

Dynamic Ecology

Multa novit vulpes

What is the typical effect size of an ecological study?

Posted on [December 17, 2019](#) by [Jeremy Fox](#)

Today's weird question: what's the typical effect size of an ecological study? Like, of *anything*? Using *any* experimental or observational method?

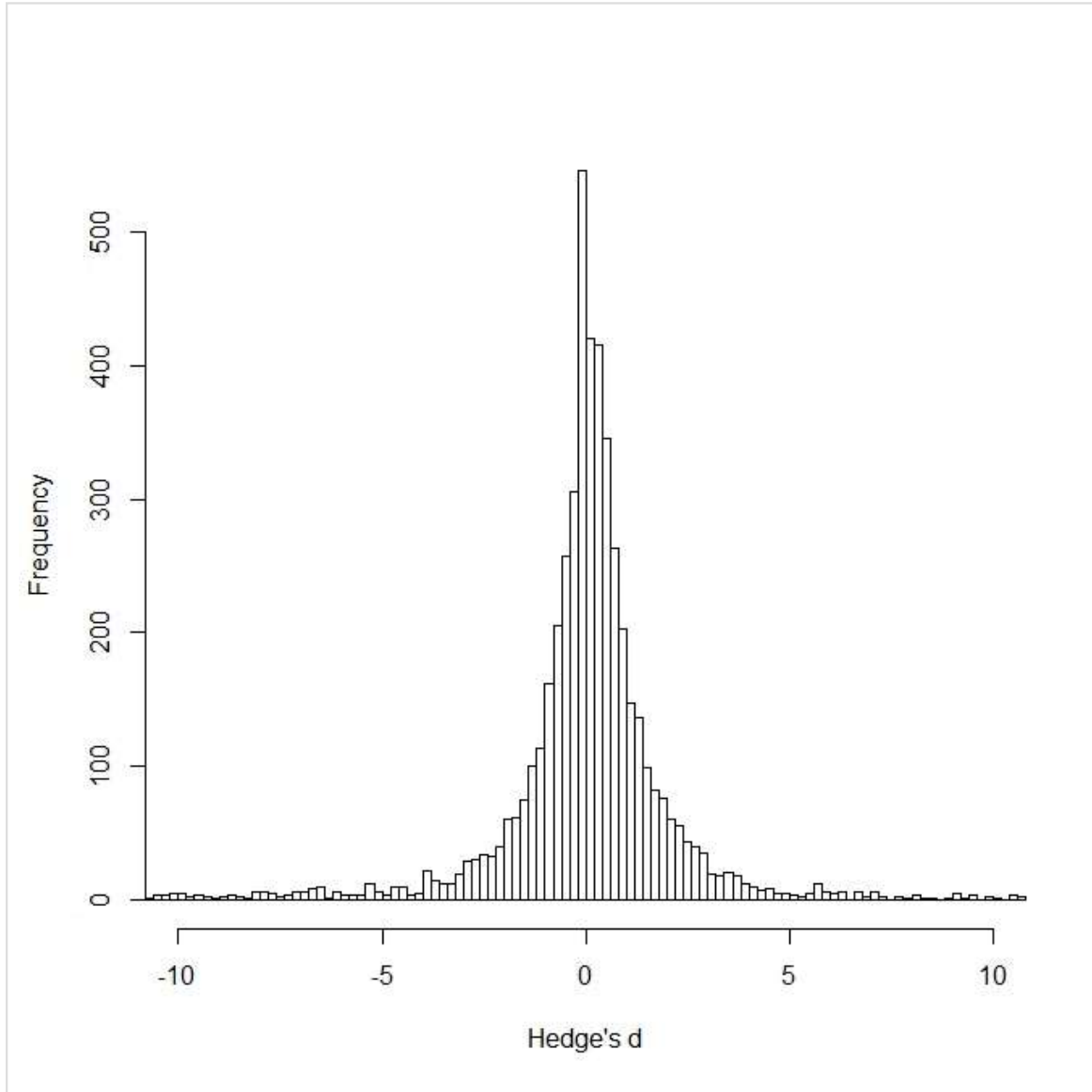
Keep reading for the answer!

