; Kozak & Cuthbert, this issue; Morris & Cuthbert, 2012), *how* to integrate these units has as of yet been unclear.

Correspondence concerning this article should be addressed to Annmarie MacNamara, Department of Psychiatry, University of Illinois at Chicago, 1747 W. Roosevelt Rd., Chicago, IL 60607. Contact: amacnamara@psych.uic.edu; 312-413-4707.

Drs. MacNamara and Phan have no conflicts of interest to declare.

Yancey and colleagues' paper (this issue) provides a much-needed, concrete example of the heterophenomenological approach. The authors report on a large adult cohort ($N > 450$) assessed for brain event-related potential (ERP), heart rate (HR), and electromyographic (EMG) reactivity during viewing of aversive stimuli, and characterized dimensionally (low, medium, and high scores) on a scale measure of threat sensitivity, the Trait Fear Inventory (TF-55; Kramer, Patrick, Krueger, & Gasperi, 2012). Their focus is on the creation of a psychoneurometric index of threat sensitivity using factor analysis to derive a latent dimension reflecting variance common to both self-report (TF-55) and three *a priori* physiological measures of threat sensitivity (HR acceleration, startle potentiation, and corrugator EMG enhancement). The third and final portion of their paper focuses on the evaluation of criterion validity for each of three indices: a) a self-report only index (TF-55 scores), b) factor scores for the physiology-only index, and c) factor scores for the psychoneurometric index. Validity was assessed in relation to a symptom-based criterion variable, reflecting fear disorder symptoms (i.e., derived from the Structured Clinical Interview for DSM-IV Axis I Disorders; First, Spitzer, Gibbon, & Williams, 2002), and using a physiological criterion variable, entailing a composite of two other facial EMG measures (general corrugator muscle tension, aversive-neutral differentiation for orbicularis EMG activity), and two brain-ERP measures (late positive potential, noise-probe P3 response).

Yancey and colleagues provide a conceptual and methodological contribution in terms of how multiple measurements from different levels of analysis can be integrated in search of more reliable and valid measures of clinically relevant processes (see also Moser, Durbin, Patrick, & Schmidt, 2015; Patrick et al., 2013; Patrick, Durbin, & Moser, 2012). Moreover, because self-report and biological measures are given equal consideration in this work, Yancey and colleagues' approach addresses criticisms of RDoC as being biologically reductionistic (Lilienfeld, 2014).

Here, we pose the following three questions to the field, in the hopes of furthering the development of the psychoneurometric approach in future iterations. In considering these questions, we call for extension and refinement of the techniques put forth by Yancey and colleagues, in the hopes that the methodological arm of RDoC will soon be a match for its conceptual arm.

First, what lessons can be learned from the scale development literature in terms of item selection and retention that can be applied to the development of a psychoneurometric index?

There are many ways to formally assess potential items for inclusion in a new scale, including differences in theoretical approach (e.g., classical test theory vs. item response theory). As such, researchers seeking to develop psychoneurometric indices have available to them many of the same methods that are commonly used in the development of self-report measures and developed over several decades of work. One of the first steps in scale creation is to select potential items. In their work, Yancey and colleagues chose four *a priori* "items" thought to measure threat sensitivity - i.e., TF-55 scores, HR, startle and corrugator EMG.

By contrast, developers of self-report measures tend to be over-inclusive when first identifying potential items. This means including even items that seem only tangentially related to the target construct, because items that are less informative will be excluded at later stages (Clark & Watson, 1995). A large pool of candidate items helps insure that a sufficient number of items and only those that are most suitable will ultimately be included in the final measure. In the development of future psychoneurometric indices, then, it is recommended that researchers begin with a broad pool of potential items.

Once a pool of potential items has been identified, it is advisable to assess the psychometric properties of these items, using a sufficiently large and heterogeneous sample. Yancey and colleagues report item-level reliability, but not other item-level psychometrics which are needed to meet the assumptions of unidimensional latent trait measurement. This typically includes insuring that the latent trait being assessed (e.g., threat sensitivity) is represented by a single real-valued number (θ); that item responses are independent after accounting for the latent trait θ ; and that items are monotonic (i.e the probability of endorsing an item increases with the trait of interest). Items that do not meet these criteria may be candidates for removal, as are items that are skewed (as they probably provide little information, correlate poorly with other items, and will fare badly in structural analyses as a result of their limited variability). At this stage, it can also be determined whether the scale is more accurate at low, middle, or high ends of the distribution – additional information which can aid in deciding whether new items should be added or existing items discarded.

