

September 2000

PROGRAM EVALUATION

Studies Helped Agencies Measure or Explain Program Performance



G A O

Accountability * Integrity * Reliability



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

Assistant Comptroller General
General Government Division

B-285377

September 29, 2000

The Honorable Fred Thompson, Chairman
Committee on Governmental Affairs
United States Senate

The Honorable Dan Burton, Chairman
The Honorable Henry A. Waxman
Ranking Minority Member
Committee on Government Reform
House of Representatives

Congressional and federal agency decisionmakers need evaluative information about how well federal programs are working, both to manage programs effectively and to help decide how to allocate limited federal resources. The Government Performance and Results Act of 1993 (GPRA) requires federal agencies to report annually on their achievement of performance goals, explain why any goals were not met, and summarize the findings of any program evaluations conducted during the year. Program evaluations are objective, systematic studies that answer questions about program performance and results. By examining a broader range of information than is feasible to monitor on an ongoing basis through performance measures, an evaluation study can explore the benefits of a program as well as ways to improve program performance.

To assist agencies in identifying how they might use evaluations to improve their performance reporting, we identified eight concrete examples of diverse ways in which agencies incorporated program evaluations and evaluation methods in their fiscal year 1999 annual performance reports. This report, which we prepared at our own initiative, discusses how the agencies used these evaluation studies to report on their achievements. Because of your interest in improving the quality of information on federal programs, we are addressing this report to you.

We selected the cases to demonstrate varied uses of evaluation on the basis of a review of several departments' fiscal year 1999 annual performance reports and consultations with agency officials. We then reviewed agency documents and interviewed agency officials to address two questions: (1) what purposes did these program evaluation studies or methods serve in performance reporting and (2) what circumstances led agencies to conduct these evaluations?

Results in Brief

The agencies used the evaluation studies in a variety of ways, reflecting differences in programs and available data, but they served two general purposes in agencies' fiscal year 1999 annual performance reports. Evaluations helped the agencies improve their measurement of program performance or understanding of performance and how it might be improved; some studies did both.

To help improve their performance measurement, two agencies used the findings of effectiveness evaluations to provide data on program results that were otherwise unavailable. One agency supported a number of studies to help states prepare the groundwork for and pilot-test future performance measures. Another used evaluation methods to validate the accuracy of existing performance data. To better understand program performance, one agency reported evaluation and audit findings to address other, operational concerns about the program. Four agencies drew on evaluations to explain the reasons for observed performance or identify ways to improve performance. Finally, three agencies compared their program's results with estimates of what might have happened in the program's absence in order to assess their program's net impact or contribution to results.

Two of the evaluations we reviewed were initiated in response to legislative provisions, but most of the studies were self-initiated by agencies in response to concerns about the program's performance or about the availability of outcome data. Some studies were initiated by agencies for reasons unrelated to meeting GPRA requirements and thus served purposes beyond those they were designed to address. In some cases, evaluations were launched to identify the reasons for poor program performance and learn how that could be remedied. In other cases, agencies initiated special studies because they faced challenges in collecting outcome data on an ongoing basis. These challenges included the time and expense involved, grantees' concerns about reporting burden, and substantial variability in states' data collection capabilities. In addition, one departmentwide study was initiated in order to direct attention to an issue that cut across program boundaries and agencies' responsibilities.

As agencies governmentwide update their strategic and performance plans, the examples in this report might help them identify ways that evaluations can contribute to understanding their programs' performance. These cases also provide examples of ways agencies might leverage their evaluation resources through

-
- drawing on the findings of a wide array of evaluations and audits,
 - making multiple use of an evaluation's findings,
 - mining existing databases, and
 - collaborating with state and local program partners to develop mutually useful performance data.

Two of the agencies discussed in this report indicated they generally agreed with it. The others either had no comments or provided technical comments.

