



POLICY FORUM: SCIENCE AND REGULATION

A Solution to Concerns Over Public Access to Scientific Data

Linda R. Cohen and Robert W. Hahn

The scientific establishment is deeply concerned over a proposed regulation that would require data to be publicly available under the Freedom of Information Act (1). The change was proposed by the Office of Management and Budget as a consequence of an amendment added by Senator Richard Shelby to the Omnibus Consolidated and Emergency Supplemental Appropriations Act, fiscal year 1999 (H.R. 4328). The regulation would apply to any federally funded data sets that have been reported in a publication and are being used in forming policy. Opponents of the regulation correctly point out that it is ambiguous in important respects and could be costly to scientists. We believe that without some change, however, a larger, more important problem is not addressed. At present, analyses used in policy-making are rarely checked carefully before big regulations are put in place. We recommend allowing greater access to information that pertains to the formulation of such regulations and propose that an agency be created to replicate key findings used to support regulations before they are finalized.

Scientists are justifiably concerned that the proposed regulation before Congress could reduce the productivity of scientists, expose them to unfair attacks by special interest groups, and place unnecessary burdens on researchers. They also argue that the rule could place severe restrictions on researchers who obtain data only by guaranteeing anonymity to subjects. Further, researchers and institutions with ties to industry fear that forced disclosure of proprietary information could jeopardize univer-

sity-industry partnerships, which frequently help spur innovation (2). To address these concerns, our proposal recommends narrowing the focus of the regulation to those areas where public access to data is likely to have the greatest social value.

The controversy over public access to data arose when the Environmental Protection Agency (EPA) finalized a regulation on particulate matter in July of 1997, which gives the agency vast new powers to

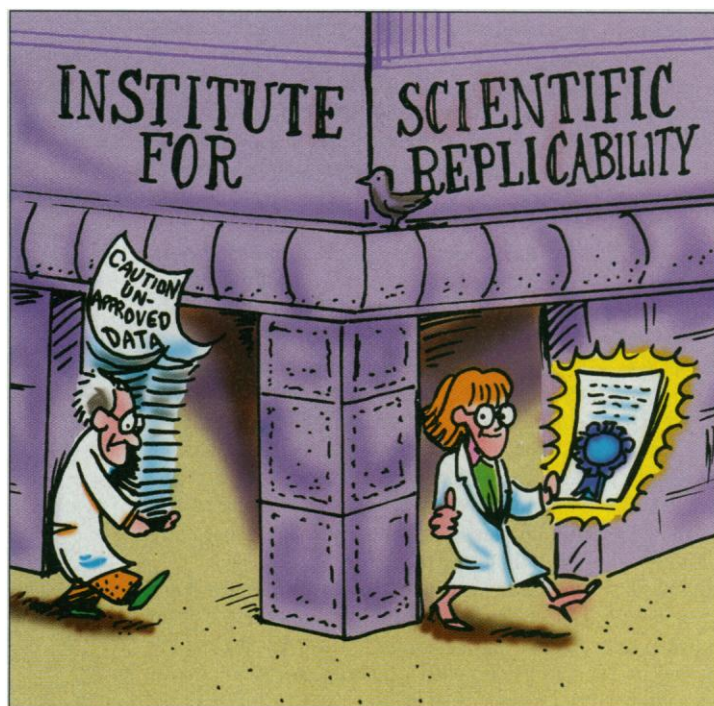
industry and the EPA, would convene an expert panel to reanalyze the data. The results of that study are not expected to be available until later this year—2 years after the regulation was made final (5). In this case, the EPA administrator argued that the Harvard study had been peer-reviewed, and that this was sufficient for using the findings in a public policy setting (6).

A closer look at the peer-review process, in general, reveals serious flaws. In fact, studies have been published that demonstrate how easily errors slip through the system. In the early 1980s, a now-famous study requested the data used in papers with statistical analyses published in the *Journal of Money, Credit and Banking*, a leading economics journal. The study authors found errors in nearly every paper that were sufficiently serious that the results could not easily be replicated. The authors also found that, notwithstanding both the general norm that data be available and the requirement of the National Science Foundation (NSF) that data be produced on NSF-funded projects, their requests for data were ignored, denied, or otherwise frustrated in a substantial number of cases (7).

