COMMENTARY

Potential Bias in Meta-Analyses of Effect Sizes in Imaging Genetics

Frieder M. Paulus[*,1], **Sören Krach**[1,2], **Anne-Grit Albrecht**[3], and **Andreas Jansen**[1]

[1]Department of Psychiatry and Psychotherapy, Philipps-University Marburg, Rudolf-Bultmann-Straße 8, 35033 Marburg, Germany; [2]Department of Neurology, Philipps-University Marburg, Rudolf-Bultmann-Straße 8, 35033 Marburg, Germany; [3]Department of Psychology, University of Mannheim, Im Schloss EO 233, 68131 Mannheim, Germany

*To whom correspondence should be addressed; Rudolf-Bultmann-Str. 8, 35033 Marburg, Germany; tel: +49-(0)6421-58 65204, fax: +49-(0)6421-58 68939, e-mail: paulusf@med.uni-marburg.de

The penetrance of genetic variation has been assumed to be higher at the level of neural phenotypes than at the level of behavioral phenotypes. One of the few attempts to validate this assumption is the study of Rose and Donohoe published in this issue. In this article, we will address 2 methodological issues we believe have to be considered for a better understanding of the present results. We briefly discuss potential solutions that might also help improve future meta-analyses of effect sizes in neuroimaging data.

Key words: meta-analysis/schizophrenia/effect sizes/ imaging genetics

Introduction

The penetrance of genetic variation has been assumed to be higher at the level of neural phenotypes than at the level of behavioral phenotypes.[1] One of the few attempts to validate this assumption is the highly relevant study of Rose and Donohoe published in this issue.[2] Here, the authors conducted meta-analyses of effect sizes to test whether neural phenotypes of a variety of schizophrenia susceptibility loci are associated with a greater population effect than behavioral phenotypes. In this article, we address 2 methodological issues we believe have to be considered in meta-analyses of neuroimaging data for a better understanding of the results. First, we show that the extraction of effect sizes is a general challenge inherent to meta-analyses of neuroimaging data entailing the potential to overestimate effects. Second, we argue that using absolute values for coding effect sizes in primary studies also bias the results, which hamper their interpretation. We conclude with a brief discussion of potential solutions.

Effect Sizes in Neuroimaging Studies

Meta-analyses require valid estimates of the observed effect sizes in primary studies. (Notably, meta-analyses fundamentally require the included effect sizes of primary studies to be independent. Especially for neuroimaging studies, the potential to accidentally include effect sizes extracted from the exact similar sample increases because samples get repeatedly analyzed and separately published with regards to different task domains and imaging modalities. Based on a brief review of the included studies, there seems to be a considerable amount of effects that might have been derived from similar or overlapping samples but are nonetheless considered as independent observations). These estimates can be derived by conversions of *t*, *z*, or *p* statistics provided in primary studies. The extraction of effect sizes from published neuroimaging studies, however, is hampered by the current practice of presenting results. In imaging genetics, specifically for those studies that use massive univariate approaches such as fMRI, the majority of published statistics represent effects that survived stringent alpha correction procedures. For meta-analyses, this is problematic insofar because the post hoc contrasts used to threshold the imaging data are nonindependent to those that are used to report the effect sizes. To illustrate this issue, current practices require first thresholding the data with a specific contrast (eg, *ZNF804A* rs1344706 AA < AC < CC) and afterward reporting, eg, *t*-values of the exact similar contrast (ie, AA < AC < CC) for the surviving clusters or peak voxels. These estimates thus reflect the effect of a gene within the sample with an inherent noise component that gets fitted in the same direction as the effect of interest. Consequently, it has been acknowledged that effects selected by nonindependent thresholding procedures overestimate the effect sizes in neuroimaging studies.[3]

In the Rose and Donohoe study, the authors extracted effects from those clusters that showed the strongest effect within a set of clusters surviving alpha correction. Notably, all estimates within these sets of clusters are affected by nonindependence bias and overestimate the effect sizes within the primary study. This bias becomes even more accentuated when selecting only the strongest effect. Furthermore, magnetic resonance imaging effect sizes usually represent the effect at the peak voxel within a cluster. This increases noise fitting and typically leads to overestimated effects for neural phenotypes. If at all, these problems do not impact behavioral phenotypes to a similar degree because of the much lower dimensionality of behavioral data. Consequently, the present comparison between neural and behavioral phenotypes is potentially biased toward stronger effect sizes for the neural phenotypes.