Background

Performance measurement under GPRA is the ongoing monitoring and reporting of program accomplishments, particularly progress toward preestablished goals. It tends to focus on regularly collected data on the level and type of program activities (process), the direct products and services delivered by the program (outputs), and the results of those activities (outcomes). For programs that have readily observable results or outcomes, performance measurement may provide sufficient information to demonstrate program results. In some programs, however, outcomes are not quickly achieved or readily observed, or their relationship to the program is uncertain. 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 individual, systematic studies that use objective measurement and analysis to answer specific questions about how well a program is working and, thus, may take many forms. Where a program aims to produce changes that result from program activities, outcome or effectiveness evaluations assess the extent to which those results were achieved. Where complex systems or events outside a program's control also influence its outcomes, impact evaluations use scientific research methods to establish the causal connection between outcomes and program activities and isolate the program's contribution to those changes. 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.

Although GPRA does not require agencies to conduct formal program evaluations, it does require them to (1) measure progress toward achieving their goals, (2) identify which external factors might affect such progress, and (3) explain why a goal was not met. GPRA recognizes the complementary nature of program evaluation and performance measurement. Strategic plans are to describe the program evaluations that were used in establishing and revising goals and to include a schedule for

future program evaluations. Agencies are to summarize the findings of program evaluations in their annual performance reports. However, in our review of agencies' 1997 strategic plans, we found that many agencies had not given sufficient attention to how program evaluations would be used in implementing GPRA and improving program performance.¹ To demonstrate the kinds of contributions program evaluations can make, this report describes examples of how selected agencies incorporated evaluation studies and methods in their fiscal year 1999 performance reports.

Scope and Methodology

To assist agencies in identifying how they might improve their performance reporting, we conducted case studies of how some agencies have already used evaluation studies and methods in their performance reports. To select these cases, we reviewed the fiscal year 1999 annual performance reports of several departments for references to program evaluations. References could be located in either a separate section on evaluations conducted during 1999 or in the detailed discussion of how the agency met its performance targets. We selected cases to represent a variety of evaluation approaches and methods without regard to whether they constituted a formally defined program evaluation study. Six of our cases consisted of individual programs, one represented an agency within a department, and another represented a group of programs within a department. All eight cases are described below.

To identify the purposes that evaluation served in performance reporting and the types of evaluation studies or methods used, we analyzed the agencies' performance reports and other published materials. We then confirmed our understandings with agency officials and obtained additional information on what circumstances led them to conduct these evaluations. Our findings are limited to the examples reviewed and thus do not necessarily reflect the full scope of these agencies' evaluation activities.

We conducted our work between May and August 2000 in accordance with generally accepted government auditing standards. We requested comments on a draft of this report from the heads of the agencies responsible for our eight cases. The Departments of Health and Human Services (HHS) and Veterans Affairs (VA) provided written comments that are reprinted in appendixes I and II. The agencies' comments are discussed at the end of this letter. The other agencies either had no comments or

¹Managing for Results: Agencies' Annual Performance Plans Can Help Address Strategic Planning Challenges (GAO/GGD-98-44, Jan. 30, 1998).

provided technical comments that we incorporated where appropriate throughout the text.

Program Descriptions

Community and Migrant Health Centers (C/MHC). Administered by the Health Resources and Services Administration (HRSA) in the Department of Health and Human Services, this program aims to increase access to primary and preventive care and to improve the health status of underserved and vulnerable populations. The program distributes grants that support systems and providers of health care in underserved areas around the country.

Hazardous Materials Transportation safety programs. Five administrations within the Department of Transportation (DOT) administer and enforce federal hazardous materials transportation law. The Research and Special Programs Administration (RSPA) has primary responsibility for issuing cross-modal safety regulations to help ensure compliance with certain packaging manufacturing and testing requirements. RSPA also collects and stores hazardous materials incident data for all the administrations. The four other administrations are largely responsible for enforcing safety regulations to gain shipper and carrier compliance in their respective modes of transportation (e.g., the Federal Aviation Administration, for the air mode).

Mediterranean Fruit Fly (Medfly) Exclusion and Detection program. This program, in the Animal and Plant Health Inspection Service (APHIS) of the U.S. Department of Agriculture (USDA), aims to control and eradicate fruit flies in the United States and in foreign countries whose exports may pose a serious threat to U.S. agriculture. The United States, Mexico, and Guatemala operate a cooperative program of detection and prevention activities to control Medfly populations in those countries.

