THE JOURNAL OF CHILD PSYCHOLOGY AND PSYCHIATRY
ACAMH THE ASSOCIATION FOR CHILD AND ADOLESCENT MENTAL HEALTH
Journal of Child Psychology and Psychiatry 58:3 (2017), pp 328–330 doi:10.1111/jcpp.12669

Commentary: Fundamental problems with candidate gene-by-environment interaction studies – reflections on Moore and Thoemmes (2016)

Richard Border,[1,2,*] and Matthew C. Keller[1,2,*]
[1]Institute for Behavioral Genetics, University of Colorado at Boulder, Boulder, CO; [2]Department of Psychology & Neuroscience, University of Colorado at Boulder, Boulder, CO, USA

Although there is long-standing agreement that environmental and genetic factors give rise to behavior through a complex interplay, the scientific study of interactions between specific genetic and environmental factors has proven difficult and controversial. In their paper, Moore and Thoemmes (2016) elaborate on one source of difficulty: how different plausible configurations of environmental influences can bias estimates of candidate gene-by-environment interactions (cG × E) when not explicitly modeled. We commend the authors' contributions in this area and observe that their results add to the body of work approaching cG × E assessment from a perspective informed by biometric models (Eaves & Verhulst, 2014). Although it is uncertain how easily investigators will be able to mitigate the sources of bias Moore and Thoemmes (2016) address, their paper highlights the complexity of the how plausible gene-environment covariation structures can affect results obtained by standard cG × E methods, and therefore encourages interpreting such findings with proper circumspection.

The authors' simulations, particularly their simulations on the common environment and evocative mediator conditions, can be viewed as instances of a broader methodological issue affecting analysis of interaction effects in general: nonlinear relationships between a predictor and an outcome variable can masquerade as interactions involving that predictor or variables related to it, even when no interactions exist (Lubinski & Humphreys, 1990). The authors' models suggest that this challenge is of acute relevance to developmental researchers, as, for instance, unmeasured environmental factors common to both parent and child have the potential to induce such nonlinear associations. Consider the following hypothetical (and likely oversimplified) example: A researcher is interested in investigating whether a particular genetic polymorphism interacts with quality of parental supervision to predict conduct problems. Although unrelated to the investigator's primary hypotheses, socioeconomic status (SES) impacts both a child's exposure to delinquent peers (perhaps as a result of living in a low income neighborhood) and parents' ability to monitor the whereabouts of their children (possibly due to decreased presence at home related to economic challenges). Supposing compromised parental monitoring exacerbates delinquent peer influence – that is, the relationship between exposure to delinquent peers and future conduct problems is moderated by quality of parental supervision – a nonlinear relationship between SES and conduct problems results. Both exposure to delinquent peers and parental supervision partially reflect SES, so their product becomes a higher order function of SES. When such a nonlinear association between SES and conduct disorder is not modeled, *any* interaction term involving SES or a variable related to it (such as parental supervision or delinquent peer exposure), regardless of whether it is testing a cG × E effect, can take on the some of the variance that would otherwise be accounted for by the nonlinear relationship between SES and conduct problems. In the presence of a true cG × E effects, estimates can become positively biased, and in the absence of any true cG × E effect, the cG × E term is more likely to attain statistical significance than expected under the null (the type-I error rate is inflated).

The issues raised by the authors are important because they focus attention on the inherent complexity surrounding analysis of interactions, but we believe these issues need to be viewed as part of a larger constellation of practical and methodological difficulties surrounding cG × E. In our opinion, the enthusiasm for cG × E effects has outpaced an understanding of the challenges facing such investigations. This has resulted in a controversial scientific literature marred by likely false positive results and a fracture between cG × E research – and candidate gene research more generally – and mainstream human genetics (Duncan, Pollastri, & Smoller, 2014). In this commentary, we hope to contextualize the problems raised by the authors within this broader array of difficulties facing cG × E investigation. Much has been written about these topics already and our goal is not to be comprehensive. Rather, our purpose is to argue that even if investigators were able to account for the sources of bias Moore and Thoemmes (2016) elaborate, it is unlikely that conventional gene-by-environment (G × E) approaches would yield reliable results.

*Joint first authorship.