Another study, published in *JAMA*, gave a paper with eight deliberate errors to 420 people to review. For the 221 reviewers that responded, the maximum number of errors detected was five, the median was two, and 16% of the respondents did not find any (8).

These findings cast doubt on the peer-review process, even for academic purposes. It is noteworthy that an increasing number of the most prestigious journals, such as *Nature*, *American Economic Review*, and *Science*, now require data availability to editors and members of the scientific community as a condition for publication. Other journals, such as *Cell*, require that data be made available for scientific scrutiny when there are disputes.

Even if the peer-review process were adequate for academic purposes, however, it is frequently not adequate for major public policy decisions, such as those involved in regulation. Making data available before passage of a regulation offers important benefits for regulatory decision-making. First, it could improve the quality of information by exposing it to widespread comment, thus leading to better decisions. If the findings of researchers were shown to



regulate a variety of emission sources ranging from power plants to lawn mowers and barbecue grills. The regulation, estimated to cost between \$9 billion and \$37 billion annually in 1990 dollars (3), was based partly on a Harvard study that suggested that reducing emissions of fine particles could lead to substantial reductions in premature mortality (4).

Several members of Congress and a number of industry organizations requested that EPA obtain the data and then release it. The EPA requested the data, but the researchers refused to turn it over. They subsequently agreed to an alternate plan, whereby the Health Effects Institute, an independent research institute funded by in-

L. R. Cohen is in the Department of Economics, University of California, Irvine, CA 92697, USA. E-mail: lrcohen@uci.edu. R. W. Hahn is director of the American Enterprise Institute-Brookings Joint Center for Regulatory Studies, Washington, DC 20036, USA. E-mail: rhahn@aei.org

be false or misleading before the development of a costly environmental regulation, such as one addressing particulate matter or toxic substances in the air, then that regulation could be withdrawn or revised. Second, public access to data ensures greater transparency, which lends legitimacy to the regulatory process. Transparency is a valuable aspect of public decision-making in a democracy.

If all regulations were good for society at large, there would be little need for concern. Research suggests, however, that more than half of the federal government's regulations would fail a strict benefit-cost test based on the government's own numbers (9), even though overall benefits are positive (10). Reanalysis of government regulations, programs, and supporting data frequently reveal that there are major problems with the initial analysis and that correcting these deficiencies yields substantially different policy conclusions (11). For example, a researcher argued that modifying some standards for lead recently proposed by the EPA could increase net benefits by more than \$20 billion (12). There is ample research showing that regulation could be significantly improved, so that more lives could be saved with fewer resources (13). One study found that a reallocation of mandated expenditures toward those regulations with the highest payoff to society could save as many as 60,000 more lives a year at no additional cost (14).

To help weed out potentially bad regulations, it is important to have key data available in a timely manner, so that policies can be analyzed carefully before they are put in place. Once a regulation is passed, it becomes more difficult to modify because constituencies grow in support of the regulation, both inside and outside government. This is true for bad regulations as well as good ones, because some constituency invariably benefits from a regulation and thus will defend it.

Because of the potential for improving public policy decisions by allowing public access to data, the government needs to develop a policy that carefully weighs the costs and benefits of the proposed rule. The costs include potentially adverse impacts on research, data development, and industry-university-researcher partnerships that help commercialize technology. Although these costs are important, the social benefits of increased public access to data under specified conditions could also be substantial. These include better public policies and increased transparency and accountability.

We believe that the initial legislation, which would have allowed access to any

publicly supported data subject only to the restrictions of the Freedom of Information Act, is too broad. However, our analysis suggests that academic norms alone provide very limited access to scientific data. The bottom line is that greater access to data is needed to enhance accountability and to improve the decision-making process. Here are five recommendations that we think would improve the process.

Recommendation 1. The data access requirements should be restricted to economically significant regulations developed by all regulatory agencies.

Targeting economically significant regulations is likely to yield the greatest economic gains for society. A reasonable cut-off point, currently used by federal regulatory agencies, is to allow access to data that affect regulations with an annual economic impact of at least \$100 million.

Recommendation 2. The data access requirements should be limited to new federally funded grants and agreements.