Montgomery GI Bill education benefits. This program in the Veterans Benefits Administration, Department of Veterans Affairs, provides educational assistance to veterans and active-duty members of the U.S. armed forces. It reimburses participants for taking courses at certain types of schools and is used by the Department of Defense as a recruiting incentive.

Occupational Safety and Health Administration (OSHA) illness and injury data. In the Department of Labor (DOL), OSHA collects incident data on workplace injuries and illnesses as part of its regulatory activities and to develop data on workplace safety and health. OSHA requires employers to keep records on these injuries and illnesses and also uses

these data to target its enforcement activities and its compliance assistance efforts.

Substance Abuse Prevention and Treatment (SAPT) block grant. The Substance Abuse and Mental Health Services Administration (SAMHSA) in HHS aims to improve the quality and availability of services for substance abuse prevention and treatment and awards block grants to states to fund local drug and alcohol abuse programs.

Upward Bound program. The Office of Postsecondary Education, in the Department of Education, administers this higher education support services program. The program aims to help disadvantaged students prepare to enter and succeed in college by providing an intense academic experience during the summer, supplemented with mentoring and tutoring over the school year in the 9th through 12th grades.

Welfare-to-Work grants. In 1998, DOL's Employment and Training Administration began administering Welfare-to-Work grants to states and localities aimed at moving "hard to employ" welfare recipients (in the Temporary Assistance for Needy Families (TANF) program administered by HHS) into lasting, unsubsidized employment and economic self-sufficiency. Formula grants go through states to local providers, while competitive grants are awarded directly, often to "nontraditional" providers outside the DOL workforce development system.

Evaluations Helped Agencies Improve Their Measurement or Understanding of Performance

In the cases we reviewed, agencies used evaluations in a variety of different ways in their performance reports, but the evaluations served two general purposes. Evaluations were used to develop or improve upon agencies' measures of program performance or to better understand performance and how it might be improved. Two of the more complex evaluations conducted multiple analyses to answer distinct questions and, thus, served several purposes in the performance report.

Program characteristics, the availability of data, and the nature of the agencies' questions about program performance influenced the designs and methods used.

- Fairly simple programs, such as the collection of workplace injury and illness data, did not require complicated study designs to learn whether the program was effective in collecting accurate, useful data.
- Programs without ready access to outcome data surveyed program participants to learn how the program had affected them. Where desired

impacts take a long time to develop, agencies tracked participants several years after they left the program.

- A few programs, to assess their net impact on desired outcomes, arranged for comparisons with what might have happened in the absence of the program.

Developing or Improving Measures of Performance

Three agencies drew on evaluations to provide data measuring achievement of their performance goals—either now or in the future. In these cases, the agencies used program evaluations to generate data on program results that were not regularly collected or to prepare to do so in the future. A fourth agency used evaluation methods to help ensure the quality of its regularly collected performance data.

The Department of Education reported results from its evaluation of the Upward Bound program to provide data on both program and departmental performance goals. Where desired impacts take a long time to develop, agencies might require data on participants' experiences years after they leave the program. This evaluation tracked a group of 13- to 19-year-old participants (low-income or potential first generation college students) for 2 years after their enrollment in the program in 1993-94 to learn about their high school courses and grades, educational expectations, high school completion, and college enrollment. The average length of participation in the program for that cohort of participants and the percentage who enrolled in college after 2 years were reported as performance data for the program for fiscal years 1996 and 1997. The report explained that this evaluation would not provide performance data on these variables for future years but that the grantee reporting requirements were being revised to make this information available in the future.

