

University of Pennsylvania Carey Law School

Penn Law: Legal Scholarship Repository

Faculty Scholarship at Penn Law

11-1-2004

Evaluating Environmental Policies

Lori Snyder Bennear

Nicholas School of the Environment and Earth Sciences, Duke University

Cary Coglianese

University of Pennsylvania Carey Law School

Follow this and additional works at: https://scholarship.law.upenn.edu/faculty_scholarship



Part of the [Environmental Law Commons](#), [Environmental Policy Commons](#), and the [Policy Design, Analysis, and Evaluation Commons](#)

Repository Citation

Bennear, Lori Snyder and Coglianese, Cary, "Evaluating Environmental Policies" (2004). *Faculty Scholarship at Penn Law*. 106.

https://scholarship.law.upenn.edu/faculty_scholarship/106

This Article is brought to you for free and open access by Penn Law: Legal Scholarship Repository. It has been accepted for inclusion in Faculty Scholarship at Penn Law by an authorized administrator of Penn Law: Legal Scholarship Repository. For more information, please contact PennlawIR@law.upenn.edu.

Evaluating Environmental Policies

Lori Snyder Bennear
Nicholas School of the Environment and Earth Sciences
Duke University
Box 90328
Levine Science Research Center
Durham, NC 27708
lori.bennear@duke.edu

Cary Coglianese
John F. Kennedy School of Government
Harvard University
79 John F. Kennedy Street, Weil Hall
Cambridge, MA 02138
cary_coglianese@harvard.edu

Evaluating Environmental Policies *

Lori Snyder Benneer and Cary Coglianese

Do environmental policies work? For many, this question may seem simple and straightforward. To anyone involved in the environmental policy process, it is certainly an important one. Yet for many environmental policies it lacks a solid answer. Decision makers often lack carefully collected evidence about what policies have accomplished in the past—and in this sense they are poorly informed about what new policies might accomplish in the future. Getting systematic answers to the question of whether environmental policies work is vital. Real resources are expended on environmental regulatory programs, and at a minimum one should expect that these programs then lead to improvements in environmental conditions.¹ While intuitions and anecdotes may provide some reason for suspecting that a given policy has made or will make a difference, the only way to be confident of such suspicions is to evaluate a policy's impact in practice. Program evaluation research provides the means to determine what has and has not worked, and thereby to decide whether to retain existing policies or adopt new or modified ones.

Academics, policy makers, activists, and business leaders do generally recognize the need for careful evaluation of existing environmental policies. Indeed, some important research has

* This paper is forthcoming in *Environment* and is also based on a report by the authors commissioned by the National Research Council's Panel on Social and Behavioral Science Research Priorities in Environmental Decision Making, copyrighted by the National Academies Press and portions reprinted here with permission. The authors are grateful for the helpful comments they received from Garry Brewer, Terry Davies, David Heath, Shelley Metzenbaum, Jennifer Nash, Paul Stern, and two anonymous reviewers.

¹ One may want to impose even higher standards on policies, such as expecting that they yield improvements in environmental quality that more than offset the costs of achieving the improvement or even that they maximize net benefits, namely they achieve the highest possible difference between the benefits from environmental improvement and the costs of compliance.

been undertaken, particularly studies of the effects of long-standing regulations such as the Clean Air Act and the Clean Water Act. Yet program evaluation research has been remarkably scarce relative to the overall volume of environmental policy decisions made at the state and federal level, as well as relative to the amount of evaluation research found in other fields, such as medicine, education, or transportation safety. A renewed and greatly expanded commitment to program evaluation of environmental policy would help move environmental decisionmaking closer to an evidence-based practice.

I. The Role of Program Evaluation in Environmental Policy

To look at this issue closely, it is useful first to define the role that program evaluation can play in policy deliberation and decisionmaking, distinguishing evaluation from other types of analysis, including risk assessment, cost-effectiveness analysis, and cost-benefit analysis. While reliance on these other types of analysis has greatly expanded over the past several decades, most of these other forms of analysis take place before decisions are made. Relatively little analysis takes place after decisions have been made and implemented, which is when program evaluation occurs. Yet anyone who takes analysis and deliberation seriously before decisions are made should also take seriously the need for research after decisions are made.

Because the overarching purpose behind environmental policies is to improve environmental conditions—and often thereby to improve human health—program evaluation can identify whether specific policies are serving this purpose and whether they are having other kinds of effects, such as reducing environmental inequities, imposing economic costs, or promoting or inhibiting technological change. In this section, we show how program evaluation

research fits into the policy process and serves an important role in environmental decisionmaking.²

A. Environmental Policymaking and Implementation

The policy process begins with the recognition of a potential environmental problem and a response by the policymaker, often the legislature.³ The response typically takes the form of a statute imposing requirements on industry or delegating authority to a regulatory agency, such as the Environmental Protection Agency (EPA) or the U.S. Fish and Wildlife Service. These agencies then create additional, more specific regulations or develop other programs to achieve legislative goals. At the federal level, for example, environmental and natural resources agencies promulgate hundreds of new regulations each year.

Regulations vary not only in their overall stringency and deadlines for compliance, but also in terms of the type of policy instrument upon which they rely. For example, technology-based instruments regulate the precise actions that regulated entities must take, while performance instruments instead compel the achievement of a specified outcome using whatever technologies or processes regulated entities choose. Performance-based regulation can take the form of requirements that all facilities achieve uniform outcomes, or they can follow a market-based approach that allows regulated entities to average or trade in outcomes such as emissions levels. Alternatively, information-based regulatory instruments require that organizations report information to the government or the public, while management-based instruments require that

² The phrase “environmental decisionmaking” here includes all policy decisions related to the environment. While most of the examples throughout this article draw on federal pollution-oriented environmental policies in the United States, the discussion applies equally to any type of environmental or natural resources policy decisionmaking, at the local, state, federal, and international levels.

³ Brewer, Garry and Peter deLeon (1983), *The Foundations of Policy Analysis* (Homewood, Illinois: The Dorsey Press).

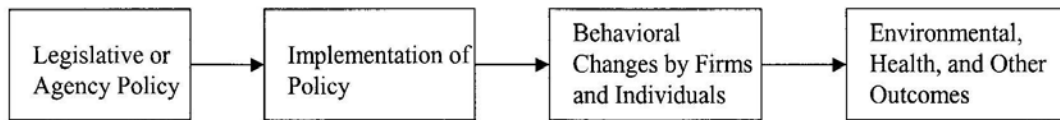
regulated entities develop a planning process or put in place a management system for addressing regulatory problems.

Policy implementation also involves choices beyond the design of standards. It can include education, licensing, and grant programs. It also can include the selection of enforcement or other strategies to ensure compliance with policies. Regulatory agencies must make decisions about how they will target firms for enforcement: (a) randomly, (b) in reaction to complaints, (c) based on past history, (d) based on size or other criteria related to the regulatory problem to be solved, or (e) some combination of these or other factors. Moreover, agency inspectors can be instructed to approach their work in an adversarial manner—that is, going “by the book” and issuing citations for any violations found—or in a more cooperative manner that seeks to encourage regulated entities to solve problems and come into compliance without a heavy use of punishment.⁴

Regulatory policies are adopted and then implemented and enforced to change the behavior of a class of businesses or individuals. The ultimate aim of policymaking and implementation is to create incentives for individuals and firms to change their behavior in ways that will solve the problems that motivated the adoption of public policy in the first place. If a policy works properly, the behavioral change it induces will in turn result in the desired changes in environmental conditions, public health, or other outcomes. A basic diagram of the environmental policy process is provided in Figure 1.

⁴ Bardach, Eugene and Robert A. Kagan (1981), *Going by the Book: The Problem of Regulatory Unreasonableness* (Philadelphia: Temple University Press); Scholz, John T. (1984) “Cooperation, Deterrence and the Ecology of Regulatory Enforcement,” *Law & Society Review* 18:601; and Hutter, Bridget (1989), “Variations in Regulatory Enforcement Styles,” *Law and Policy* 11(2):153-74.

