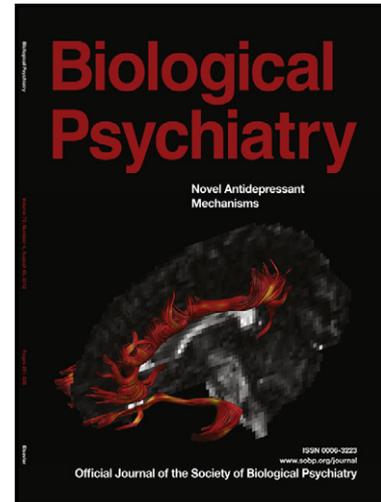


# Author's Accepted Manuscript

Appropriate approaches to summarizing GxE studies: Reply to Laoutidis and Luckhous

Amy L. Byrd MS, Aidan G.C. Wright PhD, Srijan Sen MD, PhD, Kerby Shedden PhD, Stephen B. Manuck PhD



[www.elsevier.com/locate/bps](http://www.elsevier.com/locate/bps)

PII: S0006-3223(14)00103-6  
DOI: <http://dx.doi.org/10.1016/j.biopsych.2014.02.009>  
Reference: BPS12143

To appear in: *Biological Psychiatry*

Cite this article as: Amy L. Byrd MS, Aidan G.C. Wright PhD, Srijan Sen MD, PhD, Kerby Shedden PhD, Stephen B. Manuck PhD, Appropriate approaches to summarizing GxE studies: Reply to Laoutidis and Luckhous, *Biological Psychiatry*, <http://dx.doi.org/10.1016/j.biopsych.2014.02.009>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting galley proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Appropriate approaches to summarizing GxE studies:

Reply to Laoutidis and Luckhous

Amy L. Byrd, MS<sup>1</sup>, Aidan G. C. Wright, PhD<sup>1</sup>, Srijan Sen, MD, PhD<sup>2</sup>, Kerby Shedden, PhD<sup>3</sup> & Stephen B. Manuck, PhD<sup>1</sup>

<sup>1</sup>Department of Psychology, University of Pittsburgh, Pittsburgh, PA

<sup>2</sup>Department of Psychiatry, University of Michigan, Ann Arbor, MI

<sup>3</sup>Department of Statistics, University of Michigan, Ann Arbor, MI

Corresponding Author:

Amy Byrd, M.S.  
Department of Psychology  
University of Pittsburgh  
Sterling Plaza, Suite 408  
201 N. Craig St.  
Pittsburgh, PA 15213  
Phone: 412-383-5032  
alb202@pitt.edu

Word count: 993

Figures: 0

Tables: 0

Supplemental Information: 0

Science is a cumulative enterprise. As more studies accumulate, it is important to integrate them, and meta-analysis is one approach to doing so. But what is the best way to conduct a meta-analysis? The commentary by Laoutidis and Luckhaus suggests limitations of two recent meta-analyses, one reported in this journal by Byrd and Manuck (1) and the second published elsewhere by Karg et al (2). These focus, respectively, on two reports of gene-environment (GxE) interactions, published 2002 and 2003(3-4), that have spawned over 80 replication attempts. The earlier of the two meta-analyses targeted the interaction of life stress and polymorphic variation in the serotonin transporter gene on risk for depression, and the second, the interaction of early maltreatment and analogous variation in monoamine oxidase A (*MAOA*) on later antisocial behaviors. Laoutidis and Luckhaus critique our use of the Liptak-Stouffer weighted Z-test (LST) for meta-analysis, as opposed to alternative methods involving aggregated effect sizes, and suggest any conclusions permitted by the LST are limited. Here, we show why use of the LST is appropriate, and authors from our two groups have joined in this reply.

Averaging effect sizes over multiple studies testing the same hypothesis can indeed provide meaningful and practical outcomes. For instance, if a medication's salutary effects on a consensually measured disorder (e.g., hypertension) have been tested in multiple trials of the same design, meta-analysis may yield a clinically useful estimate of therapeutic benefit. By using more information than pooled p-values, a meta-analysis predicated on effect sizes can answer to both the magnitude and reliability of observed outcomes. Still, more information does not vouchsafe a more informative conclusion, and interpretability will erode with loss of comparability among studies, whether in method or design, sampling frame, equivalence of independent and dependent variables, or fidelity of measurement. When literatures testing a common hypothesis encompass studies that vary across *all* of these dimensions, measuring different things in different ways and in different populations, a mean effect size loses meaning as a quantitative estimate of association or outcome (5).

Using the serotonin transporter/life stress literature as example, Monroe and Reid (6) illustrated substantial variability in life events assessments across just 13 studies. These varied in their ability to distinguish even acute from chronic stressors or major from minor life events; differed widely by intervals of exposure (a few months to even lifetime experience); used instruments of diverse format, content, and reliability; and all employed different procedures for quantitating participants' total event exposures. Beyond that, the actual stressors encountered vary tremendously, from financial problems and interpersonal difficulties to incident disease. In the *MAOA*/childhood adversity literature, too, indicators range widely, from maternal prenatal smoking and socioeconomic disadvantage to assault and sexual abuse, and even within the category of maltreatment, rates of exposure vary across samples. Indeed, the diversity of environmental moderators in these two GxE literatures might appear so large as to render their common elements almost an abstraction. Then also, studies may differ in outcomes and study design. In Byrd and Manuck (1), behavioral outcomes encompassed diagnostic evaluations, forensic status, and informant and self-rated antisocial acts. Studies also ranged from longitudinal investigations to opportunistic retrospective and archival reports, and early environmental exposures were variably assessed by parent- and self-report, observation, and official record. Against this immense heterogeneity of method and measurement, where each study is nearly *sui generis*, accurate or meaningful estimation of combined effect size is not

feasible. Thus, we preferred the LST approach to pooled effects meta-analysis because it is better scaled to the answerable question and less vulnerable to inferences of illusory precision.

