



**Policy Evaluation through Congressional Testimony** Secretary of the Treasury Timothy Geithner testifies before a Congressional Oversight Panel evaluating the effectiveness of the Troubled Asset Relief Program (TARP) June 22, 2010. Geithner argued that the TARP program, often criticized as the “Wall Street bailout,” was successful in stabilizing the financial community and that taxpayers were recovering a major portion of their investment. (© Benjamin J. Myers/Corbis)

# 4

## Policy Evaluation

### Finding Out What Happens After a Law Is Passed

Americans often assume that once we pass a law, create a bureaucracy, and spend money, the purpose of the law, the bureaucracy, and the expenditure will be achieved. We assume that when Congress adopts a policy and appropriates money for it, and when the executive branch organizes a program, hires people, spends money, and carries out activities designed to implement the policy, the effects of the policy will be felt by society and will be those intended. Unfortunately, these assumptions are not always warranted. The national experiences with many public programs indicate the need for careful appraisal of the real impact of public policy.

Does the government really know what it is doing? Generally speaking, no. Governments usually know how much money they spend; how many persons (“clients”) are given various services; how much these services cost; how their programs are organized, managed, and operated; and, perhaps, how influential interest groups regard their programs and services. But even if programs and policies are well organized, efficiently operated, widely utilized, adequately financed, and generally supported by major interest groups, we may still want to ask, So what? Do they work? Do these programs have any beneficial effects on society? Are the effects immediate or long range? Positive or negative? What is the relationship between the costs of the program and the benefits to society? Could we be doing something else with more benefit to society with the money and work force devoted to these programs? Unfortunately, governments have done very little to answer these more basic questions.

#### **POLICY EVALUATION: ASSESSING THE IMPACT OF PUBLIC POLICY**

Policy evaluation is learning about the consequences of public policy. Other, more complex, definitions have been offered: “Policy evaluation is the assessment of the overall effectiveness of a national program in meeting its objectives, or assessment of the relative effectiveness of two or more programs in meeting common objectives.”<sup>1</sup> “Policy evaluation research is the objective, systematic, empirical examination of the effects ongoing policies and public programs have on their targets in terms of the goals they are meant to achieve.”<sup>2</sup>

Some definitions tie evaluation to the stated “goals” of a program or policy. But since we do not always know what these “goals” really are, and because we know that some programs and policies pursue conflicting “goals,” we will not limit our notion of policy evaluation to their achievement. Instead, we will concern ourselves with all of the consequences of public policy, that is, with “policy impact.”

The impact of a policy is all its *effects on real-world conditions*, including:

- Impact on the target situation or group
- Impact on situations or groups other than the target (spillover effects)
- Impact on future as well as immediate conditions
- Direct costs, in terms of resources devoted to the program
- Indirect costs, including loss of opportunities to do other things

Ideally, all the benefits and costs, both immediate and future, should be measured.

### Measuring Impact, Not Output

“Policy impact” is not the same as “policy output.” In assessing policy impact, we cannot be content simply to measure government activity. For example, the number of dollars spent per member of a target group (per pupil educational expenditures, per capita welfare expenditures, per capita health expenditures) is not really a measure of the impact of a policy on the group. It is merely a measure of government activity—that is, a measure of *policy output*. Unfortunately many government agencies produce reams of statistics measuring outputs—such as welfare benefits paid, criminal arrests and prosecutions, Medicare payments, and school enrollments. But this “bean counting” tells us little about poverty, crime, health, or educational achievement. We cannot be satisfied with measuring how many times a bird flaps its wings; we must know how far the bird has flown. In describing public policy, or even in explaining its determinants, measures of policy output are important. But in assessing *policy impact*, we must identify changes in society that are associated with measures of government activity.

### Target Groups

The target group is that part of the population for whom the program is intended—such as the poor, the sick, the ill-housed. Target groups must first be identified and then the desired effect of the program on the members of these groups must be determined. Is it to change their physical or economic circumstances—for example, the percentage of minorities or women employed in professional or managerial jobs, the income of the poor, the infant death rate? Or is it to change their knowledge, attitudes, awareness, interests, or behavior? If multiple effects are intended, what are the priorities among different effects? What are the possible unintended effects (side effects) on target groups?

### Nontarget Groups

All programs and policies have differential effects on various segments of the population. Identifying important nontarget groups for a policy is a difficult process. For example, what is the impact of the welfare reform on groups *other* than the poor—government bureaucrats, social workers, local political figures, working-class families who are not on welfare, taxpayers, and others? Nontarget effects may be expressed as benefits as well as costs, such as the benefits to the construction industry of public housing projects.

**TABLE 4–1 Assessing Policy Impact** A rational approach to policy evaluation tries to calculate the difference between all present and future, target and nontarget, costs and benefits.

	BENEFITS		COSTS	
	Present	Future	Present	Future
Target Groups	Benefits	Benefits	Costs	Costs
Nontarget Groups	Benefits	Benefits	Costs	Costs
	Sum	Sum	Sum	Sum
	Present	Future	Present	Future
	Benefits	Benefits	Costs	Costs
	Sum All Benefits	Minus	Sum All Costs	
		=	Net Policy Impact	

### Short-Term and Long-Term Effects

When will the benefits or the costs be felt? Is the program designed for short-term emergencies? Or is it a long-term, developmental effort? If it is short term, what will prevent the processes of incrementalism and bureaucratization from turning it into a long-term program, even after the immediate need is met? Many impact studies show that new or innovative programs have short-term positive effects—for example, Operation Head Start and other educational programs. However, the positive effects frequently disappear as the novelty and enthusiasm of new programs wear off. Other programs experience difficulties at first, as in the early days of Social Security, but turn out to have “sleepers” effects, as in the widespread acceptance of Social Security today. Not all programs aim at the same degree of permanent or transient change.