Figure 1. A simple model of the environmental policy process



B. Prospective Analysis of Environmental Policy

Empirical analysis can usefully inform several stages of the policy process. During the policymaking and implementation stages, analysis can inform deliberation and decisionmaking about whether anything should be done to address an environmental problem and, if so, what set of policy instruments or strategies should be used. Currently, there are several different analytical methods used extensively during policymaking and implementation, including risk assessment, cost-effectiveness analysis, and benefit-cost analysis.⁵ Each type of analysis is used prospectively to inform the deliberative process leading up to policy decisions.

Risk assessment characterizes the health or ecological risks associated with exposure to pollution or other hazardous environmental substances or conditions.⁶ It seeks to identify the causal relationships between exposure to specific environmental hazards and specific health or ecological conditions. As such, risk assessment seeks to provide a scientific basis for

⁵ Stokey, Edith and Richard Zeckhauser (1978), "Project Evaluation: Benefit-Cost Analysis," in *A Primer for Policy Analysis* (New York, NY: W.W. Norton & Company)

⁶ National Research Council (1983), *Risk Assessment in Federal Government: Managing the Process* (Washington, D.C.: National Academy Press).

understanding the potential range of benefits that can be attained from policies that aim to reduce exposure to environmental hazards.⁷

Benefit-cost analysis seeks to help policymakers identify the benefits and costs of specific environmental policies and implementation strategies. It compares different policy or implementation alternatives based on their net benefits—that is, total benefits minus total costs.⁸ Such analysis is usually conducted in advance of policymaking to try to identify regulatory options that will be the most efficient.⁹ As such, benefit-cost analysis usually leads to estimates of expected net benefits from different alternatives.

Cost-effectiveness analysis seeks to identify the lowest cost means of achieving a specific goal.¹⁰ Unlike benefit-cost analysis, which compares alternatives in terms of both costs and benefits, cost-effectiveness analysis compares alternatives simply in terms of how much they cost to achieve a given goal—regardless of whether there will be positive net benefits from achieving this goal.

EPA's regulation phasing out lead in gasoline in the 1980s illustrates the respective roles of risk-assessment, benefit-cost analysis, and cost-effectiveness analysis. Risk assessment helped demonstrate the connection between lead exposure and cognitive development in children under the age of six, as well as demonstrated that adults suffer health consequences from lead exposure through increased blood pressure. Benefit-cost analysis quantified and monetized the benefits of

⁷ Risk assessment is not exclusively a scientific enterprise, however, as it often involves making certain policy judgments for which public deliberation may be appropriate. National Research Council (1996), *Understanding Risk: Informing Decisions in a Democratic Society* (Washington, D.C.: National Academy Press).

⁸ Arrow, Kenneth J., Maureen L. Cropper, George C. Eads, Robert W. Hahn, Lester B. Lave, Roger G. Noll, Paul R. Portney, Milton Russell, Richard Schmalensee, V. Kerry Smith, and Robert N. Stavins (1996), "Is There a Role For Benefit-Cost Analysis in Environmental, Health, and Safety Regulation?" *Science* 272: 221-222.

⁹ Viscusi, W. Kip (1996), "Regulating the Regulators," *University of Chicago Law Review* 63: 1423-1461; and Hahn, Robert W., (1998) "Government Analysis of the Benefits and Costs of Regulations," *Journal of Economic Perspectives* 12(4): 201-210.

¹⁰ Environmental Protection Agency, Office of the Administrator (2000), *Guidelines for Preparing Economic Analyses* (Washington, D.C.: U.S. Environmental Protection Agency, Document number EPA 240-R-00-003).

avoiding the health effects indicated by the risk assessment as well as the costs of complying with the lead phase-down. In the end, the benefit-cost analysis in this regulatory proceeding showed benefits in excess of three times the cost of compliance, a finding that encouraged more rapid implementation of the phase-down rule. Finally cost-effectiveness analysis suggested that costs of compliance would be greatly reduced if refineries could average their reductions over time or across facilities, through what was known as the lead trading and banking program.¹¹

As in the lead phase-down regulation, risk assessment and economic analysis of costs and benefits are typically used prospectively in the regulatory process, to aid legislatures and regulatory agencies in policy making. The prospective use of these analytic techniques has expanded greatly in the past 20 years due to 2 developments: evolving professional practices and executive orders mandating economic analysis under certain conditions. These executive orders, which call for such analysis preceding the adoption of new federal regulations that are anticipated to impose \$100 million or more in annual compliance costs, have existed under every administration since Ronald Reagan.¹² In the wake of these orders, government agencies have developed detailed guidance for conducting the required analyses.¹³

¹¹ For discussion of the role of analysis in the lead phase-down rule, see Nichols, Albert L. "Lead in Gasoline," in Morgenstern, Richard D. ed., *Economic Analyses at EPA: Assessing Regulatory Impact* (Washington, D.C.: Resources for the Future Press, 1997). For discussion of the lead trading and banking program, see Nussbaum, Barry D. "Phasing Down Lead in Gasoline in the U.S.: Mandates, Incentives, Trading, and Banking," in T. Jones and J. Corfee-Morlot, eds., *Climate Change: Designing A Tradable Permit System* (Paris, Organization for Economic Co-operation and Development, 1992); Hahn, Robert W., and Hester, Gordon L. "Marketable Permits: Lessons for Theory and Practice," *Ecology Law Quarterly* 16(2): 361-406 (1989). For other examples of the use of risk assessment and benefit-cost analysis in regulatory settings see Morgenstern, Richard D. ed., *Economic Analyses at EPA: Assessing Regulatory Impact* (Washington, D.C.: Resources for the Future Press, 1997); Hahn, Robert W. ed., *Risks, Costs, and Lives Saved* (New York, NY: Oxford Press 1996).

¹² Coglianese, Cary (2002), "Empirical Analysis and Administrative Law," *University of Illinois Law Review* 2002:1111-1137; and Hahn, Robert W. and Cass R. Sunstein (2002), "A New Executive Order for Improving Federal Regulation? Deeper and Wider Cost-Benefit Analysis," *University of Pennsylvania Law Review* 150: 1489.

¹³ Environmental Protection Agency, Office of the Administrator (2000), *Guidelines for Preparing Economic Analyses* (Washington, D.C.: U.S. Environmental Protection Agency, Document number EPA 240-R-00-003); and Office of Management and Budget, Office of Information and Regulatory Affairs (2003a) "Circular A-4: Regulatory Analysis," available at <http://www.whitehouse.gov/omb/circulars/a004/a-4.pdf>.

C. Retrospective Analysis: Program Evaluation of Environmental Policy

In contrast to the prospective use of risk assessment and benefit-cost analysis, program evaluation occurs retrospectively, as it seeks to determine the impact of a chosen policy or implementation strategy after it has been adopted. For example, MIT economist Michael Greenstone recently evaluated the effect of the Clean Air Act on sulfur dioxide concentrations, basing his analysis on nearly twenty years of government data collected after the Act was originally passed.¹⁴ Under the Clean Air Act, EPA designated certain counties as “non-attainment” with respect to ambient air quality standards for SO₂. Counties designated as “non-attainment” were subject to more stringent regulations with regard to air pollution than counties that were designated in “attainment.” The data analyzed by Greenstone showed that sulfur dioxide concentrations fell more rapidly in non-attainment counties after regulation than in attainment counties. Did this mean there was a causal connection between the nonattainment regulations and the observed decline in SO₂ concentrations? Using county-level SO₂ monitoring data together with county-level economic data, Greenstone controlled for other factors that may have been responsible for the decline. Taking these other factors into account, he estimated that non-attainment status for SO₂ was directly responsible for only a small decrease in ambient SO₂ concentrations and that this decrease was generally not statistically significant. These more careful analyses by Greenstone suggest that the Clean Air Act’s system of non-attainment regulation has had at most only a modest impact on the reductions in sulfur dioxide concentrations.

¹⁴ Greenstone, Michael (2004) “Did the Clean Air Act Cause the Remarkable Decline in Sulfur Dioxide Concentrations?,” *Journal of Environmental Economics and Management* 47:585-611.

