Supplement A: perfect reliability and Pearson's r

Figure A1 shows the bias in heterogeneity estimates in the absence of measurement unreliability in primary studies (i.e., reliability = 1).

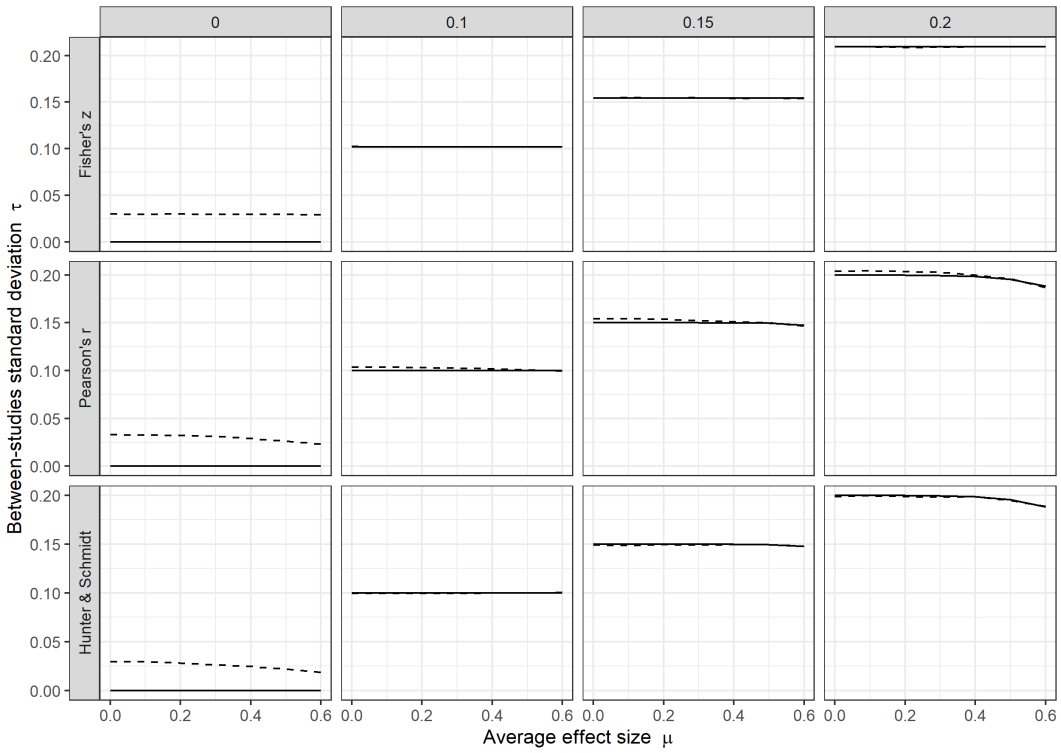

Figure A1. Bias in heterogeneity estimates in the absence of unreliability in primary studies (i.e., reliability = 1). The x-axis indicates average effect size and the y-axis estimated heterogeneity in standard deviations. Columns indicate the nominal true heterogeneity standard deviation τ. Due to truncation in Pearson's r or translation to Fisher's z these values may differ from actual true heterogeneity standard deviation (black solid lines). Code to reproduce figure: osf.io/8ygwj.

As can be seen, all three models show a positive bias when heterogeneity is zero, as heterogeneity estimates are truncated at zero. However, we also see a positive bias when heterogeneity is above zero for the Pearson's correlation using the HV method (second row). This is due to a dependancy between effect size and sampling variance for Pearson's r as normally computed. That is, the sampling variance is estimated as $\hat{\sigma}_i^2 = \frac{(1-r_i^2)^2}{n_i-1}$. We do not see this effect for the HS method or HV method using Fisher's z. In the HS method r_i has been replaced by $\bar{r}$ and for Fisher's z the sampling variance is approximated as $\hat{\sigma}_i^2 = 1/(n_i - 3)$, in both cases removing the dependency between effect size and sampling variance. To demonstrate that this is indeed the case, Figure 2 demonstrates how the bias in Pearson's r disappears if we reshuffle

the sampling variances among primary studies so that the dependency between effect size and sampling variance disappears.

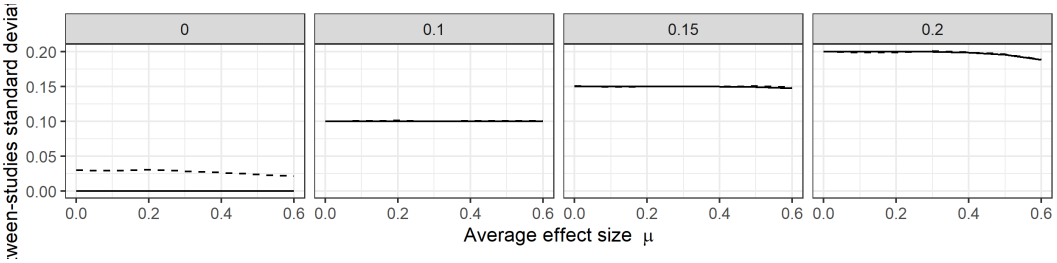

Figure A2. Removal of the dependency between effect size and sampling variance removes the positive bias in heterogeneity estimates for larger true heterogeneity when using Pearson's r. The x-axis indicates average effect size and the y-axis estimated heterogeneity in standard deviations averaged across 1e3 repetitions. Columns indicate the nominal true heterogeneity standard deviation τ. Due to truncation in Pearson's r or translation to Fisher's z these values may differ from actual true heterogeneity standard deviation (black solid lines). Code to reproduce figure: osf.io/8ygwj.