Finally, it is necessary to insure that scale items are cohesive. Yancey and colleagues used exploratory factor analysis to assess whether the a priori items indexed a common dimension. Although factor analysis is one means of assessing item cohesiveness and of narrowing down items, the assessment of internal consistency – e.g., coefficient alpha or split-half reliability – provides additional and unique information. Generally speaking, a scale should exhibit good internal consistency, balanced with scale coverage (Loevinger, 1954).

Perhaps most importantly, the methods described above can be repeated in an iterative process – with items added and excluded at each step – until the most parsimonious and reliable scale is created. Yancey and colleagues did not demonstrate all of the steps above or this iterative process; however they innovated the combination of multi-level data into a latent variable for predicting across different units of analysis. This approach and data prompts us to continue refining multi-unit trait measurement by adopting strategies borrowed from the scale construction field.

Second, how does the choice of a criterion variable affect validity estimations for a newly created psychoneurometric index?

Yancey and colleagues assess the validity of their psychoneurometric index of threat sensitivity by examining correlations with measures thought to tap a similar construct. Here, the authors illustrate the application of a classic psychometric technique - assessment of concurrent criterion validity - to the psychobiological realm (an approach espoused by

RDoC; Kozak & Cuthbert, this issue). Criterion-related validity is relatively simple to assess, but it carries with it a few caveats (Garbin, 2015).

The first caveat is that a criterion variable must exist. This can be difficult when developing a measure of a new (or redefined) construct, especially given that RDoC aims to revise the way we conceptualize mental disorders. For instance, one pitfall might be choosing a criterion variable that simply reflects diagnostic criteria, which may initially have been conceptualized inadequately (Kozak & Cuthbert, this issue). Most self-report psychopathology measures were designed to assess traits or symptoms that relate to the Diagnostic and Statistical Manual of Mental Disorders (DSM; American Psychological Association, 2013), meaning that even dimensional scores on these measures might be constrained by initial conceptions of the disorders. Furthermore, given RDoC's emphasis on the measurement of fine-grained, transdiagnostic processes, it may be necessary to identify criterion variables that measure more narrowly-defined psychological constructs such as "anhedonia" or "intolerance of uncertainty" than are usually reflected in overall symptom scores.

A related caveat is that estimations of validity will depend on the quality of the criterion variable. This is potentially problematic because few criterion variables are "gold standards" for what we want to measure (i.e., the target construct). Furthermore, even when such a measure exists, its own validity and reliability will most likely be limited, impacting validity assessment. To illustrate how the choice of criterion variable can affect estimations of validity, let's assume that a perfect measure of our target construct exists (Fig. 1). If our criterion variable is very good, it may have validity as high as $r = .8$ (sharing 64% of its variability with the perfect measure). Next, let's assume we have two new measures whose validity we want to evaluate with respect to the criterion variable. Measure 1 correlates strongly with the criterion ($r = .7$; 49% overlap), whereas Measure 2 correlates more modestly with the criterion ($r = .5$, 25% overlap). The implication is that Measure 1 is more valid, however, Measure 2 may actually have a stronger relationship with the perfect measure, but may simply look weaker because of the choice of criterion variable – e.g., the criterion variable may share more method variance with Measure 1.¹ In Yancey and colleagues' work, different estimations of validity were observed for the newly created indices, depending on whether a fear disorder composite or a physiology composite was used as the criterion variable. The highest correlations were achieved using the fear disorder composite, and in particular, between this composite and the scale-based index (i.e., TF-55 scores). This correlation may have been elevated by both higher reliability for the TF-55 (and the composite measure – more on this in the section below) and greater shared method variance between two symptom-based measures. Therefore, in the future, criterion variables themselves may need to incorporate multiple units of analysis and importantly, will need to be subject to refinement over time – e.g., via bootstrapping – as target RDoC constructs become more precisely defined.

A third and final caveat is that correlation does not imply equivalence. That is, even if a newly created measure correlates strongly with a criterion variable, it may also correlate

¹Example from Garbin (2015) and used with permission.

with other variables, raising the question - what is the new measure measuring? Although there is really no way around this (it is the nature of correlations), assessment of construct validity may be useful in further interrogating the meaning of a new measures. Moreover, predictive validity (e.g., can I test you now and predict how well you will respond to a treatment to be administered in the future?) and postdictive validity (e.g., can I test you now and get a valid measure of something that happened previously, such as childhood trauma?) may also be useful in giving meaning to a newly created psychoneurometric index.

