

STATEMENT OF

LOUIS ANTHONY (TONY) COX, JR., PH.D.

tcoxdenver@aol.com

CHIEF SCIENCES OFFICER

NEXTHEALTH TECHNOLOGIES

www.nexthealthtechnologies.com

ON

ENSURING OPEN SCIENCE AT EPA

BEFORE THE

U. S. HOUSE SCIENCE COMMITTEE

SUBCOMMITTEE ON ENVIRONMENT

FEBRUARY 11, 2014

Chairman Schweikert and Members of the Subcommittee, thank you for inviting me to discuss whether regulatory science and the data on which it rests should be made openly available. I am testifying on my own behalf today, in support of the proposed Secret Science Reform Act. I need access to data for my work on health risk assessment, and am grateful for this opportunity to explain why. I have provided the Committee members with a detailed CV describing my academic, publishing, and consulting affiliations.

We are discussing a question of great current public, policy, and scientific interest: *Is the public interest well served by requiring that data used to support policy decisions be made available to those who want to see it?* Many who argue yes believe that the very essence of good science is reproducibility of results, and sharing of the observations and data that are said to drive them (Cox, 2009, p. 5). For these people, openness to scrutiny is a hallmark of sound scientific reasoning, and a prerequisite for sound scientific process and for trustworthy conclusions. Many scientists and analysts themselves are of this persuasion. For example, a recent survey of three professional societies involved in risk assessment found that “69 percent said it was ‘very important’ to have access to the underlying raw data for the most critical studies in order to do their own independent analysis of the results.” However, “only 36 percent said that having this access was often or always the case” (Butterworth, 2013). The proposed Secret Science Reform Act will help to address that gap.

Those who oppose requiring open sharing of data used to support regulations and policies typically cite several concerns (e.g., Neutra et al., 2006; Pearce and Smith, 2011). One is that the process might be abused by unscrupulous parties. Like the tobacco industry, others might seek to “manufacture doubt” to obscure the clear implications of good science and to delay socially beneficial actions by proposing alternative, inferior analyses. A second concern is that divulging data might threaten the privacy of individuals included in study populations. A third concern is that requiring data to be shared might prove burdensome for the original investigators, exerting a chilling effect on research in the public interest.

To these three objections, taken in reverse order, it might be replied that, first, the habit of keeping well-organized and documented records, data, and lab notebooks in expectations that others will use them later to try to independently reproduce and verify important claimed findings is – or should be – part of the training of every good scientist. No extraordinary burden is imposed by such good practices, Transparency of data and methods and scrutiny of results by others, perhaps using different methods, is something that scientists should expect and welcome. There is also much that scientific journals can do, and are doing, to encourage data transparency and to facilitate making documentation of data, models, and analyses readily available to those who want to use them.

Second, the concern that making study data available could threaten the privacy of individuals rests on an important, but purely technical, statistical issue: *Do statistical data in fact allow individual attributes or facts that should be protected to be discovered?* This technical problem is best addressed by technical solutions, and many excellent ones are now available to allow statisticians to do valid analyses while protecting individual data (Reiter, 2009; Klein et al., 2013). These methods, such as multiple imputation, have already been extensively developed, tested, and successfully applied, at the Census Bureau and elsewhere. So, I think that this concern should be viewed as a bit of a red herring: appropriate technical methods to handle it are already available and are being used in other areas.