Education also reported the evaluation's estimate of Upward Bound's net impact in order to support a departmental goal that program participation will make a difference in participants' college enrollment. The study assessed the value added from participating in Upward Bound by comparing the experience of this cohort of program participants with those of a control group of similar nonparticipating students to obtain an indication of the program's contribution to the observed results. By having randomly assigned students to either participate in the program or be in the control group, the evaluation eliminated the likelihood that selection bias (affecting who was able to enter the program) could explain any difference in results between the groups. Indeed, the evaluation found no statistically significant difference between the two groups as a whole in college enrollment. The evaluation is tracking this same group of

participants and nonparticipants into their fifth year to see if there are longer term effects on their college experience. However, because no new cohorts of participants are being tracked, the evaluation will not provide data on this departmental goal for future years.

HHS reported the results of special surveys of C/MHC users and visits conducted in 1995 to provide data for its performance goals of increasing the utilization of preventive health services. Surveying nationally representative samples of centers provided national estimates for measures such as the proportion of women patients at the health centers who received age-appropriate cancer screening. HRSA proposes to repeat these surveys in fiscal year 2000 and every 5 years thereafter to provide longitudinal, if intermittent, data on these goals. HRSA used annual health center reports to provide data on the number and demographic characteristics of center users to address its performance goals related to access. Agency officials noted that they would not conduct the surveys of users and visits annually because they are intrusive, costly efforts and because yearly patient data are not needed to assess the fairly gradual trends in these variables. Agency officials suggested that some annual data on utilization of preventive services might be provided in the future by a subset of centers involved in special research initiatives on improving quality of care.

In a program new to outcome monitoring, SAMHSA is sponsoring a number of studies to lay the groundwork for a future set of treatment effectiveness performance measures for the SAPT block grant. The agency funded individual program evaluations and research studies in 19 states under the Treatment Outcomes and Performance Pilot Studies Enhancement (TOPPS II). These studies involved developing and pilot-testing measures of client status and outcomes; field-testing computerized assessment and outcome monitoring systems; determining the feasibility of linking client information with data from health, employment, and criminal justice databases; and developing data quality assurance systems. As a condition of receiving funding for the TOPPS II projects, the 19 states involved agreed to develop and monitor a core set of substance abuse treatment effectiveness measures for an interstate study. A 31-item core set of measures was adopted through consensus in fiscal year 1999. For the HHS performance report, SAMHSA has asked all states to voluntarily report data on four of these measures in their block grant applications. Agency officials told us that during fiscal year 2000, 25 states (six more than originally targeted under GPRA) reported on some of these measures.

DOL's annual performance report included the results of an OSHA data quality assurance study to attest to the accuracy of employer-provided data on workplace injuries and illnesses. Since 1997, OSHA has conducted annual, on-site audits of employer injury and illness records of nationally representative samples of the approximately 80,000 establishments in high-hazard industries. These establishments are the source of the data OSHA uses both to target its enforcement and compliance assistance interventions and to measure its performance in reducing workplace injuries and illnesses in several job sectors. The recordkeeping audits are conducted to verify the overall accuracy of the employer's source records, estimate the extent of compliance with OSHA recordkeeping requirements, and assess the consistency between the data on the employer's log (source records) and the data submitted to the agency for monitoring injuries and illnesses. Because OSHA uses these data to target its enforcement of workplace safety regulations, there were concerns that this might encourage employer underreporting. The DOL performance report notes that the audits found that the accuracy of employer recordkeeping supports OSHA's continued use of the data for targeting and performance measurement purposes.

Improving Understanding of Program Performance

Knowing whether or not a performance goal was met may not answer key questions about a program's performance, nor does it give an agency direction on how to improve program performance. Some of the agencies used evaluations to further their understanding of program performance by providing data on other aspects of performance, explaining the reasons for observed performance or why goals were not met, or demonstrating the program's net impact on its outcome goals.

Probing Other Aspects of Program Performance

DOL's performance report summarized the findings of several studies conducted of its new Welfare-to-Work grant program. These studies assessed operational concerns that were not addressed by DOL's outcome-oriented performance measure: the percentage of program terminees placed in unsubsidized employment. An evaluation and financial and performance audits were conducted to address the many questions raised about the operations of the new program. In the first phase of an effectiveness evaluation, grantees were surveyed about their organization, funding sources, participants, services, and early implementation issues—more detailed information than they would provide in their quarterly financial reports. This stage of the evaluation addressed questions such as who was served, what services were provided, and what implementation issues had emerged so far.

