Psychological Medicine (2017), **47**, 183–185.

Correspondence

Psychological Medicine, **47** (2017).
doi:10.1017/S003329171600218X
First published online 13 September 2016

Letter to the Editor

Moving science forward by increasing awareness of reporting and citation biases: a reply to Vrshek-Schallhorn *et al.* (2016)

Vrshek-Schallhorn *et al*. (2016) call for a renewed 'positive focus' on the goal of enhancing the prediction and treatment of depression. We share this goal and believe that our work examining positive focus and citation bias within the 5-HTTLPR × life stress literature (de Vries *et al.* 2016) and other work examining issues such as publication bias, analytical flexibility, and statistical power, far from being detrimental to the achievement of this goal, are an essential part of achieving it. Naturally, as with any other study, the study of bias needs to be as unbiased and transparent as possible. We have been fully transparent about our methodological choices: the article includes a supplemental table with the coding for all included papers, and the dataset, including both coding and citation frequencies, is also publicly available online (doi:10.5523/bris.z7jconxfbmdr1jj3t0w4k1hwn) to enable the reader to reach their own decisions. We are pleased that Vrshek-Schallhorn *et al.* have taken the time and effort to examine this material. However, we regret that our message, regarding the need to value negative results as highly as we do positive results, seems to be undermined by what we believe to be unjustified concerns about the rigor of our coding.

Outcome coding in observational studies

All studies inevitably involve a variety of subjective analytical choices that can affect which results are obtained, including those studies commonly regarded as objective, such as meta-analyses (Taylor & Munafò, 2016). In our study, a particularly important choice was how to handle papers that reported multiple relevant outcomes, some of which were statistically significant and some of which were not. We chose to follow previous meta-analyses by averaging the p values associated with these outcomes (Karg *et al.* 2011; Sharpley *et al.* 2014).

We agree with Vrshek-Schallhorn *et al.* that the method of averaging p values can be biased towards finding non-significant averaged p values, for instance in their example of a hypothetical paper with three findings at the $p = 0.001$ level and one finding at the $p = 0.300$ level. In reality, however, 11 out of 12 studies that provided multiple outcomes only had one statistically significant finding (and between one and six non-significant findings), and only four out of 12 studies had statistically significant findings with a p value smaller than 0.01. The alternative would be to select one of these p values for inclusion, which in our opinion is more open to conscious or unconscious subjective bias. We therefore believe our approach to be the least biased of the available options.

Vrshek-Schallhorn *et al.* consider this outcome coding approach 'atheoretical' rather than 'unbiased'. In principle we are comfortable with this alternative phrasing. However, we believe that the nature of observational studies, in which the original hypothesis and the originally intended outcome or predictor variables are rarely prospectively registered, dictates such an approach. Outcome reporting bias (changing primary outcomes or analyses based on the statistical significance of the results) is likely to be at least as common in observational studies as it is in the clinical trials literature (Dwan *et al.* 2013), but in observational studies it is impossible to detect due to the lack of prospective registration. In addition, the majority of reported gene–environment interaction (GxE) studies themselves described no clear *a priori* reason to prefer one outcome over another. Therefore, a theoretical coding approach led by each article's reported hypotheses is, in our opinion, impossible.

To ensure that our approach did not bias our results, we included a sensitivity analysis based on articles' smallest p values. In this lenient analysis, the prevalence of (partially) positive focus decreased somewhat (from 58 to 43%), but it was still common. Citation patterns did not change markedly. We believe this is a more viable alternative approach than coding articles with multiple diverging outcomes as mixed, as suggested. One of the difficulties of GxE research is the many possible combinations of outcomes, environmental measures, and genetic models, which almost inevitably leads to mixed findings and 'approximate replications' (Sullivan, 2007; Kapur *et al.* 2012). A recent study showed, for example, that 720 different 5-HTTLPR × life stress interactions could be tested in an epidemiological cohort, a number that increased to 2160 if subgroup analyses based on gender were permitted (Heininga *et al.* 2015), although we

acknowledge that this much multiplicity will not be possible in all GxE studies.

