United States General Accounting Office

GAO

Report to the Committee on Governmental Affairs, U.S. Senate

April 1998

PROGRAM EVALUATION

Agencies Challenged by New Demand for Information on Program Results

GAO/GGD-98-53

GAO

United States
General Accounting Office
Washington, D.C. 20548

General Government Division

B-277820

April 24, 1998

The Honorable Fred D. Thompson
Chairman
The Honorable John Glenn
Ranking Minority Member
Committee on Governmental Affairs
United States Senate

Congressional and agency decisionmakers need evaluative information about whether federal programs are working well or poorly, both to manage programs effectively and to help decide how to allocate limited federal resources. Increased interest in learning the results of federal programs and activities is reflected in government reforms, such as the Government Performance and Results Act of 1993 (GPRA or the Results Act), which institutes a governmentwide requirement for agencies to, among other things, report on their results in achieving their agency and program goals. However, other recent reforms, such as reducing the size and authority of the federal government while maintaining a level of services, has the potential to hinder agencies' ability to obtain this information. Because data on program results are typically more difficult and resource intensive to obtain than data on program activities, limited budget dollars mean that investing in obtaining information on results may compete with spending on program activities. The proper balance between the two spending priorities is essential, since information on program results can contribute to deciding how to allocate resources to activities to maximize program benefits.

Federal agencies are the primary source of evaluation information about their programs. In past surveys of federal agencies, we found limited (and diminishing) resources spent on formal studies of program results, that is, program evaluation.[1] Because evaluation can be vitally important in improving program results, we asked how in a context of limited federal resources and responsibility, can agencies support additional requests for program results information? This report, which we prepared under our basic legislative responsibilities, responds to that question by discussing the current status of and future needs for program evaluation in federal agencies. Because of your interest in improving the quality of information on federal programs, we are addressing this report to you. Our objectives were to identify (1) the current resources and roles for program evaluation in federal agencies, (2) the anticipated effects of governmentwide reforms

[1]Program Evaluation Issues (GAO/OCG-93-6TR, Dec. 1992) and Federal Evaluation: Fewer Units, Reduced Resources, Different Studies from 1980 (GAO/PEMD-87-9, Jan. 23, 1987).

and other initiatives on evaluation of federal programs, and (3) potential strategies for agencies to respond to the anticipated effects and provide information on program results.

Background

Program and policy decisions require a wide array of information that answers various questions. For example, descriptive information tells how a program operates—what activities are performed, who performs them, and who is reached. In contrast, evaluative information speaks to how well a program is working—such as whether activities are managed efficiently and effectively, whether they are carried out as intended, and to what extent the program is achieving its intended objectives or results. There are a variety of methods for obtaining information on program results, such as performance measurement and program evaluation, which reflect differences in how readily one can observe program results.

Performance measurement, as defined by the Results Act, is the ongoing monitoring and reporting of program accomplishments, particularly progress towards preestablished goals. It tends to focus on regularly collected data on the type and level of program activities (process), the direct products and services delivered by the program (outputs), and the results of those activities (outcomes). While performance may be defined more broadly as program process, inputs, outputs, or outcomes, results usually refer only to the outcomes of program activities. For programs that have readily observable results, performance measurement may provide sufficient information to demonstrate program results.

In some programs, however, results are not so readily defined nor measured.[2] In such cases, program evaluations may be needed, in addition to performance measurement, to examine the extent to which a program is achieving its objectives. Program evaluations are systematic studies conducted periodically to assess how well a program is working. While they may vary in their focus, these evaluations typically examine a broader range of information on program performance and its context than is feasible in ongoing performance measurement. Where programs aim to produce changes, as a result of program activities, outcome (or effectiveness) evaluations assess the extent to which those outcomes or results were achieved, such as whether students increased their understanding of or skill in the material of instruction. In cases where the program's outcomes are influenced by complex systems or events outside

[2]For example, the intended results of a statistical program—valid, reliable, and useful data—are typically measured through expert judgments. See Martin & Straf, 1992.

the program's control, impact evaluations use scientific research methods to establish the causal connection between outcomes and program activities, estimate what would have happened in the absence of the program, and thus isolate the program's contribution to those changes. For example, although outcome measures might show a decline in a welfare program's caseload after the introduction of job placement activities, a systematic impact evaluation would be needed to assess how much of the observed change was due to an improved economy rather than the new program.

In addition, a program evaluation that also systematically examines how a program was implemented can provide important information about why a program did or did not succeed and suggest ways to improve it. For the purposes of this report, we used the definition of program evaluation that is used in the Results Act, "an assessment, through objective measurement and systematic analysis, of the manner and extent to which federal programs achieve intended objectives."[3] We asked about assessments of program results, which could include both the analysis of outcome-oriented program performance measures as well as specially conducted outcome or impact evaluations.

Two government initiatives could influence the demand for and the availability and use of program evaluation information. The Results Act seeks to promote a focus on program results, by requiring agencies to set program and agency performance goals and to report annually on their progress in achieving them (beginning with fiscal year 1999). In addition to encouraging the development of information on program results for activities across the government, the Results Act recognizes the complementary nature of program evaluation and performance measurement. It requires agencies to include a schedule for future program evaluations in their strategic plans, the first of which was to be submitted to Congress by September 30, 1997. The Results Act also requires agencies to review their success in achieving their annual performance goals (which are set forth in their annual performance plans) and to summarize the findings of program evaluations in their annual program performance reports (the first of which is due by March 31, 2000). The National Performance Review (NPR) led by the Vice President's office has asked agencies to reexamine their policies, programs, and operations to find and implement ways to improve performance and service to their customers. Both of these initiatives—because of their focus on program

[3]Although the term "program evaluation" is sometimes used to refer to evaluation of program process (or implementation) without examining results, this definition, which focuses on results, is also used in the Office of Management and Budget's (OMB) requests for budget justifications.

results—could be expected to increase the demand for and the availability and use of program evaluation information.

Other recent governmentwide initiatives could have potentially conflicting effects. In several program areas, devolution of program responsibility from the federal level and consolidation of individual federal programs into more comprehensive, multipurpose grant programs has shifted both program management and accountability responsibilities toward the states. These initiatives may thus make it more difficult for federal agencies to evaluate the results of those programs. In addition, efforts to reduce the growth of the federal budget have resulted in reductions in both federal staff and program resources in many agencies. The combination of these initiatives raises a question: In an environment of limited federal resources and responsibility, how can agencies meet the additional needs for program results information?

To identify the roles and resources available for federal program evaluation, in 1996, we conducted a mail survey of offices identified by federal agency officials that were conducting studies of program results or effectiveness in 13 cabinet-level departments and 10 independent executive agencies. Detailed information on program evaluation studies refers to those conducted during fiscal year 1995 (regardless of when they began or ended). To identify how recent reforms were expected to affect federal evaluation activities and what strategies were available for responding to those changes, we interviewed external evaluation experts and evaluation and other officials at selected federal and state agencies. In this report, we use the term "agency" to include both cabinet-level departments and independent agencies.