But the most important concern, I suspect, is often not technical. It is about human behavior, and incentives, and the sociology of science. This is the concern that bad people will delay good regulations and remedial actions by misusing data and performing untrustworthy analyses to mislead the public and policy makers (Neutra et al., 2006). Such concerns have long been expressed about the use of risk analysis and technical analysis more generally (Silbergeld, 1993). To address them, I think we must candidly assess how well the scientific process delivers trustworthy results without much pressure from independent examination and reanalysis of data. It does not. We are now living in an age of catastrophic failure in the reproducibility and trustworthiness of scientific results, as witnessed by articles such as “Why most published research articles are false” (Ioannidis, 2005), “Trials and errors: Why science is failing us” (Lehrer, 2011), and “Beware the creeping cracks of bias” (Sarewitz, 2012). In the January 17th

issue of *Science* magazine this year, Editor-in-Chief Marcia McNutt noted that a worrisome proportion of peer-reviewed published results are not reproducible, and she announced plans to expand their editorial board, with advice from the American Statistical Association, “to ensure that manuscripts receive appropriate scrutiny in their methods of data analysis.” A common theme is that there is too much pressure on original investigators to use dubious statistical methods to publish results that are sensational but not necessarily true (false positives), and there is not enough encouragement for investigators to do high-quality, reproducible research, with the confident expectation that others will soon be looking over their shoulders and reanalyzing their data, perhaps using less biased methods. To fix what is manifestly broken takes more scrutiny and greater access to data, not less. As for the very legitimate fear that those who disagree with us might use open access to data and reanalyses to confuse and delay actions that we favor, this has been part of the cost and a great part of the benefit of free, democratic societies since well before John Stuart Mill wrote, in *On Liberty*, that “Wrong opinions and practices gradually yield to fact and argument: but facts and arguments, to produce any effect on the mind, must be brought before it. ... The beliefs which we have most warrant for, have no safeguard to rest on, but a standing invitation to the whole world to prove them unfounded.” The best defense against unscrupulous use or motivated interpretations of data – whether from regulators or from industry or from anyone else – is to make it openly available, so that the grounds of debate turn from who is privileged to see the facts to how one should best interpret them.

Let me end with two recent examples from my own experience in public health risk analysis. First, the public availability of the National Mortality and Morbidity Air Pollution Study (NMMAPS) data set recently allowed me to apply econometric tests for potential causality to air pollution and mortality data from 100 U.S. cities. An unexpected finding was that, although levels of air pollution are significantly associated with levels of elderly mortality rates (and both are associated with cold winter days), there is no evidence that reductions in air pollution levels have caused any reductions in mortality rates (Cox, 2012). This was a new finding from old data, using methods that other investigators had not

tried. It may be important information for policy-makers to consider. I hope that others will repeat and improve upon my analysis. Without open access to the data, that would not be possible.

Second, in 2012, Dublin extended bans on coal burning (DECLG, 2012) because of research (Clancy et al., 2002), done in part by U.S. investigators who have prominently shaped U.S. EPA beliefs, assuring them that cutting coal-burning had promptly and obviously reduced mortality rates, especially cardiovascular deaths (Harvard School of Public Health, 2002). A closer look at the data in 2013, funded by the Health Effects Institute, revealed that this was not true: these mortality rates did not decrease any faster where coal burning was banned than where it wasn't (HEI, 2013). The original investigators had not accounted for the general trend that mortality rates were coming down all over Ireland and Europe, due to better diagnosis, prevention, and treatment. Instead, they had simply misattributed that trend around Dublin to effects of the coal-ban (Cox, 2012). This mistake was ultimately fixed in 2013, after the bans had already been extended, when the Health Effects Institute paid one of the original investigators to go back and consider control groups. Although methodologists and risk analysts had already noted years ago that the fact that both pollution levels and mortality rates have declined over time does not warrant an inference that reducing one reduces the other (Wittmaack 2007; Pelucchi et al., 2009; Cox, 2012), without access to the original data, they could not quickly and easily show that the original conclusions did not follow from the data. That had to wait until the original investigators were funded by HEI to try again more carefully. And by then, Irish public policy, based on a mistaken belief about the human health benefits to be expected from extending the bans, had already been made (DECLG, 2012).

