# A conversation with Stanley Young on 03/18/13

## Participants

- Stanley Young — Assistant Director of Bioinformatics, National Institute of Statistical Science (NISS)

- Alexander Berger — Senior Research Analyst, GiveWell

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

## Summary

Stanley Young is a statistician who studies issues with multiple testing and data mining in biomedical research.

GiveWell spoke with Dr. Young as part of our investigation of opportunities to improve biomedical research. We discussed low reproducibility rates, multiple hypothesis testing, data sharing and how academic incentives are not aligned to achieve highly reproducible results.

## Factors which cause low reproducibility

It appears that about 80%-90% of published epidemiological research results don't replicate when they are subjected to randomized controlled testing, despite the fact that they are ostensibly statistically significant at the 5% level. This problem arises from a misalignment of academic incentives with careful scientific practice.

Improving reproducibility from the 10%-20% level to the 20%-40% level would greatly improve scientific productivity.

### Pressure to find statistically significant results

*The Problem*

Academic researchers are rewarded primarily for their publication record, and journals generally only publish papers that find results that are statistically significant at the 5% level.

Thus, researchers are motivated to find apparently statistically significant results, and this often leads them to *data mine*—examine many possible relationships between the variables studied—and find many relationships that are in fact present by chance.

*An idea for addressing the problem*

Lobbying journals to publish negative results and non-statistically significant results more often would reduce the pressure that scientists feel to data mine.

**The paucity of open data**

*The problem*

When scientists produce a data set, they generally want to use it to generate as many publications as possible. If they make the data set public, there is a risk that other scientists will use the data set to write papers that those who produced the data would otherwise have gotten credit for. Researchers also generally don't get academic career credit for sharing data sets.

For these reasons, biomedical researchers usually don't share the data associated with their papers. Dr. Young estimates that the frequency with which scientists don't share data associated with a given paper is 67% – 90%. He recently requested 50 data sets from papers about air pollution and was unable to obtain any of them.

If it's not possible to access the data associated with a paper, it's not possible to check to see whether the analysis was done correctly, and not possible to check for whether the authors engaged in multiple hypothesis testing or data-mining.

While the National Institutes of Health (NIH) has an official policy of requiring grantees to share their data, they don't have the legal authority to compel it. So researchers generally don't share their data sets.

*Ideas for how to address the problem*

Commissioning the creation of data sets for the scientific community to use that are paid for upon being released could incentivize speedy passage of data into the public domain.

One could also lobby the NIH and other funders of researchers to be stricter about requiring that the researchers who they fund deposit their data. The Office of Science and Technology Policy (OSTP) at the White House recently directed federal agencies that spend more than $100 million/year on research and development to require that the data associated with these papers be made public. If this policy goes into effect, it could reduce the problem.

**Lack of funding for and lack of opportunity to publish replications**

There is little funding available for researchers to replicate studies.

One reason for this is that replications are less exciting than new findings, so that funders gravitate toward funding research that leads to new findings rather than research that confirms or disconfirms existing findings. Another reason is that grant committees sometimes have members whose reputations might be damaged by discoveries that certain results don't replicate, and who therefore try to prevent replications from taking place.

Because researchers do the science that they're funded to do and which they can write papers about, researchers have little incentive to replicate published findings unless they want to use them for their own research. Even if they do find that a study doesn't replicate, they usually lack a venue to publish the finding in, because journals generally don't publish replications.

### Reputational risk of conducting replications

Scientists who publicize findings that studies don't replicate are subject to the risk developing a negative reputation in the scientific community on account of other scientists being concerned that those who publicize such findings will eventually find that their own studies don't replicate. This is especially a problem for young researchers who are at risk of experiencing reduced prospects for hiring and tenure.

### Journal editors paying insufficient attention to data mining

*The Problem*

Journal editors will sometimes be unaware of data mining as a problem, or ignore the issue owing to as reluctance to believe that the papers published in their journals are flawed.

*Ideas for a solution*

One could attempt to educate journal editors by:

- Educating statisticians about the issues of multiple hypothesis testing and data mining, for example by creating a graduate statistics course on the topic.

- Encourage statisticians (perhaps by providing financial incentives) to study whether studies used data mining, and write letters to the editors of the journals in which they're published to alert them of this when they find examples.

- Run a summer workshop about multiple hypothesis testing and data mining.

## Drug company clinical trials vs. other clinical trials

Drug companies have incentives to conduct careful clinical trials examining the efficacy of a drug. The Food and Drug Administration (FDA) requires that drug companies deposit all of the data from the trials, and that they test only one or two hypotheses (so as to guard against data mining). Doctors are often informed consumers of drugs that they prescribe, and drug companies want to do a careful job with the clinical trials so as to favorably impress them. For these reasons clinical trials that are done by drug companies are often reliable.

Clinical trials funded by the NIH are less reliable since they're typically not subject to the same kind of FDA oversight, and can therefore suffer from multiple testing problems.

## People for GiveWell to talk to

*Keith Baggerly and Kevin Coombes* — Statisticians at M. D. Anderson Cancer Center. Dr. Baggerly and Dr. Coobmes did an in-depth investigation of influential cancer research by Anil Potti, and found that it was flawed.

*Glenn Begley* — Former Head of Hematology and Oncology Research at the drug company Amgen. Dr. Begley attempted to replicate 53 landmark publications in basic cancer research and found that 47 of them did not replicate.

*All GiveWell conversations are available at* http://www.givewell.org/conversations