Supplementary Information 3: Robustness of Meta-Analytic Results

Net effects of multiple stressors in freshwater ecosystems: a meta-analysis

Jackson, M.C.[1*], Loewen, C.J.G.[2], Vinebrooke, R.D. [2], Chimimba, C.T. [1] *mjackson@zoology.up.ac.za

1. Centre for Invasion Biology, Department of Zoology and Entomology, University of Pretoria, South Africa
2. Department of Biological Sciences, University of Alberta, Canada

Publication bias identification and sensitivity analysis

A common concern with modern meta-analyses is the potential for publication bias to influence mean effect size calculations (Nakagawa & Santos, 2012). Publication bias is typically propagated by the greater likelihood of publication for experiments with significant results, which may skew meta-analytic results (Rosenthal, 1979). To test for publication bias and assess its potential impact on our results, we used the following multistep approach.

First, we evaluated our data graphically with funnel plots comparing our standardized interaction effect sizes against their pooled sample sizes (calculated as $n_p + n_o$, where n_p was calculated as $n_A + n_B$ and $n_o = n_{AB}$) and estimated precision (1/variance; Fig. S2). Here, significant asymmetry around the mean interaction effect size may indicate publication bias; however, asymmetry may also be caused by chance or by true heterogeneity in the dataset (Nakagawa & Santos, 2012).

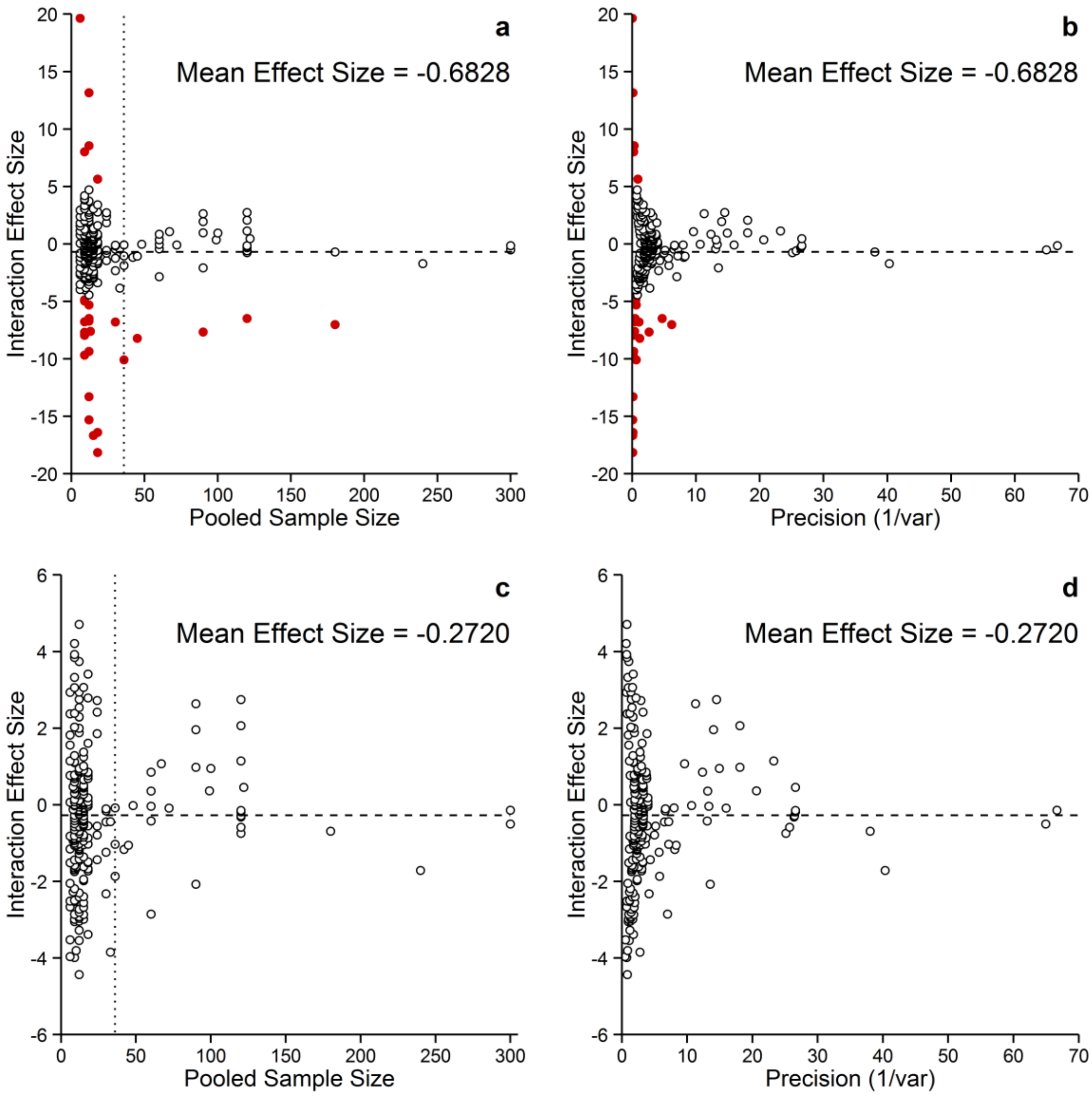
Visual assessment of our funnel plots suggested a potential bias towards negative (antagonistic) interaction effect sizes, as indicated by outlying data points (Fig. S2a,b). We followed this visual assessment with Spearman rank correlation and Eggers regression tests to statically test for data asymmetry, using MetaWin version 2.1 (Rosenberg *et al.*, 2000) and the 'metafor' package (Viechtbauer, 2010) in the R computing program (R Core Team, 2014), respectively. Statistically significant correlations/regressions may indicate publication bias towards larger effect sizes (Begg, 1994; Egger, 1997). Results of Spearman rank correlation tests indicated a significant relationship between effect size and variance (n = 286, $r_s = -0.254$, $P = <0.001$) and a non-significant relationship between effect size and pooled sample size (n = 286, r_s = 0.016, P = 0.794). Similarly, the Eggers regression tests indicated a significant relationship between effect size and standard error (n = 286, z = -10.370, P = <0.001) and a non-significant relationship between effect size and pooled sample size (n = 286, z = -1.153, P = 0.249).