In addition, the DOL's Office of the Inspector General (OIG) conducted on-site audits of both competitive and formula grant awardees to assess whether financial and administrative systems were in place. Because OIG had noted grantees' low enrollment numbers, these reviews also looked into issues surrounding the program eligibility criteria and the coordination of client outreach with HHS' TANF program. Both the interim report of the effectiveness evaluation and the OIG surveys found that grantees were slow in getting their programs under way and viewed the program eligibility criteria as too restrictive. The DOL performance report describes the operational concerns raised by these reviews and the changes made in response—both legislative changes to the eligibility criteria and the Department's provision of increased technical assistance to grantees.

Explaining the Reasons for Performance or Why Goals Were Not Met

Performance monitoring can reveal changes in performance but not the reasons for those changes. Four agencies referred to evaluation studies in their performance report to explain the reasons for their performance or the basis for actions taken or planned to improve performance. Two of the studies uncovered the reasons through examining program operations, while the other two studies examined the details of participants' outcomes.

The USDA performance report cited an APHIS evaluation completed in December 1998 to demonstrate how the agency responded when its performance suddenly declined and why it believed that it would meet its fiscal year 2000 goal. In 1998, when weekly detection reports showed a sudden outbreak of Medflies along the Mexico-Guatemala border, APHIS deployed an international team of scientists to conduct a rapid field study to learn why the program was suddenly less effective in controlling the Medfly population. The scientific team reviewed policies, practices, resources, and coordination between the two countries' detection, surveillance, control, and regulatory (quarantine) programs. This in-depth study identified causes for the outbreak and within a month recommended changes in their trapping and spraying programs. The performance report described the emergency program eradication activities under way since June 1999 in response to the evaluation's recommendations and the continuing decline in infestations throughout the year.

At VA, an evaluation study that was completed just after the performance report was issued will help explain the observed results of the Montgomery GI Bill education benefits. The program's performance measure is the extent of veterans' use of the education benefits. The evaluation's survey of program participants (both users and nonusers of the education benefits) looked at such factors as claims processes, timing

of receipt of benefits, awareness of the program, eligibility criteria, and why education benefits might not be an incentive to join the military, to understand what influences usage rates. In interviews supplementing the survey, recruiters, claims adjusters, and school officials shared their experiences on how factors such as communication about program benefits, payment schedule, and certification procedures hamper effective program administration, which in turn affects benefit usage. The study also found that lower income participants who did not complete their educational program most often cited “job responsibilities” or “ran out of money” as the reason. In addition, 41 percent of all participants reported that they would have enrolled in a different program or school if the benefit level were higher. This led the evaluators to suggest raising the benefit level.

Because analyses showed that, on the whole, the Upward Bound program had few statistically significant impacts on the evaluation’s cohort of students during their high school years, additional analyses probed whether some subgroups benefited more than others. The evaluation compared the results for subgroups of program participants with the results for subgroups of the control group. Indeed, those analyses found program impacts for students who had low expectations, were academically high-risk, or were male. The evaluation also found larger impacts for students who stayed in the program longer. This led the evaluators to suggest that the program focus more effort on increasing the length of program participation and retargeting the program to at-risk students.

DOT described the evaluation of its hazardous materials transportation safety system in fiscal year 1999 as one of the Department’s strategies to achieve its fiscal year 2001 goal to “reduce the number of serious hazardous materials incidents in transportation.” To learn how performance could be improved, DOT conducted a departmentwide study to assess how hazardous materials transportation safety was implemented in the different transportation modes and how those policies and procedures operate across the different modal administrations. A departmentwide team reviewed hazardous materials legislation and regulations; analyzed mission and function statements; reviewed internal and external reports, including the administrations’ plans and budgets; and reviewed hazardous materials industry, incident, and enforcement data. The team interviewed hazardous materials managers and field personnel and held focus groups with stakeholders in the hazardous materials community on how to improve program performance. It conducted on-site inspections of air, marine, rail, and highway freight operations and

intermodal transfer locations to observe different types of carriers and shippers and the hazards involved when a shipment's route spans different modes.