The concept of positive focus

Vrshek-Schallhorn and colleagues argue that our coding of abstracts was biased, as many authors did report mixed findings in the results section of the abstract. In stark contrast to our study, Vrshek-Schallhorn *et al.* report finding very little evidence of positive focus, as results reported in an abstract were usually consistent with the results reported in the body of an article. This is, however, not what positive focus entails. We based our definition of positive focus in observational studies on previous work regarding spin in clinical trials, where it is defined as 'specific reporting that could distort the interpretation of results and mislead readers' (Boutron *et al.* 2010). In their seminal work, Boutron *et al.* identified various strategies of spin, including 'acknowledge statistically non-significant results for the primary outcome but emphasize the beneficial effect of treatment' and 'acknowledge statistically nonsignificant results for the primary outcome but emphasize other statistically significant results'. Subsequent work showed that specific wording in an abstract can affect the interpretation of a trial and its results by clinicians, even when non-significant results are reported in the abstract and when spin is fairly mild (Boutron *et al.* 2014).

The conceptual difference with the coding by Vrshek-Schallhorn *et al.* is illustrated by the study of Eley *et al.* (2004),in which the 5-HTTLPR × stress interaction was non-significant in the full sample and only nominally significant ($p = 0.03$) in a female subgroup. The abstract describes these results as follows: 'In addition, there was a trend for an effect of 5-HTTLPR, which was significant in female subjects. Furthermore, there was a significant genotype-environmental risk interaction for 5-HTTLPR in female subjects only.' Vrshek-Schallhorn *et al.* conclude there is no positive focus, since the results in the abstract are consistent with the results in the article. However, the abstract emphasizes a non-significant trend in the full sample and a significant outcome in a subgroup analysis, which would qualify as 'spin' according to the definition by (Boutron *et al.* 2014). Moreover, the abstract continues with a claim that the results support the importance of the 5-HTTLPR × stress interaction: 'the effect being in the same direction as another recent study, reaffirming that an important source of genetic heterogeneity is exposure to environmental risk'. This positive focus is propagated to other studies within the network: 12 out of 36 studies referencing this study cite it as unconditionally positive, while the remaining 24 cite it as partially positive. No study cited it as negative, even though it clearly did not directly replicate the 5-HTTLPR × stress interaction.

The value of negative results

Our study is not a meta-analysis and was not intended to settle the question of whether 5-HTTLPR moderates the association between life stress and the development of depression. Instead, we examined whether processes that can distort the apparent strength of the evidence-base may be operating. We concluded that discussion of the 5-HTTLPR × stress interaction is more positive than warranted. Vrshek-Schallhorn *et al.* argue that there is at least a reasonable basis for concluding that the 5-HTTLPR × stress interaction is a legitimate GxE effect, and hence that discussion is not more positive than warranted when studies characterize the results of the literature positively and cite positive studies. We strongly disagree with the implication that negative results are unimportant if the effect is legitimate; on the contrary, even if a consensus on the 5-HTTLPR × stress interaction had been unequivocally established, negative studies remain crucial to inform our understanding of the size and robustness of this effect. Overemphasis of positive studies or the positive aspects of mixed findings would lead to an overestimate of the apparent importance and strength of the effect.

Moving science forward by looking backward

We believe that our approach to coding studies with multiple relevant outcomes is the best available option and that its disadvantages are adequately addressed by our sensitivity analyses. Furthermore, our approach to coding positive focus is consistent with previous work and we believe it better captures an excessive focus on positive results and unwarrantedly positive conclusions than alternative approaches. While we support the call by Vrshek-Schallhorn *et al.* to focus on the positive goal of enhancing the prediction and treatment of depression, we believe that this goal can only be achieved by critically evaluating completed studies, striving for the greatest possible methodological rigor and transparency in future studies, and valuing negative results as highly as we do positive results.