Results in Brief

Existing federal evaluation resources—at least as currently configured and deployed—are likely to be challenged to meet the increasing demand for program results information. Agencies reported devoting variable but relatively small amounts of resources to evaluating program results. Across the 13 departments and 10 independent agencies we surveyed, the resources that agencies identified as being involved in assessing their programs' results represented $194 million and 669 full-time equivalent staff (FTE) in fiscal year 1995 and were unevenly distributed across the agencies. We found 81 offices that reported conducting studies of program results. Over half of the reporting offices were small, with 18 or fewer

FTEs.[4] Similarly, a majority (56 percent) of the 81 offices reported conducting 5 or fewer studies in fiscal year 1995. The evaluation studies were diverse as well—about half were handled in-house, and almost half used existing program data.

Moreover, agencies reported that the primary role of program evaluation was internally focused on program improvement, rather than direct congressional or other external oversight. Interest in the program by high-level agency officials was most often cited as a criterion for initiating evaluation work; a small portion of studies (17 percent) were said to be conducted for a congressional committee or in response to a legislative mandate. The studies' primary audiences were reported to be program managers and higher-level agency officials. In addition, these offices reported that their evaluation activities primarily consisted of conducting the studies themselves or through contractors. Few offices reported frequent efforts to extend the use of evaluation resources by providing training in evaluation methods to others, such as federal program staff or state and local evaluators.

Some of the evaluation officials and experts we interviewed anticipated increased interest in learning the results of federal programs and policies but also additional complications in obtaining that information. For example, some said that devolving federal program responsibility in the health care and welfare systems to the states raises many new important questions about its effects, but as programs become increasingly diverse, as expected, this will make evaluating the effects of these reforms more difficult. Federal funding reductions were said by some evaluation officials not only to reduce the level of federal evaluation activity but also to diminish agency technical capacity, through the loss of some of their most experienced staff. Although federal agencies have limited experience in meeting the Results Act's reporting requirements (the first annual performance reports are not due until March 2000, for fiscal year 1999), some evaluation officials hoped the Results Act would increase the use of evaluation results in decisionmaking, while others feared that the large investments required to produce valid and reliable outcome data across whole agencies would compete for funds currently used for more in-depth evaluations of individual programs' effectiveness. We previously reported that early agency efforts to implement the Results Act have encountered

[4]We identified no offices conducting studies of program results in two departments and four independent agencies. Some of these offices noted that they assessed program outcomes by monitoring program performance, or by analyzing efficiency or compliance, but not by conducting evaluation studies per se. It is also possible that offices, including those without programs per se, did assess the effectiveness of federal activities, but they did not complete the survey for this work.

numerous analytic challenges, and agency officials found program evaluation expertise helpful in their efforts to develop performance reporting systems.[5] In addition, some evaluation officials from states with performance measurement experience noted that effectiveness evaluations would continue to be needed to assess policy impact and address problems of special interest or larger policy issues, such as the need for any government intervention at all in an area.

To meet the anticipated increase in demand for program results information as well the associated technical challenges, some evaluation officials we interviewed described efforts to leverage both federal and nonfederal resources. Suggestions included (1) adapting existing information systems to yield data on program results, (2) broadening the range of their work to include less rigorous and less expensive methods, (3) devolving program evaluation to federal (or state and local) program managers, and (4) developing partnerships with others to integrate the varied forms of performance information available on their programs. However, some agencies anticipated that major investments in their data systems would be required to produce reliable data on program outcomes; and, in a prior study, program officials were concerned that reliance on less rigorous methods would not provide an accurate picture of program effectiveness. Moreover, while some federal evaluation officials envisioned providing increased technical assistance to state and local evaluators, a few state evaluation officials suggested an alternative strategy for the federal government: (1) providing evaluation leadership through establishing a catalog of tested performance measures and (2) conducting impact evaluations to supplement the states' performance measurement information.

We drew several conclusions from our comparison of current federal evaluation resources with the anticipated challenges to meeting increased demand for information on program results. First, federal evaluation resources have important roles to play in responding to increased demand for information on program results, but—at least as currently configured and deployed—they are likely to be challenged to meet that demand. Second, in the future, carefully targeting federal agencies' evaluation resources and leveraging federal and nonfederal resources show promise for addressing key questions about program results. Some possible ways to target and leverage resources include (1) assisting program managers to develop valid and reliable performance reporting under the Results Act,

[5]Managing for Results: Analytic Challenges in Measuring Performance (GAO/HEHS/GGD-97-138, May 30, 1997).

and (2) planning evaluation studies to fill the most important information gaps—such as providing supplemental information on the reasons for observed performance or examining policy issues that extend beyond program borders. Third, one way to ensure that the results of diverse evaluation activities can be synthesized to portray programs at the national level is for federal evaluation staff to coordinate those activities in advance.

Scope and Methodology

We distributed surveys in 1996 regarding federal evaluation activities within 13 cabinet level departments and 10 independent executive agencies in the federal government. We excluded the Department of Defense from our survey of evaluation offices because of the prohibitively large number of offices it identified as conducting assessments of effectiveness. Although we asked agency officials to be inclusive in their initial nominations of offices that conducted evaluations of program results, some offices that conducted evaluations may have been overlooked and excluded from our survey. However, many offices that were initially identified as having conducted evaluations later reported that they had not done so. In our survey, we asked each office about the range of its analytic and evaluation activities and about the length, cost, purpose, and other characteristics of the program evaluation studies they conducted during fiscal year 1995. (See appendix I for more details on the scope and methodology of the survey.)

Between 1996 and 1997, we conducted interviews of program evaluation practitioners selected to represent divergent perspectives. We asked what had been or were expected to be the effects of various government changes and reforms on federally supported and related program evaluation activities and strategies for responding to those effects. We identified individuals with evaluation knowledge and expertise from a review of the survey responses, the evaluation literature, and our prior work; they were from an array of federal and state agencies and the academic and consulting communities. We then judgmentally selected 18 people to interview to reflect (1) a mix of different program types and diverse amounts of experience with program evaluation and (2) experience with some of the reforms at the state or federal level. Those selected included nine evaluation officials (six from offices in federal agencies and three from state legislative audit agencies) and seven external evaluation experts (four from private research organizations and three from universities). In addition, we interviewed an OMB official and one official from a state executive branch agency, and we also asked the

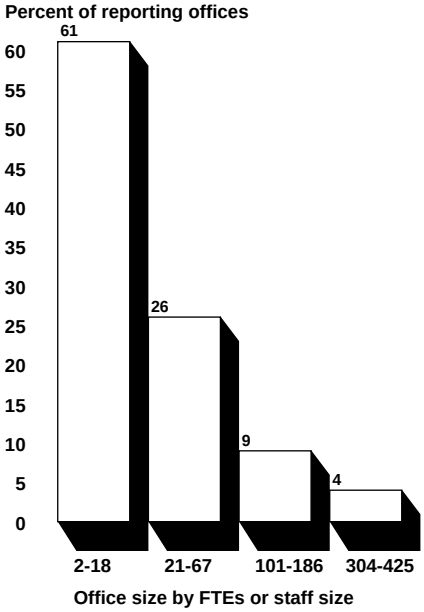
officials from the state legislative audit agencies about their experiences with state performance reporting requirements.

We conducted our review between May 1996 and July 1997 in accordance with generally accepted government auditing standards. However, we did not independently verify the types of studies conducted, other information reported by our respondents, nor information gained from interviewees.