Since the hazardous materials transportation evaluation was only recently completed, its recommended corrective actions are cited in the DOT performance report as ways DOT expects to improve program delivery, for example, by increasing emphasis on shippers, and to address data quality issues in the future. In reviewing the database on hazardous materials incidents, the evaluation team noted the need to improve the quality of incident reports and the analysis of that data in order to better understand the root causes of such incidents.

Estimating Program's Net Impact on Results

Where external events also influence achievement of a program's desired outcomes, impact evaluations are needed to isolate and assess the agency's contributions to those changes. In addition to the Upward Bound impact evaluation described above, two other cases reported on impact evaluations in their performance report. To isolate and assess the program's net impact, the two cases used different ways to estimate what might have happened in the program's absence.

HHS reported on two impact evaluations to establish what difference the health centers were having on its larger, strategic objective—reducing disparities in access to health care. To demonstrate the program's impact, HHS compared the rates at which health center users were receiving certain preventive health services, such as breast cancer screening, to the rates for other low-income patients who did not use C/MHCs. This analysis drew on HRSA's special 1995 survey of center users and visits as well as special analyses to identify a subgroup of respondents with similar income and demographics from a comparable national survey of the general population—the National Health Interview Survey. These data sets were used in a similar analysis of minority persons diagnosed with hypertension that found center users were three times as likely as a comparable national group to report their blood pressure was under control.

HHS reported on a second study that analyzed an existing medical records database to assess progress toward the performance goal of reducing health center users' hospitalizations for potentially avoidable conditions. Researchers analyzed State Medicaid Research Files, which offer data on inpatient and outpatient services, and clinical and demographic data on Medicaid beneficiaries to identify hospitalizations for a group of health center users and a similar group who used some other source of care. Researchers identified "ambulatory care sensitive conditions" (i.e., medical

conditions, such as diabetes, asthma, or hypertension, for which timely, appropriate care can prevent or reduce the likelihood of hospitalization) based on diagnostic codes used in a previous Institute of Medicine study of access to health care. These analyses found that the Medicaid beneficiaries using health centers had a lower rate of hospitalization for “ambulatory care sensitive conditions” than did Medicaid beneficiaries who relied on other sources of primary care.

The VA performance report also alerted readers that its evaluation went beyond measuring the use of education benefits to identify whether they helped GIs actually achieve their educational goals—the strategic objective of the GI Bill. To obtain this information, the VA surveyed users and compared their completion of educational programs and other outcome measures with those GIs who did not use the education benefits. The differences between the groups in employment levels, educational indebtedness, and the importance of the benefit as a service retention incentive demonstrated the effects of the educational benefits. For example, users of the education benefits had fewer difficulties in finding a job after leaving the military and were more likely to pursue 2- or 4-year academic programs.

Studies Were Initiated to Answer Questions About Program Performance

Two of the evaluations we reviewed were initiated in response to legislative provisions (e.g., to track a new program’s progress), but most studies were self-initiated to address concerns about program performance or the availability of outcome data. Several of these evaluations were initiated for reasons other than meeting GPRA requirements and thus served purposes beyond those they were designed to address.

Legislative Provisions to Assess Program Performance

Congress mandated an evaluation study to assess program performance in one of our cases, the Welfare-to-Work program, and encouraged it in another, the Upward Bound program. In the first one, Congress wanted early implementation information on a new program. In the second one, Congress challenged service providers to show evidence of program success.

Welfare reform enacted in 1996 created a new work-focused and time-limited program of Temporary Assistance for Needy Families, operated by HHS, which gave the states considerable flexibility in designing programs. In 1997, as most states focused on job search activities to move welfare clients into jobs, the Welfare-to-Work grant program was authorized to give states and localities additional resources to serve those welfare recipients who were hardest to employ. HHS, in conjunction with DOL and

the Department of Housing and Urban Development, was required to evaluate “how the grants have been used” and urged to include specific outcome measures, such as the proportion of participants placed in unsubsidized jobs and their earnings. The law required an interim report by January 1, 1999, and a final report by January 1, 2001. One of the findings in the interim report, that grantees felt the eligibility criteria were too restrictive, was addressed in legislative changes passed later that year to broaden the eligibility criteria along with other programmatic changes expected to enhance performance.