Other regulatory policies have been evaluated retrospectively, including hazardous waste cleanup laws¹⁵ and air pollution and other media-specific environmental regulations.¹⁶ Sometimes evaluations show that policies achieve significant results, while at other times they do not. A variety of innovations in environmental policy have also received retrospective study, including market-based instruments,¹⁷ planning requirements,¹⁸ information disclosure requirements such as EPA's Toxics Release Inventory (TRI),¹⁹ and various voluntary programs such as EPA's Project XL and 33/50 program.²⁰ In addition, various procedural policies have

¹⁵ Hamilton, James T. and W. Kip Viscusi (1999), *Calculating Risks?: The Spatial and Political Dimensions of Hazardous Waste Policy* (Cambridge, MA: MIT Press); and Revesz, Richard L. and Richard B. Stewart, eds. (1995), *Analyzing Superfund: Economics, Science and Law* (Washington, D.C.: Resources for the Future Press).

¹⁶ Environmental Protection Agency, Office of Air and Radiation (1997), *Final Report to Congress on Benefits and Costs of the Clean Air Act, 1970 to 1990* (Washington, D.C.: U.S. Environmental Protection Agency, Document number EPA 410-R-97-002); Davies, J. Clarence and Jan Mazurek (1998), *Pollution Control in the United States: Evaluating the System*, Washington, D.C.: Resources for the Future Press; Harrington, Winston, Richard D. Morgenstern, and Peter Nelson (2000) "On the Accuracy of Regulatory Cost Estimates," *Journal of Public Policy Analysis and Management*, 19(2): 297-322; and Chay, Kenneth Y. and Michael Greenstone (2003), "Air Quality, Infant Mortality and the Clean Air Act of 1970," National Bureau of Economic Research Working Paper Number w10053.

¹⁷ Stavins, Robert N. (1998), "What Can We Learn from the Grand Policy Experiment? Positive and Normative Lessons from SO₂ Allowance Trading," *Journal of Economic Perspectives* 12(3): 69-88.

¹⁸ Snyder, Lori D. (2004), "Are Management-Based Regulations Effective: Evidence from State Pollution Prevention Programs," in *Essays on Facility Level Response to Environmental Regulation*, Ph.D. dissertation, Harvard University.

¹⁹ Hamilton, James T. (1995), "Pollution as News: Media and Stock Market Reactions to the Toxic Release Inventory Data," *Journal of Environmental Economics and Management* 28(1): 98-113; Konar, Shameek and Mark A. Cohen (1997), "Information as Regulation: The Effect of Community Right to Know Laws on Toxic Emissions," *Journal of Environmental Economics and Management* 32(2): 109-124; Khanna, Madhu, Wilma Rose H. Quimio, and Dora Bojilova (1998), "Toxics Release Information: A Policy Tool for Environmental Protection," *Journal of Environmental Economics and Management* 36(3), 243-266; and Bui, Linda T. M. and Christopher J. Mayer (2003), "Regulation and Capitalization of Environmental Amenities: Evidence from the Toxics Release Inventory in Massachusetts," *Review of Economics and Statistics* 85(2): 396-708.

²⁰ Blackman, Allen and Janice Mazurek (2000), "The Cost of Developing Site-Specific Environmental Regulations: Evidence from EPA's Project XL," Resources for the Future Discussion Paper 99-35-REV (Washington, D.C.: Resources for the Future); Marcus, Alfred, Donald A. Geffen, and Ken Sexton (2002), *Reinventing Environmental Regulation: Lessons from Project XL* (Washington, D.C.: Resources for the Future Press); Alberini, Anna and Kathleen Segerson (2002), "Assessing Voluntary Programs to Improve Environmental Quality," *Environmental and Resource Economics* 22: 157-184; Arora, Seema and Timothy N. Cason (1995), "An Experiment in Voluntary Environmental Regulation: Participation in EPA's 33/50 Program," *Journal of Environmental Economics and Management* 28(3): 271-286; and Khanna, Madhu and Lisa A. Damon (1999), "EPA's Voluntary 33/50 Program: Impact on Toxic Releases and Economic Performance of Firms," *Journal of Environmental Economics and Management* 37(1): 1-25.

been subject to retrospective evaluation, such as the use of benefit-cost analysis²¹ and negotiated rulemaking.²² Finally, researchers have also evaluated the impact of various types of enforcement strategies.²³

Like the Greenstone study mentioned above,²⁴ such retrospective analyses have sought to ascertain what outcomes specific governmental policies or strategies have actually achieved.²⁵ Some of these outcomes are the ones the policy was intended to achieve, such as improvements in human health or the biodiversity of an ecosystem. However, program evaluation research can also consider other effects, such as whether a policy has had unintended or undesirable consequences. Has it contributed to other problems similar or related to the one the policy was supposed to solve? What kinds of costs has the policy imposed? How are the costs and benefits of the policy distributed across different groups in society? Finally, program evaluation research can also focus on other outcomes, including transparency, equity, intrusiveness, technological change, public acceptability, and conflict avoidance, to name a few.

By assessing the performance of environmental policies in terms of various kinds of outcomes, retrospective evaluation can inform policy deliberations. It becomes a vital part of

²¹ Morgenstern, Richard, ed. (1997), *Economic Analyses at EPA: Assessing Regulatory Impact* (Washington, D.C.: Resources for the Future Press); Farrow, Scott (2000), "Improving Regulatory Performance: Does Executive Office Oversight Matter?" available at <http://www.aei.brookings.org/publications/related/oversight.pdf>; and Hahn, Robert and Patrick M. Dudley (2004). "How Well Does the Government Do Cost-Benefit Analysis?" AEI-Brookings Joint Center on Regulatory Studies Working Paper 04-01.

²² Coglianesi, Cary (2001), "Assessing the Advocacy of Negotiated Rulemaking," *New York University Environmental Law Journal* 9:386-447; Langbein, Laura and Cornelius Kerwin (2000), "Regulatory Negotiation versus Conventional Rule Making: Claims, Counterclaims, and Empirical Evidence," *Journal of Public Administration Research and Theory* 10:599; Coglianesi, Cary (1997), "Assessing Consensus: The Promise and Performance of Negotiated Rulemaking," *Duke Law Journal* 46: 1255-1349.

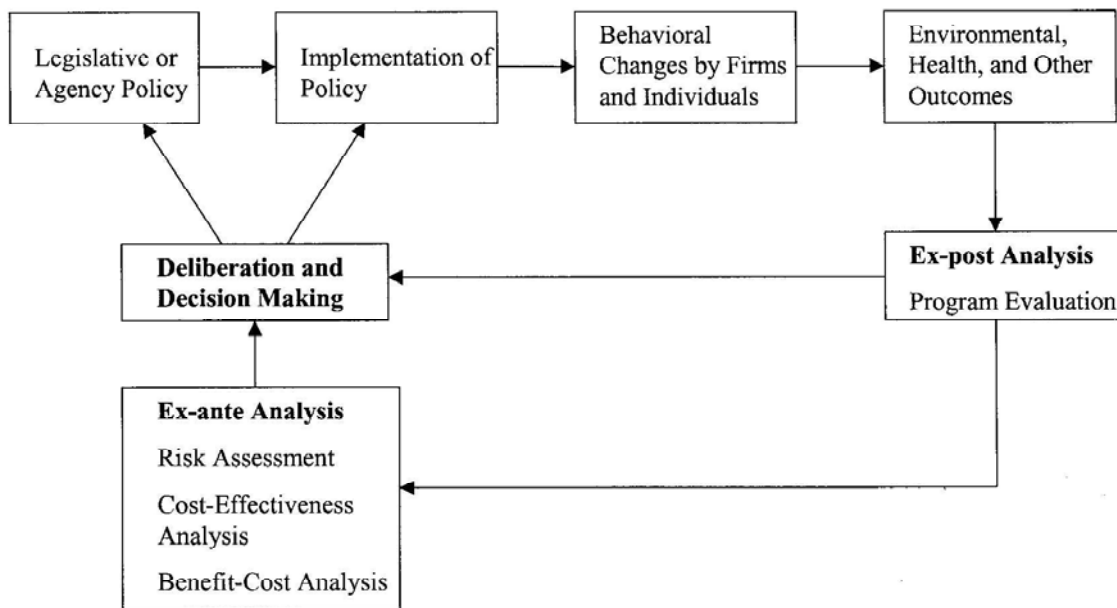
²³ Shimshack, Jay P. and Michael B. Ward (2003) "Enforcement and Environmental Compliance: A Statistical Analysis of the Pulp and Paper Industry," mimeo Tufts University; and May, Peter and Soeren Winter (2000), "Reconsidering Styles of Regulatory Enforcement: Patterns in Danish Agro-Environmental Inspection," *Law and Policy* 22:143.