Estimations of validity will need to be interpreted bearing in mind the choice of criterion variable and the limitations inherent in the meaning of a correlation. Moreover, construct validation in the age of RDoC may at times involve the iterative redefinition of the constructs themselves (Yancey et al., this issue), which may have initially been defined in subjective or psychological realms alone. Formal testing of hypothesized relationships between different latent variables and observed variables (e.g., via structural equation modeling) may be helpful in this regard.

Third, how does reliability (especially as it varies across units of analysis) affect the validity assessment of a psychoneurometric measure?

No matter how well-chosen a criterion variable, the validity of the new assessment measure will be low if either measure exhibits poor reliability (i.e., reliability places an upper limit on validity). In Yancey and colleagues' work, the internal reliability of the psychophysiological variables input into the psychoneurometric index ranged from $r = .27$ to $.58$, whereas the self-report measure input into this index (the TF-55), had an internal reliability of $r = .94$. Therefore, differences in validity estimations for Yancey and colleagues' scale-only index, physiology-only index, and combined psychoneurometric index were likely driven at least in part by differences in reliability.² Psychophysiological variables may have poorer internal reliability than self-report measures, because unlike self-report measures, they are not *designed* to have high reliability. Indeed, the reliability of psychophysiological variables may more closely resemble the reliability of a single *item* on a self-report measure than the entire measure itself (Yancey et al., this issue).

Concerted efforts are needed to improve the reliability of psychophysiological measures. For instance, because the internal reliability of psychophysiological measures can vary substantially by task (e.g., Riesel, Weinberg, Endrass, Meyer, & Hajcak, 2013), the field may wish to adopt a standardized battery of tasks that have been determined to yield the most reliable physiological measurements. An often-touted example of such an initiative is the Cognitive Neuroscience Treatment Research to Improve Cognition in Schizophrenia (CNTRCS; Barch et al., 2009) consortium, which was established for the explicit purpose of examining and *enhancing* psychometric outcome data on a variety of tasks from cognitive neuroscience. Beyond initiatives like CNTRICS, we need to demonstrate the psychometric

²It is possible to statistically "correct" for reliability-related attenuation in validity and to estimate the validity of a measure if the reliability of the measure and/or the criterion variable were improved. These corrections should be used with good judgement – e.g., not to assess the "true" validity of a measure, but to evaluate how the validity of a measure might improve, as a precursor to working to improve its reliability.

properties of neurobiological measures (Bennett & Miller, 2010; Kujawa, Klein, & Proudfit, 2013; Moran, Jendrusina, & Moser, 2013; Riesel et al., 2013; Sauder, Hajcak, Angstadt, & Phan, 2013), and to examine their reliability in comparison to other measures (e.g., behavior) derived from the same task (Kappenman, Farrens, Luck, & Proudfit, 2014; Kappenman, MacNamara, & Proudfit, 2014). In addition, a better understanding of factors influencing the reliability of neurobiological measures could lead to changes in the way such measures are collected, improving reliability (e.g., by increasing signal-to-noise ratio; Bennett & Miller, 2010; Zanto, Pa, & Gazzaley, 2014).

If psychophysiological variables are to take their place among other units of analysis (e.g., self-report variables) in RDoC inspired conceptualizations of psychopathology, they will need to demonstrate comparable reliability. In the meantime, the relative performance of measurements from these different units of analysis must be assessed with reliability in mind.

Conclusion

The goal of RDoC is to achieve coherence across psychological and biological units of analysis and Yancey and colleagues' work takes a pragmatic step towards achieving this goal. Other researchers have taken similar approaches toward combining measurements across units of analysis – e.g., behavior and self-report (Vrieze et al., 2014), and have shown improved predictive power (in treatment outcome) when using a combination of variables compared to individual variables. Thus, while practical methods for the joint analysis and combination of multilevel data are still very much in development, initial results suggest that there may be benefits to this approach.

