Response by Joe Hilgard (.htm) to DataColada[59]
April 12, 2017

In Data Colada [59], Simonsohn presents some simulations that present a grim picture for PET-PEESE. As applied, PET-PEESE appears to suffer from terrible downward bias, unable to tell the difference between delta = 0, a null effect, and delta = 0.8, a big effect.

Here are two ways you can make your PET and PEESE results a lot more helpful.

1. Don't build in a relationship between *d* and se(d).

PET, PEESE, Egger's test, funnel plots -- they all depend on the idea of a small-study effect. If there's a correlation between *d* and the *standard error of d*, there's a small-study effect, which all these techniques will consider a sign of publication bias.

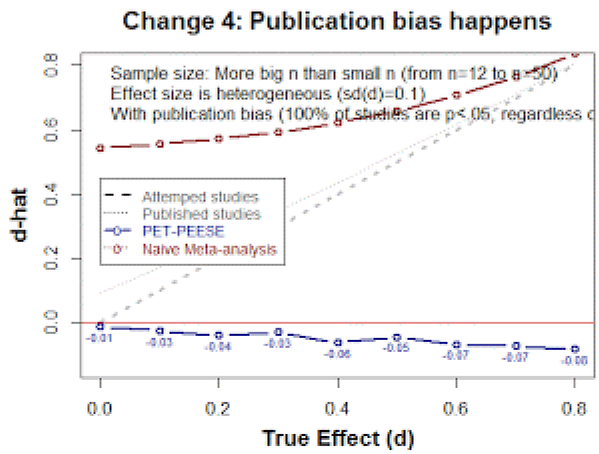
One big problem is that, for many effect size statistics, there's just such a correlation built into the math of the effect size. For example, in Simonsohn's simulation, he uses Cohen's *d*, estimating the variance as

var(d) = 2/n + d^2 / (2*(n - 3.94))

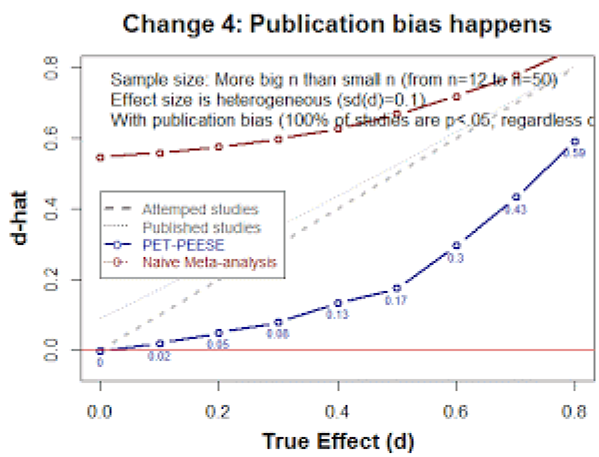
Because *d*^2 is in the numerator, using this equation leads to a positive correlation between *d* and se(*d*), even when there's no publication bias. As *d* moves away from zero, var(d) increases. This creates a spurious small-study effect, leading PET (and PEESE) to adjust downward. Scholars have written about this issue previously as a source of spurious small-study effects (Peters et al., 2006; Macaskill, et al. 2001).

You can get much better results if you use an effect size statistic that doesn't make this spurious small study effect. Here I replace d^2 with 0 in Uri's equation for the variance so that var(d) isn't related to *d*.

var(d) = 2/n



As run in Data Colada [59], with var(d) = 2/n +(d.obs^2)/(2*(n-3.94))



Run with var(d) = 2/n

As you can see, the performance of PET-PEESE is dramatically improved. Admittedly, this is kind of a ramshackle, ad hoc kludge of a fix, but the improvement in results is hard to argue with. It's for this reason that I used Fisher's Z with standard error 1/sqrt(N-3) in my meta-analysis -- it doesn't build in a spurious small-study effect.

Even after this improvement, PET-PEESE still shows downward bias. This brings me to my second suggestion...

2. Don't use the faulty logic of conditional PET-PEESE.

PET-PEESE is a two-step test composed of two estimators: PET, which is biased downward when delta > 0, and PEESE, which is biased upward when delta = 0. PET-PEESE tries to use each test when it's most effective. It first runs the PET test, then, if the PET test is significant, it switches to PEESE.

The problem is that PET, due to its downward bias, has poor statistical power. And we know that p > .05 doesn't mean that the null is true. As a result, PET-PEESE tends to inherit a lot of PET's downward bias.

My recommendation is to use PET and PEESE separately. If PET says there's a significant effect, great -- interpret the PEESE estimate. But if PET says there's no significant effect, I would still consider interpreting the PEESE estimate. Sure, maybe the effect is zero, but there's a good chance the PEESE estimate is correct. You should cover your bases and think about both.

I realize that it's dissatisfying to have two estimators and not be sure which one to use. If someone can come up with better conditional logic for PET-PEESE, I think it'd be a great manuscript.

In summary

You can make PET and PEESE perform a lot better so long as you do as I do. First, use an estimator of the standard error that is dependent only on sample size, not on effect size. Second, be aware that a non-significant PET result does not mean the effect is definitely zero. Doing these two things reduces downward bias caused by spurious small-study effects, and it keeps you from mistaking p > .05 as evidence for the null.

Big thanks to Uri for a stimulating discussion and for being patient as we send code back and forth. Science runs on criticism, and I want to see meta-analytic adjustments receive the attention and criticism they need to develop.