### Calculating Net Benefits and Costs

The task of calculating the *net* impact of a public policy is truly awesome. It would be all the benefits, both immediate and long range, minus all the costs, both immediate and future (see Table 4–1). Even if all these costs and benefits are known (and everyone agrees on what is a “benefit” and what is a “cost”), it is still very difficult to come up with a net balance.

## THE SYMBOLIC IMPACT OF POLICY

The impact of a policy may also include its symbolic effects. Its symbolic impact deals with the perceptions that individuals have of government action and their attitudes toward it. Even if government policies do not succeed in eliminating poverty, preventing crime, and so on, the failure of government to *try* to do these things would be even worse. Individuals, groups, and whole

societies frequently judge public policy in terms of its *good intentions* rather than *tangible accomplishments*. Sometimes very popular programs have little positive tangible impact.

The policies of government may tell us more about the aspirations of a society and its leadership than about actual conditions. Policies do more than effect change in societal conditions; they also help hold people together and maintain an orderly state.

Once upon a time politics was described as “who gets what, when, and how.” Today it seems that politics centers on “who *feels* what, when, and how.” What governments say is as important as what governments do. Television has made the image of public policy as important as the policy itself. Systematic policy analysis concentrates on what governments do, why they do it, and what difference it makes. It devotes less attention to what governments say. Perhaps this is a weakness in policy analysis. Our focus is primarily on activities of governments rather than their rhetoric.

## PROGRAM EVALUATION: WHAT GOVERNMENTS USUALLY DO

Most government agencies make some effort to review the effectiveness of their own programs. These reviews usually take one of the following forms:

### Hearings and Reports

The most common type of program review involves hearings and reports. Government administrators are asked by chief executives or legislators to give testimony (formally or informally) on the accomplishments of their own programs. Frequently, written annual reports are provided by program administrators. But testimonials and reports of administrators are not very objective means of program evaluation. They frequently magnify the benefits and minimize the costs of the program.

### Site Visits

Occasionally teams of high-ranking administrators, expert consultants, legislators, or some combination of these people will decide to visit agencies or conduct inspections in the field. These teams can pick up impressionistic data about how programs are being run, whether they are following specific guidelines, whether they have competent staffs, and sometimes whether or not the clients (target groups) are pleased with the services.

### Program Measures

The data developed by government agencies themselves generally cover policy output measures: the number of recipients in various welfare programs, the number of persons in work-force training programs, the number of public hospital beds available, the tons of garbage collected, or the number of pupils enrolled. But these program measures rarely indicate what impact these numbers have on society: the conditions of life confronting the poor, the success of work-force trainees in finding and holding skilled jobs, the health of the nation's poor, the cleanliness of cities, and the ability of graduates to read and write and function in society.

### Comparison with Professional Standards

In some areas of government activity, professional associations have developed standards of excellence. These standards are usually expressed as a desirable level of output: for example, the number of pupils per teacher, the number of hospital beds per one thousand people, the number of cases

for each welfare worker. Actual government outputs can then be compared with ideal outputs. Although such an exercise can be helpful, it still focuses on government outputs and not on the impact of government activities on the conditions of target or nontarget groups. Moreover, the standards themselves are usually developed by professionals who are really guessing at what ideal levels of benefits and services should be. There is rarely any hard evidence that ideal levels of government output have any significant impact on society.

### Evaluation of Citizens' Complaints

Another common approach to program evaluation is the analysis of citizens' complaints. But not all citizens voluntarily submit complaints or remarks about governmental programs. Critics of government programs are self-selected, and they are rarely representative of the general public or even of the target groups of government programs. There is no way to judge whether the complaints of a vocal few are shared by the many more who have not spoken up. Occasionally, administrators develop questionnaires for participants in their program to learn what their complaints may be and whether they are satisfied or not. But these questionnaires really test public opinion toward the program and not its real impact on the lives of participants.

### Surveys of Public Opinion

Occasionally governments undertake to survey citizens about their satisfaction or dissatisfaction with various programs and services. This is more common at the local level of government. Yet even polls focused on federal government services can be instructive (see Table 4–2).

## PROGRAM EVALUATION: WHAT GOVERNMENTS CAN DO

None of the common evaluative methods just mentioned really attempts to weigh costs against benefits. Indeed, administrators seldom calculate the ratio of costs to services—the dollars required to train one worker, to provide one hospital bed, to collect and dispose of one ton of garbage. It is even more difficult to calculate the costs of making specific changes in society—the dollars required to raise student reading levels by one grade, to lower the infant death rate by one point, to reduce the crime rate by one percent. To learn about the real impact of governmental programs on society, more complex and costly methods of program evaluation are required.

Systematic program evaluation involves comparisons—comparisons designed to estimate what changes in society can be attributed to the program rather than nonprogram factors. Ideally, this means comparing what “actually happened” to “what would have happened if the program had never been implemented.” It is not difficult to measure what happened; unfortunately too much program evaluation stops there. The real problem is to measure what would have happened without a program and then compare the two conditions of society. The difference must be attributable to the program itself and not to other changes that are occurring in society at the same time.

### Before Versus After Comparisons

There are several common research designs in program evaluation. The most common is the before-and-after study, which compares results in a jurisdiction at two times—one before the program was implemented and the other some time after. Usually only target groups are examined.

**TABLE 4–2 Public Satisfaction/Dissatisfaction with Federal Government Programs** Polls can reflect general satisfaction or dissatisfaction with federal programs. Often the military ranks at or near the top of public esteem; the public is decidedly less satisfied with energy policy, health care, poverty programs, and the nation’s finances.