Despite the appeal of multi-level trait conceptualization, however, there may be times at which coherence across different units of analysis may not be the desired outcome. One such instance is measuring risk in the context of multifinality, in which similar conditions result in different outcomes (e.g., monozygotic twins who share a biological predisposition to a disorder but are nonetheless discordant on symptoms; Lilienfeld, 2014). As another example, because of differences in resolution, it may not be possible to always achieve concurrence between measurements from different levels of analysis. For example, distinctions that are evident at one level of analysis (e.g., brain) may not be evident at another, less fine-grained level of analysis such as self-report (e.g., fear and worry; Sharp, Miller, & Heller, 2015). Further, it remains untested whether variation in specific brain structure and function units (e.g., amygdala and associated “fear” nodes) referred to by Kozak and Cuthbert (this issue) as “circuits” will further validate the threat sensitivity/acute threat construct. Therefore, while a psychoneurometric bridge across units of analysis will be informative for understanding some psychopathological constructs, there may be exceptions and unanticipated complexity to this rule. The challenges for RDoC going forward will thus include not only refining and developing new approaches to multilevel construct conceptualization, but also knowing when and when not to apply these techniques.

Acknowledgments

Author Note

Psychophysiology. Author manuscript; available in PMC 2017 March 01.

AM is supported by NIMH K23MH105553. KLP is supported by NIMH R01MH101497. Thank you to Dan Foti, Autumn Kujawa, Leah H. Rubin, and Anna Weinberg for their comments on an earlier version of this commentary.

References

- Barch DM, Carter CS, Arnsten A, Buchanan RW, Cohen JD, Geyer M, ... Heinssen R. Selecting paradigms from cognitive neuroscience for translation into use in clinical trials: proceedings of the third CNTRICS meeting. *Schizophrenia Bulletin*. 2009; 35(1):109–114. <http://doi.org/10.1093/schbul/sbn163>. [PubMed: 19023126]
- Bennett CM, Miller MB. How reliable are the results from functional magnetic resonance imaging? *Annals of the New York Academy of Sciences*. 2010; 1191(1):133–155. <http://doi.org/10.1111/j.1749-6632.2010.05446.x>. [PubMed: 20392279]
- Clark LA, Watson D. Constructing Validity: Basic Issues in Objective Scale Development. *Psychological Assessment*. 1995; 7(3):309–19.
- Cuthbert, BN. Research Domain Criteria (RDoc). Audiovisual presentation, Anxiety and Depression Association of America; 2013.
- Cuthbert BN. The RDoc framework: facilitating transition from ICD/DSM to dimensional approaches that integrate neuroscience and psychopathology. *World Psychiatry*. 2014a; 13(1):28–35. <http://doi.org/10.1002/wps.20087>. [PubMed: 24497240]
- Cuthbert BN. Translating intermediate phenotypes to psychopathology: The NIMH Research Domain Criteria. *Psychophysiology*. 2014b; 51(12):1205–1206. <http://doi.org/10.1111/psyp.12342>. [PubMed: 25387702]
- Dennett D. Who's On First? Heterophenomenology Explained. *Journal of Consciousness Studies*. 2003; 10(9–10):19–30.
- First, MB.; Spitzer, RL.; Gibbon, M.; Williams, J.B.W. Structured Clinical Interview for DSM-IV Axis I Disorders, Research Version. New York: Biometrics Research, New York State Psychiatric Institute; 2002. Non-patient Edition (SCID-I/NP)
- Garbin, CP. Validity. 2015. Retrieved September 14, 2015, from <http://psych.unl.edu/psycrs/971/scale/validity.PDF>
- Insel TR, Quirion R. Psychiatry as a clinical neuroscience discipline. *JAMA: The Journal of the American Medical Association*. 2005; 294:2221–2224.
- Kappenman ES, Farrens JL, Luck SJ, Proudfit GH. Behavioral and ERP measures of attentional bias to threat in the dot-probe task: poor reliability and lack of correlation with anxiety. *Frontiers in Psychology*. 2014; 5:1368. <http://doi.org/10.3389/fpsyg.2014.01368>. [PubMed: 25538644]
- Kappenman, ES.; MacNamara, A.; Proudfit, GH. Electroocortical evidence for rapid allocation of attention to threat in the dot-probe task. *Social Cognitive and Affective Neuroscience*. 2014. <http://doi.org/10.1093/scan/nsu098>
- Kozak MJ, Cuthbert BN. The NIMH Research Domain Criteria Initiative: Background, Issues, and Pragmatics. *Psychophysiology*. 2015
- Kramer MD, Patrick CJ, Krueger RF, Gasperi M. Delineating physiologic defensive reactivity in the domain of self-report: phenotypic and etiologic structure of dispositional fear. *Psychological Medicine*. 2012; 42(6):1305–1320. <http://doi.org/10.1017/S0033291711002194>. [PubMed: 22008475]
- Kujawa A, Klein DN, Proudfit GH. Two-year stability of the late positive potential across middle childhood and adolescence. *Biological Psychology*. 2013; 94(2):290–296. <http://doi.org/10.1016/j.biopsycho.2013.07.002>. [PubMed: 23872165]
- Lilienfeld SO. The Research Domain Criteria (RDoC): an analysis of methodological and conceptual challenges. *Behaviour Research and Therapy*. 2014; 62:129–139. <http://doi.org/10.1016/j.brat.2014.07.019>. [PubMed: 25156396]
- Loevinger J. The attenuation paradox in test theory. *Psychological Bulletin*. 1954; 51(5):493–504. <http://doi.org/10.1037/h0058543>. [PubMed: 13204488]
- Moran TP, Jendrusina AA, Moser JS. The psychometric properties of the late positive potential during emotion processing and regulation. *Brain Research*. 2013; 1516:66–75. <http://doi.org/10.1016/j.brainres.2013.04.018>. [PubMed: 23603408]

