Basing Rules on Empirical Evidence: Transparency in Law Making

David S. Levine
School of Law
Elon University
Fellow
CITP

Victoria Stodden
Graduate School of Library and Information Science
University of Illinois at Urbana-Champaign

Center for Information Technology Policy Seminar
Princeton University
February 24, 2015
Agenda

1. Framing the question: the rise of evidence-based policy making

2. What standards for scientific evidence are permitted? required?

3. What are the issues? What’s being done?

4. Examples
Evidence-Based Policy Making

• We assert that the use of scientific findings is and will be increasingly used to inform law, policy, and regulation creation in government.

• Our questions: Should standards adhere to scientific findings used as the basis for policy making? What should they be?
Current Federal Standards

• Administrative Procedure Act

• Data Quality Act

• Freedom of Information Act

• Proposed Legislation

• Public Access to Data from Federally Funded Research: Provisions in OMB Circular A-110
Guiding Principles

• Disclosure of use: no “unknown unknowns”

• Transparency of the role findings played in rulemaking

• Scientific reasoning transparency; Public data / code availability

• Credibility of results / Independent verification

• Germaneness / Recency

• Completeness (representing the entire body of research)

• Peer review / publication

• [Public input?]
Public Data / Code: Disclosure

- Disclosure: What should be made available?
  - article (ideal: article markup with exact findings used)
  - protocols
  - experimental design
  - raw data
  - processing and inference steps (perhaps in software)
Public Data / Code: Access

Access: How should it be made available?

- Transparency from the Office of Science and Technology policy
- Transparency from Science Funding Agencies
- Transparency from Scientists
Open Science from the Whitehouse

- Feb 22, 2013: Executive Memorandum directing federal funding agencies to develop plans for public access to data and publications.

- May 9, 2013: Executive Order directing federal agencies to make their data publicly available.

Executive Memorandum: “Expanding Public Access to the Results of Federally Funded Research”

- “Access to digital data sets resulting from federally funded research allows companies to focus resources and efforts on understanding and exploiting discoveries.”

- “digitally formatted scientific data resulting from unclassified research supported wholly or in part by Federal funding should be stored and publicly accessible to search, retrieve, and analyze.”

- “digital recorded factual material commonly accepted in the scientific community as necessary to validate research findings”

- “Each agency shall submit its draft plan to OSTP within six months of publication of this memorandum.”
Executive Order: “Making Open and Machine Readable the New Default for Government Information"

• “The Director … shall issue an Open Data Policy to advance the management of Government information as an asset”

• “Agencies shall implement the requirements of the Open Data Policy”

• “Within 30 days of the issuance of the Open Data Policy, the CIO and CTO shall publish an open online repository of tools and best practices”
Request for Input: “Strategy for American Innovation”

- “to guide the Administration's efforts to promote lasting economic growth and competitiveness through policies that support transformative American innovation in products, processes, and services and spur new fundamental discoveries that in the long run lead to growing economic prosperity and rising living standards.”

- “(11) Given recent evidence of the irreproducibility of a surprising number of published scientific findings, how can the Federal Government leverage its role as a significant funder of scientific research to most effectively address the problem?”
Legal Barriers to Disclosure

- Copyright
- Patents / Bayh-Dole
- HIPPA
- FERPA
- ...
Legal Barriers: Copyright

- “To promote the Progress of Science and useful Arts, by securing for limited Times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries.” (U.S. Const. art. I, §8, cl. 8)

- Original expression of ideas falls under copyright by default (papers, code, figures, tables..)

- Copyright secures exclusive rights vested in the author to:
  - reproduce the work
  - prepare derivative works based upon the original
  - limited time: generally life of the author +70 years

- Exceptions and Limitations: Fair Use.
Responses Outside the Sciences 1: Open Source Software

- Software with licenses that communicate alternative terms of use to code developers, rather than the copyright default.