*Next we are going to name some major areas the federal government handles. For each one please say whether you are satisfied or dissatisfied with the work the government is doing.*

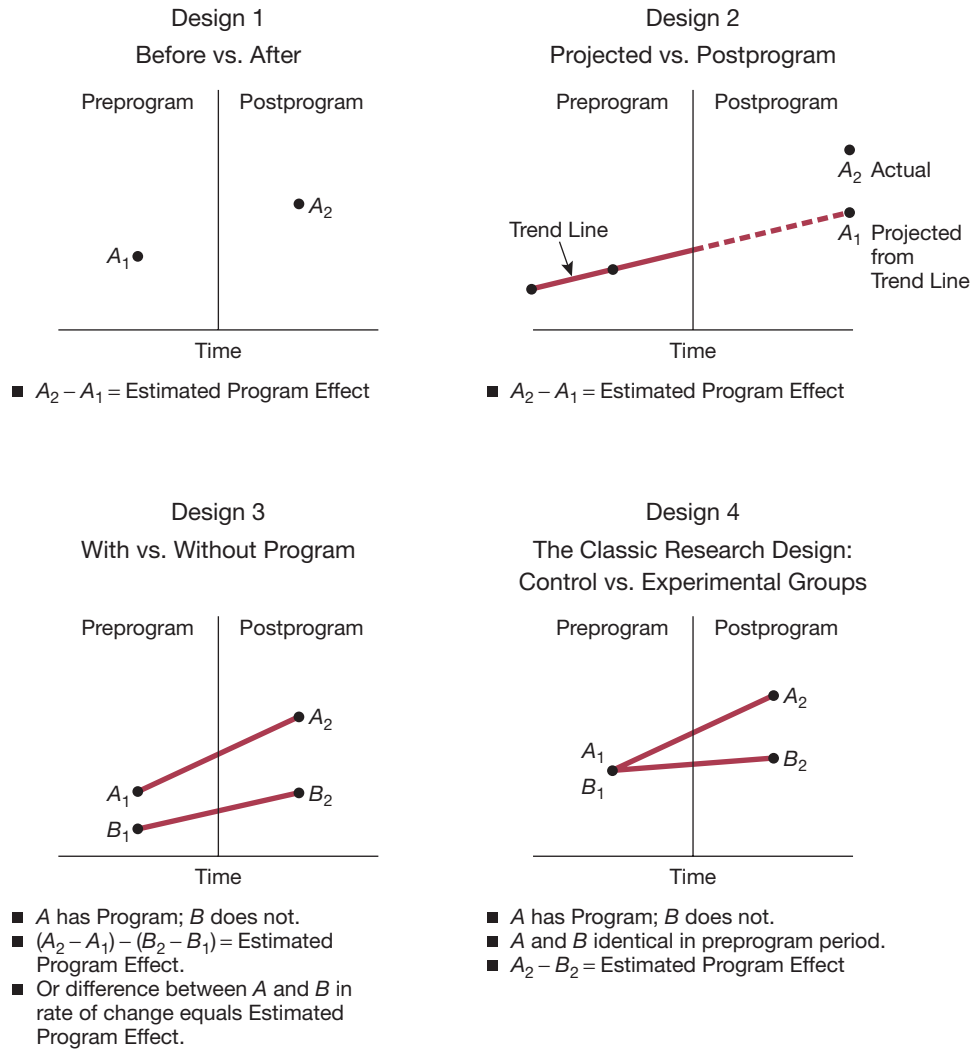
	Satisfied	Dissatisfied	Unsure
National parks	71	27	2
Military and national defense	59	40	1
Agriculture and farming	56	38	5
Transportation	56	42	2
Homeland security	50	49	1
Environmental issues	48	51	1
Public housing/urban development	47	49	4
Criminal justice	47	52	1
Labor and employment issues	44	54	2
Foreign affairs	41	58	1
Education	41	59	0
Job creation/economic growth	39	60	1
Responding to natural disasters	33	66	1
Energy policy	27	71	2
Health care	24	75	1
Poverty programs	24	75	1
The nation’s finances	23	76	1

SOURCE: *The Polling Report*, accessed January 2011, [www.pollingreport.com](http://www.pollingreport.com).

These before-and-after comparisons are designed to show program impacts, but it is very difficult to know whether the changes observed, if any, came about as a result of the program or as a result of other changes that were occurring in society at the same time (see Design 1, Figure 4–1).

**Projected Trend Line Versus Postprogram Comparisons**

A better estimate of what would have happened without the program can be made by projecting past (preprogram) trends into the postprogram time period. Then these projections can be compared with what actually happened in society after the program was implemented. The difference between the projections based on preprogram trends and the actual postprogram data can be attributed to the program itself. Note that data on target groups or conditions must be obtained for several time periods before the program was initiated, so that a trend line can be established (see Design 2, Figure 4–1). This design is better than the before-and-after design, but it requires more effort by program evaluators.



**FIGURE 4-1 Policy Evaluation Research Designs** Policy evaluation can utilize a variety of research designs.

Consider, for example, efforts at evaluating welfare reform (see the section “Evaluation: Is Welfare Reform Working?” in Chapter 7). To date, most evaluations of welfare reform have followed the trend line research design. If the goal of the reform is to reduce welfare rolls, there is ample evidence that the program has contributed to that goal (see Figure 7-6). The “target group” (recipients of cash welfare payments) has been reduced by over half since the ending of the federal cash entitlement program, Aid to Families with Dependent Children, and its substitution with the federally aided state program, Temporary Assistance for Needy Families, in 1996. But it is not clear exactly what proportion of this reduction is due to the policy itself and what proportion is due to other economic factors. All we really know is that the welfare rolls declined.

