Letters to the Editor

Large Sample Sizes Cannot Compensate for Mismeasured Environments in Gene-by-Environment Research

TO THE EDITOR: We respect Border et al.'s ambitious undertaking (1), which was published in the May 2019 issue of the *Journal*. We endorse greater scrutiny of novel effects within the literature on gene-by-environment (G×E) interactions, including replication using highly powered methods. We argue, however, that the approach used here to measure the environment—several dichotomized questionnaire items—is insufficient to draw conclusions, despite probing measurement error. Evidence for the importance of the environment for depression is well established (i.e., approximately 63% of liability; [2]), and vulnerability-stress models of depression are widely endorsed and long precede G×E research. However, G×E research has failed to invest in measuring the environment (3); we urge adoption of "E" assessment that matches the rigor of "G" assessment.

Before G×E research, stress interview measures emerged as gold-standard measures among researchers studying stress and depression. Indeed, one 1998 review indicated, "today, use of life event checklists designed to rate the presence or absence of a finite number of events has largely been abandoned" (4, p. 301). With the advent of G×E research, a need for quick measures resurrected questionnaire measurement. Although a full rationale is beyond this letter (5), appropriate measures maximally disambiguate stress exposure from response, account for investigator-rated severity, distinguish interpersonal from noninterpersonal stress, attend to events' depressogenic time-frame (<3 months), and establish temporal precedence.

Why is it insufficient to use a large N and estimate the impact of measurement error? First, even in their "catastrophic" simulations, the authors vastly underestimate the amount of random error introduced by inadequate stress measures. One estimate suggests that questionnaires accounted for only 16% of variance of interview measures (6), even without artificially dichotomizing questionnaires as done here. Would the field tolerate such poor validity measurement of genotypes? Second, large samples address random error but not systematic error. Specifically, the authors' approach does not account for findings that different types of stress confer significantly different unique variance for depression (7), are poor indicators of each other (major interpersonal compared with noninterpersonal events, r=0.04 and r=0.32, respectively; [7, 8]), and can produce G×E effect sizes in opposite directions. For example, in a simultaneous model, a significant interpersonal major event G×E effect was the opposite direction from the trending noninterpersonal major event G×E effect (9). Similarly, a serotonergic multilocus score produced significantly stronger G×E effects for major interpersonal events than other event categories (full results available from the authors).

Further, differential susceptibility theory (10) suggests that some genetic variants—including many in the present study—confer sensitivity to the environment for better and worse, meaning the same genetic variant may have opposite effects depending on environmental conditions. Differential susceptibility renders the effects of many variants uninterpretable without the accurate capture of environmental effects. The field has attempted to address inconsistent findings with ever-growing sample sizes. To the extent that they create the need to rely on inadequate "envirotyping," dazzling sample sizes not only fail to increase the likelihood of the detection of real G×E effects for differential susceptibility variants—they may completely obscure them.

REFERENCES

1. Border R, Johnson EC, Evans LM, et al: No support for historical candidate gene or candidate gene-by-interaction hypotheses for major depression across multiple large samples. Am J Psychiatry 2019; 176:376–387
2. Sullivan PF, Neale MC, Kendler KS: Genetic epidemiology of major depression: review and meta-analysis. Am J Psychiatry 2000; 157: 1552–1562
3. Monroe SM, Reid MW: Gene-environment interactions in depression research: genetic polymorphisms and life-stress polyprocedures. Psychol Sci 2008; 19:947–956
4. Mazure CM: Life stressors as risk factors in depression. Clin Psychol Sci Pract 1998; 5:291–313
5. Harkness KL, Monroe SM: The assessment and measurement of adult life stress: basic premises, operational principles, and design requirements. J Abnorm Psychol 2016; 125:727–745
6. McQuaid JR, Monroe SM, Roberts JR, et al: Toward the standardization of life stress assessment: definitional discrepancies and inconsistencies in methods. Stress Med 1992; 8:47–56
7. Vrshek-Schallhorn S, Stroud CB, Mineka S, et al: Chronic and episodic interpersonal stress as statistically unique predictors of depression in two samples of emerging adults. J Abnorm Psychol 2015; 124:918–932
8. Starr LR, Huang M: HPA-axis multilocus genetic variation moderates associations between environmental stress and depressive symptoms among adolescents. Dev Psychopathol (Epub ahead of print, Nov 5, 2018)
9. Vrshek-Schallhorn S, Mineka S, Zinbarg RE, et al: Refining the candidate environment: interpersonal stress, the serotonin transporter polymorphism, and gene-environment interactions in major depression. Clin Psychol Sci 2014; 2:235–248
10. Belsky J, Pluess M: Beyond diathesis stress: differential susceptibility to environmental influences. Psychol Bull 2009; 135:885–908

Suzanne Vrshek-Schallhorn, Ph.D.
Gail M. Corneau, M.A.
Lisa R. Starr, Ph.D.