In 1991, during consideration of Upward Bound’s legislative reauthorization, there were concerns about improving college access and retention for low-income and first-generation students. The administration proposed to replace this and two other college-based programs with a formula-driven state block grant program. In contrast, the grantee service providers encouraged legislation to maintain the existing program structure and require ongoing evaluations to identify effective practices. Congress passed legislation that, to improve the operations of the program, encouraged Education to evaluate the effectiveness of the various Upward Bound programs and projects, describe the programs or practices that were particularly effective, and share these results with other providers. Education’s Program Evaluation Service, in conjunction with the program office, has conducted a series of effectiveness and impact studies that followed a cohort of program participants.

Agency Concerns About Program Performance

Some evaluations were initiated by the agencies in response to specific concerns about program performance and helped identify how to improve performance.

In our most dramatic case, when APHIS program officials received monitoring reports of the most serious Medfly outbreak since the pest was eradicated from Mexico in 1982, the agency quickly deployed a study team to learn the causes. A multinational team led by APHIS was charged with assessing the effectiveness of current operations and the appropriateness of current methods and with recommending specific technical interventions to address the current situation and a strategy for the future. The evaluation recommended specific changes in program strategy and a quick infusion of resources. Implementation of these changes appears to have improved the situation remarkably the following year.

Agency officials told us that senior DOT leadership made the commitment to evaluate the Department’s hazardous materials transportation policies in their Strategic Plan to meet corporate management as well as mission-

oriented goals. They said that they were looking for a crosscutting issue that would address the Secretary's goal of having the different modal administrations in the Department work better together. Hazardous materials transportation surfaced as a promising area for such an evaluation because it involved a key strategic goal—safety—and the Department had wrestled for several years with the disparate ways in which its hazardous materials programs had been implemented by the administrations. Since the performance report was released, agency officials reported that the Department had implemented the recommendation to create a centralized DOT-wide institutional capacity to both coordinate hazardous materials programs and implement the report's remaining recommendations.

The DOL's Office of the Inspector General audited Welfare-to-Work grantees under two broad initiatives. First, postaward surveys of competitive grants were conducted immediately upon awarding the grants because these grants aimed to reach nontraditional faith-based and welfare organizations and others that were new to DOL's grant management and reporting requirements. Lacking that experience, these organizations were considered to be at risk of not having the financial, organizational, or management systems needed to meet the grant requirements. Second, after the grantees' financial status and management reports showed that state formula grantees were not drawing down funds at the expected rate, OIG assumed that they were probably having difficulties implementing the program and examined a sample of grantees to identify the extent and causes of any difficulties. OIG's findings reiterated the problems with the eligibility criteria and client outreach found by an HHS evaluation, which were later addressed in legislation. In response to some of the grant management problems identified, agency officials described increasing their oversight and providing grantees with intensive training and technical assistance in fiscal year 2000.

Challenges to Collecting Outcome Data

Several of the studies we reviewed were initiated to address concerns about the quality or availability of outcome data. Some agencies faced considerable challenges in obtaining outcome information. Some states and service providers had limited data collection capabilities or incompatible data systems, while federal officials reported pressures to reduce data collection costs and the burden on service providers.

SAMHSA and the states have been working together for several years to develop common state data on the effectiveness of substance abuse treatment programs funded by the SAPT block grant. In 1995, HHS requested that the National Research Council convene a panel to report on

the technical issues involved in establishing performance measures in 10 substantive public health areas, including substance abuse treatment, to support a proposed Performance Partnership Grants program. The expert panel concluded that few data sources were available that would effectively support the development of performance monitoring systems because data were not comparable across the states. Therefore, the panel recommended that HHS assist states in standardizing both health outcome measures and methods for collecting data.

