JOURNAL OF Evolutionary Biology

doi: 10.1111/jeb.12946

COMMENTARY

Graphic illustration of a potential problem: a commentary on Morrissey (2016)

M. D. JENNIONS

Evolution, Ecology & Genetics, Research School of Biology, Australian National University, Canberra, ACT, Australia

Morrissey (2016) is an enjoyable but challenging read that highlights misapplication of meta-analysis to questions in evolutionary biology. The problems highlighted in the three case studies all arise when estimating the mean magnitude rather than the mean value of a relationship (i.e. using absolute rather than signed effect sizes). A statistical maven speaks, but the language remains technical, and the message might be lost, or worse, misunderstood. I therefore focused my efforts on summarizing some key messages in a form that I could use to teach students. My commentary is directed to such readers. The result is a cartoon (Fig. 1). I hope it provides accessible insights into the problems Morrissey raised. We can note the following:

1 Biased estimates of the mean magnitude of an effect arise whenever the estimated effect in a study is not in the same direction as the true effect (shown by the grey part of the sampling variance bar). This still contributes a positive estimate of the *absolute* effect size. The dark part of the bar and the bar above each line (which is the same length as the grey bar) shows the extent to which this creates an asymmetric in estimates of the *absolute* effect size.

2 Weighting studies by the inverse of their sampling variance, which is often closely linked to sample size (e.g. for Fisher's z transformation of r, it is $1/[N\text{-}3]$), is useful. It reduces bias in estimates of the mean magnitude of the effect. Compare effect A with B, or C with D. The effect that is estimated with a smaller sampling variance is less likely to cross the zero boundary such that the distribution of estimated absolute values is biased upwards. Consequently, if studies are weighting by their sample variance, the bias in the estimated mean is reduced. I do not think this insight is obvious from Morrissey's review.

3 With greater variance in true effect sizes, there is a lower likelihood that the sampling variance will produce estimates either side of the zero boundary that inflate the estimated mean magnitude of an effect. That is, for distribution I, far fewer of the true effects are greater than or equal to C or D than is the case for distribution II.

4 The underlying statistics for commonly implemented meta-analyses assume that (i) the true distribution of effect sizes is symmetric and (ii) that the sampling variance is symmetric. Assumption (i) is false for absolute effect sizes when the distribution of true effects includes zero (compare, say, I and III). Although not illustrated, in I, the distribution of absolute effect sizes is an asymmetric folded normal distribution; for III, it is not (ignoring the very few true effects below zero). Obviously, as the situation moves from III towards I, the problem increases. Assumption (ii) is incorrect when the sampling variance includes values opposite in direction to the true effect (most likely for case A and least likely for case C).

None of the above qualifiers negate Morrissey's insight that transforming then analysing observed effect sizes inflates the estimated mean magnitude of an effect. The technical validity of Morrissey's analyse-then-transform mixed model approach to resolve the problem is beyond me, but it makes sense because it uses the appropriate variances. Ultimately, Fig. 1 simply illustrates that variances are being misspecified for meta-analysis of absolute values. In hindsight, the problem is fairly obvious, but in what other situations do problems arise? Morrissey suggests that *'many*

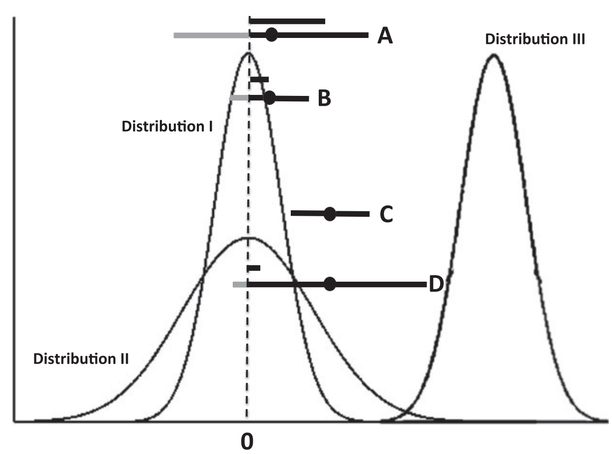

Fig. 1 The mean true effect is zero. Three distributions of the true effect size are shown (I–III). I has a lower variance (smaller standard deviation) than II. Cases A to D represent studies in which the true effect size (indicated by a solid circle) is either a value close to, or far from, zero. For simplicity, only cases with positive true effects are shown. The symmetric bars around the true mean indicate the sampling variance. For simplicity, we can think of these as the range of estimates that will be obtained 95% of the time for a given study. Studies with a smaller bar have a lower sampling variance (i.e. a larger sample size).