²⁴ Greenstone, note 14 above.

²⁵ Sometimes program evaluation researchers distinguish between the "outcomes" and "outputs" of a program. For example, a new enforcement initiative might increase the number of enforcement actions that a regulatory agency brings (an output), but the program evaluation researcher would want to ask whether this new initiative (and the corresponding increase in enforcement actions) actually reduced pollution (an outcome).

what is sometimes called an adaptive management approach to environmental and natural resources policy.²⁶ Policymakers revisit regulatory standards periodically, whether due to statutory requirements, industry or environmental group petitions, or a commitment to adaptive management principles. In addition, existing policies are often used as model solutions for new environmental problem. For this reason, program evaluation provides critical information for prospective analysis of new policy initiatives. By knowing what policies have accomplished in other contexts, prospective analyses—such as benefit-cost analysis—can be grounded in experience as well as theory and forecasting. The accuracy of the estimation strategies used in prospective analyses can also be refined by comparing ex-ante estimates with the actual outcomes discerned through program evaluation research. Figure 2 on page 02 illustrates the role of program evaluation in the policy process.

Figure 2. Program evaluation in the policy process



²⁶ Holling, C.S., ed. (1978), *Adaptive Environmental Assessment and Management* (London: John Wiley); Lee, Kai N. (1993), *Compass and Gyroscope: Integrating Science and Politics for the Environment* (Washington, D.C.: Island Press); Iles, Alastair T. (1996), "Adaptive Management: Making Environmental Law and Policy more Dynamic, Experimentalist and Learning," *Environmental and Planning Law Journal* 13:288; Clark, William C. (2002), "Adaptive Management, Heal Thyself," *Environment* 44:2.

II. Methods of Program Evaluation

The goal of program evaluation is to ascertain the causal effect of a program on one or more outcomes, that is, the change in outcomes that would not have occurred but for the program. Even if an environmental policy is correlated with a particular environmental or social outcome, this does not necessarily mean that there is a causal relationship between the policy initiative and the change in outcomes. By using state-of-the-art evaluation methods, however, researchers can isolate the effects of specific policy interventions and thereby inform environmental decisionmaking.

Program evaluation methods aim to identify the causal impact of a treatment on an outcome or outcomes. In the field of environmental policy, the treatment will often include government-mandated regulations, such as technology and performance standards,²⁷ market-based instruments,²⁸ information disclosure policies,²⁹ or management-based policies.³⁰ The treatment could also consist of a variety of implementation strategies, ranging from different types of enforcement strategies, grant requirements, or public recognition and waiver programs.³¹ The treatment could even include international treaties such as the Montreal Protocol or nongovernmental initiatives, such as trade association self-regulatory efforts like the chemical

²⁷ Coglianese, Cary, Jennifer Nash, and Todd Olmstead (2003), "Performance-Based Regulation: Prospects and Limitations in Health, Safety, and Environmental Regulation," *Administrative Law Review* 55: 741-764.

²⁸ Stavins, Robert N. (2003), "Experience with Market-Based Environmental Policy Instruments," in Maler, Karl-Goran and Jeffrey R. Vincent eds. *Handbook of Environmental Economics, Volume 1 Environmental Degradation and Institutional Responses* (Amsterdam: North-Holland Press).

²⁹ Kleindorfer, Paul, and Eric Orts (1998), "Informational Regulation of Environmental Risks," *Risk Analysis* 18:155-70.

³⁰ Coglianese, Cary and David Lazer (2003), "Management-Based Regulation: Prescribing Private Management to Achieve Public Goals," *Law and Society Review* 37(4): 691-730.

³¹ For a review of a number of such programs, see Bruijn, Theo de and Vicki Norberg-Bohm (2001), "Voluntary, Collaborative, and Information-Based Policies: Lessons and Next Steps for Environmental and Energy Policy in the United States and Europe," Regulatory Policy Program, Center for Business and Government, RPP Report No. 2 (23 November 1988).

industry's Responsible Care program³² or the wood and paper industry's Sustainable Forestry Initiative.³³

For each treatment to be evaluated, the researcher must obtain reliable measures of outcomes. Outcome measures used in evaluations of environmental policies can include measures of facility or firm environmental performance (for example, emissions of pollutants, energy use); human health impacts (such as days of illness or mortality or morbidity rates); or overall environmental impacts (such as acres of wetland or ambient air quality). When the ultimate outcome of concern cannot be directly measured, proxies must be used to assess the impact of a policy. For example, if one wants to measure the effectiveness of a program designed to reduce risk from exposure to toxic chemicals, the ultimate outcome of interest would be health effects from toxics. But measuring the health effects that stem from toxic emissions is complicated. Toxic emissions translate to different concentrations in the air and water based on geographic and geologic factors. Exposure to these concentrations varies based on age, activity levels, and other factors. The health effects of such exposure also can vary with age and current health status, among other factors. Even though this complexity may sometimes make it infeasible to measure directly the impact of a program on health, it will often be possible to assess the impact on some other measurable proxy for health risk, such as toxic emissions or ambient concentrations.

A. Isolating the Causal Effects of Treatments on Outcomes

The goal of program evaluation is to go beyond simple correlation to estimate the causal effect of the treatment on the outcomes selected for study. A treatment and outcome may be correlated, but the treatment can properly be said to be effective only if it has a causal effect on

³² See *Responsible Care*, <http://www.responsiblecare-us.com> (accessed 12 August 2004).

³³ See *Sustainable Forestry Initiative*, <http://www.aboutsfi.org/core.asp> (accessed 12 August 2004).

the outcome. For example, in his study of non-attainment designations under the Clean Air Act, Michael Greenstone noted that there was a strong correlation between regulation and subsequent decreases in sulfur dioxide. Areas that were designated as non-attainment areas experienced more rapid decreases in sulfur dioxide concentrations than areas that were designated as attainment. This correlation, though, proved insufficient to conclude that the Clean Air Act *caused* the decline in sulfur dioxide concentrations, as other factors instead accounted for the observed change.

How can researchers establish a causal link between policies and outcomes? In an ideal (but completely imaginary) world, the researcher would be able to manipulate policies and observe resulting outcomes, almost as in a chemistry experiment. For example, ideally the research would be able to designate a county as “non-attainment,” subject facilities in that county to more stringent regulations, and observe the sulfur dioxide concentrations that result. Then the ideal researcher would travel back in time, and replay history only this time the areas would be designated as “attainment” and facilities would not be subject to the more stringent regulations. If the researcher could actually observe both sets of outcomes for each county (that is, the level of sulfur dioxide concentrations in each with and without the non-attainment regulations), then the causal effect of the program would be a straightforward difference between these concentration levels.

Of course, the problem of causal inference arises because researchers cannot travel back in time and reassign counties from non-attainment to attainment and observe the resulting difference in concentrations. In reality, researchers never observe both potential outcomes for any given area. At any single point in time, the researcher can only observe the concentration levels of regulated areas, given that they were regulated, and the pollution levels for

nonregulated areas, given that they were not regulated. The challenge for program evaluation researchers is to use observable data to obtain valid estimates of the inherently unobservable difference in *potential* outcomes between the treatment and nontreatment (or control) groups.

B. Methods for Drawing Causal Inferences

How can researchers meet this challenge and draw reliable inferences about the causal effects of environmental policies?³⁴ If possible, the best approach would be to conduct a policy experiment and rely on random assignment of the treatment. If regulated entities subject to a treatment are assigned at random, then other factors that determine potential outcomes are also likely to be randomly distributed between the treatment and the control group. For example, with random assignment, there should not be systematic differences in the treatment and control groups in terms of such things as industry characteristics, size of firms, or publicly traded versus privately held ownership. In the case of random assignment, any differences in outcomes between the two groups of entities could be attributed to the treatment.

