Commentary

Authors’ reply to the commentary on “Establishing norms for error-related brain activity during the arrow Flanker task among young adults”

Annmarie MacNamara a,*, Michael J. Imburgio a, Kaylin E. Hill b, Iulia Banica c, Anna Weinberg c, Dan Foti b

a Department of Psychological and Brain Sciences, Texas A&M University, 4235 TAMU, College Station, TX, USA
b Department of Psychological Sciences, Purdue University, West Lafayette, IN, United States
c Department of Psychology, McGill University, Montreal, QC, Canada

A R T I C L E   I N F O

Keywords:
Norms
Error-related negativity (ERN)
Flanker task
Event-related potential (ERP)
Assessment
Research Domain Criteria (RDoC)

A B S T R A C T

In their commentary on our article, “Establishing norms for error-related brain activity during the arrow Flanker task among young adults” (Imburgio et al., 2020), Clayson and colleagues (2021) voiced their concerns about our development of norms for an event-related potential measure of error monitoring, the error-related negativity (ERN). The central flaw in their commentary is the idea that because we don’t know all the factors that can affect the ERN, it should not be normed. We respond to this idea, while also reiterating points made in our original manuscript: a) at present, the reported norms are not intended to be used for individual clinical assessment and b) our norms should be considered specific to the procedures (i.e., recording and processing parameters) and task used (i.e., arrow Flanker). Contrary to Clayson and colleagues’ claims, we believe that information about the distribution of the ERN (i.e., our norms) in a large sample representative of those used in much of the ERN literature (i.e., unselected young adults) will be useful to the field and that this information stands to increase, not decrease, understanding of the ERN.

We have norms for many different measures, without fully understanding their variability. But Clayson and colleagues (Clayson et al., 2021) argue that because we do not know all the factors that can affect the error-related negativity (ERN), it should not be normed. One point they make is that we should not norm the ERN because the reasons for individual differences in the ERN are sometimes unclear. It is certainly true that the ERN reflects contributions from multiple, overlapping phenotypes (Hajcak et al., 2019) — meaning that bivariate correlations (rather than, for example, regression-based models) are likely to fail at painting a cohesive picture of how the ERN relates to psychopathology. Therefore, pointing to conflicting or null zero-order correlations is not particularly useful at this point in time. Along these lines, the National Institute of Mental Health’s (NIMH) Research Domain Criteria (RDoC) initiative (and contemporary thinking on psychopathology more broadly) suggest that psychopathology should be conceptualized along multiple dimensions, rather than categorically. As such, the utility of between-group norms (e.g., norms for clinically anxious patients and healthy controls) is at present unclear. Our sample was not selected to be psychiatrically healthy and we would expect that it provides normative variation in phenotypes related to the ERN.

Despite overlapping associations between the ERN and various dimensional phenotypes, the ERN has, nonetheless, been robustly associated with clinical anxiety (e.g., Cavanagh et al., 2017; Moser et al., 2013; Riesel et al., 2011; Saunders and Inzlicht, 2020; Stern et al., 2010; Weinberg et al., 2012). Indeed, this is reflected in our norms. For example, among datasets that used the same task and data collection/processing procedures as in our norms, participants with clinical anxiety (i.e., generalized anxiety disorder) were found to have ERN difference scores that fell in the 75th–90th percentile range according to our norms (−9.24 μV in Weinberg et al., 2010, −8.69 μV in Weinberg et al., 2012), whereas psychiatrically healthy individuals from the same datasets fell around the 50th percentile (−5.20 μV and −5.38 μV, respectively). This demonstrates the utility of our norms for anchoring and contextualizing group results, which was not possible before. Nonetheless, categorization at the individual level (e.g., for psychiatric decision-making) involves other considerations (e.g., sensitivity and specificity) and is not advisable at the current time.