Federal Agencies Devote Variable but Small Amounts of Resources to Evaluating Program Results, Primarily for Internal Use

The resources allocated to conducting systematic assessments of program results (or evaluation studies) were small and unevenly distributed across the 23 agencies (departments and independent agencies) we surveyed. We found 81 offices that reported expending resources—funds and staff time—on conducting program effectiveness studies in fiscal year 1995. Over half of those offices had 18 or fewer full-time equivalent staff FTEs, while only a few offices had as many as 300 to 400 FTEs. (See figure 1.)

Figure 1: Offices That Conducted Program Evaluations in FY 1995 (Office Size in FTEs)



Note: No data on seven offices.

Source: GAO survey on program evaluation in federal agencies governmentwide.

Moreover, about one-third of the offices reported spending 50 percent or more of their time on evaluation activities (including development of performance measures and assessments of program effectiveness, compliance, or efficiency), since program evaluation was only one of these offices' many responsibilities. (See survey question 9 in appendix I.) Two of the 3 largest offices (over 300 FTEs) spent about 10 percent of their staff time on program evaluation activities. Thus, the estimated staff and budget resources that the 81 offices actually devoted to evaluation activities totaled 669 FTEs and at least $194 million across the 23 agencies surveyed.[6]

In addition, most (61 of 81) offices reported also conducting management analysis activities; the most frequent activities were conducting

[6]Since several offices did not report either their 1995 funding level or the proportion of funding (or staffing resources) spent on evaluation, this may underreport actual expenditures. By comparison, we conducted a similar survey in 1984 that found that evaluation resources represented roughly comparable proportions of agency funding and staff resources. See Federal Evaluation: Fewer Units, Reduced Resources, Different Studies from 1980 (GAO/PEMD-87-9, Jan. 23, 1987).

management studies, developing strategic plans, and describing program implementation. Of those offices that could estimate their staff time, about half reported spending less than 25 percent of their time on management analysis. Similarly, many offices reported conducting policy planning and analysis, but most of them reported spending less than 25 percent of their time on it. Thus, a majority of the offices (45 of the 81 identified) conducted few evaluation studies (5 or less in fiscal year 1995), while 16 offices—representing 7 agencies—accounted for two-thirds of the 928 studies conducted. (See table 1.)

Table 1: Distribution of Program Evaluations Conducted in FY 1995 Across the Offices Surveyed

Number of studies reported per office	Number of offices	Studies conducted	
		Number	Percent
1-5	45	111	12%
6-17	20	193	21
20-79	16	624	67
Total	**81**	**928**	

Source: GAO survey.

Finally, 6 of the 23 agencies we surveyed did not report any offices conducting evaluation studies in fiscal year 1995. A few of these agencies indicated that they analyzed program accomplishments or outputs or conducted management reviews to assess their programs' performance but did not conduct an evaluation study per se. Some of the 6 agencies also reported conducting other forms of program reviews that focused on assessing program compliance or efficiency rather than program results.

Offices conducting program evaluations were located at various levels of federal agencies. A few of the 81 offices were located in the central policy or administrative office at the highest level of the organization (5 percent) or with the Inspector Generals (5 percent); many more were located in administrative offices at a major subdivision level (43 percent) or in program offices or technical or analytic offices supporting program offices (30 and 16 percent, respectively). (See table 2.)

Table 2: Organizational Location of Offices Conducting Evaluations for the Total Sample and Offices Conducting 20 or More Studies, in FY 1995

Organizational location	Percent of offices in each location	Number of offices that reported conducting 20 or more studies
Inspector General	5%	2
Central administrative office	5	2
Central administrative office for a major operating subdivision of department or agency	43	6
Program office	30	4
Technical or analytical unit supporting a program office	16	2
Other	1	0
Total offices	**81**	**16**

Source: GAO survey.

Four of the 23 agencies surveyed had offices at all 3 levels (agency, division, and program), and over half the agencies (14 of 23) had conducted evaluations at the program level. The 16 offices conducting 20 or more studies were more likely to be centralized at the agency or division level than at the program level.

A diverse array of evaluation studies were described in the surveys. Just over half of the studies for which we have such information were conducted in-house (51 percent), and 27 percent lasted under 6 months. But the studies that were contracted out tended to be larger investments—almost two-thirds of them took over a year to complete, and over half cost between $100,000 and $500,000. Moreover, almost a third of all the studies lasted more than 2 years, reflecting some long-term evaluations. (See table 3.)

Table 3: Reported Percentage of Duration and Cost of Evaluation Studies Conducted in FY 1995, by Study Locus[a]

	Study Locus			
	In-house	Contracted out	Jointly with another federal, state, or local agency	Total
Duration				
Less than 6 months	49%	5%	2%	27%
6 to 12 months	29	31	<1	25
13 to 24 months	10	28	11	17
Over 2 years	12	36	86	31
Total number of studies	**418**	**281**	**124**	**823**
Cost[b]				
Less than $100,000	65%	14%	23%	39%
Between $100,000 and $499,000	30	57	33	41
Between $500,000 and $999,000	2	13	21	9
Over $1 million	3	16	24	11
Total number of studies	**319**	**263**	**106**	**688**

Note: Percents may not total 100 due to rounding.

[a]Respondent information on study length and cost was incomplete. Information was provided on study length for 89 percent of the 928 studies respondents reported were conducted in fiscal year 1995, while information was provided on cost for only 74 percent of the 928 studies. Many studies from offices reporting no cost information at all lasted less than 6 months.

[b]Study cost was the total cost regardless of funding source or fiscal year in which funds were obligated. For contracted-out studies, the amount of the contract or grant itself as well as staff costs associated with issuing and monitoring the contract or grants were included.

Source: GAO survey.

For example, a study of the impact of a medical treatment program, which used an experimental design with a complex set of Medicare program and clinical data from thousands of patients on numerous outcomes (for both patients and program costs), took over 2 years and cost over $1 million to complete.[7]

[7]For comparison, our 1984 survey of offices conducting program evaluations found that half of all the studies lasted under 6 months, and the majority of the studies that were contracted out lasted 1 year or less.

Many of the 1995 studies reportedly used relatively simple designs or research methods, and many relied on existing program data. The two most commonly reported study designs were judgmental assessments (18 percent) as well as experimental designs employing random assignment (14 percent). (See table 4 for a list of designs ranging from the most to least amount of control over the study conditions.)

Table 4: Research Methods Used for the Evaluation Studies Conducted in FY 1995

Research method	Number of studies using method[a]	Percent of all studies[b]
Experimental and control groups randomly assigned: after-intervention measures	129	14%
Some experimental control: before- and after-intervention outcome measures with some statistical controls	66	7
Time series: many repeated measures taken before and after intervention	56	6
Matched pairs or cross-sectional studies: outcomes measured after intervention with statistical controls	39	4
Panel studies: several repeated measures taken during and after intervention	21	2
Statistical modeling or simulation: program outcomes compared to predicted (control) results	59	6
Simple before- and after- studies: outcomes measured before and after intervention	114	12
One time survey of outcomes	96	10
Judgmental assessment: use of outside expert, program administrator, or participant judgment of program effects	170	18
Other[c]	79	8

[a]An individual study could have been reported as using more than one research method.

[b]These percents are based on all 928 studies that respondents reported were conducted in fiscal year 1995; however, responses to this question were not provided on all studies. For example, 6 offices did not answer this question for any of the 117 studies they reported having conducted. Therefore, these percents may underestimate the actual level of use of these methods.

[c]The other methods used included citation analysis, cost benefit analysis, and regression analysis.

Source: GAO survey.