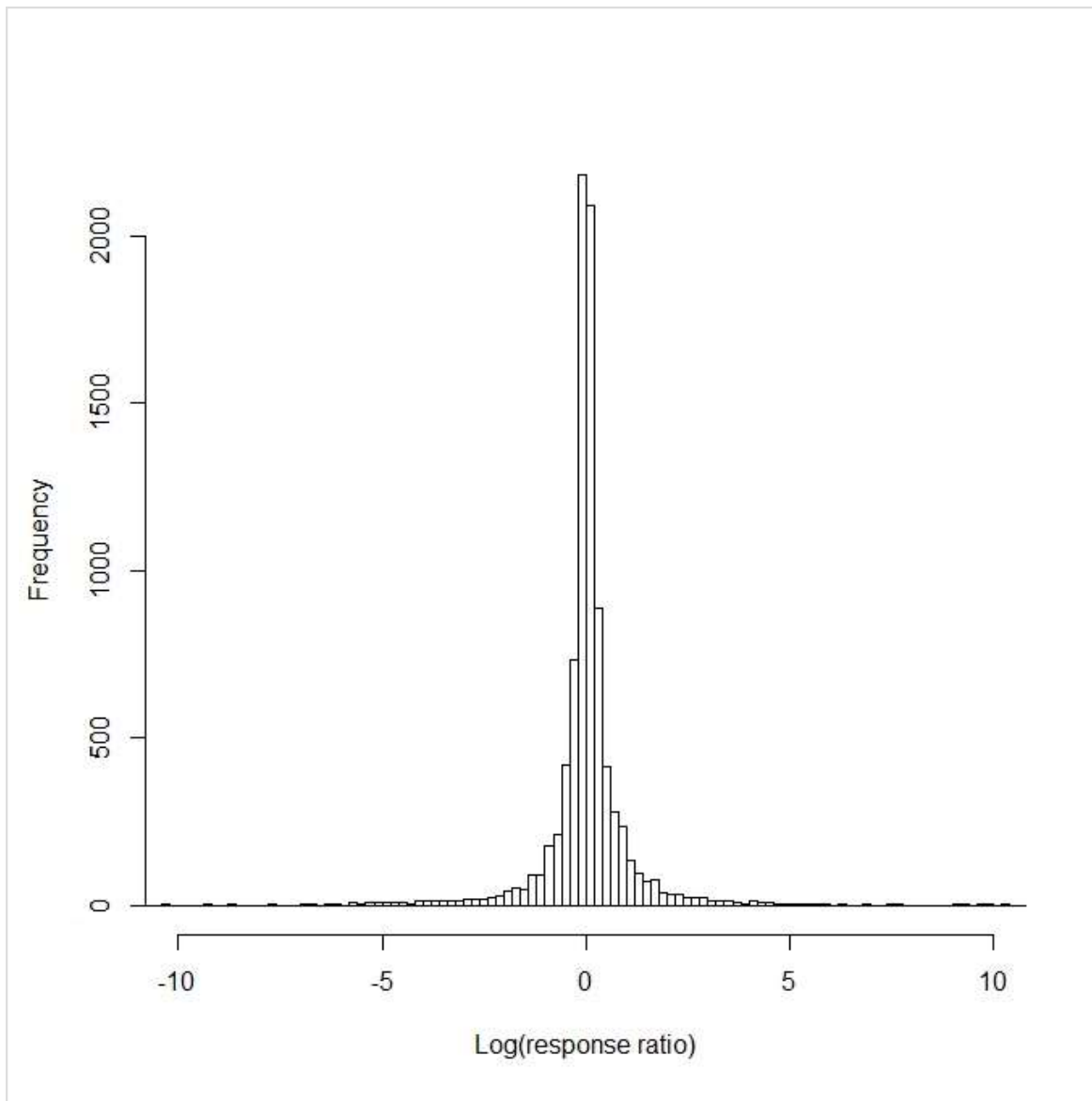
As part of a side project, I am compiling effect size data from many different ecological meta-analyses. I don't care at all about the topic or the methods. All I care about is that it was a meta-analysis about something ecological, and that the authors published a table of all the effect sizes included in their meta-analysis.*