Instead, the norms will be useful in different ways. As exemplified above, researchers with ERN data (who used the same, widely used task and processing stream) will now have an idea of how their data aligns with that of a large sample representative of those frequently used in ERP research (i.e., unselected young adults). This will be useful not just

* Corresponding author.
E-mail addresses: amacnamara@tamu.edu, annmarie.macnamara@gmail.com (A. MacNamara).

https://doi.org/10.1016/j.neuroimage.2021.117908
Received 9 October 2020; Received in revised form 18 February 2021; Accepted 20 February 2021
Available online 27 February 2021
1053-8119/© 2021 Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)
for clinical samples, but also to ascertain where ERN data from any new sample falls, relative to the population. By analogy, when assessing associations between depression symptoms and reaction time, it is useful to know whether the observed depression scores represent only a narrow segment or the full array of scores that are possible in the broader population. In addition to benchmarking study results, another way in which our norms may be useful is to facilitate the recruitment or stratification of research samples according to the size of the ERN. As described in our original manuscript, and as advocated for by the NIMH (e.g., Insel et al., 2010), research participants can now be recruited according to their scores on neurophysiological variables that correspond to variation along dimensions of relevance to psychopathology. By flipping the traditional research design on its head (i.e., dependent variables become independent variables), it may be possible to ascertain greater understanding of both normal and abnormal functioning. Before the publication of our norms, such an approach was not feasible for the ERN.

Clayson and colleagues also raise the issue of nuisance factors. Doubtlessly, multiple factors can affect the magnitude of an ERN component (Frodl et al., 2001; McCarthy and Wood, 1985; Urbach and Kutas, 2002), though skull thickness appears to have a different effect than described by Clayson and colleagues. In contrast to Clayson and colleagues, MRI-EEG work indicates that skull thickness has an additive (not multiplicative) effect on ERPs (Frodl et al., 2001), which would preserve the validity of difference scores, consistent with our recommendation (in the original manuscript) that difference scores may be more appropriate for normative comparisons. Generally speaking, however, any measure in the fields of psychology and neuroscience will be subject to some contamination by nuisance variables. For instance, performance on a cognitive test may be lower when an individual is sleep-deprived — yet despite this, norms for those tests are still thought to be informative. Indeed, sample sizes for norms are large for exactly this reason — so that the contribution of individuals who are not representative of the population is attenuated.

Clayson and colleagues also note that there are differences in the way researchers measure the ERN. The pooled data presented in our norms was derived using only the arrow version of the Flanker task with stimuli, timing and instructions consistent across datasets. Here, we reiterate again that our norms should only be assumed to hold for this (widely used) arrow version of the Flanker task. We also note that the ERN from this task has been shown to have high internal consistency (Clayson, 2020; Foti et al., 2013; Olvet and Hajcak, 2009; Riesel et al., 2013), rendering Clayson and colleague’s point about varied internal consistency in the Clayson meta-analysis (Clayson, 2020) irrelevant. We did find that ERN values varied somewhat with the length of the task, though it was not clear that this mattered much overall (Imburgio et al., 2020).

To emphasize how different ways of quantifying the ERN can affect resulting values, Clayson and colleagues point to the difference in amplitude between mean raw ERN (i.e., not the difference score) in 326 males in our norms (3.18 μV) and in 429 males in Fischer and colleagues (−5.37 μV; 2016), who used a different means of scoring the ERN. Using the error minus correct difference score (as recommended in our manuscript; Imburgio et al., 2020), our mean amplitude of −6.75 μV is actually quite similar to Fischer et al. (2016), −6.80 μV. Nonetheless, we agree that measurement decisions, such as what task to use, how to clean the data and what time windows to use for scoring the ERN can affect amplitudes (Sandre et al., 2020) and as such, our norms cannot be assumed to hold for data measured with a different processing pipeline. That is, we established a set of norms to be used for interpreting the ERN in a young adult sample (similar to the samples often used in ERN research), in the context of a particular task and set of processing parameters. Our norms should be used only for data collected within these same parameters, just as norms from one cognitive test are not used to benchmark the results of a different cognitive test.

In closing, we believe that work towards a more standardized ERN processing pipeline can be pursued simultaneously with the goal of understanding how the ERN is distributed in the population. Science is by definition a process of revision, and we are open to the idea that different norms may be published in the future (e.g., using a standardized processing pipeline shared by most ERP researchers). For now, however, when abiding by the limitations reiterated above (i.e., don’t use the norms for individual clinical assessment at this point in time; apply the norms to studies using the same task and processing parameters as well as a similar sample), we believe that our norms stand to increase, not decrease understanding of this widely studied ERP component.