Although our funnel plot assessments and regression/correlation-based tests both indicated significant asymmetry in our dataset, these results do not necessarily indicate significant publication bias as data asymmetry may also be attributable to chance or true heterogeneity in the dataset (Nakagawa & Santos, 2012). Thus, these results require careful interpretation with consideration of the total amount of heterogeneity present and whether publication bias is in fact the most reasonable mechanism to explain the specific asymmetries observed. For instance, our results suggest a potential bias towards negative interaction effects sizes (and antagonisms); however, we would intuitively expect that interactions involving synergies would have a greater likelihood of being over-reported in the literature since they represent more severe or dramatic scenarios. Further, we compare a diverse set of measurements from stressor experiments on different organism groups and across levels of biological organization in our global meta-analysis, and thus the considerable heterogeneity observed in this analysis (Q_{total} = 912.70) was expected. To explain this variance we used a series of mixed effects categorical meta-analyses using biologically relevant moderators. Indeed, we found that levels of these moderators varied in their mean effect sizes (see Supporting Information 2), reflecting the variable responses of the groups to multiple stressor impacts.

Nevertheless, we conducted a series of tests to assess the sensitivity of our global meta-analysis to effect size outliers (potential publication bias). First, we used Rosenthal's method to estimate a fail-safe number, which is the number of non-significant or missing observations that would be needed to change the significance of our findings (Rosenthal, 1979). A fail-safe number larger than 5(n) + 10 (where n is the number of studies in the meta-analysis) is generally considered to be robust against publication bias (Rosenthal, 1979). We used MetaWin version 2.1 (Rosenberg *et al.*, 2000) to estimate a fail-safe number of 24,797, which far exceeds the minimum recommended number based on our sample size (24,797 > 1,440). This suggests that our estimates are reliable, even with the observed data asymmetry. Secondly, we used the 'metafor' package (Viechtbauer, 2010) in the R computing program (R Core Team, 2014) to conduct a trim and fill analysis. Similar to the fail-safe number, this approach is used to assess the impact of potentially missing observations on the meta-analytic results (Nakagawa & Santos, 2012). However, trim and fill analysis failed to identify any missing studies needed to restore symmetry (missing studies = 0), though it performed poorly (P = 0.50 that the model estimated the correct number of missing studies), likely owing to the considerable heterogeneity in our dataset.

