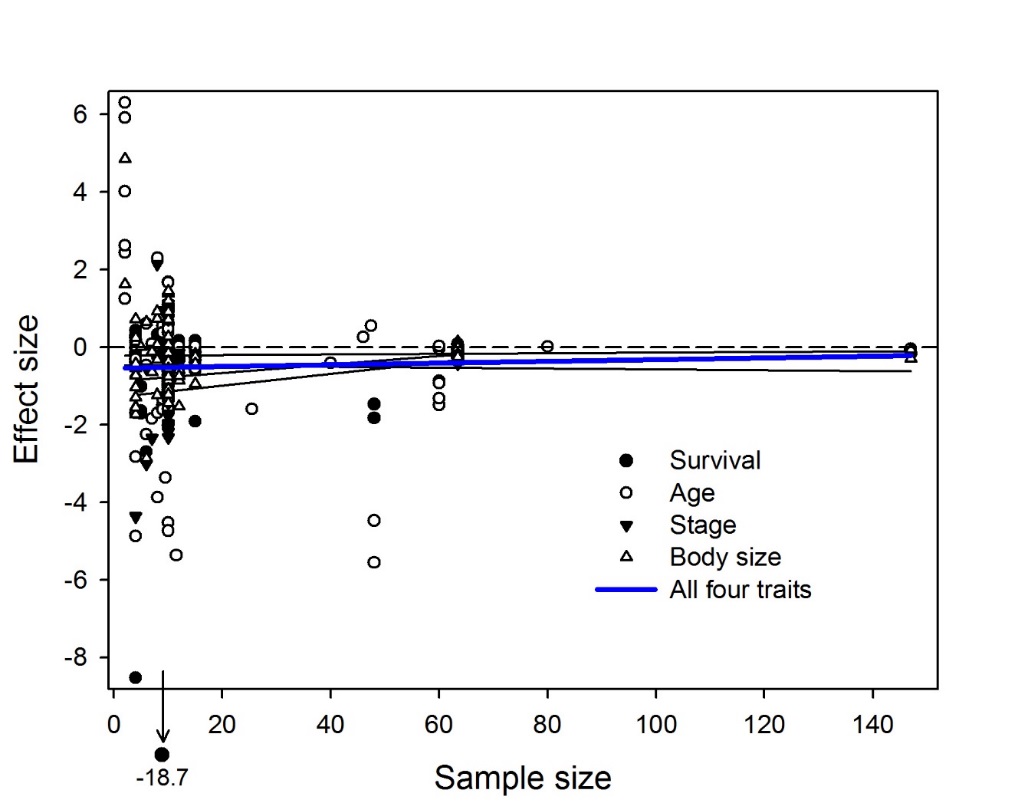
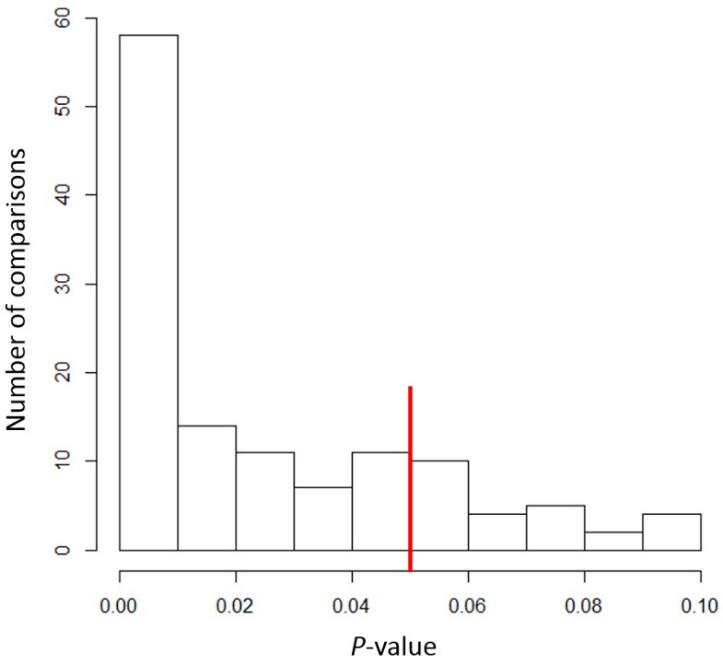
**Electronic Appendix C: Publication bias and *P*-hacking**

The data set contained no strong evidence for publication bias. The funnel plot below shows that the slope of effect size on sample size (average of the two treatments) was positive but not significant (b = 0.0110, *F*1,297 = 0.39, *P* = 0.53). The trait-by-sample size interaction was also not significant because trends in effect size were about the same for all four traits (*F*3,297 = 0.83, *P* = 0.48). High variation among observations at low sample size is expected, but the weak decline in (negative) effect size with increasing sample size suggests that the “file drawer effect” is not important. That is, studies observing non-significant effects of predation risk on amphibian traits at hatching are not badly underrepresented in the data set (Rosenthal, 1979).



There is little cause for concern about inflation bias (or *P*-hacking; Head et al., 2015), judging from the distribution of *P*-values for all four response variables. Inflation bias occurs when researchers collect data only until an effect becomes significant, attempt multiple approaches to analysis and report the one that reaches significance, or record multiple response variables and report only those that are significant. Head et al. (2015) illustrated how two tests can detect inflation bias in meta-analysis data sets. The first, a test for evidential value in the data, asks whether the distribution of *P*-values below 0.05 is right-skewed (79 observations in the bin 0 ≤ *P* < 0.025, 22 in the bin 0.025 ≤ *P* < 0.05; binomial test, *P* < 0.0001). This reflects strong evidence for an effect of embryonic exposure to predators on the four responses. The second, a test for over-representation of *P*-values just below 0.05, compares the number of values in the bin 0.04 ≤ *P* < 0.045 with the number in the bin 0.045 ≤ *P* < 0.05 (6 and 5, respectively; binomial test, *P* = 1). This result suggests no evidence for inflation bias. There were insufficient data to perform these tests separately on the four response variables.



**Literature Cited**

Head ML, Holman L, Lanfear R, Kahn AT, Jennions MD. 2015. The extent and consequences of p-hacking in science. PLoS Biology 13:e1002106.

Rosenthal R. 1979. The file drawer problem and tolerance for null results. Psychological Bulletin 86:638-641.