Data and code availability statement

There are no associated data or code.

References

Cavanagh, J.F., Meyer, A., Hajcak, G., 2017. Error-specific cognitive control alterations in generalized anxiety disorder. Biol. Psychiatry: Cogn. Neurosci. Neuroimaging 2 (5), 413–420. doi:10.1016/j.bpsc.2017.07.004.

Clayson, P.E., 2020. Moderators of the internal consistency of error-related negativity scores: a meta-analysis of internal consistency estimates. Psychophysiology 57 (8), e13583. doi:10.1111/psyp.13583.

Clayson, P.E., Kappenberg, E.S., Gehring, W.J., Miller, G.A., & Larson, M.J. (2021). A commentary on establishing norms for error-related brain activity during the arrow Flanker task among young adults. Neuronimaging 213, 116694. doi:10.1016/j.neuroimage.2021.116694.

Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D.S., Quinn, K., Sanislow, C., Wang, P., 2010. Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. Am. J. Psychiatry 167 (7), 748–751. doi:10.1176/appi.ajp.2010091379.

McCarthy, G., Wood, C.C., 1985. Scalp distributions of error-related potentials: an ambiguity associated with an artifact of variance models. Electroencephalogr. Clin. Neurophysiol. 62 (3), 203–208. doi:10.1016/0168-5593(85)90015-2.

Moser, J.S., Moran, T.P., Schroder, H.S., Donnellan, M.B., Yeung, N., 2013. On the relationship between anxiety and error monitoring: a meta-analysis and conceptual framework. Front. Hum. Neurosci. 7, 389. doi:10.3389/fnhum.2013.00366.

Olvet, D.M., Hajcak, G., 2009. The stability of error-related brain activity with increasing trials. Psychophysiology 46 (5), 957–961. doi:10.1111/j.1469-8986.2009.00848.x.

Riesel, A., Endrass, T., Kaufmann, C., Kathmann, N., 2011. Overactive error-related brain activity as a candidate endophenotype for obsessive-compulsive disorder: evidence from unaffected first-degree relatives. Am. J. Psychiatry 168 (3), 317–324. doi:10.1176/appi.ajp.2010030416.

Riesel, A., Weinberg, A., Endrass, T., Meyer, A., Hajcak, G., 2013. The ERN is the ERN? Convergent validity of error-related brain activity across different tasks. Biol. Psychol. 93 (3), 377–385. doi:10.1016/j.biopsycho.2013.04.007.

Sandre, A., Banica, I., Riesel, A., Flase, J., Klawohn, J., Weinberg, A., 2020. Comparing the effects of different methodological decisions on the error-related negativity and its association with behaviour and gender. Int. J. Psychophysiol. 156, 18–39. doi:10.1016/j.ijpsycho.2020.06.016.

Saunders, B., Inzlicht, M., 2020. Assessing and adjusting for publication bias in the relationship between anxiety and the error-related negativity. Int. J. Psychophysiol. Off. J. Int. Organ. Psychophysiol. 155, 87–98. doi:10.1016/j.biopsycho.2020.05.008.

Stern, E.R., Liu, Y., Gehring, W.J., Lister, J.J., Yin, G., Zhang, J., Fitzgerald, K.D., Himle, J.A., Abelson, J.L., Taylor, S.F., 2010. Chronic medication does not affect hyperactive error responses in obsessive-compulsive disorder. Psychopharmacology 207 (5), 913–920. doi:10.1007/s00213-010-2098-x.

Urbach, T.P., Kutas, M., 2002. The intractability of scaling scalp distributions to infer neuroelectric sources. Psychophysiology 39 (6), 791–808. doi:10.1111/1469-8986.34084.

Weinberg, A., Klein, D.N., Hajcak, G., 2012. Increased error-related brain activity distinguishes generalized anxiety disorder with and without comorbid major depressive disorder. J. Abnorm. Psychol. 121, 885–896. doi:10.1037/a0028270.

Weinberg, A., Olvet, D.M., 2010. Increased error-related brain activity in generalized anxiety disorder. Biol. Psychol. 85 (3), 472–480. doi:10.1016/j.biopsycho.2010.09.011.