Many of the studies (over 70 of the 129) that used experimental designs were evaluations of state demonstration programs, which were required

by law to use such methods, and were conducted out of one office. Experimental designs and designs using statistical controls are used to identify a program's net impact on its objectives where external factors are also known to affect its outcome. However, without knowing the circumstances of many of the programs being evaluated, it is impossible for us to determine the adequacy of the designs used to assess program effectiveness.

At least 40 percent of the studies employed existing program records in their evaluations, while about one-quarter employed special surveys or other ad hoc data-collection methods specially designed for the studies. Just under half (40 percent) of the studies used data from program administrative records that were produced and reported at the federal level; almost a third (28 percent) used data from routinely produced, but not typically reported, program records; 5 percent of the studies used data from administrative records of other federal agencies; and 14 percent used administrative records from state programs. Some studies may have used many types of data sources, which would suggest a heavy reliance on administrative and other program-related data. (See table 5.)

Table 5: Data Sources Reported for the Evaluation Studies Conducted in FY 1995

Data sources	Number of studies[a]	Percent of studies[b]
Generated by the program		
Program administrative records routinely produced and reported federally	369	40%
Program administrative records routinely produced but not typically reported or aggregated centrally	257	28
Recurring surveys of program participants	111	12
Special surveys or other ad hoc data collections on all or part of the program or in selected locales	254	27
Generated by another agency		
Administrative records or routinely aggregated data from other federal programs (e.g., Medicare Claims and Utilization Files)	46	5
Administrative records or routine data from state programs or agencies (e.g., Federal Accident Reporting System (FARS))	133	14
Federally sponsored multipurpose national surveys (e.g., the decennial census)	44	5
Other[c]	73	9

[a]An individual study could have been reported as using more than one data source.

[b]These percents are based on all 928 studies that respondents reported were conducted in fiscal year 1995; however, responses to this question were not provided on all studies. For example, 5 offices did not answer this question for any of the 81 studies they reported having conducted. Therefore, these percentages may underestimate the actual level of use of these data sources.

[c]The other data sources included private industry, medical records abstracts, and special studies generated by other agencies.

Source: GAO survey.

Studies Primarily Served Internal Program Improvement Purposes, Not Congressional Oversight

The primary reported purpose of the studies was to evaluate ongoing programs, either on an office's own initiative or at the request of top agency officials. In the survey, most officials conducting evaluations reported having a formal and an ad hoc planning process for deciding what evaluation work they would do. Many criteria were indicated as being used to select which studies to do (such as a program office request, congressional interest, or continuation or follow-up of past work), but the criterion most often cited was the interest of high-level agency officials in the program or subject area. Moreover, about one-fourth of the studies were requested by top agency officials. About one-fourth of the studies were indicated to be self-initiated. Most offices were not conducting studies for the Congress or as the result of legislative mandates; only

17 percent of the studies were reported to have been requested in those ways. (See table 6.)

Table 6: Sources of Request or Mandate for Evaluation Studies Conducted in FY 1995

Source of mandate or request	Number of reported studies[a]	Percent of all studies[b]
Self-initiated	224	24%
Top agency officials	224	24
Program personnel	186	20
Legislation or congressional committee	154	17
OMB or executive order	26	3
Other[c]	125	13

[a]An individual study could have been reported as being requested or mandated by more than one source.

[b]These percents are based on all 928 studies that respondents reported were conducted in fiscal year 1995; however, responses to this question were not provided on all studies. For example, 6 offices did not answer this question for any of the 105 studies they reported having conducted. Therefore, these percents may underestimate the actual level of study sources.

[c]Other sources of study requests included advisory boards, reengineering directives, NPR efforts, and state governments.

Source: GAO survey.

For those offices reporting that they conducted studies, about half of the 570 studies for which we have information evaluated ongoing programs.[8] Ongoing programs of all sizes were evaluated, ranging in funding from less than $10 million to over $1 billion. About one-third of these studies evaluated demonstration programs and many of them cost less than $10 million. In contrast, few reported evaluations of new programs and many of these new programs reportedly were small (with funding under $10 million).

Program evaluation was reported to be used more often for general program improvement than for direct congressional oversight. Their primary uses most often were said to be to improve program performance (88 percent), assess program effectiveness (86 percent), increase general knowledge about the program (62 percent), and guide resource allocation decisions within the program (56 percent). (See table 7.) Accordingly, these offices overwhelmingly (over three-fourths of respondents) reported

[8]Information was provided on program dollar size and maturity for only 61 percent of the 928 studies that respondents reported were conducted. Eighteen offices did not specifically answer this question for any of the 280 studies they reported having conducted.

program managers and higher-level agency officials as the primary audience of their studies. (See table 8.)

Table 7: Primary Uses Reported for the Results of Evaluation Studies Conducted in FY 1995

Uses of study results	Percent of offices reporting as primary[a]
Identify opportunities to improve program performance	88%
Ascertain the extent of the program's effectiveness	86
Increase general knowledge about the program or topic	62
Guide decisions on resource allocation within the program	56
Support program budget requests	31
Ascertain the success of corrective actions	30
Support program reauthorization	20
Other	14

Note: Percents do not add to 100 because respondents could give multiple responses.

[a]Percents are based on a total of 81 offices reporting that they conducted evaluation studies in fiscal year 1995. However, responses to this survey question were not provided by all offices, so these percents may somewhat underrepresent the various uses.

Source: GAO survey.

Table 8: Primary Audience Reported by Offices for the Results of Evaluation Studies Conducted in FY 1995

Primary audience	Percent of offices[a]
Program managers	85%
Higher-level agency officials	78
Agency evaluation staff	37
Office of Management and Budget	25
Congress, in general	32
Appropriations committees	22
Legislative committees	20
Program partners in private sector	26
Outside professional audience (experts, researchers, or analysts)	28
General public or others	25

Note: Percents do not add to 100 because respondents could give multiple responses.

[a]Percents are based on a total of 81 offices reporting that they conducted evaluation studies in fiscal year 1995. However, responses to this survey question were not provided by all offices so these percents may somewhat underrepresent the various uses.

Source: GAO survey.

About one-third of the offices reported support for budget requests as a primary use and one-third reported congressional audiences were primary users for their studies. Fewer respondents (20 percent) reported program reauthorization as a primary use of the study results. (See tables 7 and 8.)

Program evaluation was not the primary responsibility for most of these offices and the offices often reported "seldom, if ever" performing the program evaluation roles we asked about. The role most likely to be characterized as 'most often performed' was conducting studies of programs administered elsewhere in their agency. (See table 9.)

Table 9: Roles Offices Played in Conducting Evaluation Studies in FY 1995

	Percent Reported					
Offices' roles in evaluations	Seldom, if ever	Sometimes	Often (about 1/2 the time)	Very often	Always or almost always	Number reported
Conduct evaluation studies of programs that are run by your office	38%	34%	10%	7%	10%	68
Conduct evaluation studies of programs administered elsewhere in agency	26	29	10	17	17	69
Design studies to be performed by others (other than your own contractors)	67	25	5	3	0	60
Monitor evaluations conducted by others	34	45	3	8	9	64
Provide technical or design assistance to others' evaluations	24	50	8	12	6	66
Approve plans for studies conducted by others	59	23	3	10	5	61
Conduct joint or cooperative studies with other units	22	52	13	9	4	69
Train other units (e.g., federal, state, or local government, private agencies or firms) in research or evaluation methods	61	26	3	7	3	61
Other activity/role	17	33	17	8	25	12

Source: GAO survey.