The great majority of published cG × E findings rely on a variation of the following statistical model: in the context of a generalized linear model, the outcome of interest is regressed on a function of a linear combination of genotype, a measure of environment, their product, and any additional covariates. Despite the apparent simplicity of such an approach, there are many potential statistical pitfalls, several of which Moore and Thoemmes (2016) noted in their manuscript. We highlight three here. First, the parameter estimates associated with product terms depend on the measurement scales of the outcome/environmental variable. Transformations between additive and multiplicative scale (e.g. logarithmic scales), which are often implicit in the selection of link function, can generate or eliminate evidence for multiplicative interaction terms. Techniques for properly accounting for measurement models in cG × E research are in their infancy, and publications reporting cG × E effects rarely address this issue (Eaves & Verhulst, 2014). Second, controlling for the additive effects of covariates does not control for their effects on interactions terms; rather one must also include covariate-gene and covariate-environment product terms, a crucial issue that has been almost wholly neglected in the cG × E literature to date (Keller, 2014). Third, in the presence of nonlinear relationships between predictors and the outcome of interest, failure to model these relationships (e.g. through the inclusion of higher order predictors) can generate spurious multiplicative effects (Lubinski & Humphreys, 1990). As discussed above, Moore and Thoemmes' (2016) results highlight that this source of statistical artifacts might be particularly relevant to developmental research, where synergistic interactions between child and caregiver behaviors over time can effectively induce nonlinear association between parenting behaviors and child outcomes.

Beyond the statistical problems inherent to analysis of interactions, perhaps the most salient issue with cG × E research conducted on behavioral traits to date is that it has relied on a candidate gene approach that has yielded very little in terms of scientific progress. With the exception of relatively simple and biologically well-understood systems (e.g. drug metabolism and pharmacokinetics), it was, in hindsight at least, perhaps overly optimistic to believe that scientists could successfully guess which of tens of millions of common polymorphisms would be related to behavioral outcomes given our inchoate understanding of the underlying biological mechanisms. Indeed, with few exceptions (e.g. nicotine dependence, alcohol metabolism), none of the most studied candidate polymorphisms show evidence of being associated at levels above chance with behavioral outcomes in large genome-wide association studies (GWAS; see, e.g. Farrell et al., 2015). These 'historical' candidate gene polymorphisms (e.g. *5-HTTLPR*, *COMT Val/Met*, etc.) were originally hypothesized to have main effects on complex behavioral outcomes, but have since been coopted to investigate G × E hypotheses. In our opinion, there is no compelling reason to believe that these same polymorphisms should be related to environmental sensitivity to a behavioral outcome, the complex 'trait' under investigation in cG × E studies, although these same polymorphisms comprise the majority of polymorphisms investigated in such studies to date. Nor, in our opinion, is there compelling reason to believe that the odds ratios of any common polymorphism, much less the few continually investigated in most candidate gene research, will have odds ratios much bigger than 1.3, about the largest effect sizes seen for behavioral outcomes in GWAS. Thus, even if one of these polymorphisms truly affected environmental sensitivity to a specific behavioral outcome, previous cG × E studies have been woefully underpowered to detect the effect in light of our current understanding of genetic effect sizes on complex traits and low statistical power for detecting interactions (Duncan & Keller, 2011).

The above level of skepticism seems at odds with a literature replete with reports of statistically significant cG × E findings accounting for, by modern genetic standards, exceptionally large (e.g. >1%) amounts of variation. However, the two factors detailed above – low prior probability of cG × E hypotheses being correct and low statistical power to detect them – would necessarily lead to high false-positive rates. This occurs even in the absence of publication bias, but to the degree this occurs, it would further exacerbate the proportion of false cG × E findings. The fact that almost all first reports of a particular cG × E finding are statistically significant, that larger replication attempts are less, not more, likely to replicate the original cG × E finding, and that positive cG × E findings appear to require smaller sample sizes than negative findings to be deemed publishable, are all consistent with the possibility of widespread publication bias (Duncan & Keller, 2011). Thus, as scientists have learned from candidate gene main effect findings, results in the published literature can be misleading. Highly powered genome-wide interaction studies have been proposed but are not yet widely employed; it would be surprising if the performance of previously studied candidate gene polymorphisms fare any better in tests of interactions than they have in tests of main effects.

