Mozart Effect, Effect Size and Small-Study Effect

François Léonard[1], Ezio Tirelli[1]

[1] Department of psychology, ULiège

Introduction

The so-called "Mozart Effect" is the increase of visuo-spatial intelligence after the audition of Mozart's sonata K448. This effect was published by Rauscher et al. (1993).

Since 1993, his existence has been debated in the literature. In 2010, a Random Effect Meta-Analysis (REMA) found a small statistically significant pooled effect size (PES) of $d = 0.37$ (95%CI$[0.23, 0.52]$) (Pietschnig et al., 2010). Here, we reanalyzed this meta-analysis using three Small-Study Effect (SSE) correction methods. We also assessed False Discovery Rate (FDR) of the Mozart effect literature.

Methods

Effect sizes (Cohen's d) and standard errors of individual studies were extracted from Pietschnig et al. (2010). With these data, we recalculated a REMA and a Trim-and-Fill PES. The restricted maximum likelihood estimator was used to compute between-study heterogeneity and we used the Knapp-Hartung adjustments to calculate the confidence interval of the pooled effect size. We corrected the REMA PES using the PET-PEESE (Stanley & Doucouliagos, 2014) and the limit method (Rücker et al., 2011).

We used these four PES as hypothetical effect sizes to calculate the prospective power of individual studies. These prospective powers were then used to generate FDR curves (Ioannidis, 2005).

Results

All pf the pooling methods we used (REMA, Trim-and-Fill, PET-PEESE and limit) yielded small, statistically significant effects (Table 1).

Table 1: Pooled effect sizes

Method	Effect size	95% CI
Random effect	0.327	[0.225, 0.520]
Trim & fill	0.283	[0.101, 0.466]
PET-PEESE	0.398	[0.152, 0.645]
Limit	0.233	[0.013, 0.453]

The median prospoective statistical powers computed from these four pooled effect sizes range from 0.179 to 0.381 (Table 2 and Figure 1).

Table 2: Median prospective powers for the pooled effect sizes

Effect size	Mediane	IQR
Random effect (d = 0.327)	0.381	[0.311, 0.68]
Trim & fill (d = 0.283)	0.243	[0.201, 0.456]
PET-PEESE (d = 0.398)	0.425	[0.348, 0.738]
Limit (d = 0.233)	0.179	[0.15, 0.329]

The median sample size computed from the selected reports (n = 41) tells us that the smallest effect size detectable with a statistical power of 0.80 using a two-sided two-samples t-test is d = 0.897. For a two-sided paired-sample t-test the smallest detectable effect size is d= 0.448.