Acknowledgements

Marcus R. Munafò is a member of the United Kingdom Centre for Tobacco and Alcohol Studies, a UKCRC Public Health Research: Centre of Excellence. Funding from British Heart Foundation, Cancer Research UK, Economic and Social Research Council, Medical

Research Council, and the National Institute for Health Research, under the auspices of the UK Clinical Research Collaboration, is gratefully acknowledged.

Declaration of Interest

None.

References

Boutron I, Altman DG, Hopewell S, Vera-Badillo F, Tannock I, Ravaud P (2014). Impact of spin in the abstracts of articles reporting results of randomized controlled trials in the field of cancer: the SPIIN randomized controlled trial. *Journal of Clinical Oncology* **32**, 4120–4126.

Boutron I, Dutton S, Ravaud P, Altman DG (2010). Reporting and interpretation of randomized controlled trials with statistically nonsignificant results for primary outcomes. *Journal of the American Medical Association* **303**, 2058–2064.

de Vries YA, Roest AM, Franzen M, Munafò MR, Bastiaansen JA (2016). Citation bias and selective focus on positive findings in the literature on 5-HTTLPR, life stress, and depression. *Psychological Medicine*. Published online: 12 August 2016. doi:10.1017/S0033291716000805.

Dwan K, Gamble C, Williamson PR, Kirkham JJ (2013). Systematic review of the empirical evidence of study publication bias and outcome reporting bias - an updated review. *PLoS ONE* **8**, e66844.

Eley TC, Sugden K, Corsico A, Gregory AM, Sham P, McGuffin P, Plomin R, Craig IW (2004). Gene–environment interaction analysis of serotonin system markers with adolescent depression. *Molecular Psychiatry* **9**, 908–915.

Heininga VE, Oldehinkel AJ, Veenstra R, Nederhof E (2015). I just ran a thousand analyses: benefits of multiple testing in understanding equivocal evidence on gene-environment interactions. *PLoS ONE* **10**, e0125383.

Kapur S, Phillips AG, Insel TR (2012). Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it? *Molecular Psychiatry* **17**, 1174–1179.

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.

Sharpley CF, Palanisamy SKA, Glyde NS, Dillingham PW, Agnew LL (2014). An update on the interaction between the serotonin transporter promoter variant (5-HTTLPR), stress and depression, plus an exploration of non-confirming findings. *Behavioural Brain Research* **273**, 89–105.

Sullivan PF (2007). Spurious genetic associations. *Biological Psychiatry* **61**, 1121–1126.

Taylor AE, Munafò MR (2016). Triangulating meta-analyses: the example of the serotonin transporter gene, stressful life events and major depression. *BMC Psychology* **4**, 23.

Vrshek-Schallhorn S, Sapuram V, Avery BM (2016). Bias in the measurement of bias. Letter regarding 'Citation bias and selective focus on positive findings in the literature on the serotonin transporter gene (5-HTTLPR), life stress, and depression'. *Psychological Medicine*. doi:10.1017/S0033291716002178.

Y. A. DE VRIES[1*], A. M. ROEST[1], M. FRANZEN[2], M. R. MUNAFÒ[3,4] AND J. A. BASTIAANSEN[1]

[1] *Department of Psychiatry, Interdisciplinary Center Psychopathology and Emotion Regulation, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands*

[2] *Department of Psychology, University of Groningen, Groningen, The Netherlands*

[3] *MRC Integrative Epidemiology Unit (IEU) at the University of Bristol, Bristol, UK*

[4] *UK Centre for Tobacco and Alcohol Studies, School of Experimental Psychology, University of Bristol, Bristol, UK*

*Address for correspondence: Y. A. de Vries, Department of Psychiatry, University Medical Center Groningen, Hanzeplein 1, 9713 GZ Groningen, The Netherlands. (Email: y.a.de.vries@umcg.nl)