About one-half of those who responded reported "sometimes" providing technical or design assistance to others or conducting joint studies, while a few offices saw their role as training others in research or evaluation methods. One office dealing with an evaluation mandate conducted work sessions with state and local program managers and evaluators as well as provided training to enhance state evaluation capabilities. Two-thirds of the offices seldom, if ever, designed evaluations conducted by other

offices or agencies, trained others in research or evaluation methods, or approved plans for studies by others.

Reforms May Increase Interest In, But Complicate The Process of, Obtaining Results

Some of our interviewees thought that recent governmentwide reforms would increase interest in learning the results of federal programs and policies but would also complicate the task of obtaining that information. Devolution of federal program responsibility in the welfare and health care systems has increased interest in evaluation because the reforms are major and largely untested. However, in part because programs devolved to the states are expected to operate quite diversely across the country, some evaluation officials noted that evaluating the effects of these reforms was expected to be more difficult. In addition, federal budget reductions over the past few years were said by some not only to have reduced the level of federal evaluation activity but also to have diminished agency technical capacity through the loss of some of their most experienced staff.

Because implementation of the Results Act's performance reporting requirements is not very far along (the first annual reports on program performance are not due until March 2000), several of our interviewees thought it was too early to estimate the effect of the Results Act. Some hoped the Act would increase the demand for results information and expand the role of data and analysis in decisionmaking. One interviewee thought it would improve the focus of the evaluations they now conduct. A few evaluation officials were concerned that a large investment would be required to produce valid and reliable outcome (rather than process) data. A few also noted that resources for obtaining data on a greatly expanded number of program areas would compete for funds used for more in-depth evaluations of program impact. Other evaluators noted that changes in the unit of analysis for performance reporting from the program level to budget account or organization might make classic program evaluation models obsolete.

As we previously reported,[9] the federal program officials who have already begun implementing performance measurement appeared to have an unusual degree of program evaluation support and found it quite helpful in addressing the analytic challenges of identifying program goals, developing measures, and collecting data. Many of these program officials said they could have used more of such assistance; but, when asked why they were not able to get the help they needed, the most common response

[9]GAO/HEHS/GGD-97-138.

was that it was hard to know in advance that evaluation expertise would be needed. In addition to using program evaluation techniques to clarify program goals and develop reliable measures, several of these program officials saw the need for impact evaluations to supplement their performance data. Their programs typically consisted of efforts to influence highly complex systems or events outside government control, where it is difficult to attribute a causal connection between their program and its desired outcomes. Thus, without an impact evaluation or similar effort to separate the effects of their programs from those of other external events or factors, program officials from the previous study recognized that simple examination of outcome measures may not accurately reflect their programs' performance.

Some states' experiences with performance measurement suggested that performance measurement will take time to implement, and the federal experience suggests that it will not supplant the need for effectiveness evaluations. Two state officials described a multiyear process to develop valid and reliable measures of program performance across the state government. While performance measures were seen as useful for program management, some state agency and legislative staff also saw a continuing need for evaluations to assess policy impact or address problems of special interest or "big-picture" concerns, such as whether a government program should be continued or privatized.

NPR was seen by several of those we interviewed as not having much of an effect on efforts to evaluate the results of their programs beyond increasing the use of customer surveys. This may have been because it was seen as primarily concerned with internal government operations, or because, as one agency official reported, its effect was most noticeable in only a few areas: regulatory programs and other intergovernmental partnerships. However, one agency official said that NPR had a big impact on reorienting their work toward facilitating program improvement, while two others felt that it reaffirmed changes they had already begun.

Agencies Suggested Varied Strategies for Obtaining Program Results Information

Leveraging Limited Resources

Given constraints on federal budgets, some officials we interviewed in general did not expect federal evaluation resources to rise to meet demand, so they described efforts to leverage and prioritize available resources. While an evaluation official reported supplementing his evaluation teams with consultants, concern was also expressed that staff reductions in their unit had left the technical expertise too weakened to competently oversee consultants' work. Another evaluation official explained that they responded to the increasing demand for information by narrowing the focus and scope of their studies to include only issues with major budget implications or direct implications for agency action. Both a state official and two external evaluation experts felt that states grappling with new program responsibilities would have difficulty evaluating them as well, so that continued federal investment would be needed. A federal official, however, noted that private foundations could fund the complex rigorous studies needed to answer causal questions about program results.

Some of the evaluators we interviewed expected that fewer impact studies would be done. Some expected that the range of their work may broaden to rely on less rigorous methods and include alternatives such as monitoring program performance and customer satisfaction. From our interviews, we learned that a few agencies have devolved responsibility for evaluations to the program offices, which may have more interest in program improvement. Another agency reported that it had built evaluation into its routine program review system, which provides continuous information on the success of the program and its outcomes, noting that it thereby reduced the need for special evaluation studies. One evaluation official reported that by having redefined evaluation as part of program management, program evaluation became more acceptable in his agency because it no longer appeared to be overhead.

A few agencies reported that they were adapting the elements of their existing program information systems to yield information on program results. But in other agencies, evaluation officials and external experts

thought that their systems were primarily focused on program process, rather than results. The evaluation official said that structural changes to, and a major investment in, their data systems will be required to provide valid and meaningful data on results.

Creating New Evaluation Designs and Partnerships

As program responsibility shifts to state and local entities, evaluation officials and others we interviewed described the need for study designs that can handle greater contextual complexity, new ways to measure outcomes, and the need to build partnerships with the programs' stakeholders. One of the officials saw classical experimental research designs as no longer feasible in programs, which, due to increased state flexibility in how to deliver services, no longer represented a discrete national program or were unlikely to employ rigorous evaluation techniques that entailed random assignment of particular program services to individuals. Others noted the need to develop evaluation designs that could reflect the multiple levels on which programs operate and the organizational partnerships involved. To address some of these complexities, federal offices with related program interests have formed task groups to attempt to integrate their research agendas on the effects of major changes in the health and welfare systems. Similarly, a few federal evaluation officials reported an interest in consulting with their colleagues in other federal offices to share approaches for tackling the common analytic problems they faced.

In other strategies, federal evaluation officials described existing or planned efforts to change the roles they and other program stakeholders played in conducting evaluations. One agency has arranged for the National Academy of Sciences to work with state program officials and the professional communities involved to help build a prototype performance measurement system for federal assistance to state programs. One evaluation office expects to shift its role toward providing more technical assistance to local evaluators and synthesizing their studies' results. Another federal office has delegated some evaluation responsibility to the field while it synthesizes the results to answer higher level policy questions, such as which types of approaches work best.

The Results Act recognizes and encourages the complementary nature of program evaluations and performance measures by asking agencies to provide a summary of program evaluation findings along with performance measurement results in their annual performance reports. One federal evaluation official said his agency had efforts under way to

"align" program evaluation and performance measurement through, for example, planning evaluations so that they will provide the performance data needed. But, the official also expressed concern about how to integrate the two types of information. Officials from states that had already begun performance measurement and monitoring said they would like to see the federal government provide more leadership by (1) providing a catalog of performance measures available for use in various program areas and (2) funding and designing impact evaluations to supplement their performance information.

