

One-page Summary Young

Testimony of Committee on Science, Space and Technology, 5 March 2013

Scientific Integrity and Transparency

S. Stanley Young, PhD, FASA, FAAAS

Integrity and transparency are two sides of the same coin. Transparency leads to integrity. Transparency means that study protocol, statistical analysis code and data sets used in papers supporting regulation by the EPA should be publicly available as quickly as possible and not just going forward. Some might think that peer review is enough to ensure the validity of claims made in scientific papers. Peer review only says that the work meets the common standards of the discipline and on the face of it, the claims are plausible, Feinstein, Science, 1988. Peer review is not enough.

The current state of science is arguably very poor. For medical observational studies over 80% of initial claims failed to replicate, Ioannidis, JAMA, 2005, Young and Karr, Significance, 2011. Environmental epidemiology studies are likely no better. Recent evidence indicates that experimental studies also fail to replicate close to 90% of the time, Begley and Ellis, Nature, 2012. Scientific fraud is common in retracted science papers, Fang et al., PNAS, 2012. So the evidence is that science claims usually fail to replicate and that fraud is being committed. Again, the current state of science is arguably poor. Promoting transparency is a key solving problems of validity and integrity.

Three things make sense. Going back in time, key regulations and the papers used to support them should be identified. The EPA should secure copies of data used in those papers and make the data public. For example, a number of papers on air pollution and mortality use the ACS CPS II database. Data used in the Harvard Six Cities study is not public. Over 100 key air pollution data sets were identified and a request for data was sent to 50 randomly selected authors. Most of these papers were funded by government grants. These authors provided no data sets, 0 for 50. Without data sets, there is essentially no opportunity for scientific oversight and there is opportunity for fraud.

The EPA should secure a copy of the ACS CPS II, Harvard and other key data sets and make them public. Where data sets are not available, claims in those papers are essentially "trust me" science. The EPA should not be relying on trust me science. The recent OSTP memorandum, Holdren, 22 February 2013, on data access should be supported with legislation: Regulatory agencies can not cite or use papers to support rules or regulations without the underlying data available. Basically, EPA funded authors should follow the guidelines for "reproducible research".

Going forward the EPA should fund collection and analysis of data separately. One group should have the incentive to provide a high quality data set. The analysis group can then provide a high-quality analysis of the data, knowing that others have an opportunity to do the same, oversight. If data building and analysis are funded together, there is a natural tendency authors not to share the data until the last ounce of information is extracted. Without public access to the data, transparency, initial flawed analysis or even outright fraud can not reliably be detected or corrected. The public, Congress, and the EPA should all want an efficient science process to support sound regulations. It can take years to overturn a claim that is wrong. Making data available by implementing these steps would be big steps toward improving the science process at the EPA and promoting scientific integrity.