Correspondence: Michael D. Jennions, Evolution, Ecology & Genetics, Research School of Biology, Australian National University, Canberra, ACT 2601, Australia.
Tel.: +61 (0) 431 546 390; fax: +61 6125 3540;
e-mail: michael.jennions@anu.edu.au

quantities of potential meta-analytic interest might best be obtained by modeling the distribution of quantities that are reported in the literature' but which?

Primary studies ('the literature') can report findings in ways that violate underlying model assumptions and bias estimates. For example, a publication bias towards statistically significant results generates an asymmetric distribution of effect sizes that biases mean estimates upwards for nonzero true effects (Jennions *et al.*, 2013). A similar problem arises for 'quantities' that tend to go unreported when negative, such as heritability. Also, some reported 'quantities' already have distributions that violate assumptions underlying standard meta-analyses (Mengersen & Gurevitch, 2013). Unfortunately, Morrissey's three case studies all seem to vary on the 'absolute value' problem. A longer list of problematic quantities (whether reported in the primary literature or 'literature reports') could help to identify broader categories of concern. In my view, highlighting '*quantities that do not depend on the dispersion of the values reported in the literature'* is unhelpful. The follow-up suggestion to be cautious if '*the quantity of interest is an aspect of the dispersion*' is intriguing, and I do not dispute it, but the underpinning reasoning is opaque.

Morrissey's case studies are excellent reminders that conceptual problem are often associated with an incorrect or even unstated null hypothesis. The sexual antagonism case study is a great example. Whenever estimates are imprecise, secondary relationship will contain spurious pairings. Morrissey cleverly illustrates this by simulating pairs of estimated selection gradients where there is no selection on either sex. Estimates of sexual antagonism arose in 50% of cases (his fig. 3c). Simulations are indeed valuable, but you do not always need a formal simulation. Here, simply consider what happens when you toss a coin twice – in 50% of cases, you get a head (positive) and a tail (negative). It is a short leap to work out what happens with a coin that has a side bias.

Morrissey concludes with a cautionary note that meta-analysis is reducing the use of qualitative synthesis (i.e. narrative reviews). No one can dispute that individual studies can be deeply insightful. However, it is always perilous to extrapolate. Textbooks are littered with nonreplicable studies that once seemed solid. There is no alternative to quantitatively synthesizing data from multiple studies. Perhaps we should refine our inclusion criteria (based on study design *not* outcome), but that merely means we should conduct better meta-analyses. Misapplication of many statistical analyses is rife, but we do not abandon them. If so, where to for mixed models? The same reasoning holds for meta-analysis.

References

Jennions, M.D., Lortie, C.J., Rosenberg, M.S. & Rothstein, H. 2013. Publication and related biases. In: *Handbook of Meta-analysis in Ecology & Evolution* (J. Koricheva, J. Gurevitch & K. Mengersen, eds), pp. 207–236. Princeton University Press, Princeton.

Mengersen, K. & Gurevitch, J. 2013. Using other metrics of effect sizes in meta-analysis. In: *Handbook of Meta-analysis in Ecology & Evolution* (J. Koricheva, J. Gurevitch & K. Mengersen, eds), pp. 72–75. Princeton University Press, Princeton.

Morrissey, M.B. 2016. Meta-analysis of magnitudes, differences and variances in evolutionary parameters. *J. Evol. Biol.* **29**: 1882–1904.

Received 16 April 2016; accepted 20 April 2016