Young One-page Summary

Testimony of Committee on Science, Space and Technology
Fostering Quality Science at EPA
S. Stanley Young, PhD, FASA, FAAAS

Openness and transparency are at the core of the scientific process. Science and technolgy have been the major drivers for social change, improvements in living standards and improved health for humans over the last 250 years.

To support openness, data sets used in papers supporting regulation by the EPA should be publicly available as quickly as possible. It is just good science to have data used in papers public.

Some might think that peer review is enough to ensure the validity of claims made in scientific papers. Peer review is not enough. Peer review only says that the work meets the common standards of the discipline and on the face of it, the claims are plausible. Scientists doing peer review essentially never ask for data sets and subject the paper to the level of examination that is possible by making data electronically available. Also, the evidence is that many claims made in peer reviewed journals don't hold up. For medical observational studies over 80% of initial claims failed to replicate. Environmental epidemiology studies are likely no better. The scientific process is back and forth. This process would be faster and more efficient with data sets publicly available.

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. The EPA should secure a copy of this data set and make it 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. Using taxpayer dollars, the EPA supports current research. As papers are published on this research, authors should provide, at time of publication, three things: the study protocol, the statistical analysis code, and an electronic copy of the data set used in the publication. Basically, EPA funded authors should follow the guidelines for "reproducible research". Finally, going forward the EPA should fund data colletion and analysis separately. Data collection and staging, the building of data sets, is a distinct area requiring different skill sets from data analysis. Each could be done more efficiently if done separately. If data building and analysis are together, there is a natural tendency authors not to share the data until the last ounce of information is extracted. It would be better to open up the analysis to multiple teams of scientists.

The public, Congress, and the EPA should all want an efficient science process to support sound regulations. Making data available by implementing these three steps would be big steps toward improving the science process at the EPA.