The government should abide by the terms of existing grants and agreements with researchers or at least should not impose major additional costs on researchers without compensation. The terms of new federally funded grants and agreements that fall under the new regulation should be restricted to data used in published research in refereed journals that are directly related to the grant (15).

Recommendation 3. The researcher should be required to provide as full a rendering of the data set as possible.

There is a natural tendency in some research fields to report the results that are statistically significant or that will increase the chances for publication, even if they tell only part of the story (16).

Recommendation 4. The new rule about data access, if implemented, should be evaluated after 5 years by an expert panel selected by the National Academy of Sciences.

The panel should include individuals who can evaluate the economic, social, and scientific impacts of the regulation. It should offer recommendations for improving the regulation, if needed.

Recommendation 5. Congress should create an agency to replicate findings for economically significant regulations that have an annual economic impact of at least \$100 million. Government should be allowed to use those research findings in developing regulations only after the agency has replicated the results or has certified that the results have been independently replicated.

Replication is a key to ensuring the quality of results. Replication should require a finding by the newly created agen-

cy that the basic conclusions drawn from the data are supported by the data. The replication exercise could be defined narrowly in terms of reproducing the results of the initial research or policy analysis. We would prefer to define it a little more broadly, although that could make it harder to define the conditions under which the data actually support the results (17).

The requirement for replication before promulgation of the regulation is critical. Because of the difficulty attached to changing a poor regulation once it is already in place, the benefits of such replication for improving regulation are likely to be large.

The concerns about data access raised by Congress deserve a serious response. Taken together, our recommendations would help lay the foundation for a regulatory system that is more accountable and has more scientific integrity.

References and Notes

1. <http://www.aaas.org/spp/dspp/sfrl/projects/omb/ombltr.htm>.
2. National Science Board, press statement (23 February 1999).
3. Environmental Protection Agency, *Regulatory Impact Analysis for the Particulate Matter and Ozone National Ambient Air Quality Standards and Proposed Regional Haze Rule* (Innovative Strategies and Economics Group, Office of Air Quality Planning and Standards, Research Triangle Park, NC, 1997).
4. J. Kaiser, *Science* **277**, 467 (1997).
5. Although the regulation was finalized in July 1997, it has not been implemented.
6. M. Fumento, *Reason* (August 1998). Available at: <http://www.reasonmag.com/9708/fe.fumento.html>
7. W. Dewald, J. Thursby, R. Anderson, *Am. Econ. Rev.* **76**, 587 (1986).
8. F. Godlee, C. Gale, C. Martyn, *JAMA* **280**, 237 (1998).
9. R. Hahn, in *Risks, Costs, and Lives Saved: Getting Better Results from Regulation*, R. Hahn, Ed. (Oxford University Press and American Enterprise Institute Press, New York and Washington, DC, 1996), chap. 10.
10. Office of Management and Budget, *Report to Congress on the Costs and Benefits of Federal Regulations* (1998).
11. L. Lave, *Science* **212**, 893 (1981); W. Viscusi and J. Hamilton, American Enterprise Institute (AEI)-Brookings Joint Center for Regulatory Studies, Working Paper 99-2, April 1999.
12. R. Lutter, AEI-Brookings Joint Center for Regulatory Studies, Working Paper 99-5, May 1999.
13. J. Morral, *Regulation* **10**, 25 (1986); W. Viscusi, *J. Econ. Perspect.* **10**, 119 (1996).
14. T. Tengs and J. Graham, in *Risks, Costs, and Lives Saved: Getting Better Results from Regulation*, R. Hahn, Ed. (Oxford University Press and American Enterprise Institute Press, New York and Washington, DC, 1996), chap. 8.
15. For more information, see L. Cohen and R. Hahn, AEI-Brookings Joint Center for Regulatory Studies, *Regulatory Analysis* 99-1, May 1999.
16. E. Leamer, *Am. Econ. Rev.* **73**, 31 (1983).
17. A broader replication effort would consider the appropriateness of the policy framework and the sensitivity of the policy conclusions to changes in key assumptions.
18. We would like to thank I. Chan for valuable research assistance and G. Charney, R. Crandall, C. DeMuth, D. Elliott, T. Gorman, R. Litan, J. G. Sidak, and M. Spitzer for helpful comments.