In summary, we believe that the pitfalls in cG × E research are numerous: published cG × E investigations to date have generally employed inadequate analytic procedures, have relied on samples orders of magnitude too small to detect plausible effects, and have relied on a particular candidate gene approach that has been unfruitful and largely jettisoned in mainstream genetic analyses of complex traits. It is unlikely that utilizing these procedures in

developmental studies will overcome these shortcomings. Given this, we make the following recommendations for studying GxE effects:

- History has shown that small exploratory studies have typically had limited utility in psychiatric genetics. Investigators interested in the exploration of cG × E hypotheses should work to incorporate their data into large consortia and collaboratively select their measures with this goal in mind.
- Investigators should publicly preregister the measures and analytic procedures they will use, be adequately powered to detect effect sizes typical of complex traits, and seek to mitigate statistical artifacts to the greatest extent possible.
- Above all, there is no longer a justifiable reason for restricting GxE investigations to one or a few of the most studied, and now largely abandoned, candidate gene polymorphisms. Genome-wide arrays are now as cheap as custom arrays, and the ability to impute genome-wide arrays to a very high marker density means that investigators can take an unbiased look at the moderating effects of millions of polymorphisms across the genome, including the vast majority of historical candidate polymorphisms. In cases where in-depth analysis of candidate polymorphisms is deemed necessary, the choice of the candidates should be based on compelling statistical evidence from well-powered genome-wide studies, not on historical precedent.
- Given the general consensus that genetic effects on complex traits (and probably therefore genetic effects on environmental sensitivity) are likely to be small, it might be more productive for researchers to focus on conglomerate effects of genes rather than individual ones. For example, GREML (genome-wide restricted maximum likelihood) estimation implemented in, for example, the GCTA software (Yang, Lee, Goddard, & Visscher, 2011) can test whether there is evidence for environmental moderation of the type discussed above across all SNPs genome-wide, taken together. GREML, as well as polygenic risk score approaches, can also test a related but distinct form of genetic moderation, in which the overall heritability of a trait changes across the levels of an environmental moderator. Neither approach is free from limitations, but they do not suffer from the problems associated with the low prior probability of choosing relevant genetic polymorphisms or the likely small effect sizes associated with individual polymorphisms.

We are in the midst of an exciting era – investigators now have the ability to cost-effectively examine individual differences at the molecular level across the genome. This new era demands new practices; analytic procedures must evolve to meet the challenges that the genetic architecture of complex traits presents, and investigators must collaborate on grander scales if we hope to begin to understand how specific genes and environments combine to affect behavior.

Acknowledgements

This work was supported by NIMH grant T32MH016880 (PI Hewitt) to R.B. and MH100141 to M.C.K. The authors are grateful to Scott Vrieze, John Hewitt, and Marissa Ehringer for helpful comments on drafts of this manuscript. This commentary was invited by the Editors of JCPP and has been the subject of internal review. The authors have declared that they have no potential or competing conflicts of interest in relation to this article.

Correspondence

Matthew C. Keller and Richard Border, Institute for Behavioral Genetics, University of Colorado at Boulder, Boulder, CO 80303, USA; Email: matthew.c.keller@gmail.com; Richard.Border@colorado.edu

References

Duncan, L.E., & Keller, M.C. (2011). A critical review of the first 10 years of candidate gene-by-environment interaction research in psychiatry. *American Journal of Psychiatry*, *168*, 1041–1049.

Duncan, L.E., Pollastri, A.R., & Smoller, J.W. (2014). Mind the gap: Why many geneticists and psychological scientists have discrepant views about gene-environment interaction (GxE) research. *American Psychologist*, *69*, 249–268.

Eaves, L., & Verhulst, B. (2014). Problems and pit-falls in testing for G × E and epistasis in candidate gene studies of human behavior. *Behavior Genetics*, *44*, 578–590.

Farrell, M.S., Werge, T., Sklar, P., Owen, M.J., Ophoff, R.A., O'Donovan, M.C., … & Sullivan, P. F. (2015). Evaluating historical candidate genes for schizophrenia. *Molecular Psychiatry*, *20*, 555–562.

Keller, M.C. (2014). Gene-by-environment interaction studies have not properly controlled for potential confounders: The problem and the (simple) solution. *Biological Psychiatry*, *75*, 18–24.

Lubinski, D., & Humphreys, L.G. (1990). Assessing spurious "moderator effects": Illustrated substantively with the hypothesized ("synergistic") relation between spatial and mathematical ability. *Psychological Bulletin*, *107*, 385–393.

Moore, S.R., & Thoemmes, F. (2016). What is the biological reality of gene–environment interaction estimates? An assessment of bias in developmental models. *Journal of Child Psychology and Psychiatry*, *57*, 1258–1267.

Yang, J., Lee, S.H., Goddard, M.E., & Visscher, P.M. (2011). GCTA: A tool for genome-wide complex trait analysis. *American Journal of Human Genetics*, *88*, 76–82.

Accepted for publication: 18 October 2016