Conclusions

Seeking to improve government performance and public confidence in government, the Results Act has instituted new requirements for federal agencies to report on their results at the same time that other management reforms may complicate the task of obtaining such information. Comparison of current federal program evaluation resources with the anticipated challenges leads us to several conclusions.

First, federal agencies' evaluation resources have important roles to play in responding to increased demand for information on program results, but—as currently configured and deployed—they are likely to be challenged to meet these future roles. It is implausible to expect that, by simply conducting more program evaluation studies themselves, these offices can produce data on results across all activities of the federal government. Moreover, some agencies reported that they had reduced their evaluation resources to the point that the remaining staff feel unable to meet their current responsibilities. Lastly, the devolution of some program responsibilities to state and local governments has increased the complexity of the programs they are being asked to evaluate, creating new challenges.

Second, in the future, carefully targeting and reshaping the use of federal evaluation resources and leveraging federal and nonfederal resources show promise for addressing the most important questions about program results. In particular, federal evaluators could assist program managers to develop valid and reliable performance reporting by sharing their expertise through consultation and training. Early agency efforts to meet the Results Act's requirements found program evaluation expertise helpful in managing the numerous analytical challenges involved, such as clarifying program goals and objectives, developing measures of program outcomes, and collecting and analyzing data. In addition, because performance measures will likely leave some gaps in needed information,

strategic planning for future evaluations might strive to fill those gaps by focusing on those questions judged to have the most policy importance. In many programs, performance measures alone are not sufficient to establish program impact or the reasons for observed performance. Program evaluations can also serve as valuable supplements to program performance reporting by addressing policy questions that extend beyond or across program borders, such as the comparative advantage of one policy alternative to another.

Finally, without coordination, it is unlikely that the increasingly diverse activities involved in evaluating an agency's programs will efficiently supplement each other to meet both program improvement and policymaking information needs. As some agencies devolve some of the evaluations they conducted in the past to program staff or state and local evaluators, they run the risk that, due to differences in evaluation resources and questions, data from several studies conducted independently may not likely be readily aggregated. Thus, in order for such devolution of evaluation responsibility to better provide an overall picture of a national program, those evaluations would have to be coordinated in advance. Similarly, as federal agencies increasingly face common analytic problems, they could probably benefit from cross-agency discussion and collaboration on approaches to those problems.

Agency Comments

The Director of OMB commented on a draft of this report and generally agreed with our conclusions. OMB noted that other countries are experiencing public sector reforms that include a focus on results and increasing interest in program evaluation. OMB also provided technical comments that we have incorporated as appropriate throughout the text. OMB's comments are reprinted in appendix II.

We are sending copies of this report to the Chair and Ranking Minority Member of the House Committee on Government Reform and Oversight, the Director of OMB, and other interested parties. We will also make copies available to others on request.

Please contact me or Stephanie Shipman, Assistant Director at (202) 512-7997 if you or your staff have any questions. Major contributors to this report are listed in appendix III.

Sincerely yours,

Susan S. Westin

Susan S. Westin
Associate Director, Advanced Studies
and Evaluation Methodology

Contents

Figure

Abbreviations

DOD	Department of Defense
FTE	full-time equivalent
FY	fiscal year
NPR	National Performance Review
OMB	Office of Management and Budget

Scope and Methodology of Survey

Departments and Independent Agencies Surveyed

The 23 federal executive agencies (13 cabinet-level departments and 10 independent agencies) that we surveyed are listed as follows.

Departments	Independent Agencies
Agriculture	Agency for International Development
Commerce	Environmental Protection Agency
Education	Federal Emergency Management Agency
Energy	General Services Administration
Health and Human Services	National Aeronautics and Space Administration
Housing and Urban Development	National Science Foundation
Interior	National Research Council
Justice	Office of Personnel Management
Labor	Small Business Administration
State	Social Security Administration
Transportation	
Treasury	
Veterans Affairs	

These represent 23 of the 24 executive agencies (we excluded the Department of Defense) covered by the Chief Financial Officer's Act. The 24 represent about 97 percent of the executive branch's full-time staff and cover over 99 percent of the federal government's outlay for fiscal year 1996.

Identifying Offices Conducting Evaluation Studies

To identify the roles and resources expended on federal program evaluation, we surveyed all offices (or units) in the 23 executive branch departments and independent agencies that we identified as conducting evaluation in fiscal year 1995. We defined evaluation as systematic analysis using objective measures to assess the results or the effects of federal programs, policies, or activities. To identify these evaluation offices, we (1) began with the list of evaluation offices that we surveyed in 1984 (2) added offices based on a review of office titles implying analytical responsibilities and discussions with experts knowledgeable about evaluation studies, and (3) talked with our liaison staff and other officials

in the federal departments and agencies to ensure broad yet appropriate survey coverage.

In some instances, the survey was distributed to offices throughout an agency by agency officials, while in other instances we sent the survey directly to named evaluation officials. We attempted to survey as many evaluation offices as possible; however, in some cases, we may not have been told about or directed to all such offices. Therefore, we cannot assume that we have identified all offices that conducted program evaluation studies in fiscal year 1995. Overall, we received about 160 responses, of which 81 were from offices that conducted such studies.

The survey was directed toward results-oriented evaluation studies, such as formal impact studies, assessments of program results, and syntheses or reviews of evaluation studies. We sought to exclude studies that focused solely on assessing client needs, describing program operations or implementation, or assessing fraud, compliance, or efficiency. However, we allowed the individual offices to (1) define "program" since a federal program could be tied to a single budget account, represent a combination of several programs, or involve several state programs and (2) determine whether or not they did this type of study and, if not, they could exempt themselves from completing the survey. We did not verify the accuracy of the responses provided by evaluation units. We also had some information on fiscal year 1996 activities but did not report those results since they were comparable to the fiscal year 1995 results.

Respondent Reporting

Some respondents were unable to complete different parts of the survey. About one-third of the respondents did not report either the office's budget, its number of full-time equivalent staff (FTE), cost information about studies, or the sources of data used in the studies. For some questions, respondents were asked to answer in terms of the number of studies conducted, and we used the total number of studies indicated by all respondents to the question as the denominator when computing percents. However, when the level of nonresponse to individual survey questions was above 20 percent or was unclear due to incomplete information on how many studies had been reported on, we used the full complement of 928 studies to provide a conservative estimate.

Selected Survey Questions

The questions for which we reported results are reproduced on the following pages.

1. Where is your **unit located** within your department or agency? (Please circle the most appropriate response.)

A. Inspector General
B. Central administrative office, e.g., assistant secretary for budget, administration, policy planning or analysis
C. Central administrative office for a major operating subdivision of the department/agency
D. Program office
E. Technical or analytic unit supporting a program office
F. Other (Please describe) _____

3. How many staff, in terms of full-time equivalent personnel **(FTEs)** were assigned to your unit in fiscal year (FY95) ?

FY95 FTE's _____

4. Please indicate the dollar amount of your unit's **gross budget outlays** in FY95.

FY95 $ _____

9. Below is a list of **analysis and evaluation-related activities**. By program evaluation we mean studies and activities that use objective measures to assess program results. Evaluation-related activities are tasks conducted by the federal agency that are associated with conducting an evaluation study, such as assisting others in conducting evaluation studies or developing evaluation plans.

