Brewin Meta Analysis

Andy P. Field

2023-12-15

## Methods

## Analysis plan

All analyses were conducted using R 4.3.1 (2023), using the packages metafor (Viechtbauer 2010), meta (Balduzzi, Rücker, and Schwarzer 2019), and tidyverse (Wickham et al. 2019) for general data wrangling.

### Effect Sizes

Effect sizes were estimated for all self-report or judge-rated variables identified by study authors as assessing incoherence or disorganisation are included, along with all available contemporaneous measures of ASD/PTSD.[[1]](#footnote-21) Effects that re-expressed existing effects in an alternative form were omitted. For example, where difference between ASD/PTSD and control group mean disorganisation was accompanied by the correlation between the ASD/PTSD and disorganization measures, only the effect size for the group means would be included because these effects carry essentially the same information. Effect sizes from computer-scored measures were not included because these have been criticised (rightly) as lacking any evidence of validity, face or any other kind.

The standardised mean difference, (Hedges 1981), was used as the effect size metric. This is an unbiased estimator of Cohen’s (Cohen 1988) calculated using the number of participants in each group ( and ), the mean score for each group ( and ) and the standard deviation for each group ( and ):

The sampling variance of is

Of the 80 effect sizes, 36 could be computed from reported means and standard deviations, 11 were converted from reported *t*-statistics of group differences, and 33 were converted from reported correlation coefficients. When a correlation coefficient had to be used, where the relevant descriptive information was available the formulae for the effect size and its standard error in Mathur and VanderWeele (2020) was use ( = 6),

which expresses the effect in terms of the change in disorganization resulting from an increase in the continuous measures of PTSD of units. To make the resulting comparable to studies for which represents differences between PTSD groups, was set to be the difference between mean PTSD scores in PTSD and control groups. To apply this formula, the standard deviation of the PTSD measure () for the entire sample is required. Articles that did not report descriptive information for the PTSD measure for the entire sample did report it for PTSD and control subgroups allowing the whole sample variance () to be estimated using the standard formula (see O’Neill 2014):

in which , and are the sample size, mean and standard deviation of the PTSD measure for the PTSD group and , and are the corresponding values for the control group. Where this information was not available the formulae to convert a point biserial correlation was used ( = 27),

In effect, this is the same as applying the equation in Mathur and VanderWeele (2020) but with . In other words, it will yield the same value as the equation used for a continuous measure if PTSD and control groups had a mean difference on a continuous PTSD measure of about 2 standard deviations. [[2]](#footnote-22)

Estimates of are biased in small samples so were adjusted using a correction , (Hedges 1981),

such that

Finally, not FOA studies used different types of measures of disorganization, with some scales directly measuring disorganisation (i.e., high scores reflect greater disorganization) and others measuring organisation/coherence (i.e., low scores reflect less organization and by implication greater disorganization). The direction of effect sizes based on scales where a low score represented disorganization were reversed. Therefore, a positive always equates to greater disorganization.

### Statistical model

Most, but not all, studies contributed more than one effect size estimate; therefore, a random effects meta-analysis was conducted using a three level multilevel model in which effect sizes are nested within studies. The model fitted was

In which is an estimate of effect size nested within study , is the overall average population effect, is the within-study variability in effect sizes, and is the between-study variability in effect sizes. All models used cluster-robust tests and confidence intervals based on a sandwich-type estimator (see Colin Cameron and Miller 2015; Viechtbauer 2010) using restricted maximum likelihood (REML).

Effect sizes could be classified according to whether they came from a study classified as FOA or one that was not FOA. Within FOA studies, effect sizes could quantify measures that use either detailed or global coding. Within not FOA studies effect sizes could quantify measures of disorganisation (disorganisation is represented by a high score) or a lack of organisation (disorganisation is represented by a low score). Therefore, each effect size could be classified into one of four categories: (1) FOA (detailed coding); (2) FOA (global coding); (3) not FOA (disorganisation); or (4) not FOA (organisation). The substantive hypothesis was that effects would vary across these four categories. A secondary hypothesis was that effect sizes would vary by age (adult vs. youth sample). To test these moderators, the above model was expanded to incorporate one or more fixed effect predictors () and their associated parameters ()

Specifically, the substantive hypothesis was tested by fitting a model that represented the four effect size classifications as three dummy variables with FOA (detailed coding) as the reference category. This coding resulted in parameter estimates representing the difference in effect sizes between FOA (detailed coding) and each of FOA (global coding), not FOA (disorganisation) and not FOA (organisation).

Publication bias was modelled using a step function selection model. Let denote the relative likelihood of selection given the *p*-value of a study, then the selection function is:

That is, denotes the likelihood of selection. There were too few studies to model moderate and severe publication biases using the s suggested by Vevea and Woods (2005) so instead one-tailed moderate and severe publication bias were modelled with fewer cut points:

## Results

### Overall effects

[Figure 1](#fig-fp) shows the aggregated effect size within each study organised by study classification. For studies classified as FOA, when detailed coding is used effect sizes are generally above zero but more than half of studies have confidence intervals that contain zero. When global coding is used studies have positive effect sizes and confidence intervals to not cross zero. For studies classified as not FOA, when measures of disorganisation are used effect sizes are generally above zero but half of studies have confidence intervals that contain zero. When measures of organisation are used effect sizes hover around zero with all confidence intervals contain zero.

