A conversation with Ken Witwer on 02/28/13

Participants

- Ken Witwer — Johns Hopkins School of Medicine, Assistant Professor
- Alexander Berger — GiveWell, Senior Research Analyst

Note: This set of notes was compiled by GiveWell and gives an overview of the major points made by Ken Witwer. Ken Witwer was representing himself and his statements are not meant to represent his employer.

Summary

Ken Witwer has studied the frequency with which authors of microarray papers deposit the data that the papers report on, as the standards state they should.

GiveWell spoke with Ken Witwer as a part of our investigation of the cause of meta-research. The subjects discussed included data sharing in science, journals and peer review, and replications.

Data sharing

Biomedical researchers often don't deposit the data from their experiments. Furthermore, when researchers ask other researchers for their data, the researchers who are asked for their data often unable to present it because they've lost it. When Dr. Witwer requests data from other researchers, they give it to him less than 50% of the time.

Dr. Witwer published a study in *Clinical Chemistry* examining 127 papers that had been published in between July 2011 and April 2012 in journals that ostensibly require that researchers deposit their microarray data. He found that the data was not submitted for almost 60% of papers, and that data for 75% of papers were not in a format suitable for replication .

It would be desirable for journals to publish researchers' data sets and classify a data set as a publication distinct from the paper reporting on the data set. This could increase the prestige associated with publishing data sets.

Alternatively, databases could try to act more like journals. If databases were to include interviews with the researchers who complied data sets and issue press releases about useful and impressive data sets, then researchers who compiled the data sets would get visibility from publishing data sets, and the prestige associated with publishing data sets might increase.

Journals and peer review

The paucity of information on journals' replication rates

Prestigious journals appear to have low replication rates, and their replication rates may even be lower than average. There don't appear to be any systematic efforts to determine which journals have higher replication rates than others.

If a journal's prestige were tied to the frequency with which papers that it publishes are reproducible, that would improve researchers' incentives to produce correct findings.

It would be hard to research different journals' replication rates, because it would cost a lot to replicate even a small fraction of studies that have been done.

Open peer review vs. closed peer review

Because peer reviewers are anonymous, they don't have external incentives to do a careful job with their reviews. Sometimes a reviewer will even assign the review to a postdoc or graduate student who's not qualified to do the review. If peer review were open, reviewers would have a stronger incentive to do careful reviews. It would be interesting for journals to experiment with having some papers being reviewed by a publicly known reviewer.

A downside of open peer review is that a reviewer might refrain from writing an accurate (negative) review if the reviewer is intimidated by the paper's authors.

Thoroughness of the review

While some reviewers don't look at the data from the study that the paper reported, some do. The reviewers who look at the data may not go through all of the steps that the authors took, but they sometimes do plausibility checks of whether the analysis is internally consistent. For example, the reviewer can check whether the number of subjects that the authors report on in the paper is the same as the number of subjects in the data. Sometimes these two figures are very different, and catching such instances can help reduce the number of inaccurate papers that are in the literature.

Open access journals vs. non-open access journals

The United States government has required that papers in biomedical research be put into the public domain within 12 months of publication. For this reason, it's not as important for journals to be open access as one would expect.

On average, non-open access journals seem to have a better peer review process

and more careful staff than open access journals do, but this is not always the case.

Replications of biomedical research findings

The number of replications that are done

The fraction of research findings that have never been replicated is very large, often simply because of a lack of effort to do so.

Recently some journals, such as PLOS ONE, have been willing to consider publishing papers that report on replications. Most journals used to reject such papers out of hand.

Incentives to do replications

It's currently the case that scientists' reputation is based on their ability to come up with ideas and do experiments that provide evidence for them. Doing replications of other researchers' studies has low prestige.

If funders were to give positive weight to whether researchers had done replications when deciding whom to give grants to, the frequency with which researchers do replications might increase as a result. Because the National Institutes of Health (NIH) funds the majority of biomedical research, in order for this to have a substantial impact, the NIH would probably have to do it.