(Please indicate which tasks your unit performed in **FY95**. Please check all relevant boxes in column 1. For each of the activity categories (in bold letters), please indicate in columns 2 and 3, the estimated percentage (%) of professional staff time and budget the different categories represent. You may use general figures such as 10, 20, 50, or 75 percent, and the percentages should total 100%.

Activities	1). FY95 (check)	2). Staff time (%)	3). Budg-et (%)
MANAGEMENT ANALYSIS			
Budget formulation or analysis			
Development of agency strategic plans or mission statements			
Management studies (e.g., quality management initiatives, NPR streamlining efforts)			
Descriptions of program implementation or operations			
Fraud, waste or abuse investigations, program efficiency audits			
Other			
POLICY PLANNING AND ANALYSIS			
Prospective analyses of probable costs and benefits of proposed policies			
Policy analyses or research summaries not directly concerning the effects of existing programs			
Needs or threat assessments			
Other			

Activities	1). FY95 (check)	2). Staff time (%)	3). Budg-et (%)
PROGRAM EVALUATION ACTIVITIES AND STUDIES			
Assessment of program compliance or quality			
Studies assessing a program's coverage of its target population or problem			
Evaluation of the quality of existing program data or information systems			
Development of agency evaluation plans			
Collection of data to serve as a baseline or to assist in evaluation design			
Development or field testing of new program evaluation methodologies			
Studies of program efficiency			
Results-Oriented Program Evaluation Activities and Studies			
Development or analysis of program performance measures			
Testing of weapons systems or other products			
Studies of the outcomes, results, effects or the impacts of ongoing programs			
Synthesis or further analysis of findings from previous evaluations			
Other Activity Representing 10% or more of Staff Time or Budget			

13. Which **roles did your unit play** in the conduct of program evaluation studies in FY95? And how often? (Please rank these roles in order of frequency by checking the appropriate column. If your unit played some other role, please describe it briefly and indicate its frequency.)

Unit's roles in evaluations	FY95 Frequency				
	Seldom-if ever	Some-times	Often (about 1/2 the time)	Very often	Always or almost always
1. Conduct evaluation studies of programs that are run by your office					
2. Conduct evaluation studies of programs administered elsewhere in agency					
3. Design studies to be performed by others (other than your own contractors)					
4. Monitor evaluations conducted by others					
5. Provide technical or design assistance to others' evaluations					
6. Approve plans for studies conducted by others					
7. Conduct joint or cooperative studies with other units					
8. Train other units (e.g., federal, state or local government, private agencies or firms) in research or evaluation methods					
9. Other activity/role:_____					

20. Did your unit conduct results-oriented **program evaluation studies** at any time in FY95? (If yes, continue. If no, check option d, and go to the last page of survey. You have completed this survey. Thank you!) NOTE: The surveys with option d checked were dropped from our analysis.

Studies	Number
On-going studies	_____(a)
Completed studies	_____(b)
TOTAL Studies	_____(c)
None	_____(d)

NOTE: The rest of the questions were answered only by those offices that conducted effectiveness studies.

24. Please select the option below that most closely describes your unit's **planning process** for deciding which program evaluation studies are to be performed in this or the next few fiscal years. (Please circle the number of the one answer most relevant to your unit.)

1. There is a formal planning process that determines the mix of mandated, requested, unit self-initiated and ad hoc work. And it is often followed.

2. There is a planning process, but ad hoc requests often override our planned mix of studies

3. There is no formal planning process, congressionally-mandated studies or requests from the agency head account for most of our work (Skip to question 25)

4. Other _____

25. What **criteria** were used in the planning process to **select** programs to be evaluated? (Please circle the number for all relevant answers.)

1. Program office request
2. Program or component has not been evaluated in a long time
3. Continuation or follow-up to past or related work
4. Need to determine and document program improvements
5. High-level agency interest expressed in program or subject area
6. Congressional interest or specific request
7. Other (Please describe.) _____
8. Don't know

27. Please indicate how many of your FY95 program evaluation studies were **intramural, extramural or conducted under a joint agreement.**

(If a program evaluation or a major component was conducted by your unit with consultant services and your unit provided day-to-day direction, list it as intramural.)

Study locus	Number of studies conducted
1. Intramural (in-house)	
2. Extramural (by grant or contract)	
3. Jointly under federal agency cooperative agreement	
4. Jointly conducted with state or local entity	
5. TOTAL	

28. Please indicate the **duration** of the intramural and other evaluation studies conducted during FY95 by noting the **number of studies** in each duration category.

How/where studies were conducted	Duration of study			
	Under 6 months	6 - 12 months	13 - 24 months	More than 2 years
1. Intramural				
2. Extramural (by grant or contract)				
3. Jointly under federal agency cooperative agreement				
4. Jointly conducted with state or local entity				

29. Please indicate the **cost** of the intramural and other evaluation studies conducted in FY95 by noting the **number of studies** in each cost category.

(Please use the total cost regardless of funding source or fiscal year in which funds were obligated. When estimating total costs of external program evaluations, please include staff costs associated with issuing and monitoring the contract or grants, as well as the amount of the contract or grant itself.)

Where/how studies conducted	Cost categories			
	Under $100,000	$100,000 - 499,000	$500,000 - 999,000	$1,000,000 or over
1. Intramural				
2. Extramural (by grant or contract)				
3. Jointly under federal agency cooperative agreement				
4. Jointly conducted with state or local entity				

30. How many FY95 studies were requested by the **various sources** listed below? Include all evaluations started, on-going, and completed during FY95. The study request may have been made at an earlier date, and there may have been more than 1 requestor per study.

Sources of mandate or request	Number of studies requested
1. Top agency officials	
2. Program personnel	
3. Self-initiated	
4. OMB or executive order	
5. Legislation or congressional committee	
6. Other _____	

31. How many of your unit's FY95 studies used each of the following **study design and tools**? They range from the most costly and resource intensive to the least.

Study design and tools		Number of studies for which tool was used
1.	Experimental and control groups randomly assigned: after-intervention measures	
2.	Some experimental control: before- and after-intervention outcome measures with some statistical controls	
3.	Time series: many repeated measures taken before and after intervention	
4.	Matched pairs or cross-sectional studies: outcomes measured after intervention with statistical control of matching	
5.	Panel studies: several repeated measures taken during and after intervention	
6.	Statistical modelling or simulation: implemented program outcomes compared to predicted (control) results	
7.	Simple before- and after- studies: outcomes measured before and after intervention	
8.	One time survey of outcomes	
9.	Judgmental assessment: use of outside expert, program administrator, or participant judgement on program effects	
10.	Other (Describe) _____ _____	

32. How many of your unit's FY95 evaluation studies used each of the following **data sources**? A study may have used more than 1 data source. (Please indicate the number of studies using each data source option.)

Data sources used	Number of studies in which used
Program Generated Information/Data	
1. Program administrative records routinely produced and reported federally	
2. Program administrative records routinely produced but not typically reported or aggregated centrally	
3. Recurring surveys of program participants	
4. Special surveys or other ad hoc data collections on all or part of the program or in selected locales. (Please name source, also.)	
Other Agency Generated Information/Data	Please also name the source for each category below
5. Administrative records or routinely aggregated data from other federal programs (e.g. Medicare Claims and Utilization Files)	
6. Administrative records or routine data from state programs or agencies (e.g., Federal Accident Reporting System (FARS))	
7. Federally-sponsored multi-purpose national surveys (e.g. the decennial census, or National Household Survey of Drug Abuse)	
8. Other (Please describe.) _____	

35. What were the **primary uses** of the **results** of the program evaluation studies conducted in FY95? (Circle the number for all appropriate answers).