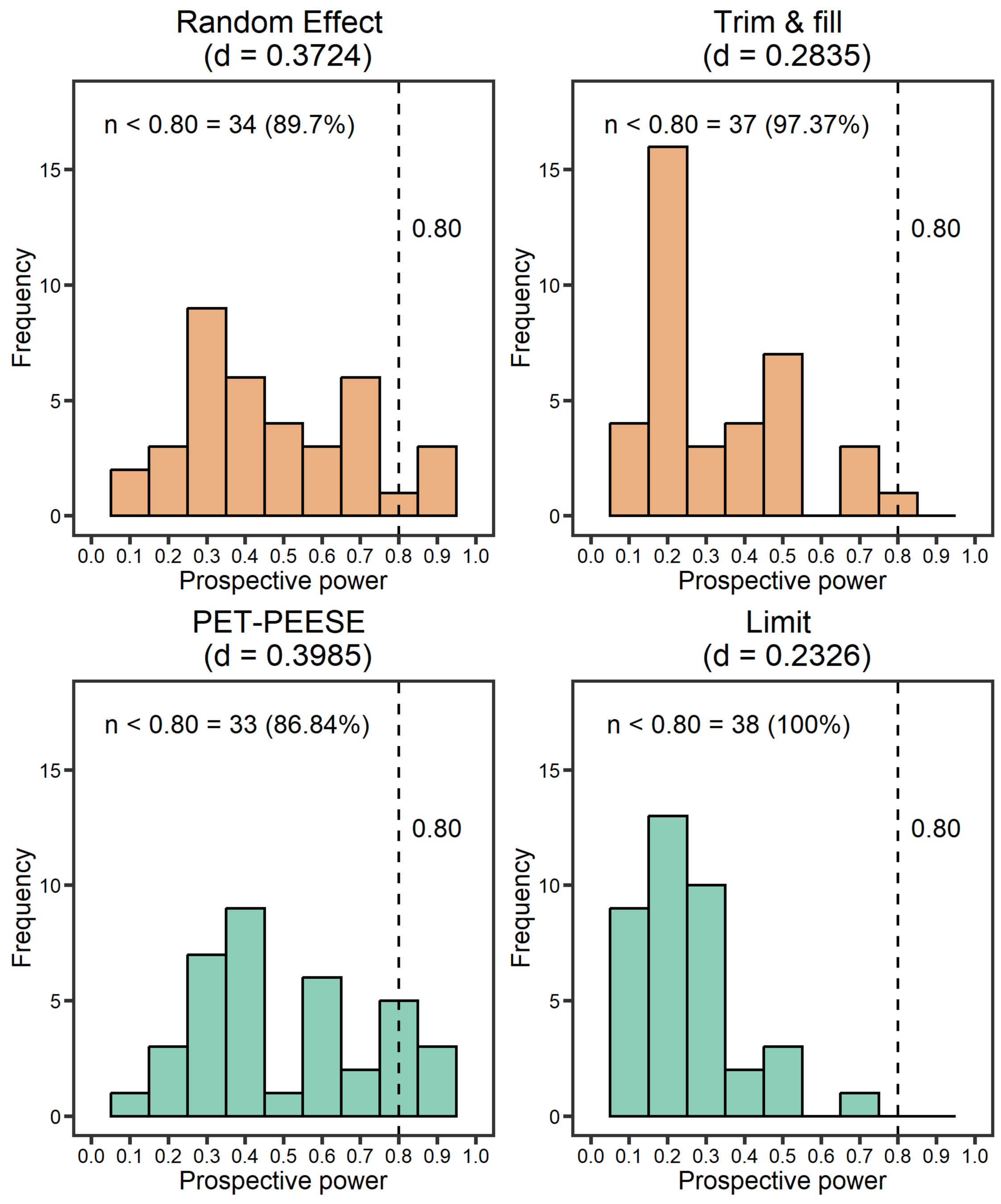

Figure 1: Histograms of prospective statistical powers for four PES. The dashed vertical line indicates to the recommended level of power (80%). Top left panel: prospective powers using the PES found with a REMA (d = 0.37). Top right panel: prospective powers found using the PES found with the Trim-and-Fill method (d = 0.28). Bottom left panel: prospective powers found using the PES of the PET-PEESE method (d = 0.4). Bottom right panel: prospective powers found using the PES of the limit method (d = 0.23).

The FDR was computed for our four pooled effect sizes (on significant study effect only). With a H_1 prior probability of 0.1 range from 0.51 to 0.71, indicating that more than half of the significant results in the literature may be false positives. With a H_1 prior probability of 0.5 the FDR range from 0.10 to 0.21, indicating (at worst) that one fifth of the significant results in the literature may be false positives (Figure 2).