True random experimental designs are rare, if not entirely nonexistent, in environmental policy. Regulation, voluntary program participation, and other treatments of interest are almost never randomly assigned. Instead, regulatory status is frequently determined by factors that also correlate with potential outcomes—such as the size of the facility, the facilities' pollution levels, and the age of the facility. For environmental policy analysis, researchers will generally be forced to use observational study designs, which are also referred to as quasi-experimental

³⁴ A comprehensive answer to the question is, of course, beyond the scope of this paper. For an extensive discussion of the methods of program evaluation research, see Cook, Thomas D. and Donald T. Campbell (1979), *Quasi-Experimentation: Design and Analysis Issues for Field Settings* (Boston, MA: Houghton Mifflin Company). For a thorough treatment of qualitative methods of causal inference, see King, Gary, Robert O. Keohane, and Sidney Verba (1994), *Designing Social Inquiry* (Princeton, NJ: Princeton University Press). For discussion of the uses of evaluation methods in the policy process, see Rossi, Peter H. and Howard E. Freeman (1993), *Evaluation: A Systematic Approach*, 5th ed. (Newbury Park, CA: Sage Publications).

designs.³⁵ Because assignment to treatment is not random in observational studies, and treatment may be correlated with other determinants of potential outcomes, more sophisticated methods are required to isolate the causal effect of the treatment.

In observational studies where strict random assignment does not hold, there may be random assignment conditional on other observable variables. For example, imagine that one state's legislature passes a new regulation on hazardous waste while another state's does not. If the two states were quite similar—that is, they had the same types of facilities and the same socioeconomic and demographic variables—then the conditions of random assignment may be effectively met. If the states are not identical (that is, there are some differences in the types of facilities or community demographics), then observed differences in environmental performance across the states may be due to the difference in regulation *or* to the differences in these other variables. One state, for instance, may simply have larger or older industrial facilities which will affect how much hazardous waste they produce.

Variables that are correlated with the treatment and also with outcomes are called confounders—the presence of these variables confounds researchers' ability to draw causal inferences from a simple difference in average outcomes. If the confounders can be quantified with available data, however, then they are “observable.”³⁶ If all of the confounders are observable, then the causal effect of regulation could be estimated by examining the difference in outcomes, conditional on the confounding variables. In our hypothetical two-state example, a researcher could estimate the causal effect of the treatment by controlling for confounders such

³⁵ For a detailed description of a wide range of observational study designs, see Rosenbaum, Paul R. (2002) *Observational Studies* (New York: Springer-Verlag). Joshua Angrist and Alan Krueger offer an excellent summary of program evaluation methods, as applied to labor policies, including substantially more detail on each of the estimation methods discussed here. Angrist, Joshua D. and Alan B. Krueger (1999), “Empirical Strategies in Labor Economics,” in Ashenfelter, Orley and David Card, eds., *The Handbook of Labor Economics, Volume 3* (Amsterdam: Elsevier Science).

³⁶ The terms “observable” and “unobservable” mean what is observable and unobservable from the perspective of the researcher.

as the size or age of the facilities in both states. The researcher would essentially be comparing the environmental performance of facilities in the two states that have the same size, age, and other characteristics related to the generation of hazardous wastes.

Program evaluation researchers find analytic techniques such as regression and matching estimators to be useful when conditional random assignment holds. Regression analysis estimates a relationship between the outcome measure and a set of variables that may explain or be related to the outcome. One of these explanatory variables is the treatment variable, and the others are the confounders (also called control variables). Regression analysis isolates statistically the effect of the treatment holding all of the control variables constant.

To illustrate, imagine that Massachusetts passes a new law designed to lower pollution levels at all electronics plants. Connecticut also has many electronics plants, but these plants are not subject to the Massachusetts law. Plants in the two states are very similar except that plants in Massachusetts tend to be larger than plants in Connecticut. A regression of pollution levels on a variable that designates whether the plant is in Massachusetts and on another variable that measures plant size will yield an estimate of the effect of the Massachusetts regulation on pollution levels, holding the size of the plant fixed. If size were to be the only confounder, then this regression would yield a valid estimate of the causal effect of the Massachusetts regulation on pollution levels in electronics plants. For example, the Greenstone study seeks to isolate the causal effect of “non-attainment” status on county-level ambient levels of sulfur dioxide. To isolate the effect of non-attainment status, Greenstone uses regression analysis and controls for other variables that may also explain the decrease in sulfur dioxide concentrations including per capita income, total employment, and total population, among others.

An alternative statistical technique would be to use a matching estimator. For each observation that is subject to the treatment (such as an industrial facility subject to a regulation) the researcher finds a “matching” observation that is not subject to the treatment. To illustrate, let us return to the hypothetical Massachusetts regulation. To implement a matching estimator in this case, the researcher would take each facility in Massachusetts and find a facility in Connecticut of the same size. The researcher would then calculate the difference in pollution levels for the Massachusetts facility and its matching facility in Connecticut. The average of these differences for all Massachusetts plants is the average effect of the regulation on pollution.

Finding a “match” is relatively easy when there is only one confounder (size of the plant in our example). But what if it is important to control not just for size, but also for age of the facility and socio-economic characteristics of the community, such as the percent employed in manufacturing, population density, median household income, and so forth? To employ a matching estimator in this case, for each facility in Massachusetts the research would need to identify a facility in Connecticut of the same size, age, and with the same socio-economic characteristics. This may not be possible. This problem is often referred to as the “curse of dimensionality” because the number of dimensions (characteristics) on which facilities must be matched is large. One estimation technique that avoids the curse of dimensionality is matching on the propensity score.³⁷ The propensity score is simply the probability of being treated conditional on the control variables. Observations are then matched on the basis of their propensity to receive treatment, rather than on each individual control variable. In his study of the Clean Air Act, Greenstone also used matching on the propensity score to assess the effect on non-attainment designations on changes in sulfur dioxide concentrations. He first estimated the

³⁷ Rosenbaum, Paul R. and Donald B. Rubin (1983) “The Central Role of the Propensity Score in Observational Studies for Causal Effects,” *Biometrika*, 70(1):41-55.

likelihood that each county would be designated “non-attainment” based on the control variables. Then counties were matched based on this likelihood of treatment. The average difference between changes in treated counties’ sulfur dioxide concentrations and the changes in their corresponding match was the estimated causal effect of the Clean Air Act’s nonattainment regulations.

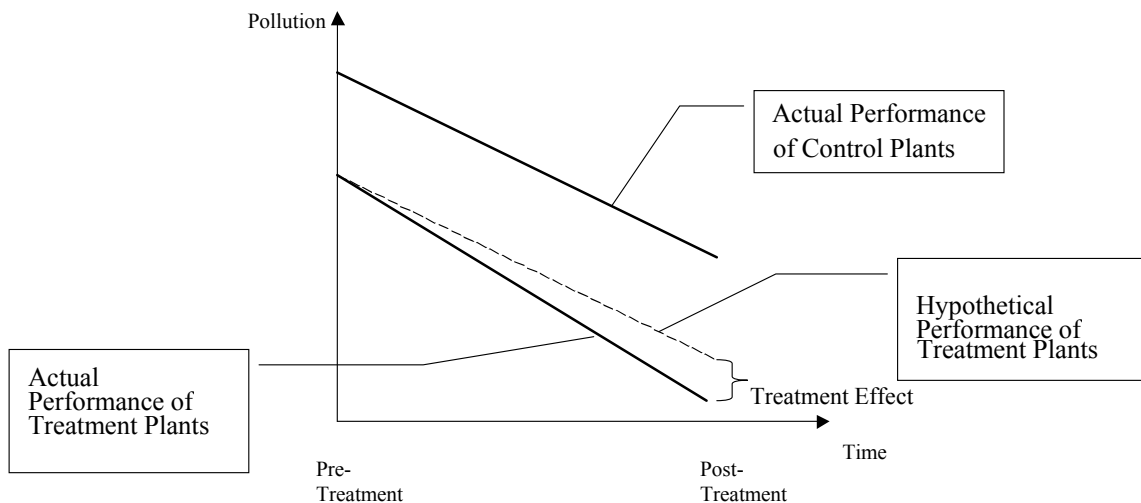
Regression and matching estimates assume that all of the confounders are observable. However, there are frequently cases when there are unobservable factors that are correlated with the treatment as well as potential outcomes. For example, facilities whose managers have a strong personal commitment to the environment may be more likely to participate in certain types of treatment, such as voluntary or so-called “beyond compliance” programs established by government agencies. However, the managers’ commitment, which will likely be unobservable to the researcher, is also likely to be correlated with the facility’s environmental performance regardless of participation in the program.³⁸ When there are unobservable confounders, standard regression and matching estimators will fail to provide a fully valid estimate of the causal effect of the treatment. In voluntary programs, for example, an ordinary regression estimate will be biased because it will be showing not only the effect of the voluntary program but also the effect of managers’ personal commitment to the environment, without being able to separate the level of impact of the two causal factors.