### Comparisons Between Jurisdictions With and Without Programs

Another common evaluation design is to compare individuals who have participated in programs with those who have not, or to compare cities, states, or nations which have programs with those that do not. Comparisons are sometimes made in the postprogram period only; for example, comparisons of the job records of those who have participated in work-force training programs with those who have not, or comparisons of homicide rates in states that have the death penalty with the homicide rates in states without the death penalty. But so many other differences exist between individuals or jurisdictions that it is difficult to attribute differences in their conditions to differences in government programs. For example, persons who voluntarily enter a work-force training program may be more motivated to find a job or have different personal characteristics than those who do not. States with the death penalty may tend to be rural states, which have lower homicide rates than urban states, which may or may not have the death penalty.

Some of the problems involved in comparing jurisdictions with and without programs can be resolved if we observe both kinds of jurisdictions before and after the introduction of the program. This enables us to estimate differences between jurisdictions before program efforts are considered. After the program is initiated, we can observe whether the differences between jurisdictions have widened or not (see Design 3, Figure 4-1). This design provides some protection against attributing differences to a particular program when underlying socioeconomic differences between jurisdictions are really responsible for different outcomes.

### Comparisons Between Control and Experimental Groups Before and After Program Implementation

The classic research design involves the careful selection of control and experimental groups that are identical in every way, the application of the policy to the experimental group only, and the comparison of changes in the experimental group with changes in the control group after the application of the policy. Initially, control and experimental groups must be identical, and the preprogram performance of each group must be measured and found to be the same. The program must be applied only to the experimental group. The postprogram differences between the experimental and control groups must be carefully measured (see Design 4, Figure 4-1). This classic research design is preferred by scientists because it provides the best opportunity of estimating changes that derived from the effects of other forces in society.

## EXPERIMENTAL POLICY RESEARCH

Many policy analysts argue that policy experimentation offers the best opportunity to determine the impact of public policies. This opportunity rests on the main characteristics of experimental research: the systematic selection of experimental and control groups, the application of the policy under study to the experimental group only, and the careful comparison of differences between the experimental and the control groups after the application of the policy. But government-sponsored experimental policy research raises a series of important questions.

### A Bias Toward Positive Results

First, are government-sponsored research projects predisposed to produce results supportive of popular reform proposals? Are social scientists, whose personal political values are generally liberal and reformist, inclined to produce findings in support of liberal reform measures? Moreover,

successful experiments—in which the proposed policy achieves positive results—will receive more acclaim and produce greater opportunities for advancement for social scientists and administrators than will unsuccessful experiments—in which the policy is shown to be ineffective. Liberal, reform-oriented social scientists expect liberal reforms to produce positive results. When reforms appear to do so, the research results are immediately accepted and published; but when results are unsupportive or negative, the social scientists may be inclined to go back and recode their data, redesign their research, or reevaluate their results because they believe a “mistake” must have been made. The temptation to “fudge the data,” “reinterpret” the results, coach participants on what to say or do, and so forth will be very great. In the physical and biological sciences, the temptation to “cheat” in research is reduced by the fact that research can be replicated and the danger of being caught and disgraced is very great. But social experiments can seldom be replicated perfectly, and replication seldom brings the same distinction to a social scientist as does the original research.

### The Hawthorne Effect

People behave differently when they know they are being watched. Students, for example, generally perform at a higher level when something—anything—new and different is introduced into the classroom routine. This “Hawthorne effect” may cause a new program or reform to appear more successful than the old, but it is the newness itself that produces improvement. The term is taken from early experiments at the Hawthorne plant of Western Electric Company in Chicago in 1927. It was found that worker output increased with *any* change in routine, even decreasing the lighting in the plant.<sup>3</sup>

### Generalizing Results to the Nation

Another problem in policy research is that results obtained with small-scale experiments may differ substantially from those that would occur if a large-scale nationwide program were adopted. For example, years ago a brief experiment involving a small number of families purported to show that a government-guaranteed income did not change the work behavior of recipients; they continued to behave as their neighbors did—searching for jobs and accepting employment when it was offered.<sup>4</sup> Subsequent studies of the effects of a guaranteed government income challenged even these experimental group findings but also predicted that a *nationwide* program would produce much more dramatic changes in working behavior. If everyone in the nation were guaranteed a minimum annual income, cultural standards might be changed nationwide; the resulting work disincentives might “seriously understate the expected cost of an economy-wide program.”<sup>5</sup>

### Ethical and Legal Issues

Experimental strategies in policy impact research raise still other problems. Do government researchers have the right to withhold public services from individuals simply to provide a control group for experimentation? In the medical area, where giving or withholding treatment can result in death or injury, the problem is obvious and many attempts have been made to formulate a code of ethics. But in the area of social experimentation, what are we to say to control groups who are chosen to be similar to experimental groups but denied benefits in order to serve as a base for comparison? Setting aside the legal and ethical issues, it will be politically difficult to provide services for some people and not others.



**Evaluation as a Partisan Activity** House Budget Committee member Paul Ryan (R.-WI) points to the 2000-plus page Obama health care reform bill still in markup binders. The bill became the Patient Protection and Affordable Care Act of 2010 on strictly partisan votes in the House and Senate. Few legislators knew what was in the bill when they voted on it. Full implementation, including the mandate that every American obtain health insurance, occurs in 2014. (Washington Post/Getty Images)

### Political Interpretations of Results

Finally, we must acknowledge that the political milieu shapes policy research. Politics helps decide what policies and policy alternatives will be studied in the first place. Politics can also affect findings themselves, and certainly the interpretations and uses of policy research are politically motivated.

Despite these problems, the advantages of policy experimentation are substantial. It is exceedingly costly for society to commit itself to large-scale programs and policies in education, welfare, housing, health, and so on without any real idea about what works.