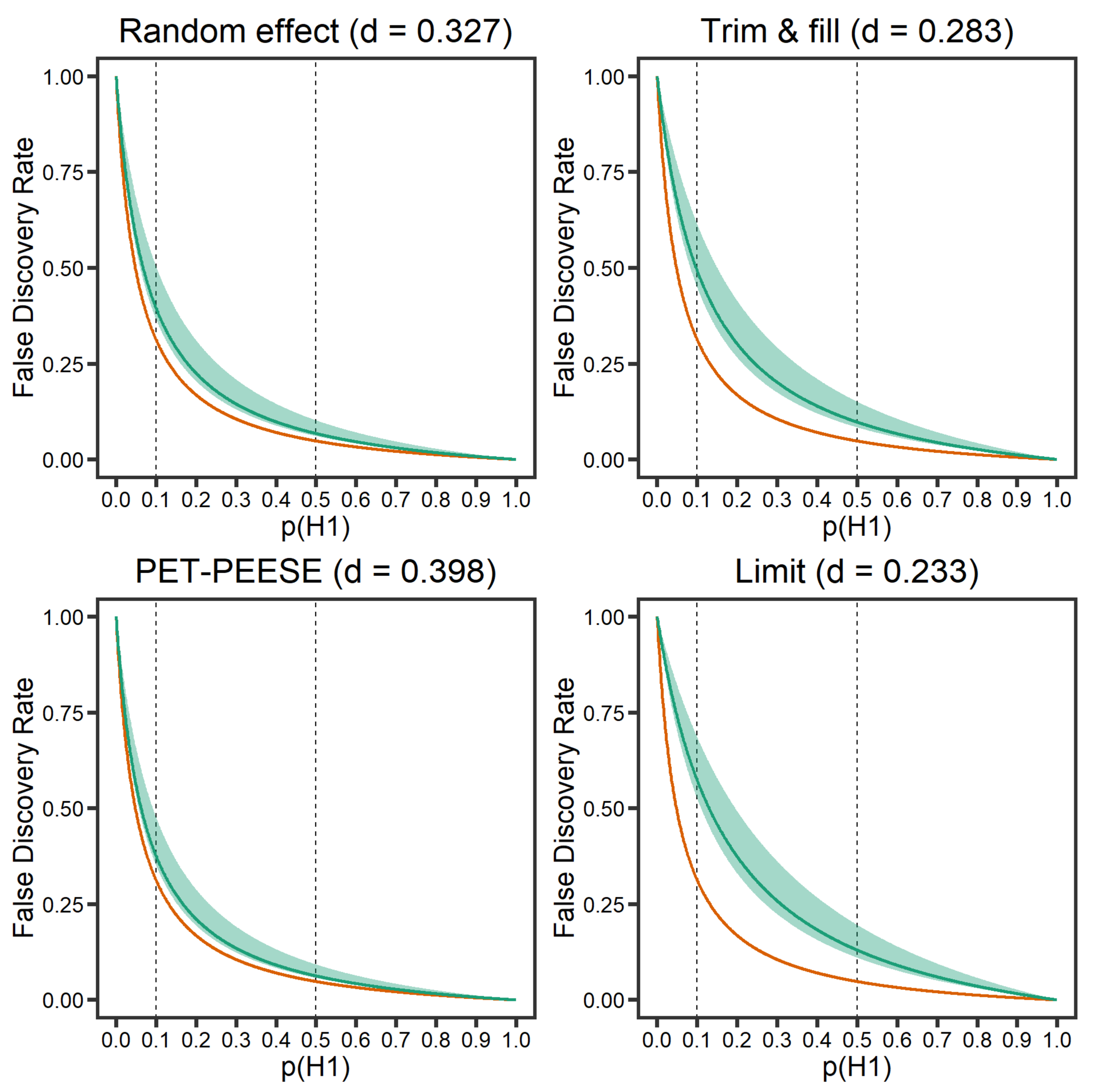

Figure 2: Relationship between FDR and the probability of H1. FDR curves (green) with interquartile range (green area) for each PES in contrast of a FDR curves of a statistical power of 0.99 (orange).Top left panel: FDR curve calculated with the median prospective power using the PES of the REMA (power = 0.69, IQR = [0.44, 0.78]) . Top right panel: FDR curve computed with the median prospective power calculated with the PES of the trim and fill (power = 0.46, IQR = [0.28, 0.55]). Bottom left panel: FDR curve computed with the median the prospective power calculated with the PES of the PET-PEESE method (power = 0.75, IQR = [0.49, 0.83]). Bottom right panel: FDR curve computed with the median the prospective power calculated with the PES of the limit method (power = 0.33, IQR = [0.20, 0.40]).

Discussion

Despite the Limit method being a severe correction, and resulting in a reduction of the effect size, the Mozart effect remained statistically significant. However, the majority of the included studies had low statistical power (1-beta < 80%), and only one study, with a statistical power close to 80%, reported a non-significant effect size of $d = 0.2$ (95% $CI[-0.07; 0.47]$) (Steele et al., 1999). Therefore, our findings suggest that the effect sizes reported in other significant studies may have been overestimated.

Based on the results of our analysis, we found that low power increases the FDR, especially at low plausibility. Considering the low prior probability of H_1 in the case of the Mozart effect and the low power observed in most studies, our findings suggest that at least half of the significant studies in the Mozart effect literature are likely to be false positives.

Pietschnig et al's meta-analysis includes almost only low quality studies, so the effect sizes are probably inflated. The "garbage-in garbage-out" problem of meta-analysis reminds us that pooled effect sizes are upper bound estimates. Furthermore, high-powered replications of studies found, on average, effect sizes three times smaller than the pooled effect size of the meta-analysis (Kvarven et al., 2020).

Therefore, the true Mozart effect is smaller than the computed one and is likely to be of negligible significance and scientific importance.

References

Ioannidis, J. P. A. (2005). Why Most Published Research Findings Are False. *PLoS Medicine*, *2*(8), 6. https://doi.org/10.1371/journal.pmed.0020124

Kvarven, A., Strømland, E., & Johannesson, M. (2020). Comparing meta-analyses and preregistered multiple-laboratory replication projects. *Nature Human Behaviour*, *4*(4), 423–434. https://doi.org/10.1038/s41562-019-0787-z

Pietschnig, J., Voracek, M., & Formann, A. K. (2010). Mozart effectShmozart effect: A meta-analysis. *Intelligence*, *38*(3), 314–323. https://doi.org/10.1016/j.intell.2010.03.001

Rauscher, F. H., Shaw, G. L., & Ky, C. N. (1993). Music and spatial task performance. *Nature*, *365*(6447), 611–611. https://doi.org/10.1038/365611a0

Rücker, G., Schwarzer, G., Carpenter, J. R., Binder, H., & Schumacher, M. (2011). Treatment-effect estimates adjusted for small-study effects via a limit meta-analysis. *Biostatistics*, *12*(1), 122–142. https://doi.org/10.1093/biostatistics/kxq046

Stanley, T. D., & Doucouliagos, H. (2014). Meta-regression approximations to reduce publication selection bias. *Research Synthesis Methods*, *5*(1), 60–78. https://doi.org/10.1002/jrsm.1095

Steele, K. M., Brown, J. D., & Stoecker, J. A. (1999). Failure to Confirm the Rauscher and Shaw Description of Recovery of the Mozart Effect. *Perceptual and Motor Skills*, *88*(3), 843–848. https://doi.org/10.2466/pms.1999.88.3.843