In such cases, alternative estimation strategies need to be used. An estimator known as the differences-in-differences estimator can yield a valid estimate of causal effects if the unobservable differences between the treated and nontreated entities remain constant over time. For example, imagine that the researcher has data on two sets of facilities: one set that

³⁸ Coglianese, Cary and Jennifer Nash (2001), *Regulating from the Inside: Can Environmental Management Systems Achieve Policy Goals?* (Washington, D.C.: Resources for the Future Press).

participates in a voluntary environmental program and one that does not. However, these two sets of facilities do not have identical indicators of environmental performance before the program is created. In fact, suppose the facilities that participate in the program have, on average, lower pollution levels even before participation. This is depicted graphically in Figure 3 on page 03. It is clear from the figure that it would be incorrect to characterize the difference in environmental performance after the program as the causal effect of the regulation, because some of that difference existed before program came into existence. The differences-in-differences estimator assumes that, in the absence of treatment, the difference in environmental performance would have been the same between the two sets of facilities. The dashed line represents the hypothetical pollution levels of the treatment plants if they never participated in the program. The causal effect of the program is correctly estimated as the incremental decrease in pollution in the post-program period, labeled “treatment effect” in Figure 3.

Figure 3. Graphical illustration of the differences-in-differences estimator



As noted, Figure 3 assumes that the unobservable differences have remained constant over time, but at times there may be good reason to think that they did not. Similarly, the differences-in-differences estimator requires data on at least two time periods -- one pre-treatment and one post-treatment -- which sometimes do not exist for environmental policies. In either of these situations, alternative estimation methods will be required.

One frequently used estimation technique in such circumstances is the instrumental variables method. To illustrate how this method works, return to the example of a voluntary program where participation is determined, in part, by facility managers' personal commitment to the environment, something which we assume is generally unobservable to the researcher. For sake of illustration, imagine that the regulatory agency administering the voluntary program sent letters inviting facilities to participate and did so to a completely random sample of facilities. Further, assume that, on average, facilities that receive the letter are more likely to participate than facilities that do not receive the letter, but that the correspondence between receipt of the letter and participation is not perfect. In other words, some facilities that receive the letter do not participate and some facilities that do not receive a letter nonetheless choose to participate. In such a circumstance, the participation decision is not randomly assigned and traditional statistical estimates of the effect of participation on outcome measures will be biased by the unobservable differences between participants and nonparticipants. What the instrumental variables estimator would do is capitalize on the fact that the government agency randomly assigned facilities to receive the invitation letter. In other words, some set of facilities would participate if they received a letter and would not participate if they did not receive a letter.³⁹ For these facilities

³⁹ In the parlance of the instrumental variables literature, these facilities are labeled compliers. This contrasts with always-takers (facilities that would have participated regardless of whether or not they received the letter), never takers (facilities that would not have participated regardless of whether they received the letter), and defiers (facilities that would have participated if they did not receive a letter, but would not have participated if they did

only, participation would be randomly assigned, because the letters were randomly sent. The statistical technique of instrumental variables estimation could isolate the effect of participation for those whose participation decisions were determined by whether they received a letter.

While the preceding discussion only briefly highlights the primary methods for estimating causal effects, it is clear that these methods are fairly well developed and available for use in evaluating the impacts of environmental policies. Indeed researchers have already used these methods to evaluate some environmental policies and programs. Yet other environmental programs remain significantly under- investigated. It is important to encourage more research using these kinds of methods so that reliable inferences can be drawn about the causal effects of environmental policies.

C. Data Availability and Program Evaluation of Environmental Policies

All of the program evaluation methods we have reviewed depend on valid and reliable data on environmental outcomes and other non-policy determinants of environmental outcomes (such as economic and technological factors). In other fields of public policy, researchers have available to them longstanding national surveys such as the Current Population Survey,⁴⁰ the National Longitudinal Survey of Youth,⁴¹ and the Panel Study of Income Dynamics.⁴² For the most part, these kinds of independent longitudinal data sets have not existed for environmental program evaluation.

receive a letter). The instrumental variables method provides a valid estimate of the causal effect of the treatment for compliers. Angrist, Joshua D., Guido W. Imbens, and Donald B. Rubin (1996), "Identification of Causal Effects Using Instrumental Variables," *Journal of the American Statistical Association*, 91(434): 444-472.

⁴⁰ U.S. Bureau of Labor Statistics and U.S. Census Bureau, *Current Population Survey*, <http://www.bls.census.gov/cps/cpsmain.htm> (accessed 12 August 2004).

⁴¹ U.S. Department of Labor, Bureau of Labor Statistics, *National Longitudinal Surveys*, <http://www.bls.gov/nls/home.htm> (accessed 12 August 2004).

⁴² University of Michigan Institute for Social Research, Panel Study of Income Dynamics, <http://psidonline.isr.umich.edu/> (accessed 12 August 2004).

Much of the data collected on environmental performance are built into the regulations themselves. For example, researchers have data on releases of toxic chemicals available from the Toxics Release Inventory (TRI); however, these data are available only for facilities that are subject to TRI regulations and only for the years during which TRI has been in effect. Similarly, data are reported by regulated facilities on their air emissions, water discharges, and hazardous waste generation, but these data exist only for the facilities that are regulated under the relevant statutes and for the years in which the regulations have been in effect. This close connection between data and regulation necessarily limits researchers' ability to evaluate the effects of these regulations as a treatment, because the mandated data are not available for unregulated facilities (the control group). However, these data can be used to evaluate the impact of other policies (such as voluntary programs or enforcement strategies) by comparing the outcomes for regulated firms subject to the treatment with outcomes for other regulated firms not subject to the treatment.

There are some instances where longitudinal data are available. Yet often these data have to do with ambient environmental conditions (such as air quality), and it is extremely difficult to pinpoint the effects of specific policy changes using these indicators. In most cases, it is impossible to use ambient data to identify the effects on individual firms or facilities. Researchers seeking longitudinal data on individual facility performance have often used TRI data because they are readily accessible for many (but by no means all) regulated firms since the late 1980s. But these data have their limitations too. Most obviously, they do not capture all the impacts firms have on the environment, as the data only cover releases of certain toxic pollutants. Furthermore, these data are self-reported, not adjusted for risk, and only reported by

facilities that exceed the established reporting thresholds.⁴³ All these factors can affect the valid use of TRI data as outcome measures for policy evaluation.

Researchers have sometimes used other measures of environmental impacts, such as total suspended solids levels or biological oxygen demand in water⁴⁴ or levels of water usage.⁴⁵ However, obtaining these measures has generally required intensive collection efforts which have so far limited the use of these data. To a large extent, the future of program evaluation in environmental policy will therefore be married to the future of environmental reporting and performance measurement.⁴⁶ This remains an important area for future research and funding.

III. The Future of Program Evaluation of Environmental Policy

The idea of subjecting policies to program evaluation research is certainly not new. At about the same time that environmental issues emerged on the federal policy agenda in the 1960s and early 1970s, the federal government also began to emphasize the use of performance

⁴³ Graham, Mary and Catherine Miller (2001) "Disclosure of Toxic Releases in the United States," *Environment* 43(8): 9-20; Snyder, Lori D. (2004b) "Are the TRI Data Valid Measures of Facility-Level Environmental Performance?: Reporting Thresholds and Truncation Bias" in *Essays on Facility Level Response to Environmental Regulation*, Ph.D. dissertation, Harvard University. A more recent concern is that government may begin to restrict access to such data due to concerns about its potential use by terrorists. For now at least, TRI data continue to be publicly available despite these concerns.

⁴⁴ Gunningham, Neil, Robert A. Kagan, and Dorothy Thornton (2003), *Shades of Green: Business, Regulation, and Environment* (Palo Alto: Stanford University Press).

⁴⁵ Olmstead, Sheila M., W. Michael Hanemann, and Robert N. Stavins (2003) "Does Price Structure Matter?: Household Water Demand Under Increasing-Block and Uniform Prices," mimeo, Yale School of Forestry and Environmental Studies.