For the overall model ignoring predictors, the between study variability was = 0.052 and the within study variability was = 0.048. There was significant heterogeneity, (79) = 208.81, *p* < .001. The overall effect was a strong positive effect that was significant = 0.32 [0.17, 0.46], *t*(17) = 4.59, *p* < .001; the 95% confidence interval suggests that if this sample is one of the 95% that capture the population value then disorganization in PTSD groups could be anything between 0.17 and 0.46 standard deviations higher than in control groups.

### The effect of FOA

[Table 1](#tbl-foa) shows that the effect was large and significantly different from 0 for FOA studies using detailed coding ( = 0.47), FOA studies using global coding ( = 0.76) and not FOA that measured disorganisation ( = 0.39), but virtually zero in not FOA studies that measured organisation ( = 0.01).

The classification of effect sizes was dummy coded such that each category of effect size was compared to ones from FOA studies that quantified measures that used detailed coding. In the model including these dummy variables the between study variability was = 0.004 and the within study variability was = 0.031. There was significant heterogeneity, (76) = 114.33, *p* = 0.003. The overall moderation effect was significant, (3, 14) = 28.96, *p* < .001, indicating that effect sizes were significantly different across the four categories. [Table 2](#tbl-foapars) shows the parameter estimates of this model. Compared to effect sizes relating to FOA studies that use detailed coding, effect sizes were (1) significantly larger in FOA studies that use global ratings; (2) significantly smaller in not FOA studies that measured organisation; and (3) not significantly different to not FOA studies that measured disorganisation.

Table 1: Individual meta-analyses of each study type

| FOA |  |  |  |  |  |  |  | 95% CI |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| FOA (Detailed) | 19 | 0.07 | 0.00 | 32.97 | 18 | 0.017 | 0.47 | [0.23, 0.72] | 3.76 | 0.00 |
| FOA (Global) | 15 | 0.00 | 0.08 | 33.53 | 14 | 0.002 | 0.76 | [0.57, 0.95] | 7.93 | 0.00 |
| Not FOA (Disorganisation) | 13 | 0.02 | 0.00 | 12.03 | 12 | 0.444 | 0.39 | [0.23, 0.56] | 4.66 | 0.00 |
| Not FOA (Organisation) | 33 | 0.00 | 0.01 | 35.81 | 32 | 0.294 | 0.01 | [-0.09, 0.11] | 0.17 | 0.86 |

Table 2: Parameter estimates for comparisons between study types

| Comparison |  | 95% CI |  |  |
| --- | --- | --- | --- | --- |
| Intercept | 0.46 | [0.22, 0.70] | 4.14 | 0.001 |
| FOA (Global) vs. FOA (Detailed) | 0.27 | [0.01, 0.54] | 2.24 | 0.042 |
| Not FOA (Disorganisation) vs. FOA (Detailed) | -0.06 | [-0.37, 0.24] | -0.44 | 0.665 |
| Not FOA (Organisation) vs. FOA (Detailed) | -0.45 | [-0.71, -0.18] | -3.64 | 0.003 |

### The effect of age group

In the model including age group as a moderator, the between study variability was = 0.055 and the within study variability was = 0.048. There was significant heterogeneity, (78) = 202.20, *p* < .001. The overall moderation effect was close to zero and not significant, = -0.07 [-0.36, 0.23], *t*(16) = -0.49, *p* = 0.630. [Table 3](#tbl-age) shows that the effects were very similar in ( = 0.33) and youth ( = 0.24) samples.

Table 3: Individual meta-analyses by age of sample

| Age group |  |  |  |  |  |  |  | 95% CI |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Adult | 64 | 0.06 | 0.03 | 155.41 | 63 | <.001 | 0.33 | [0.17, 0.49] | 4.04 | 0.00 |
| Youth | 16 | 0.02 | 0.09 | 46.79 | 15 | <.001 | 0.24 | [-0.00, 0.49] | 1.93 | 0.05 |

### The effect of the type of disorganization measure

An analysis was done to see whether effect sizes could be predicted by whether the measure of disorganisation measured disorganisation (disorganisation is represented by a high score) or a lack of organization (disorganisation is represented by a low score) instudies from the by the Rubin/Bernsten laboratory.

In this model, the between study variability was = 0.000 and the within study variability was = 0.006. There was not significant heterogeneity, (23) = 25.41, *p* = 0.329. The effect of the type of scale was strong but only just significant (*k* is small), = 0.35 [0.00, 0.69], *t*(4) = 2.79, *p* = 0.049. [Table 4](#tbl-reverse_rubin) shows that the effect was large and significantly different from 0 when a measure of disorganization was used ( = 0.26) but was close to zero and non-significant for studies that used measures of organization/coherance ( = -0.08).