So far I have 49 meta-analyses, on topics ranging from the effect of plant-soil feedback on plant diversity (Crawford et al. 2019) to the effects of sedimentation on various features of marine organisms (Magris & Ban 2019) to phenotypic selection on flowering phenology (Munguia-Rosas et al. 2011), and more. These 49 meta-analyses were published in a bunch of different journals, from Nature to Ecology Letters to Ecosphere and more. Most of them used one of two measures of effect size: Hedge's d , or the log-transformed response ratio. Hedge's d (which is actually the same thing as Cohen's d , right?) is the difference between two means, such as a treatment mean and a control mean, divided by the pooled standard deviation. The log(response ratio) is the log of the ratio of two means, usually a treatment mean divided by a control mean. The sign of the effect obviously depends on the (sometimes arbitrary) decision about which mean to designate as the "treatment" mean, and on the effect of the treatment, if any. But for both effect size measures, values near 0 indicate small effects in terms of absolute magnitude, and values far from zero indicate large effects. A typical ecological meta-analysis will include dozens to hundreds of measures of effect size, because it will include many papers, many of which will report multiple effect sizes (e.g., because the same study was repeated on multiple study species, or in multiple habitats, or etc.).

Obviously there are complicated issues of hierarchical data structure and non-independence here. But we don't need to worry about them to get a first-pass answer to our question.

Here are histograms of all values of Hedge's d , and all log-transformed response ratios, from all meta-analyses I've looked at so far. Note that the x-axis on each panel is scaled so as to exclude a few extreme outliers:





The first thing to notice about both those histograms is the peak at zero. The modal effect size of ecological studies is no effect. However, the second thing to notice in both histograms is the wide x-axis scale. The modal value is not *typical*; both of those histograms include a lot of big effects! Cohen (1988) famously suggested a rough rule of thumb that an absolute value of $d=0.2$ was a “small” effect, $d=0.5$ was a “medium” effect, and $d=0.8$ was a “large” effect. That rule of thumb is from psychology, and there are various issues with its interpretation (see this old [post](#) for discussion of how to define “small” effects). But FWIW, 80% of the d values in this dataset are >0.2 in absolute magnitude, 60% are >0.5 , and the median is 0.67. Turning to the log response ratios, 54% of them have absolute values >0.182 , indicating that one mean is at least 20% larger than the other. 21% of the log response ratios are >0.69 in absolute value, indicating that one mean is at least twice as large as the other.

In sum, effect size distributions in ecology are leptokurtic. They have a sharp peak at zero, but very heavy tails.** Which means that the typical effect size of published ecological studies is moderate to large in absolute magnitude, for one standard definition of “moderate to large”.

Looking forward to discussion of what to make of these results. Is it heartening to know that ecological studies often find moderate to large effects? Or is that fact just meaningless, because it's stripped of so much ecological context? In particular, one could argue that effect sizes in experiments reflect the size of the manipulation. We probably shouldn't be surprised that experimenters typically "kick the study system hard enough to hear it yell."

*If you think this sounds like question-free data-dredging, well, you're right, though I prefer the term "exploratory". 😊 Like I said, it's only a side project.

**Note that effect sizes reported in meta-analyses may slightly overstate the kurtosis for all ecological studies, because meta-analyses of experiments with several treatments (e.g., several different temperatures) often only include the effect size for the most extreme pair of treatments (e.g., lowest vs. highest temperature).

This entry was posted in [Uncategorized](#) by [Jeremy Fox](#). Bookmark the [permalink](#) [<https://dynamicecology.wordpress.com/2019/12/17/what-is-the-typical-effect-size-of-an-ecological-study/>] .



About Jeremy Fox

I'm an ecologist at the University of Calgary. I study population and community dynamics, using mathematical models and experiments.

[View all posts by Jeremy Fox](#) →

20 THOUGHTS ON "WHAT IS THE TYPICAL EFFECT SIZE OF AN ECOLOGICAL STUDY?"



James Palmer

on [December 17, 2019 at 8:16 am](#) said:

Thanks, just getting a sense of the range is interesting and useful. Do you know of similar investigations in other disciplines?



[Jeremy Fox](#)

on [December 17, 2019 at 8:18 am](#) said:

By Hattie, in education.