⁴⁶ Esty, Daniel C. (2001), "Toward Data Driven Environmentalism: The Environmental Sustainability Index," *The Environmental Law Reporter News and Analysis* 31:10,603; Metzenbaum, Shelley (1998), "Making Measurement Matter: The Challenge and Promise of Building a Performance-Focused Environmental Protection System," Brookings Institution Center for Public Management, Report No. CPM-92-2. In addition to requiring good metrics on outcomes (that is, environmental performance) for the treatment and control groups, policy evaluation also requires data on other potential determinants of environmental performance. These include key variables describing the regulated entities (such as production processes, production levels, or market characteristics). The behavior of firms also remains an issue in need of further research, despite important work on corporate management that has begun to emerge. Andrews, et al., (2003), *Environmental Management Systems: Do They Improve Performance?*, Final Report of the National Database on Environmental Management Systems, University of North Carolina; Prakash, Aseem (2000), *Greening the Firm: The Politics of Corporate Environmentalism* (Cambridge: Cambridge University Press); and Reinhardt, Forest (2000), *Down to Earth: Applying Business Principles to Environmental Management* (Boston, MA: Harvard Business School Press).

evaluations as part of the budgetary process. These early attempts to encourage program evaluations of government programs certainly have spilled over into the field of environmental policy from time to time. Yet compared with other types of government programs, environmental policies have generated relatively little systematic program evaluation research.⁴⁷ Nevertheless, the present time is an especially ripe one for expanding program evaluation of environmental policies. Recent developments in policy innovation, government management, and data availability all contribute to a more conducive climate for expanded program evaluation on environmental policy issues.

First, numerous environmental policy innovations have been implemented in the past 15 years, and they are now ripe for evaluation. After an initial round of environmental policymaking in the early 1970s established the main framework for environmental regulation in the United States, there followed an extended period of concentrated efforts to implement these framework laws. By the late 1980s and early 1990s, however, a variety of factors led to a burst of innovative projects and policies implemented on the federal level and in the states. This later time period saw the introduction of EPA's well-known "bubble policy" that allowed facilities to average air emissions across multiple smokestacks or pipes, the Toxics Release Inventory that required the disclosure of information on releases of hazardous chemicals, and state pollution prevention laws that required planning to reduce the use of toxic chemicals. In addition, government has had considerable experience with a host of innovative, so-called voluntary or partnership programs,

⁴⁷ EPA is not the only regulatory agency to lag behind in program evaluation. In part due to the requirements under the Government Performance and Results Act, there is renewed research interest in understanding the institutional barriers and facilitators of program evaluation in different government agencies. For more information, see U.S. General Accounting Office (2003) Program Evaluation: An Evaluation Culture and Collaborative Partnerships Help Build Agency Capacity (Washington, D.C.: General Accounting Office Report # GAO-03-454); Mendeloff, John (2004), "Evaluation in Two Safety Regulatory Agencies," AEI-Brookings Working Paper (Regulatory Analysis 04-05).

such as EPA's Project XL and the Department of Interior's and National Marine Fisheries Service's Habitat Conservation Planning program.

Many of these programs have been in place for a long enough time so that their results can be estimated through sustained research efforts. Importantly, many of these programs apply selectively to a subset of all facilities within an industry or sector. Thus, these policies often make it feasible to compare the behavioral responses of participants and nonparticipants (the treatment and control groups). Of course, that does not imply that isolating the causal effect of these policies will be straightforward. The causal effect of voluntary programs is almost always confounded by differences in facilities that explain the decision to participate in the program in the first place—so called selection effects which, as discussed above, may include confounding factors. Some initial efforts are already underway to address these issues,⁴⁸ but more research on voluntary programs that takes such selection effects into account would be highly valuable.

The second factor favoring environmental program evaluation at the present time is a general climate in Washington and state capitals that supports performance evaluation of government programs. The Government Performance and Results Act (GPRA) requires all federal agencies to devise specific performance goals and report on their achievement of these goals using performance measures.⁴⁹ This focus on performance measures -- rather than on administrative measures, such as numbers of inspections or numbers of voluntary participants -- increases the need for outcome-based evaluation.⁵⁰ Furthermore, the Office of Management and Budget has developed the Performance Assessment Rating Tool (PART) and required that

⁴⁸ For a survey of evaluation of voluntary environmental programs that provides detailed references for evaluations addressing issues of selection bias, see Alberini, Anna and Kathleen Segerson (2002), "Assessing Voluntary Programs to Improve Environmental Quality," *Environmental and Resource Economics*, 22: 157-184.

⁴⁹ 31 U.S.C. § 1101 and 39 U.S.C. § 2801 (1993).

⁵⁰ Sparrow, Malcolm (2000), *The Regulatory Craft: Controlling Risks, Solving Problems, Managing Compliance* (Washington, D.C.: Brookings Institution Press).

government programs begin to use this tool to evaluate whether they are resulting in significant progress toward public goals.⁵¹ Just as executive orders on the ex ante use of economic analysis for major regulations have given greater prominence to those analytic tools within government agencies, GPRA and PART may increase demand within environmental and natural resources agencies for program evaluation research.

Finally, while data limitations remain the greatest barrier to program evaluation in the environmental field, these data are increasingly available online and agencies are taking other steps that make it easier to use these data for evaluation research.⁵² While EPA has collected data on air emissions, water discharges, hazardous waste generation, and toxics releases for several decades, in the past these data were collected and maintained separately by the respective program offices within the agency. As a result, each office generated its own metadata and, importantly, its own numbering system for identifying facilities. Thus, the same facility was assigned an AIRS (Aerometric Information Retrieval System) identifier for the air office, a Permit Compliance System identifier for the water office, and a TRI identifier for the office of information. Researchers hoping to combine data from more than one source were forced to match facilities by hand—usually by name and address. Recently, however, EPA has instituted a common Facility Registry System (FRS) identifier. This identifier has been added to all existing EPA databases, allowing researchers more easily to match data on a facility from multiple sources.

⁵¹ OMB, *Program Assessment Rating Tool (PART)*, <http://www.whitehouse.gov/omb/part/>. See also Office of Management and Budget, Office of Information and Regulatory Affairs (2003b), Testimony of the Honorable Clay Johnson III, Deputy Director for Management, before the Committee on Government Reform, U.S. House of Representatives, September 18, 2003, available at http://www.whitehouse.gov/omb/legislative/testimony/cjohnson/030918_cjohnson.html.

⁵² See Esty, Daniel C. (2004) “Environmental Protection in the Information Age,” *New York University Law Review* 79: 115-209.

Another recent development that is likely to improve environmental policy evaluation is EPA's Risk Assessment Environmental Indicators (RSEI) model. The RSEI model combines data on toxics releases from TRI with scientific indicators of the effect of these releases on health risks. By weighting TRI data by risk, RSEI allows researchers to draw inferences about the health effects of policy interventions. Obviously the value of the RSEI model is married to the value of the underlying TRI data and the toxicity weights, both of which are open to criticism. But developing pollution-risk models is an important step in promoting more meaningful risk measures that can be used for evaluation purposes.

EPA has also expanded data on regulatory compliance. The Integrated Data for Enforcement Analysis (IDEA), Enforcement and Compliance History Online (ECHO), and Online Tracking Information System (OTIS) provide researchers with easier access to certain kinds of data on enforcement and compliance behavior. EPA has gone still further in integrating enforcement, compliance, and environmental performance data through the Sector Facility Indexing Project (SFIP). For five industry sectors—automobile assembly, pulp and paper, petroleum refining, iron and steel production, and metal smelting and refining—SFIP provides one-stop access to data on the number of inspections, compliance with federal regulations, enforcement actions, toxic release levels, and spills. The SFIP database also provides information about the facility, including production capacity and demographic characteristics of the surrounding area.

While there is much more work to be done to develop and categorize meaningful metrics,⁵³ recent developments appear headed in a valuable direction for the future of program

⁵³ Metzenbaum, Shelley (2003), "More Nutritious Beans," *Environmental Forum*, April/May: 18-41.

evaluation research.⁵⁴ Table 1 on page 01 provides information on key types of data currently available for program evaluation of environmental policies. Improvements in data quality and data access, combined with the ripeness of a variety of innovative regulatory instruments and the managerial pressure to evaluate the effectiveness of government programs, suggest that the coming years could be much more promising for program evaluation research on environmental policy.