Table 4: Individual meta-analyses of each study type (Rubin/Berntsen lab only)

| Disorganization |  |  |  |  |  |  |  | 95% CI |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Low score | 17 | 0.00 | 0.02 | 19.88 | 16 | 0.226 | -0.08 | [-0.22, 0.07] | -1.04 | 0.30 |
| High score | 8 | 0.00 | 0.00 | 5.53 | 7 | 0.595 | 0.26 | [0.09, 0.44] | 3.02 | 0.00 |

## Publication bias

[Table 5](#tbl-pubbias) shows the results of modelling moderate and severe one-tailed selections bias. These models should be interpreted as a sensitivity analysis in which the unadjusted effect size estimate () is compared to the effect size estimate adjusted for selection bias (). For all types of studies and measures moderate publication bias makes little difference to the estimated effect size, whereas severe publication bias dramatically reduces the estimated effect size in all cases except for studies classified as FOA that use global coding.

Table 5: Selection models of publication bias

| Domain | Selection model |  |  | 95% CI |  |
| --- | --- | --- | --- | --- | --- |
| FOA (Detailed) | Moderate | 0.47 | 0.39 | [0.12, 0.67] | 0.006 |
| FOA (Detailed) | Severe | 0.47 | 0.25 | [-0.15, 0.65] | 0.217 |
| FOA (Global) | Moderate | 0.70 | 0.68 | [0.48, 0.88] | <.001 |
| FOA (Global) | Severe | 0.70 | 0.67 | [0.46, 0.87] | <.001 |
| Not FOA (Disorganisation) | Moderate | 0.37 | 0.32 | [0.15, 0.49] | <.001 |
| Not FOA (Disorganisation) | Severe | 0.37 | 0.26 | [0.08, 0.44] | 0.006 |
| Not FOA (Organisation) | Moderate | -0.04 | -0.07 | [-0.22, 0.08] | 0.360 |
| Not FOA (Organisation) | Severe | -0.04 | -0.18 | [-0.32, -0.04] | 0.010 |

## Figures

|  |
| --- |
| Figure 1: Forest plot |

## References

Balduzzi, Sara, Gerta Rücker, and Guido Schwarzer. 2019. “How to Perform a Meta-Analysis with {R}: A Practical Tutorial.”

Cohen, J. 1988. *Statistical Power Analysis for the Behavioural Sciences*. 2nd ed. New York: Academic Press.

Colin Cameron, A., and Douglas L. Miller. 2015. “A Practitioner’s Guide to Cluster-Robust Inference.” *Journal of Human Resources* 50 (2): 317–72. <https://doi.org/10.3368/jhr.50.2.317>.

Hagenaars, Muriel A., Agnes van Minnen, Cees A.L. Hoogduin, and Marc Verbraak. 2009. “A Transdiagnostic Comparison of Trauma and Panic Memories in PTSD, Panic Disorder, and Healthy Controls.” *Journal of Behavior Therapy and Experimental Psychiatry* 40 (3): 412–22. <https://doi.org/10.1016/j.jbtep.2009.04.001>.

Hedges, L. 1981. “Distribution Theory for Glass’s Estimator of Effect Size and Related Estimators.” *Journal of Educational Statistics* 6: 107–28.

Mathur, Maya B., and Tyler J. VanderWeele. 2020. “A Simple, Interpretable Conversion from Pearson’s Correlation to Cohen’s <Em>d</Em> for Continuous Exposures.” *Epidemiology* 31 (2): e16–17. <https://doi.org/10.1097/EDE.0000000000001111>.

O’Neill, B. 2014. “Some Useful Moment Results in Sampling Problems.” *The American Statistician* 68 (4): 282–96. <https://doi.org/10.1080/00031305.2014.966589>.

R Core Team. 2023. “R: A Language and Environment for Statistical Computing.” <https://www.R-project.org/>.

Vevea, J. L., and C. M. Woods. 2005. “Publication Bias in Research Synthesis: Sensitivity Analysis Using a Priori Weight Functions.” *Psychological Methods* 10 (December): 428–43.

Viechtbauer, Wolfgang. 2010. “Conducting Meta-Analyses in {R} with the {Metafor} Package.” <https://doi.org/10.18637/jss.v036.i03>.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D’Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. “Welcome to the {Tidyverse}.” <https://doi.org/10.21105/joss.01686>.

1. The original effect size from Hagenaars et al. (2009) was implausibly large (). The authors confirmed our suspicion that the reported standard deviations were standard errors so the reported values converted to standard deviations. [↑](#footnote-ref-21)
2. For the = 6 effect sizes where descriptive information about the PTSD measure was available, the average mean difference was 19.30 with an average standard deviation of 11.81 yielding a ratio of 1.63. The implication being for the = 27 effect sizes where descriptive information about the PTSD measure was not available, the effect sizes from the point biserial formula are a reasonable approximation of the values that would have been obtained had enough information been available to use the formula to convert *r* based on a continuous measure. [↑](#footnote-ref-22)