### FEDERAL EVALUATION: THE OFFICE OF MANAGEMENT AND BUDGET

Among its many responsibilities, the Office of Management and Budget (OMB) in the Executive Office of the President undertakes program evaluations and encourages executive agencies to do so as well. It advises executive agencies to “embrace a culture where performance measurement and evaluation are regularly used.” It further recommends:

Rigorous evaluations using experimental or quasi-experimental methods that identify the effects of programs in situations where doing so is difficult using other methods; and rigorous qualitative evidence that complement what can be learned from empirical evidence and provide greater insight into the contexts where programs and practices are implemented more or less successfully.<sup>6</sup>

OMB funds some “rigorous program evaluations” through a competitive review process. But it emphasizes the development of agency infrastructure for undertaking their own program evaluations.

### Benefit-Cost Analysis of Federal Regulations

As early as 1936, Congress required the Army Corps of Engineers to undertake benefit-cost analysis in their flood control projects, to ensure that projects would produce benefits in excess of costs. Subsequently, the Corps of Engineers led the way in the development of government benefit-cost analysis. The Reagan Administration was the first to establish a broad commitment to benefit-cost analysis in regulatory decision making. Agencies were ordered to undertake regulatory action only on the basis of “reasoned determination” that benefits justify the costs and that the regulatory action maximized net societal benefits (benefits minus costs). Subsequent presidential administrations have reaffirmed their commitment to applying benefit cost analysis to federal regulatory actions.

Ideally, government agencies should (a) propose or adopt a regulation only upon a reasoned determination that its benefits justify its costs (recognizing that some benefits and costs are difficult to quantify); (b) tailor its regulations to impose the least burden on society consistent with obtaining regulatory objectives; (c) select, in choosing among alternative regulatory approaches, those approaches that maximize net benefits (including potential economic, environmental, public health, safety, and equity impacts; (d) identify and assess available alternatives to direct regulation, including providing economic incentives to encourage the desired behavior; and (e) use the best available techniques to quantify anticipated present and future benefits and costs as accurately as possible. These criteria are set forth in various presidential executive orders.

### Value of a Statistical Life

Among the controversies in benefit-cost analysis is the valuation of a human life, often required in the design of evaluations of health and safety regulations. Among the agencies that have developed a value for a statistical life are the Environmental Protection Agency, Department of Transportation, Food and Drug Administration, Occupational Safety and Health Administration, and Department of Homeland Security. Recent valuations have fluctuated between \$5 million and \$7 million. These calculations are required in monetarizing mortality risks; they do not suggest the value of any individual's life.

## FEDERAL EVALUATION: THE GENERAL ACCOUNTABILITY OFFICE

The General Accountability Office (GAO) is an arm of Congress. It has broad authority to audit the operations and finances of federal agencies, to evaluate their programs, and to report its findings to Congress. For most of its history, the GAO confined itself to financial auditing and management

and administrative studies. Over time, however, it has increasingly undertaken evaluative research on government programs.

The GAO was established by Congress as an independent agency in 1921, in the same Budget and Accounting Act that created the first executive budget; its authority to undertake evaluation studies was expanded in the Congressional Budget and Impoundment Control Act of 1974, the same act that established the House and Senate Budget Committees and the Congressional Budget Office (see Chapter 7). The GAO is headed by the Comptroller General of the United States. Most GAO reports are requested by Congress, although the office can also undertake studies on its own initiative.

According to the GAO, “Program evaluation—when it is available and of high quality—provides sound information about what programs are actually delivering, how they are being managed, and the extent to which they are cost-effective.”<sup>7</sup> The GAO believes that evaluation efforts by federal agencies fall woefully short of what is required for rational decision making. It has been especially critical of the Defense Department for failing to test weapons systems adequately, to monitor defense contractors and their charges, or to adjust its future plans to expected reductions in defense spending (see Chapter 15). The GAO has criticized the Environmental Protection Agency for measuring its own success in terms of input measures—numbers of inspections performed and enforcement actions undertaken—rather than actual improvements in environmental conditions, such as in water quality or air quality (see Chapter 13). The GAO has also reported on the Social Security trust fund and the dangers of spending trust fund money on current governmental operations (see Chapter 7). It has reported on the high and growing cost of medical care in the United States, especially Medicaid and Medicare, and noted the lack of correlation between medical spending and measures of the nation’s health (see Chapter 8). It has undertaken to assess the overall impact of drug control policies (see Chapter 6), and it has studied the default rate on student loans and recommended collection of overdue loans by withholding tax refunds (see Chapter 9). In short, the GAO has been involved in virtually every major policy question confronting the nation.<sup>8</sup>

## PROGRAM EVALUATION: WHY IT FAILS SO OFTEN

Occasionally, government agencies attempt their own policy evaluations. Government analysts and administrators report on the conditions of target groups before and after their participation in a new program, and some effort is made to attribute observed changes to the new program itself. Policy experimentation is less frequent; seldom do governments systematically select experimental and control groups of the population, introduce a new program to the experimental group only, and then carefully compare changes in the conditions of the experimental group with a control group that has not benefited from the program. Some of the problems confronting policy evaluation include:

- The first task confronting anyone who wants to evaluate a public program is to determine what the goals of the program are. What are the target groups, and what are the desired effects? But governments often pursue incompatible goals to satisfy diverse groups. Overall policy planning and evaluation may reveal inconsistencies in public policy and force reconsideration of fundamental societal goals. Where there is little agreement on the goals of a public program, evaluation studies may engender a great deal of political conflict.

Government agencies generally prefer to avoid conflict, and hence to avoid studies that would raise such questions.