We need not repeat such experiences here. We can choose to make the data available and to invite methodologists to take a look. Whether reducing current and recent past levels of air pollution should be expected to cause any reductions in mortality rates, and if so by how much, remains a great unanswered question – unanswered, that is, by sound science and statistical analysis of data. Today, answers are often simply assumed, without sound factual support, for purposes of regulatory benefits calculations (Cox, 2012, Chapter 7). It is possible and desirable to do much better. To do so requires

making original data open for others to analyze, and not to wait until policy has been made and changes enacted before allowing the public to find out whether better analyses would have led to different results.

Thank you for your attention.

REFERENCES

- Butterworth, T. Politics, Environmentalism Beating Out Science In Regulating Risk Say Experts. <http://www.forbes.com/sites/trevorbutterworth/2013/12/19/politics-environmentalism-beating-out-science-in-regulating-risk-say-experts/>. See also http://www.isrtp.org/GMU%20WEBINAR_DEC_2013/GMU%20Study%20Document4.pdf
- Cox, LA, Jr. 2009. *Risk Analysis of Complex and Uncertain Systems*. Springer, New York.
- Cox, LA Jr. 2012. *Improving Risk Analysis*. Springer, New York. www.amazon.com/Improving-Analysis-International-Operations-Management/dp/1461460573
- DECLG Department of the Environment Community and Local Government, March 9, 2012. New Smoky Coal Ban Regulations will bring Cleaner Air, Fewer Deaths and can help efficiency <http://www.environ.ie/en/Environment/Atmosphere/AirQuality/SmokyCoalBan/News/MainBody.31034.en.htm>. Last Retrieved 1 February 2014.
- Ioannidis JPA. Why most published research findings are false. 2005. PLoS Med 2(8): e124. doi:10.1371/journal.pmed.0020124
- Klein M, Mathew T, Sinha B. 2013. [A Comparison of Statistical Disclosure Control Methods: Multiple Imputation Versus Noise Multiplication](http://www.census.gov/srd/papers/pdf/rrs2013-02.pdf). Center for Statistical Research & Methodology Research Report Series (Statistics #2013-02). U.S. Census Bureau. Available online at <http://www.census.gov/srd/papers/pdf/rrs2013-02.pdf>
- Lehrer J. Trials and errors: Why science is failing us. Wired. January 28, 2012. <http://www.wired.co.uk/magazine/archive/2012/02/features/trials-and-errors?page=all>
- Mill, JS. 1869. On Liberty. <http://www.bartleby.com/130/2.html>
- Neutra RR, Cohen A, Fletcher T, Michaels D, Richter ED, Soskolne CL. 2006. [Toward guidelines for the ethical reanalysis and reinterpretation of another's research](http://www.euro.who.int/en/ViewArticle.aspx?pid=3267). Epidemiology. May;17(3):335-8.
- Pearce N, Smith AH. 2011. [Data sharing: not as simple as it seems](http://www.elsevier.com/locate/S092464601100107). Environ Health. Dec 21;10:107.
- Pelucchi C, Negri E, Gallus S, Boffetta P, Tramacere I, La Vecchia C. [Long-term particulate matter exposure and mortality: a review of European epidemiological studies](http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2804448/). BMC Public Health. 2009 Dec 8;9:453.
- Reiter, JP. 2009. [Multiple imputation for disclosure limitation: Future research challenges](http://www.springerlink.com/openurl.do?fulltext=10.1007/s11464-009-9123-1). Journal of Privacy and Confidentiality. 1(2): 223-233
- Sarewitz D. 2012. [Beware the creeping cracks of bias](http://www.nature.com/news/2012/0510/briefing0510149a). Nature. 10 May 485:149
- Silbergeld EK. 1993 [Risk assessment: the perspective and experience of U.S. environmentalists](http://www.elsevier.com/locate/S0924646093900104). Environ Health Perspect. June; 101(2): 100-104.
- Wittmaack K. [The big ban on bituminous coal sales revisited: serious epidemics and pronounced trends feign excess mortality previously attributed to heavy black-smoke exposure](http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1844448/). Inhal Toxicol. 2007 Apr;19(4):343-50.