

01 Testimony

Words : 628

Title: Make data used in federally funded research publicly available
S. Stanley Young, PhD, FASA, FAAAS
National Institute of Statistical Sciences

“As an abstract principle, the sharing of research data is a noble goal and meets with little opposition. However, when data sharing is attempted in a particular circumstance, the conflicting interests of the parties can thwart the exchange.” so said Joe Cecil of the Justice Department in 1985. What is the current status of science in general and data availability in particular? First, where are we with science claims? In 2005, John Ioannidis published two papers of interest. In one he asserted that 90% of the claims made in science papers are wrong in the sense that they are not expected to replicate. In another he noted that 5/6 papers based on observational data failed to replicate. I published a paper in 2011 and showed that of 52 hypotheses suggested from observational studies none replicated in the expected direction and five were statistically significant, but in the opposite direction. Begley and Ellis (2012) reported that 47/53 claims made in major science journals failed to usefully replicate.

Where are we with data sharing? Ioannidis (2011) selected 10 papers each from the 50 highest impact journals, NEJM, Nature, Science, etc. and asked, Is the data used in these papers publicly available? “Overall, only 47 of 500 papers (9%) deposited full primary raw data online. None of the 149 papers not subject to data availability policies made their full primary data publicly available.” I report on two personal experiences. Dr. Beate Ritz, UCLA, made a claim in Environmental Health Perspectives (2012) that air pollution in LA county leads to low birth weights. Dr. Frederica Perera (2009) of Columbia University asserted in the journal Pediatrics that air pollution decrease IQ in children. NIEHS provided funding for both studies. In both cases I asked for the data sets from the authors and also asked for help from NIEHS and resorted to FOI. I received neither data set. Recently, I was informed that NIH/NIEHS does not have the legal authority to compel an author to provide data that was funded by them. Operationally NIH funding, the Shelby amendment, etc. mean very little with respect to data availability. Mostly authors do not provide data sets used in their publications.

It is technically very easy to share data used in publications. Others will discuss “Reproducible Research,” provide study protocol, statistical analysis code and an electronic copy of data sets use in the paper. There are technical methods for dealing with de-identifying people.

Just why are we in this situation, where most claims do not replicate and authors will not make data sets available? In a long and illustrious career, W. Edwards Deming made the point that if a system is failing it is not the workers’ fault. The fault is with management, in this case funding agencies and journal editors. For over 30 years, workers have been admonished to do their work in better ways and to make their data sets available. It was

reported in Science in 1988 that there were serious problems with observational studies. Nothing has changed in 25 years.

Congress, funding agencies and journal editors need to step up and manage the scientific process. They should require authors to deposit study protocol, statistical analysis code and data sets on publication of their paper. Funding of data set construction and analysis should be separate. They should require data analysis strategies that demonstrate reproducibility. For example, any claim should be replicated in a separate data set before publication. Remember current scientific claims only replicated only 10 to 20% of the time.

John Holdren on 22Feb2012 of the Office of Science and Technology Policy issued a memorandum, "Expanding Public Access to the Results of Federally Funded Research." This memorandum should be supported legislatively by requiring data availability for papers cited in support of rule-making.

Appendix I: Proposed laws

1. Use of Science Transparency Act

Any federal agency proposing rule-making or legislation shall specifically name each document used to support the proposed rule-making or legislation and provide all data used in said document for viewing by the public.

2. Federal Study Transparency Act

If federal funds are provided for a study, all data relating to the reporting of results of said study must be provided for scrutiny by the public at the time of publication.