Laoutidis and Luckhous additionally point to ongoing discussion in the statistical community bearing on interpretation of the LST. While framed as a debate about whether a significant meta-analytic outcome only implies that at least one included study reflects a “real” association versus the “average association being real,” it is not nearly as distinct as they convey. As a practical matter, a pooled p-value can prove significant by the LST even when all individual studies are not, and a pooled p-value can be non-significant even when one or more of the individual values are significant.<sup>1</sup> Also, this technique mirrors findings by other methods. For example, the childhood maltreatment x *MAOA* interaction is confirmed by LST when applied to the same 8 studies included in an earlier meta-analysis that used a variant of the Hedges-Olkin method (7) ( $z_w = 4.02$ ;  $p = 0.0006$ ). And Karg et al. (2) replicated an *absence* of interaction between stressful life events and serotonin transporter genotype when applying the LST to only those studies examined in two previous, negative meta-analyses (8)(9).

On the more technical issue, methods of combining p-values, including the LST, do operate under a null hypothesis that all included effects are simultaneously non-significant, suggesting that at minimum a single positive finding could reject the null. In practice, different computational methods for combining p-values differ in their sensitivity to patterns of significant and non-significant findings that might cause the null hypothesis to be rejected. While some techniques are indeed sensitive to a single significant value, others achieve global significance optimally when considerably more of the individual findings are significant (10-11). In this regard, the LST performs best when evidence against the null hypothesis is *not* limited to a single small p-value, but spans more than a small fraction of the combined values (10-11). This is exactly the pattern seen in both of our reviews and is further supported by sensitivity analyses showing no individual studies to disproportionately influence aggregate findings. This point is buttressed further by the several stratified analyses reported in each review, all showing significant pooled effects across multiply segmented subsets of the included studies.

In sum, different methods of combining data have utility in different situations. With the complex and heterogeneous sets of studies addressing the GxE hypotheses tested in our two meta-analyses, we believe the LST approach is an appropriate choice and should be considered in future meta-analyses of similarly comprised literatures.

**Acknowledgments:** None

**Conflicts of Interest:** The authors report no biomedical financial interests or potential conflicts of interest.

#### FOOTNOTE

1. As hypothetical examples, consider respectively: a) 5 studies, each with  $N=500$  and 1-tailed  $p=0.08$ ; for this set,  $z_w = 3.13$ ,  $p=0.0017$ ; and b) again 5 studies of  $N=500$  each, where 4 studies have a 1-tailed  $p=0.45$  and the fifth a 1-tailed  $p=0.01$ ; here  $z_w = 1.27$  and is not significant ( $p=0.20$ ).

#### References

1. Byrd AL, Manuck SB (2014): MAOA, Childhood Maltreatment, and Antisocial Behavior: Meta-analysis of a Gene-Environment Interaction. *Biological Psychiatry*. 75:9-17.
2. Karg K, Burmeister M, Shedden K, Sen S (2011): The serotonin transporter promoter variant (5-HTTLPR), stress, and depression meta-analysis revisited: Evidence of genetic moderation. *Archives of General Psychiatry*. 68:444-454.
3. Caspi A, McClay J, Moffitt TE, Mill J, Martin J, Craig IW, et al. (2002): Role of genotype in the cycle of violence in maltreated children. *Science*. 297:851-854.
4. Caspi A, Sugden K, Moffitt TE, Taylor A, Craig IW, Harrington H, et al. (2003): Influence of Life Stress on Depression: Moderation by a Polymorphism in the 5-HTT Gene. *Science*. 301:386-389.
5. Manuck SB, McCaffery JM (2014): Gene-environment interaction. *Annual Review of Psychology*. 65:41-70.
6. Monroe SM, Reid MW (2008): Gene-Environment Interactions in Depression Research Genetic Polymorphisms and Life-Stress Polyprocedures. *Psychological Science*. 19:947-956.
7. Taylor A, Kim-Cohen J (2007): Meta-analysis of gene-environment interactions in developmental psychopathology. *Dev Psychopathol*. 19:1029-1037.
8. Munafò MR, Durrant C, Lewis G, Flint J (2009): Gene  $\times$  environment interactions at the serotonin transporter locus. *Biological Psychiatry*. 65:211-219.
9. Risch N, Herrell R, Lehner T, Liang KY, Eaves L, Hoh J, et al. (2009): Interaction between the serotonin transporter gene (5-HTTLPR), stressful life events, and risk of depression: A meta-analysis. *JAMA: The Journal of the American Medical Association*. 301:2462-2471.
10. Loughin TM (2004): A systematic comparison of methods for combining p-values from independent tests. *Computational Statistics & Data Analysis*. 47:467-485.
11. Owen AB (2009): Karl Pearson's meta-analysis revisited. *The Annals of Statistics*. 3867-3892.