- Many programs and policies have primarily symbolic value. They do not actually change the conditions of target groups but merely make these groups feel that the government “cares.” A government agency does not welcome a study that reveals that its efforts have no tangible effects; such a revelation itself might reduce the symbolic value of the program by informing target groups of its uselessness.
- Government agencies have a strong vested interest in “proving” that their programs have a positive impact. Administrators frequently view attempts to evaluate the impact of their programs as attempts to limit or destroy the programs or to question the competence of the administrators.
- Government agencies usually have a heavy investment—organizational, financial, physical, psychological—in current programs and policies. They are predisposed against finding that these policies do not work.
- Any serious study of policy impact undertaken by a government agency would involve some interference with ongoing program activities. The press of day-to-day business generally takes priority over study and evaluation. More important, the conduct of an experiment may necessitate depriving individuals or groups (control groups) of services to which they are entitled under law; this may be difficult, if not impossible, to do.
- Program evaluation requires funds, facilities, time, and personnel, which government agencies do not like to sacrifice from ongoing programs. Policy impact studies, like any research, cost money. They cannot be done well as extracurricular or part-time activities. Devoting resources to studies may mean a sacrifice in program resources that administrators are unwilling to make.

## HOW BUREAUCRATS EXPLAIN NEGATIVE FINDINGS

Government administrators and program supporters are ingenious in devising reasons why negative findings about policy impacts should be rejected. Even in the face of clear evidence that their favorite programs are useless or even counterproductive, they will argue that:

- The effects of the program are long range and cannot be measured at the present time.
- The effects of the program are diffuse and general in nature; no single criterion or index adequately measures what is being accomplished.
- The effects of the program are subtle and cannot be identified by crude measures or statistics.
- Experimental research cannot be carried out effectively because to withhold services from some persons to observe the impact of such withholding would be unfair to them.
- The fact that no difference was found between persons receiving the services and those not receiving them means that the program is not sufficiently intensive and indicates the need to spend *more* resources on the program.
- The failure to identify any positive effects of a program is attributable to inadequacy or bias in the research itself, not in the program.

Political scientist James Q. Wilson formulated two general laws to cover all cases of social science research on policy impact:

*Wilson's First Law:* All policy interventions in social problems produce the intended effect—if the research is carried out by those implementing the policy or by their friends.

*Wilson's Second Law:* No policy intervention in social problems produces the intended effect—if the research is carried out by independent third parties, especially those skeptical of the policy.

Wilson denies that his laws are cynical. Instead he reasons that:

Studies that conform to the First Law will accept an agency's own data about what it is doing and with what effect; adopt a time frame (long or short) that maximizes the probability of observing the desired effect; and minimize the search for other variables that might account for the effect observed. Studies that conform to the Second Law will gather data independently of the agency; adopt a short time frame that either minimizes the chance for the desired effect to appear or, if it does appear, permits one to argue that the results are "temporary" and probably due to the operation of the "Hawthorne Effect" (i.e., the reaction of the subjects to the fact that they are part of an experiment); and maximize the search for other variables that might explain the effects observed.<sup>9</sup>

## WHY GOVERNMENT PROGRAMS ARE SELDOM TERMINATED

Government programs are rarely terminated. Even when evaluative studies produce negative findings; even when policymakers themselves are fully aware of fraud, waste, and inefficiency; even when highly negative benefit-cost ratios are reported, government programs manage to survive. Once policy is institutionalized within a government, it is extraordinarily difficult to terminate.

Why is it so difficult for governments to terminate failed programs and policies? The answer to this question varies from one program to another, but a few generalizations are possible.

### Concentrated Benefits, Dispersed Costs

Perhaps the most common reason for the continuation of inefficient government programs and policies is that their limited benefits are concentrated in a small, well-organized constituency, while their greater costs are dispersed over a large, unorganized, uninformed public. Although few in number, the beneficiaries of a program are strongly committed to it; they are concerned, well-informed, and active in their support. If the costs of the program are spread widely among all taxpayers, no one has a strong incentive to become informed, organized, or active in opposition to it. Although the costs of a failed program may be enormous, if they are dispersed widely enough so that no one individual or group bears a significant burden, there will be little incentive to organize an effective opposition. (Consider the case of a government subsidy program for peanut growers. If \$300 million per year were distributed to 5,000 growers, each would average \$60,000 in subsidy income. If each grower would contribute 10 percent of this subsidy to a political fund to reward friendly legislators, the fund could distribute \$30 million in campaign contributions. If the costs of the program could be dispersed evenly among 300 million Americans, each would pay only \$1. No one would have a sufficient incentive to become informed, organized, or active in opposition to the subsidy program. So it would continue, regardless of its limited benefits and extensive costs

to society.) When program costs are widely dispersed, it is irrational for individuals, each of whom bears only a tiny fraction of these costs, to expend the time, energy, and money to counter the support of the program's beneficiaries.

### Legislative and Bureaucratic Interests

Among the beneficiaries of any government program are those who administer and supervise it. Bureaucratic jobs depend on a program's continuation. Government positions with all of their benefits, pay, prerequisites, and prestige are at stake. Strong incentives exist for bureaucrats to resist or undermine negative evaluations of their programs, to respond to public criticism by making only marginal changes in their programs, or even by claiming that their programs are failing because not enough is being spent on them.

Legislative systems, both in Congress and in state capitals, are structured so that legislators with the most direct control over programs are usually the most friendly to them. The committee system, with its fragmentation of power and invitation to logrolling ("You support my committee's report, and I'll support yours") favors retention of existing programs and policies. Legislators on committees with jurisdiction over the programs are usually the largest recipients of campaign contributions from the organized beneficiaries of the programs. These legislators can use their committee positions to protect failed programs, to minimize reform, and to block termination. Even without the incentives of bureaucratic position and legislative power, no public official wants to acknowledge failure publicly.