Brian McGill

on [December 17, 2019 at 11:45 am](#) said:

If you want to take a data-driven view to education (Primarily K-12 but some college studies in there) then https://www.amazon.com/Visible-Learning-Teachers-Maximizing-Impact/dp/0415690153/ref=sr_1_1?keywords=john+hattie&qid=1576601029&s=books&sr=1-1 is the best re-

source around. It pulls together over 500 individual meta-analyses on almost 100 different interventions in education.



James Palmer

on **December 17, 2019 at 12:06 pm** said:

Thanks Brian. These are more in line with Cohen's guidelines. See these summary charts:

<https://visible-learning.org/hattie-ranking-influences-effect-sizes-learning-achievement/hattie-ranking-student-effects/>

<https://visible-learning.org/hattie-ranking-influences-effect-sizes-learning-achievement/>

Looking for other examples.



Raphaël Proulx

on **December 17, 2019 at 9:45 am** said:

Thank you Jeremy,

Brute force data-dredging is like comfort food, it's not always healthy but it makes you feel better ;-).

More seriously, your conclusions are in agreement with a former study by Low-Décarie who showed that R-squared values in > 18,000 studies from ecology journals are typically above 0.5. Yet, between 1960 and 2010, the average R-squared value across studies has fallen from ca. 0.75 to ca. 0.5. Maybe you can show the same kind of trend with your data? The low-hanging fruits are long gone! Here is the link:

https://esajournals.onlinelibrary.wiley.com/doi/pdf/10.1890/130230?casa_token=qX-ve18gaycAAAAA:liV7zZTZK0QMLHZnGE3fpy7seRp-ATd48tIYrg2rrX1RYI_VdExs0gTb68yE-kQahw_1EF3XnZ8C6yJ

-Raphaël



Jeremy Fox

on **December 17, 2019 at 10:01 am** said:

“Brute force data-dredging is like comfort food, it's not always healthy but it makes you feel better ;-).”

That's a very seasonally-appropriate comment. 😊

I'll need to think more about how the results in this post relate to Low-Decarie et al. Just offhand, I'm not sure they do? Low-Decarie et al. is about R^2 for model fits. The effect sizes in this post aren't from model fits. And plenty of the effect sizes in this post are reasonably large; there's no sign in this post that all the low-hanging fruit has been picked. Finally, I didn't show it here, but only in 2 of these 50 meta-analyses was there even a hint of the mean effect size declining over time.

p.s. The declining R^2 trend that Low-Decarie et al. found is real, but there's a *lot* of variation around the trend. The declining trend in mean R^2 only explains a tiny fraction of the variation in R^2 values. I find the declining R^2 trend intriguing, but I'm not convinced it's something ecologists need to worry about. I don't think all the low-hanging fruit has been picked.



James Palmer

on **December 17, 2019 at 3:48 pm** said:

It could be due to the file-drawer effect.



Jeremy Fox

on **December 17, 2019 at 10:22 am** said:

At the request of a correspondent, here are the meta-analyses with the largest effect sizes on average (absolute magnitude of the unweighted mean effect size):

Hedge's d : Zhou & Staver 2019 Ecology (effects of plant invasion on soil nutrient enzymes), Magris & Ban 2019 Global Ecol Biogeog (sedimentation effects on various marine organisms, Jackson et al. 2016 Global Change Biol (effects of multiple stressors in freshwater ecosystems).

log(response ratio): Kiaer et al. 2013 Journal of Ecology (combined effects of root and shoot competition in herbaceous plants).

Just offhand, there's no obvious pattern in terms of where large effects tend to show up. It's not that effect sizes tend to run larger in experiments than in observational studies, or larger in studies that were motivated by formal mathematical theory, or whatever. At least, not to any obvious extent; maybe you'd find something if you dug into the data more.



SB

on [December 17, 2019 at 11:00 am](#) said:

If the control mean is near zero, then log(RR) becomes problematic. You record a big effect size log(RR) whenever the control is near zero. Perhaps Kiaer et al's control is near zero? Hedges-D (like the simple t-value) seems overall a better measure. Not sure, if Hedge-D can be interpreted as if it were same as Cohen-D.



