

Freakophenia

Ben Klemens

6 August 2008

Let me start with an example. You may have read in the New York Times that obesity is contagious¹, in the sense that you're more likely to be obese if your friends are. The linked article is reporting a publication [Christakis and Fowler, 2007] from the New England Journal of Medicine (NEJM), one of the most well-regarded journals around, which retains its high regard via a press office that puts out press releases on notable articles in each issue (as do many other journals). I made a point of not closely reading the article, and not critiquing the methods; I'm fine with believing that they were good enough to pass peer review, made honest use of the data, and that the statistical significance claimed is a correct read of the data.

But from this subsequent rebuttal [E and Fletcher, 2008]: “We replicate the NEJM results using their specification and a complementary dataset. We find that point estimates of the ‘social network effect’ are reduced and become statistically indistinguishable from zero once standard econometric techniques are implemented.” That is, the results were basically an artifact of the original authors' data and methods, and statistical significance disappears upon replication.

So it goes. Maybe another study will come by and re-replicate. But right now it seems that the initial proposal was a matter of what I'd been discussing entry #005: if you have enough researchers staring at one data set—and we know there's a critical mass of researchers working on obesity and on social networks—then eventually one researcher will verify any given hypothesis.

This isn't improper behavior of any sort on the part of the authors of the original study, the NEJM, the NYT, or the many people who re-reported the results after they appeared in newspapers. But the publication system is built around the new and exciting, which is by definition the stuff that hasn't been replicated or seriously verified. After all, *New study verifies results of study that's already been out for a year* just doesn't count as news. Because of the novelty premium, it's easy to publish—and publicize—a study that seems statistically significant but happened to work out only because of luck and the volume of researchers staring at the problem.

There are some ways by which non-results can get published. In the example above, we saw a null-result rebuttal to a paper that found a positive result. That is, once a positive result appears, null results become newsworthy (in the academic sense. I don't think the NYT published anything about the failure to replicate the obesity headline). There is finally a Journal of Articles in Support of the Null Hypothesis² aimed at deal-

¹<http://www.nytimes.com/2007/07/25/health/25cnd-fat.html>

²<http://www.jasnh.com/>

ing with this very problem (which they call the “file drawer problem,” because a study that gets significant results gets published, and a study that fails to reject the null winds up in the file drawer).

In medicine, there is the funnel plot, which plots all of the p -values for a hypothesis from several studies, and then draws a theoretical symmetric funnel around the points; the gaps in the ideal funnel are assumed to be missing (i.e., unpublished) papers. This is done in medicine and not other fields because medicine has enough studies on a single question that you could do this sort of thing.

But Freakonoscience doesn't have that luxury: it's all about quirky one-off studies that have zero attempts at replication. So we don't have funnel plots, or any other easy tools to tell us what confidence to place in the results trumpeted in the headlines. As in prior episodes, even though the reported confidence levels are correct for the researcher's context, they are not correct for the reader's larger context, which should include both this one study and all those others that may or may not have been published. In the larger context, we basically have nothing.

In an episode or two, some notes on how we can respond to this problem.

References

Nicholas A Christakis and James H Fowler. The spread of obesity in a large social network over 32 years. *N Engl J Med*, 357(4):370–379, 2007.

Cohen-Cole E and J M Fletcher. Is obesity contagious? social networks vs. environmental factors in the obesity epidemic. *Journal of Health Economics*, 27(5):1382–7, September 2008.