- Hundreds of open source software licenses:
  - GNU Public License (GPL)
  - (Modified) BSD License
  - MIT License
  - Apache 2.0 License
  - ... see http://www.opensource.org/licenses/alphabetical
Responses Outside the Sciences 2: Creative Commons


- Adapts the Open Source Software approach to artistic and creative digital works.
Response from Within the Sciences

The Reproducible Research Standard (RRS) (Stodden, 2009):

- A suite of license recommendations for computational science:
  1. Release media components (text, figures) under CC BY,
  2. Release code components under Modified BSD or similar,
  3. Release data to public domain or attach attribution license.

- Remove copyright’s barrier to reproducible research and,
- Realign the IP framework with longstanding scientific norms.
Copyright and Data

- Copyright adheres to raw facts in Europe.

- In the US raw facts are not copyrightable, but the original “selection and arrangement” of these facts is copyrightable. (Feist Publns Inc. v. Rural Tel. Serv. Co., 499 U.S. 340 (1991)).

- the possibility of a residual copyright in data (attribution licensing or public domain certification).

- Law doesn’t match reality on the ground: What constitutes a “raw” fact anyway?
Example 1: Fracking

Hydraulic Fracturing

Hydraulic fracturing, or “fracking,” involves the injection of more than a million gallons of water, sand and chemicals at high pressure down and across into horizontally drilled wells as far as 10,000 feet below the surface. The pressurized mixture causes the rock layer, in this case the Marcellus Shale, to crack. These fissures are held open by the sand particles so that natural gas from the shale can flow up the well.
Large areas of the United States long considered geologically stable with little or no detected seismicity have recently become seismically active. The increase in earthquake activity began in the mid–continent starting in 2001 (1) and has continued to rise. In 2014, the rate of occurrence of earthquakes with magnitudes ($M$) of 3 and greater in Oklahoma exceeded that in California (see the figure). This elevated activity includes larger earthquakes, several with $M > 5$, that have caused significant damage (2, 3). To a large extent, the increasing rate of earthquakes in the mid–continent is due to fluid–injection activities used in modern energy production (1, 4, 5). We explore potential avenues for mitigating effects of induced seismicity. Although the United States is our focus here, Canada, China, the UK, and others confront similar problems associated with oil and gas production, whereas quakes induced by geothermal activities affect Switzerland, Germany, and others.
Example 2: EPA

Reanalysis of the Harvard Six Cities Study and the American Cancer Society Study of Particulate Air Pollution and Mortality

Report #: Special Report, Publication Type: Special Reports
Author: Health Effects Institute, 2000-01-01 Order this Publication

About This Publication A Special Report of the Institute’s Particle Epidemiology Reanalysis Project.
The overall objective of this project was to conduct a rigorous and independent assessment of the findings of the Harvard Six Cities and American Cancer Society Studies of air pollution and mortality. This objective was met in two parts. In Part I: Replication and Validation, the Reanalysis Team led by Dr. Daniel Krewski sought to replicate the original studies via a quality assurance audit of a sample of the original data and to validate the original numeric results. In Part II: Sensitivity Analyses, they tested the robustness of the original analyses to alternate risk models and analytic approaches.

Table of Contents:
- HEI Statement
- Preface
- Contributors to the Project
- Health Review Committee’s Commentary
- Original Investigators
FDA debates trial-data secrecy

US drug regulator weighs up merits of disclosing preliminary results.

Heidi Ledford

29 July 2014

Despite a trend towards increased transparency in clinical-trial data, the US Food and Drug Administration (FDA) is asking whether there are times when participants and researchers should be kept in the dark. As pharmaceutical companies push for studies that first justify a drug’s approval, then monitor safety once it reaches the market, the agency fears that publicizing the early data could bias the final results.

In raising the matter, the FDA could energize the debate about a long-standing clinical conundrum, says Iain Chalmers, coordinator of the James Lind Initiative, a group based in Oxford, UK, that aims to improve clinical trials. “There hasn’t been much discussion about this,” he says. “There needs to be much more.”

On 11 August, the FDA will hold a public hearing in Silver Spring, Maryland, to discuss situations in which preliminary results from clinical trials should be kept confidential. The FDA is obliged to release a summary of the data that it uses to approve a drug. But the public rarely sees the data given to safety committees to decide whether a trial should continue. Even if those data are not definitive but lean one way or another, making them public may spook study participants or bias investigators towards a particular outcome, the agency fears.