Jeff Houlahan

on [December 17, 2019 at 11:18 am](#) said:

I would have guessed that sample sizes would have explained some of that variability – with large effects being associated with small sample sizes. Partly, because to get statistical significance with a small sample size will require a large effect. Also, because small sample sizes will lead to the poorest estimates of the difference between two groups and, IF, true effects tend to be small to moderate we are most likely to see large observed effects in studies with small sample sizes. Further, small sample sizes are likely to give us poorer estimates of the true variability and in the cases where the study gets the variability wrong on the low side, the true effect size will be overestimated even if the data capture the true difference in means.



Jeremy Fox

on [December 17, 2019 at 11:21 am](#) said:

Yes, sampling variation is definitely part of what's going on in the graphs in the post.



Brian McGill

on [December 17, 2019 at 11:46 am](#) said:

I don't know if you captured the data so this would be hard to plot or not, but a cross-meta-analysis funnel plot (effect size on y-axis vs sample size N on

the x-axis) would be a cool plot to see.



Jeremy Fox

on **December 17, 2019 at 11:56 am** said:

I could do it, but it would require me to go back and pull all the sample size data first...

And surely it would be...funnel shaped?



Brian McGill

on **December 17, 2019 at 3:04 pm** said:

I would hope it would be funnel shaped (or the laws of statistics will have broken down)! But it would contextualize e.g. "60% are >0.5 , and the median is 0.67" as I bet most of those >0.5 are at small sample sizes (i.e. $N=50$ is a spindle lying right on effect size=0 it makes you realize all those $d>0.5$ might not mean a lot.

But probably not worth your time if the sample sizes aren't already captured.



Jeremy Fox

on **December 17, 2019 at 3:32 pm** said:

I did compile the variances of individual effect size estimates, when the meta-analysis authors provided them.

If you just look at all 12,000ish estimates of Hedge's d for which I have variances, a plot of the variance of Hedge's d vs. Hedge's d does indeed show that yes, many big effect sizes are indeed imprecise estimates with high sampling variances.

But even if you restrict attention to the 1300ish Hedge's d values with variance <0.2 (an arbitrary choice; that's the most precise ~15% of all Hedge's d estimates), the mean absolute value of Hedge's d is 0.55 and the median absolute value is 0.4. So even restricting attention to the most precisely-estimated effect sizes, you still find a goodly percentage of moderate-to-large

effect sizes. And those precisely-estimated effects are from various meta-analyses of various topics, so this isn't an artifact of studies of one particular topic happening to all be very precise estimates of very large effects.



James Palmer

on [December 17, 2019 at 3:54 pm](#) said:

It may be useful to know that Hedges' g has a correction for small samples that may or may not be used.



Emilie Champagne

on [December 18, 2019 at 10:24 am](#) said:

Oh interesting, I wonder if you have my meta-analysis and if so, I hope the data was easy to use.

But, is it the best way to look at Hedge's d ? Shouldn't you use absolute values?



Jeremy Fox

on [December 18, 2019 at 10:32 am](#) said:

I don't have your meta-analysis but would be happy to add it. 😊 Champagne et al. 2016 Ecosphere, right? Just had a quick glance, looks like your txt file has everything I need in easy-to-extract form. Thanks!

Re: absolute values of Hedge's d , you can eyeball that distribution reasonably well from the histogram, since the histogram is symmetrical around 0.



Emilie Champagne

on [December 18, 2019 at 10:48 am](#) said:

Yes, that's the one!

Indeed, the symmetry is quite good.



Gerard

on **January 7, 2020 at 7:48 pm** said:

Hi Jeremy.

Great post. “We probably shouldn’t be surprised that experimenters typically “kick the study system hard enough to hear it yell.”. This pretty much sums up many of the issues with current stress-response ecology and slow adoption of more reliable methods that are common and discussed in other fields such as ecotoxicology. For single laboratory studies, these being: i) lack of regression analyses that allows trends to be modelled and interpolated ii) lack of useful effect-size metrics derived from the regression analysis, such as ‘effect concentrations’, and iii) not quantitatively (using percentiles) relating the effect-size metric to field measurements of the stressor.

This site uses Akismet to reduce spam. [Learn how your comment data is processed.](#)