- Morris SE, Cuthbert BN. Research Domain Criteria: cognitive systems, neural circuits, and dimensions of behavior. *Dialogues Clin Neurosci*. 2012; 14(1):29–37. [PubMed: 22577302]
- Moser JS, Durbin CE, Patrick CJ, Schmidt NB. Combining neural and behavioral indicators in the assessment of internalizing psychopathology in children and adolescents. *Journal of Clinical Child and Adolescent Psychology*. 2015; 44(2):329–340. <http://doi.org/10.1080/15374416.2013.865191>. [PubMed: 24484405]
- Patrick CJ, Durbin CE, Moser JS. Reconceptualizing antisocial deviance in neurobehavioral terms. *Development and Psychopathology*. 2012; 24(Special Issue 03):1047–1071. <http://doi.org/10.1017/S0954579412000533>. [PubMed: 22781871]
- Patrick CJ, Venables NC, Yancey JR, Hicks BM, Nelson LD, Kramer MD. A construct-network approach to bridging diagnostic and physiological domains: Application to assessment of externalizing psychopathology. *Journal of Abnormal Psychology*. 2013; 122(3):902–916. <http://doi.org/10.1037/a0032807>. [PubMed: 24016026]
- Riesel A, Weinberg A, Endrass T, Meyer A, Hajcak G. The ERN is the ERN is the ERN? Convergent validity of error-related brain activity across different tasks. *Biological Psychology*. 2013; 93(3): 377–385. <http://doi.org/10.1016/j.biopsycho.2013.04.007>. [PubMed: 23607999]
- Sauder CL, Hajcak G, Angstadt M, Phan KL. Test-retest reliability of amygdala response to emotional faces. *Psychophysiology*. 2013; 50(11):1147–1156. <http://doi.org/10.1111/psyp.12129>. [PubMed: 24128307]
- Sharp, PB.; Miller, GA.; Heller, W. Transdiagnostic dimensions of anxiety: Neural mechanisms, executive functions, and new directions. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology*. 2015. <http://doi.org/10.1016/j.ijpsycho.2015.07.001>
- Vrieze E, Demyttenaere K, Bruffaerts R, Hermans D, Pizzagalli DA, Sienaert P, ... Claes S. Dimensions in major depressive disorder and their relevance for treatment outcome. *Journal of Affective Disorders*. 2014; 155:35–41. <http://doi.org/10.1016/j.jad.2013.10.020>. [PubMed: 24210628]
- Zanto TP, Pa J, Gazzaley A. Reliability measures of functional magnetic resonance imaging in a longitudinal evaluation of mild cognitive impairment. *NeuroImage*. 2014; 84:443–452. <http://doi.org/10.1016/j.neuroimage.2013.08.063>. [PubMed: 24018304]

Figure 1.

An illustration of the extent to which construct-criterion overlap can influence estimations of validity. Here, the choice of criterion variable makes Measure 1 look more valid (greater overlap with the criterion) than Measure 2 (less overlap with the criterion, but more overlap with a “perfect” measure of the construct, were one to exist). Adapted from Garbin (2015) and used with permission.