### Incrementalism at Work

Governments seldom undertake to consider any program as a whole in any given year. Active consideration of programs is made at the margin—that is, attention is focused on proposed changes in existing programs rather than on the value of programs in their entirety. Usually this attention comes in the budgetary process, when proposed increases or decreases in funding are under discussion in the bureaucracy and legislature. Negative evaluative studies can play a role in the budgetary process—limiting increases for failed programs or perhaps even identifying programs ripe for budget cutting. But attention is almost always focused on changes or reforms, increases or decreases, rather than on the complete termination of programs. Even mandating "sunset" legislation, used in many states (requiring legislatures periodically to reconsider and reauthorize whole programs), seldom results in program termination.

## POLITICS AS A SUBSTITUTE FOR ANALYSIS

Policy analysis, including systematic policy evaluation, is a rational process. It requires some agreement on what problems the government should undertake to resolve; some agreement on the nature of societal benefits and costs and the weights to be given to them; and some agreement on the formulation of a research design, the measurement of benefits and costs, and the interpretation of the results. Value conflicts intrude at almost every point in the evaluation process, but policy analysis cannot resolve value conflicts.

Politics is the management of conflict. People have different ideas about what the principal problems confronting society are and about what, if anything, the government should do about

them. Value conflicts explain why policymakers rely so little on systematic policy analysis in the formulation, selection, or evaluation of policy. Instead, they must rely on political processes.

A political approach to policy analysis emphasizes:

- The search for common concerns that might form the basis for identification of societal problems
- Reasonable trade-offs among conflicting values at each stage of the policymaking process
- The search for mutually beneficial outcomes for diverse groups; attempting to satisfy diverse demands
- Compromise and conciliation and a willingness to accept modest net gains (half a loaf) rather than suffer the loss of more comprehensive proposals
- Bargaining among participants, even in separate policy areas, to win allies (“I’ll support your proposals if you support mine.”)

At best, policy analysis plays only a secondary role in the policymaking process. But it is an important role, nonetheless. Political scientist Charles E. Lindblom explains “the intelligence of democracy”:

Strategic analysis and mutual adjustment among political participants, then, are the underlying processes by which democratic systems achieve the level of intelligent action that they do....

There is never a point at which the thinking, research, and action is “objective,” or “unbiased.” It is partisan through and through, as are all human activities, in the sense that the expectations and priorities of those commissioning and doing the analysis shape it, and in the sense that those using information shape its interpretation and application.

Information seeking and shaping must intertwine inextricably with political interaction, judgment, and action. Since time and energy and brainpower are limited, strategic analysis must focus on those aspects of an issue that participating partisans consider to be most important for persuading each other. There is no purely analytical way to do such focusing; it requires political judgments: about what the crucial unknowns are, about what kind of evidence is likely to be persuasive to would-be allies, or about what range of alternatives may be politically feasible.<sup>10</sup>

## THE LIMITS OF PUBLIC POLICY

Never have Americans expected so much of their government. Our confidence in what governments can do seems boundless. We have come to believe that they can eliminate poverty, end racism, ensure peace, prevent crime, restore cities, clean the air and water, and so on, if only they will adopt the right policies.

Perhaps confidence in the potential effectiveness of public policy is desirable, particularly if it inspires us to continue to search for ways to resolve societal problems. But any serious study of public policy must also recognize the limitations of policy in affecting these conditions.

1. Some societal problems are incapable of solution because of the way in which they are defined. If problems are defined in *relative* rather than *absolute* terms, they may never

be resolved by public policy. For example, if the poverty line is defined as the line that places one-fifth of the population below it, poverty will always be with us regardless of how well off the “poor” may become. Relative disparities in society may never be eliminated. Even if income differences among classes were tiny, tiny differences may come to have great symbolic importance, and the problem of inequality would remain.

2. Expectations may always outrace the capabilities of governments. Progress in any policy area may simply result in an upward movement in expectations about what policy should accomplish. Public education never faced a dropout problem until the 1960s, when for the first time a majority of boys and girls were graduating from high school. At the turn of the century, when high school graduation was rare, there was no mention of a dropout problem.
3. Policies that solve the problems of one group in society may create problems for other groups. In a plural society, one person’s solution may be another person’s problem. For example, solving the problem of inequality in society may mean redistributive tax and spending policies, which take from persons of above-average wealth to give to persons with below-average wealth. The latter may view this as a solution, but the former may view it as creating serious problems. There are *no* policies that can simultaneously attain mutually exclusive ends.
4. It is quite possible that some societal forces cannot be harnessed by governments, even if it is desirable to do so. It may turn out that the government cannot stop urban location patterns of whites and blacks, even if it tries to do so. Whites and blacks may separate themselves regardless of government policies in support of integration. Some children may not be able to learn much in public schools no matter what is done. In other words, governments may not be *able* to bring about some societal changes.
5. Frequently, people adapt themselves to public policies in ways that render the policies useless. For example, we may solve the problem of poverty by government guarantees of a high annual income, but by so doing we may reduce incentives to work and thus swell the number of dependent families beyond the fiscal capacities of government to provide guarantees. The possibility always exists that adaptive behavior may frustrate policy.
6. Societal problems may have multiple causes, and a specific policy may not be able to eradicate the problem. For example, job training may not affect the hardcore unemployed if their employability is also affected by chronic poor health.
7. The solution to some problems may require policies that are more costly than the problem. For example, it may turn out that certain levels of public disorder—including riots, civil disturbances, and occasional violence—cannot be eradicated without the adoption of very repressive policies—the forceable breakup of revolutionary parties, restrictions on the public appearances of demagogues, the suppression of hate literature, the addition of large numbers of security forces, and so on. But these repressive policies would prove too costly in democratic values—freedom of speech and press, rights of assembly, freedom to form opposition parties. Thus, a certain level of disorder may be the price we pay for democracy. Doubtless, there are other examples of societal problems that are simply too costly to solve.