Coding of Effect Sizes

When sampling data out of a population, variability of observed effect sizes is an expected phenomenon. The variability of the observed effects in primary studies depends on the errors made by sampling the population. It is therefore perfectly reasonable to observe negative effects if the population effect is small or the sampling error relatively large. Aggregating primary studies in meta-analyses needs to take into account the heterogeneity of effect sizes by recoding observed effects so that their signs convey similar meaning (eg, AA < AC < CC in case of *ZNF804A* rs1344706) to obtain valid population estimates.[4] By using absolute values for coding effect sizes in primary studies, however, meta-analyses overestimate the magnitude of the underlying population effect. In a hypothetical scenario with the population effect being zero, meta-analyses of absolute values estimate a positive effect in the population and increasingly so with greater sampling error. Moreover, confidence intervals are bound to be smaller because the variability of observed effects is reduced. Additionally, analyses of publication bias will get corrupted because the distribution of observed absolute effect sizes turns out to be asymmetric.

These issues raise concerns toward using absolute values of effect sizes in meta-analyses per se; nonetheless, it is reasonable to assume heterogeneous population effects across different genetic risk variants and phenotypes.[5] However, using absolute values to compensate the heterogeneity of signs of the underlying population effects also has consequences for comparing the absolute magnitude of population effects across domains, as might be illustrated with a simplistic example: Consider 2 sets of 100 primary studies, each having similar effect size and sampling error. Further assume that within each set, 50 show a positive and 50 a negative effect. In

such a case meta-analytical aggregation of effect sizes may estimate different absolute magnitudes of population effects depending on the nature of the primary studies. First, if all 100 effect sizes represent the effect of the same genetic variant on the same phenotype (eg, *ZNF804A* rs1344706 effects on brain volume), the meta-analytic estimate of the population effect would be zero. Second, if the 50 positive effects represent a qualitatively different phenotype compared with the 50 negative ones (eg, rs1344706 effects on intelligence vs neuroticism), meta-analyses would be able to identify the differences in the directionality of the underlying population effects and indicate an effect of the gene in absolute terms. For both sets of studies, meta-analyses of absolute values (ie, 100 positive effects) would estimate an incorrectly high and also similar effect even if the true underlying population effects are entirely different. Simply put, using absolute values for coding effect sizes of primary studies on different genes and phenotypes irretrievably confounds heterogeneity of signs due to sampling error with differences in the underlying population effects. Thus, the present comparison of population means for brain and behavioral phenotypes should not stand without careful reconsiderations on the included genes and exact phenotypes in each domain.

Conclusion

Practices in publishing results in imaging genetics are designed to verify the presence of an effect and avoid false-positive findings. As we have pointed out, the results of such analyses should not be used for meta-analyses of effect sizes without further considerations. Doing so will overestimate population effects and, in particular, bias the comparison between brain and behavioral phenotypes. Publishing effect size maps for whole-brain volumes in primary studies could help to achieve population estimates less biased by the inclusion of effect sizes that sample idiosyncratic features of the primary data.[6,7]

Moreover, in order to estimate the absolute magnitude of effects, meta-analyses need to correctly deal with heterogeneous effects obtained in primary studies. The usage of absolute values not only corrupts nearly any aspect of the analyses but also differentially affects study domains as a function of sampling error, the underlying population effect size, and qualitative differences in the sets of primary studies. If one is interested in estimating the absolute magnitude of effect sizes, one could run separate meta-analyses for each risk variant and phenotype and compare population means and confidence intervals.[5,8] This, however, would require a sufficient amount of primary studies and repeated efforts to replicate findings for each single gene and phenotype.

Acknowledgment

References

1. Meyer-Lindenberg A. From maps to mechanisms through neuroimaging of schizophrenia. *Nature*. 2010;468:194–202.
2. Rose EJ, Donohoe G. Brain vs behavior: an effect size comparison of neuroimaging and cognitive studies of genetic risk for schizophrenia [published online ahead of print April 12, 2012]. *Schizophr Bull*. doi:10.1093/schbul/sbs056.
3. Kriegeskorte N, Lindquist MA, Nichols TE, Poldrack RA, Vul E. Everything you never wanted to know about circular analysis, but were afraid to ask. *J Cereb Blood Flow Metab*. 2010;30:1551–1557.
4. Hunter JE, Schmidt FL. *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings*. Newbury Park, CA: Sage Publications; 2004:616.
5. Mier D, Kirsch P, Meyer-Lindenberg A. Neural substrates of pleiotropic action of genetic variation in COMT: a meta-analysis. *Mol Psychiatry*. 2010;15:918–927.
6. Paulus FM, Krach S, Bedenbender J, et al. Partial support for ZNF804A genotype-dependent alterations in prefrontal connectivity. *Hum Brain Mapp*. 2013;34:304–313.
7. Poldrack RA, Fletcher PC, Henson RN, Worsley KJ, Brett M, Nichols TE. Guidelines for reporting an fMRI study. *Neuroimage*. 2008;40:409–414.
8. Munafò MR, Brown SM, Hariri AR. Serotonin transporter (5-HTTLPR) genotype and amygdala activation: a meta-analysis. *Biol Psychiatry*. 2008;63:852–857.