Conclusion

Program evaluation research provides valuable information for policy decisionmaking. Decision makers in state and federal regulatory agencies, legislatures, and other oversight bodies (such as OMB) need to design and implement policies that work to achieve public goals. With information from retrospective evaluations of policies, policymakers will be better able to determine what policies to adopt (and how to design them) in the future. Policy evaluation research can also help identify ways to change existing policies to make them more beneficial.

To be sure, when research shows that policies having intuitive appeal do not yield the anticipated or desired results, some decisionmakers may remain faithful to their intuitions rather than to what the evidence shows. Resistance to research findings can also occur when actors in the policy process have interests at stake in certain policies. Although these are real considerations, it should be noted that the same was (and even still is to a certain extent) true in

⁵⁴ In addition to developments in EPA's data management, promising nongovernmental efforts to study and improve different kinds of environmental metrics have also emerged in recent years. O'Malley, Robin, Kent Cavender-Bares, and William C. Clark (2003) "Providing 'Better' Data: Not as Simple as it Might Seem," *Environment* 45:8-18 (May); The H. John Heinz III Center for Science, Economics and the Environment, (2003), *The State of the Nation's Ecosystems: Measuring the Lands, Waters, and Living Resources in the United States* (Cambridge: Cambridge University Press); Esty, Daniel and Peter K. Cornelius, *Environmental Performance Measurement: The Global Report, 2001-2002* (Oxford: Oxford University Press); and National Academy of Engineering (1999), *Industrial Environmental Performance Metrics* (Washington, D.C.: National Academy Press).

other areas like medicine or education that have more fully adopted principles of evidence-based decisionmaking. The value of evidence-based practice is only made more compelling when one acknowledges the strength of the biases that can and do affect decisionmaking.

More program evaluation research should help counteract the skeptical responses to research in the policy process. If a single study demonstrates that a program is effective or ineffective, those who are predisposed to think otherwise may be quick to dismiss the findings. With multiple program evaluation studies on environmental policies, such dismissals will become more difficult to sustain. If several studies reach consistent results, then over time the preponderance of the empirical evidence will be more likely to affect the decisions of policymakers.

Moreover, the reality is that some regulatory officials are receptive to research that can tell them about what works and what does not work. For example, EPA has recently released a strategy document on environmental management systems that gives priority to the need for careful program evaluation of initiatives in this area.⁵⁵ Consistent with this priority, EPA has even recently sponsored research efforts on management-based strategies for improving environmental performance.⁵⁶

Only with more efforts to give priority to program evaluation research will decisionmaking over environmental policy be able to become based more on careful deliberation than on rhetorical and political contestation. It is doubtful that program evaluation research will end political conflict altogether or immunize policymakers from all error. But it can help sharpen

⁵⁵ Environmental Protection Agency EMS Permits and Regulations Workgroup (2004), *EPA's Strategy for Determining the Role of Environmental Management Systems in Regulatory Programs* (April 12).

⁵⁶ Coglianese, Cary and Jennifer Nash (2001), *Regulating from the Inside: Can Environmental Management Systems Achieve Policy Goals?* (Washington, D.C.: Resources for the Future Press); and Coglianese, Cary and Jennifer Nash (2004), "Leveraging the Private Sector: Management-Based Strategies for Improving Environmental Performance," Regulatory Policy Program Report No. 6, John F. Kennedy School of Government.

the focus of policy deliberation as well as inform government's choices about how to allocate scarce resources more effectively. Making program evaluation of environmental policy a priority will be a necessary step toward an evidence-based approach to environmental decisionmaking.

Table 1. Data Sources for Program Evaluation of Environmental Policy

Topic	Data	Source	Description	Types of Facilities Covered
Data on Outcomes				
Toxics and Hazardous Waste	Toxics Release Inventory (TRI)	Self-reported by facilities	Contains data on pounds of chemicals released to air, water, land, underground injection, and transferred offsite. Also includes data on pollution prevention activities and recycling.	Manufacturing facilities that meet certain thresholds.
	CERCLIS		Contains data on Superfund sites, including whether they are on the National Priority List (NPL), ownership information, dates and descriptions of actions taken.	Superfund sites
	Record of Decisions (ROD)		Provides *.pdf files of decisions regarding Superfund sites	Superfund sites
	Resource Conservation and Recovery Act Information (RCRAInfo)		Contains data on hazardous waste generation for large quantity generators of hazardous waste and disposal information for all treatment, storage and disposal facilities (TSDs). Replaces two previously maintained databases, the Biennial Reporting System (BRS) and the Resource Conservation and Recovery Information System (RCRIS).	Generators of Hazardous Waste and Hazardous Waste Treatment Storage and Disposal Facilities

Table 1 (continued)

Topic	Data	Source	Description	Types of Facilities Covered
Data on Outcomes (Continued)				
Water	Permit Compliance System (PCS)	Discharge data is self-reported by facilities. Other information entered and maintained by either U.S. EPA or the states.	Contains data on permit limits, discharge levels, enforcement, and inspection activities.	All National Permit Discharge and Elimination System (NPDES) permit holders
	Safe Drinking Water Information System	Maintained by U.S. EPA or designated states.	Contains data on drinking water contaminant violations and enforcement actions.	Public drinking water systems
Air	Aerometric Information Retrieval System/AIRS Facility Subsystem (AIRS/AFS)	Self-reported by facilities	Contains data on permits, emissions, inspection, and compliance with air quality standards.	All air permit holders
Compliance and Enforcement	ECHO	Combined enforcement and compliance data from PCS, AIRS, and RCRAInfo	Contains data on inspection and compliance for water, air, and hazardous waste permit holders.	Same as underlying PCS, AIRS, and RCRAInfo databases
	Integrated Data for Enforcement Analysis	Combined enforcement and compliance data from PCS, AIRS, and RCRAInfo	Contains data on inspection and compliance for water, air, and hazardous waste permit holders.	Same as underlying PCS, AIRS, and RCRAInfo databases

Table 1 (continued)

Topic	Data	Source	Description	Types of Facilities Covered
Data on Covariates				
Firm Data	Compustat	Standard and Poor's	Contains income, balance sheet, and cash flow data.	Publicly held companies. Data is available by subscription.
	Dunn and Bradstreet Million Dollar Database	Dunn and Bradstreet	Contains data on sales, employment, industry, and ownership.	1.6 million U.S. and Canadian companies, both private and public. Data is proprietary and available by subscription only.
Plant Data	Dunn and Bradstreet Million Dollar Database	Dunn and Bradstreet	Contains employment information at plant and firm level.	1.6 million U.S. and Canadian companies, both private and public. Data is proprietary and available by subscription only.
	Longitudinal Research Database	U.S. Census Bureau	Contains data from the Census of Manufacturers and the Annual Survey of Manufacturers. Data include employment, product classes, and shipments.	Available only by approved proposal at one of eight regional data centers.

Author Note

Lori Snyder Benneer is an assistant professor of Environmental Economics and Policy at the Nicholas School of the Environment and Earth Sciences at Duke University. She received her Ph.D. in Public Policy from Harvard University. Her research focuses on estimating the effect of different regulatory innovations on measures of facility-level environmental performance, such as pollution levels, chemical use, and technology choice. Her recent work has focused on measuring the effectiveness of management-based regulations, which require each regulated entity to develop its own internal rules and initiatives to achieve reductions in pollution, as well as the effectiveness of regulations that mandate public reporting of toxic emissions. She can be reached at (919) 613-8083 or lori.benneer@duke.edu.

Cary Coglianese is associate professor of Public Policy and chair of the Regulatory Policy Program at the John F. Kennedy School of Government at Harvard University and Visiting Professor of Law at the University of Pennsylvania Law School. He received his J.D., M.P.P., and Ph.D. from the University of Michigan. His interdisciplinary research focuses on regulatory policy and administrative law, with a particular emphasis on the empirical evaluation of alternative and innovative regulatory strategies. His recent work has focused on the application of management-based and performance-based regulation to health, safety, and environmental problems; the role of science, economics, and information in the regulatory process; and the effects of consensus-building on regulatory policymaking. He can be reached at (617) 495-1402 or cary_coglianese@harvard.edu.