Department of Psychology, University of North Carolina at Greensboro, Greensboro (Vrshek-Schallhorn, Corneau); Department of Psychology, University of Rochester, Rochester, N.Y. (Starr).

Send correspondence to Dr. Vrshek-Schallhorn (suzanne.schallhorn@uncg.edu) and Dr. Starr (lisa.starr@rochester.edu).

The authors report no financial relationships with commercial interests.

Accepted May 6, 2019.

Am J Psychiatry 2019; 176:667–668; doi: 10.1176/appi.ajp.2019.19040374

Measurement Error Cannot Account for Failed Replications of Historic Candidate Gene-by-Environment Hypotheses: Response to Vrshek-Schallhorn et al.

TO THE EDITOR: Vrshek-Schallhorn et al. dispute our conclusion that historic candidate gene-by-environment (G×E) hypotheses were incorrect, criticizing the measures of environmental stressors we employed. We appreciate the opportunity to respond to their points.

We agree with their point that measures of environment are important, as we emphasized in our article (1). Our results demonstrate that both interpersonal and noninterpersonal environmental measures influence depression liability (see Table S6 in the online supplement to our article). The additive effects of these environmental measures replicate previous findings that a variety of stressors affect depression liability. However, Vrshek-Schallhorn et al. suggest that the measures we used are completely obscuring real candidate G×E effects. Below, we argue that this cannot be the case and that several of their comments are inaccurate.

They refer to the use of artificially dichotomized questionnaires, implying that we chose arbitrary cutoff points for continuous measures. In actuality, the binary stress measures we examined were inherently dichotomous, indicating whether or not participants endorsed one of a handful of events. For example, exposure to trauma in childhood was coded affirmatively if participants stated that they had been subject to sexual or physical abuse in childhood. On the other hand, we did not dichotomize the Townsend deprivation index, a continuous measure of socioeconomic hardship. In addition, Vrshek-Schallhorn et al. state that our worst-case-scenario measurement error simulations "vastly underestimate the amount of random error introduced by inadequate stress measures," suggesting that a noisy stress questionnaire measure might account for as little as 16% of the variance of its adequately measured analogue. In response, we introduced this degree of error variance in both the depression and stress measures in our simulations. Even with this extreme degree of error, we still observed greater than 90% power to detect even modest interaction effects by candidate gene standards (odds ratio≥1.26).

Systematic error can influence results, as Vrshek-Schallhorn et al. assert. However, the several severe systematic measurement error regimes we examined (see section S4.3.3 in the online supplement to our article) demonstrate that misclassification of environment measures cannot account for the lack of candidate gene or candidate G×E replication. In truth, we are unable to construct a plausible measurement error model that reconciles the validity of previously reported candidate gene findings with our observations that every stressor we examined evidenced substantial, highly significant effects on every depression measure (see section S6 in the online supplement to our article) but that no candidate gene polymorphism or stressor-by-polymorphism interaction had detectable effects, despite ~100% power across a broad array of measurement error scenarios. This is not to say that measurement error is unimportant. However, with respect to the large effects reported in the candidate gene literature in small samples, measurement error cannot account for the lack of support for historical candidate gene hypotheses in our study (1) or in other large, collaborative studies (2) that have investigated the genetic underpinnings of depression, even in carefully phenotyped studies specifically testing genome-wide G×E hypotheses (3). Instead, the most plausible explanation for these failures to replicate is that the original candidate gene findings were false positives.

Vrshek-Schallhorn et al. also state that we failed to distinguish between weakly correlated types of stress that may interact in different ways with genetic variants. The great majority of scenarios wherein either or both of these types of stress interact with a candidate genetic variant should still produce detectable genetic variant main effects—as noted, none were detected (see section S7 in the online supplement to our article). In the unlikely case of a complete crossover interaction where the variant has no main effect, any interactions detectable in candidate G×E samples as small as those cited by our critics would in turn induce detectable differences in variance across genotypes in large samples, even allowing for extreme measurement error. We found no evidence for such heteroscedasticity (results available upon request).

The trajectory of increasing complexity in candidate gene research follows a pattern of reactions to repeated replication failures. In the 1990s, it was hypothesized that specific, common polymorphisms within serotonergic and other neurotransmitter genes would explain substantial variation in depression liability. In the 2000s, it was hypothesized that moderation of genetic effects by environmental stressors would explain inconsistent main effect findings. Now, Vrshek-Schallhorn et al. suggest that our null findings can be explained by different types of stressors that produce interaction effects in opposite directions, or by catastrophic measurement error, and that their hypotheses cannot be adequately tested in existing well-powered samples. We instead suggest that these lines of inquiry are fundamentally flawed; neither the notion that common variants have large