Finally, to demonstrate the robustness of our global meta-analytic findings in spite of any publication bias, we reanalysed our dataset after omitting potential effect size outliers. We identified 41 potential effect size outliers based on a visual assessment of our funnel plots (red data points in Fig. S2a,b). Even after removing these data points, our random effects model found a significantly antagonistic mean net effect (d = -0.2720 with upper and lower bootstrapped confidence intervals of -0.4602 and -0.0877, respectively; Fig. S2c,d). Together, these results suggest that our meta-analytic findings are robust to data asymmetry, regardless of whether the observed pattern reflects publication bias or true heterogeneity.

Fig. S2 Funnel plots of standardized interaction effect sizes (Hedge's *d*) against pooled sample sizes (**a** and **c**) and precision estimates (**b** and **d**). Plots (**a**) and (**b**) present our entire dataset (n = 286; 13 data points not shown where effect sizes >20 (n = 4) or <20 (n = 9)) and plots (**c**) and (**d**) present our dataset reduced to test for sensitivity to outlying effect sizes and potential publication bias (n = 245). Effect sizes omitted for our publication bias sensitivity analysis are coloured red (**a** and **b**); dashed horizontal lines indicate weighted mean interaction effect sizes; and dotted vertical lines indicate a pooled sample size of 26 (used as a cut-off to reduce our dataset for our sample size sensitivity analysis (see below)).



Sample size sensitivity analysis

Similar to the publication bias sensitivity analysis, we explored the sensitivity of our analyses to variation in study sample size. Specifically, observations with larger sample sizes are expected to have lower variance and thus carry more weight in our meta-analytic models. To test the robustness of our dataset to such variations, we explored how observations with large sample sizes (control replication of ≥ 12 or pooled sample size ≥ 26) may skew our results by reanalysing our data with these points (n = 28; Fig. S2a,c) omitted. We found that the mean effect size did not change significantly overall, or in any categorical grouping (where at least three studies were omitted; Table S3). Since the mean effect sizes were similar to those calculated based on our entire dataset, and none of the interaction type assignments changed (Table S3), this further demonstrates the robustness of our meta-analytic results.

Table S3. Mean interaction effect sizes for levels of moderators where more than three studies were omitted after excluding those studies with large sample sizes (replication ≥ 12 or pooled sample ≥ 26). N = group sample size; *d* = mean effect size; CI = 95% bootstrapped confidence intervals; and Interaction = mean interaction type.

Analysis / Level	**Full Dataset**			**Reduced Dataset**			**Interaction**
	N	***d***	**CI**	**N**	***d***	**CI**	
Global Analysis	230	-0.65	-0.95 to -0.33	202	-0.69	-1.04 to -0.36	Antagonistic
Level of Organization							
Population	70	-0.60	-1.21 to 0.08	55	-0.66	-1.04 to 0.16	Additive
Organism	23	-0.91	-1.47 to -0.41	10	-1.43	-2.63 to -0.53	Antagonistic
Stressor Pair							
Contamination x Habitat Alteration	19	-0.25	-1.07 to 0.86	16	-0.09	-1.04 to 1.26	Additive
Contamination x Warming	33	-0.87	-1.69 to -0.18	18	-1.34	-2.43 to -0.39	Antagonistic
Response Level							
Vertebrate	56	-0.62	-1.15 to -0.10	37	-0.75	-1.32 to -0.16	Antagonistic

References

Begg CB (1994) Publication bias. In: *The Handbook of research synthesis* (eds Cooper H, Hedges LV), pp. 399–409, Russell Sage Foundation, New York.

Egger M, Smith GD, Schneider M, Minder C (1997) Bias in meta-analysis detected by a simple, graphical test. British Medical Journal, **315**, 629–634.

Nakagawa S, Santos ESA (2012) Methodological issues and advances in biological meta-analysis. Evolutionary Ecology, **26**, 1253–1274.

R Core Team (2014) R: A language and environment for statistical computing. R Foundation for Statistical Computing: Version 3.0.3, Vienna, Austria. URL: http://www.R-project.org/.

Rosenberg MS, Adams DC, Gurevitch J (2000) MetaWin: Statistical software for meta-analysis: Version 2.1.

Rosenthal R (1979) The “file drawer problem” and tolerance for null results. Psychological Bulletin, **86**, 638–641.

Viechtbauer W (2010) Conducting meta-analyses in R with the metaphor package. Journal of Statistical Software, **36**, 1–48.