8. The political system is not structured for completely rational decision making. The solution of societal problems generally implies a rational model, but government may not be capable of formulating policy in a rational fashion. Instead, the political system may reflect group interests, elite preferences, institutional forces, or incremental change, more than rationalism. Presumably, a democratic system is structured to reflect mass influences, whether these are rational or not. Elected officials respond to the demands of their constituents, and this may inhibit completely rational approaches to public policy.

## SUMMARY

Policy evaluation is learning about the consequences of public policy.

1. Policy evaluation involves assessing the impact of policy on target and nontarget groups, future as well as immediate impacts, and direct as well as indirect costs.
2. Government agencies themselves usually report policy output measures, rather than the effects of these outputs on societal conditions.
3. Systematic policy evaluation may involve before and after comparisons, projected trend line versus post program comparisons, and comparisons of governments with and without programs.
4. The classic research design involves comparisons between control and experimental groups both before and after program implementation.
5. The Office of Management and Budget (OMB) and the General Accountability Office (GAO) undertake evaluations of federal programs.
6. There are many political and bureaucratic obstacles to effective policy evaluation.

## MySearchLab® EXERCISES

Apply what you learned in this chapter on MySearchLab ([www.mysearchlab.com](http://www.mysearchlab.com)).

## NOTES

1. Joseph S. Wholey, et al., *Federal Evaluation Policy* (Washington, DC: Urban Institute, 1970), p. 25.
2. David Nachmias, *Public Policy Evaluation* (New York: St. Martin's Press, 1979), p. 4.
3. See David L. Sills, ed., *International Encyclopedia of the Social Sciences*, vol. 7 (New York: Free Press, 1968), p. 241.
4. David Kershaw and Jerelyn Fair, eds., *Final Report of the New Jersey Graduated Work Incentive Experiment* (Madison: University of Wisconsin, Institute for Research on Poverty, 1974).
5. John F. Cogan, *Negative Income Taxation and Labor Supply: New Evidence from the New Jersey–Pennsylvania Experiment* (Santa Monica, CA: Rand Corporation, 1978). See also SRI International, *Final Report of the Seattle–Denver Income Maintenance Experiment* (Washington, DC: U.S. Government Printing Office, 1983), for experimental results suggesting that government job training has no effect on a person's subsequent earnings or employment and that a guaranteed income significantly lowers earnings and hours of work and contributes to marital dissolutions.
6. Office of Management and Budget, *Budget of the United States Government 2012*, Analytical Perspectives, Chapter 8 Program Evaluation, p. 83.
7. *Federal Evaluation Issues* (Washington, DC: General Accounting Office, 1989).

8. *Annual Index of Reports Issued* (Washington, DC: General Accountability Office, annually).
9. James Q. Wilson, "On Pettigrew and Armor," *The Public Interest* 31 (Spring 1973), pp. 132–134.
10. Charles E. Lindblom and Edward J. Woodhouse, *The Policy-Making Process*, 3rd ed. (Englewood Cliffs, NJ: Prentice Hall, 1993), pp. 31–32.

## BIBLIOGRAPHY

- AMMONS, DAVID N. *Tools for Decision-Making: A Practical Guide for Local Government*. Washington, DC: CQ Press, 2002.
- BARDACH, EUGENE. *A Practical Guide for Policy Analysis*, 3rd ed. Washington, DC: CQ Press, 2008.
- BINGHAM, RICHARD D., and CLAIRE L. FELBINGER. *Evaluation in Practice*, 2nd ed. Washington, DC: CQ Press, 2002.
- HEINEMAN, ROBERT A., et al. *The World of the Policy Analyst*, 3rd ed. Washington, DC: CQ Press, 2001.
- O'SULLIVAN, ELIZABETHANN, et al. *Research Methods for Public Administrators*, 5th ed. New York: Longman, 2008.
- POSAVAC, EMIL J. *Program Evaluation*, 8th ed. New York: Pearson, 2011.
- PRESSMAN, JEFFREY L., and AARON WILDAVSKY. *Implementation*. Berkeley: University of California Press, 1974.
- WEIMER, DAVID, L., and AIDAN R. VINING. *Policy Analysis: Concepts and Practices*, 5th ed. New York: Longman, 2011.
- WILDAVSKY, AARON. *Speaking Truth to Power*. New York: John Wiley, 1979.

## WEB SITES

- GENERAL ACCOUNTABILITY OFFICE. The GAO is the investigative and evaluative arm of the Congress. Its purpose is to hold the executive branch accountable to the Congress. Its reports cover a wide variety of issues. [www.gao.gov](http://www.gao.gov)
- OFFICE OF MANAGEMENT AND BUDGET. In addition to preparing the Budget of the United States, OMB performs management studies, including reports on financial management and regulatory matters. [www.whitehouse.gov/omb](http://www.whitehouse.gov/omb)
- ASSOCIATION FOR PUBLIC POLICY ANALYSIS AND MANAGEMENT. Academic organization that publishes the *Journal of Policy Analysis and Management*, with articles on public administration and management. [www.appam.org](http://www.appam.org)
- AMERICAN SOCIETY FOR PUBLIC ADMINISTRATION. Organization of academic and professional public administrators, with news, job listings, and publication—*Public Administration Review*. [www.aspanet.org](http://www.aspanet.org)
- RAND CORPORATION. Originally devoted almost exclusively to research on defense and weapons systems, RAND studies now include space research, Internet technology, information protection, and assessments of government programs across a wide spectrum. [www.rand.org](http://www.rand.org)
- CATO INSTITUTE, *Regulation* magazine. Articles assessing the costs and effectiveness of government programs, especially regulatory programs and agencies. [www.cato.org/pubs/regulation](http://www.cato.org/pubs/regulation)