1. Identify opportunities to improve program performance
2. Ascertain the extent of the program's effectiveness
3. Guide decisions on resource allocation within the program
4. Ascertain the success of corrective actions
5. Support program budget requests
6. Support program reauthorization
7. Increase general knowledge about the program or topic
8. Other
9. Don't know

36. Who was the **primary audience** for the **results** of the evaluation studies conducted in FY95? (Circle the number for all appropriate answers).

1. Program managers
2. Higher level agency officials
3. Agency evaluation staff
4. Office of Management and Budget (OMB)

5. Congress, in general
6 appropriations committees
7. Legislative committees

8. Program partners in private sector
9. Outside professional audience (experts, researchers, or analysts)
10. General public, others
11. Don't know

Comments From the Office of Management and Budget

EXECUTIVE OFFICE OF THE PRESIDENT
OFFICE OF MANAGEMENT AND BUDGET
WASHINGTON, D.C. 20503

FEB - 5 1998

Susan Westin
Associate Director
Advanced Studies and Evaluation Methodology
U.S. General Accounting Office
441 G Street NW
Washington, D.C. 20548

Dear Ms. Westin:

Thank you for your January 16, 1998 letter to the Director seeking comments on the draft report entitled Program Evaluation: Agencies Challenged by New Needs for Information on Program Results.

We agree with the General Accounting Office that program evaluation is of considerable value both to agency managers, external decision-makers, and other stakeholders. Evaluation is particularly important in a result-oriented environment, because it provides feedback on the performance of public programs. Moreover, evaluation is a useful complement to other performance management approaches, such as ongoing performance measurement and reporting, and service quality initiatives, and in this way, contributes to improved management

It is important to note that a focus on results is a central element in recent public sector reforms not only in the United States, but in other democratic, market economies. According to the Organization for Economic Cooperation and Development (OECD), there are signs that interest in using program evaluation as a management tool are increasing. What is new, compared with previous evaluation, are more realistic expectations, more widespread acceptance of less rigorous methodologies, and greater understanding about utilization in organizations. As here in the United States, there is also a strong emphasis on more systematic, outcome-oriented evaluation with linkages to the budget process.

For these reasons, we have been an active participant in work by the OECD's Public Management Committee to learn how evaluation is used in other countries as a decision-making tool. One product of this work will be a "policy brief" presenting best practice guidelines for evaluation. An accompanying background report, *"Promoting the Use of Program Evaluation,"* will present a more detailed discussion of evaluation, its role in public management and ways of enhancing its use, and explore evaluation practice in selected OECD countries. We will be happy to share these with you when they are completed later this Spring.

Sincerely,

G. Edward DeSeve
Acting Deputy Director for
Management

Major Contributors

General Government Division

Elaine Vaurio, Evaluator-in-Charge
Joseph Wholey, Senior Advisor for Evaluation Methodology

Bibliography

Committee on Governmental Affairs, United States Senate. "Government Performance and Results Act of 1993." Report No. 103-58, June 16, 1993.

Evaluation Practice. "Past, Present, Future Assessments of the Field of Evaluation." Entire Issue. M.F. Smith, ed., Vol. 15, #3, Oct. 1994.

Martin, Margaret E., and Miron L. Straf (eds.). Principles and Practices for a Federal Statistical Agency. Washington, D.C.: National Academy Press, 1992.

National Performance Review. "Mission-Driven, Results-Oriented Budgeting." Accompanying Report of the National Performance Review of the Office of the Vice President, Sept. 1993.

New Directions for Program Evaluation. "Evaluation in the Federal Government: Changes, Trends, and Opportunities." Entire issue. C.G. Wye and R. Sonnichsen, eds. #55, Fall 1992.

New Directions for Program Evaluation. "Progress and Future Directions in Evaluation: Perspectives on Theory, Practice, and Methods." Entire issue. Debra Rog and Deborah Fournier, eds. #76, Winter 1997.

Office of Evaluation and Inspections. Practical Evaluation for Public Managers: Getting the Information You Need. Washington, D.C.: Office of Inspector General, Department of Health and Human Services, 1994.

Public Law 103-62, Aug. 3, 1993, "Government Performance and Results Act of 1993."

Wargo, Michael J. "The Impact of Federal Government Reinvention on Federal Evaluation Activity." Evaluation Practice, 16(3) (1995), pp. 227-237.

Related GAO Products

The Results Act: An Evaluator's Guide to Assessing Agency Annual Performance Plans (GAO/GGD-10.1.19, Mar. 1998).

Balancing Flexibility and Accountability: Grant Program Design in Education and Other Areas (GAO/T-GGD/HEHS-98-94, Feb. 11, 1998).

The Government Performance and Results Act: 1997 Governmentwide Implementation Will Be Uneven (GAO/GGD-97-109, June 2, 1997).

Managing for Results: Analytic Challenges in Measuring Performance (GAO/HEHS/GGD-97-138, May 30, 1997).

Block Grants: Issues in Designing Accountability Provisions (GAO/AIMD-95-226, Sept. 1995).

Program Evaluation: Improving the Flow of Information to the Congress (GAO/PEMD-95-1, Jan. 30, 1995).

Management Reform: Implementation of the National Performance Review's Recommendations (GAO/OGC-95-1, Dec. 5, 1994).

Public Health Service: Evaluation Set-Aside Has Not Realized Its Potential to Inform the Congress (GAO/PEMD-93-13, Apr. 1993).

Program Evaluation Issues (GAO/OCG-93-6TR, Dec. 1992).

"Improving Program Evaluation in the Executive Branch." A Discussion Paper by the Program Evaluation and Methodology Division (GAO/PEMD-90-19, May 1990).

Program Evaluation Issues (GAO/OCG-89-8TR, Nov. 1988).

Federal Evaluation: Fewer Units, Reduced Resources, Different Studies from 1980 (GAO/PEMD-87-9, Jan. 23, 1987).

(966704/973810)

Ordering Information

The first copy of each GAO report and testimony is free. Additional copies are $2 each. Orders should be sent to the following address, accompanied by a check or money order made out to the Superintendent of Documents, when necessary. VISA and MasterCard credit cards are accepted, also. Orders for 100 or more copies to be mailed to a single address are discounted 25 percent.

Orders by mail:

U.S. General Accounting Office
P.O. Box 37050
Washington, DC 20013

or visit:

Room 1100
700 4th St. NW (corner of 4th and G Sts. NW)
U.S. General Accounting Office
Washington, DC

Orders may also be placed by calling (202) 512-6000 or by using fax number (202) 512-6061, or TDD (202) 512-2537.

Each day, GAO issues a list of newly available reports and testimony. To receive facsimile copies of the daily list or any list from the past 30 days, please call (202) 512-6000 using a touchtone phone. A recorded menu will provide information on how to obtain these lists.

For information on how to access GAO reports on the INTERNET, send an e-mail message with "info" in the body to:

info@www.gao.gov

or visit GAO's World Wide Web Home Page at:

http://www.gao.gov

United States
General Accounting Office
Washington, D.C. 20548-0001

Bulk Rate
Postage & Fees Paid
GAO
Permit No. G100

Official Business
Penalty for Private Use $300

Address Correction Requested