SAMHSA subsequently created the TOPPS II collaborative partnership program with the states to further performance measurement development through obtaining consensus on and pilot-testing treatment outcome measures. SAMHSA officials indicated to us that the greatest barriers to obtaining outcome data were poor infrastructure for data collection in some states (funding, people, software, and hardware), lack of standardized definitions and training to use them, and lack of buy-in from the treatment providers who are the original source of the data. Agency officials suggested that states are more likely to get buy-in from treatment providers if they consider them as partners and share the data on client results as useful feedback to help providers modify their own programs.

To obtain outcome data on its GI Bill educational benefits program, VA conducted an impact evaluation that was also used to help understand program use and operations. VA recognized that the program's performance goal—increasing usage of the education benefits—provided little information about the Department's strategic goal of assisting veterans to achieve their educational and career goals. Because the program is one of the VA's major benefits to veterans and a Department of Defense recruiting incentive, VA officials said they needed to better understand what influenced veterans' use of the benefit as well as its effectiveness. They said that understanding the program's efficacy is also important to strategic planning on how to respond to changes in the veteran population and their educational needs. The study integrated an assessment of program administration and effectiveness and might lead to program design changes, such as increasing the tuition benefit level.

VA officials stated that the extensive resources involved in obtaining primary data posed a challenge to collecting outcome data, noting that it was expensive and time-consuming to track, locate, and interview eligible program participants. They said that they could not conduct an evaluation like this annually, but could use this study to provide baseline data and identify performance measures for use in the future, when they expect to

augment their current process-oriented measures with more outcome-oriented ones.

The evaluations of the C/MHCs are part of a multiyear effort to obtain improved performance data for GPRA reporting. Officials noted that they attempt to balance their need to have complete and useful information for performance monitoring with the importance of minimizing reporting burden on grantees. The agency described a three-part strategy to improve program data while not overburdening grantees.

First, HRSA created a uniform data system to collect annual aggregate administrative, demographic, financial, and utilization data from each funded organization. Second, it fielded sample surveys of center users and visits in 1995 to obtain data on patient care. These are parallel to two recurring national surveys of the general population that HHS used to set the Healthy People 2000 and 2010 objectives. A comparable survey of center users and visits is being fielded in 2000. Third, HRSA funded evaluations that analyze previously collected research data to compare center users with similar populations of nonusers to assess performance goals related to reducing disparities in access to care.

In addition, HRSA plans collaborative arrangements with a limited number of centers to conduct focused studies on selected diseases. While the agency might use this last type of information to assess health status improvements, officials said that it would primarily be used by provider sites to document quality of care improvements.

Even when an agency has performance data, assessing the accuracy, completeness, and consistency of those data is important to ensuring their credibility.² OSHA initiated a formal data validation process soon after developing a new source of performance data. In 1995, OSHA implemented a system to gather and compile occupational injury and illness information from employers for use in both targeting its enforcement activities and measuring its effectiveness. In 1997, audits of employer recordkeeping were instituted to ensure the accuracy of the data for both of those uses. Concern was expressed that employers might underreport injuries or lost workdays if they believed that those reports might lead them to be targeted for enforcement. OSHA officials told us that the Office of Management and Budget required OSHA, as part of the agency's request for permission to collect this information from employers, to assess the

²Performance Plans: Selected Approaches for Verification and Validation of Agency Performance Information (GAO/GGD-99-139, July 30, 1999).

quality of these data each year that it collects them. From the findings of these reviews, OSHA has made improvements to its review protocol, piloted an automated assessment of records to streamline the review process, and revised the recordkeeping regulation to help improve the quality of the records. Additional audit improvements and outreach efforts are expected to further improve record quality.

Agency Capability to Gather and Use Performance Information

Over the last several years, we have noted that, governmentwide, agencies' capability to gather and use performance information has posed a persistent challenge to making GPRA fully effective. Our reviews of agencies' performance plans for fiscal years 1999 and 2000 found that the plans provided limited confidence in the credibility of their performance information. Agencies provided little attention to ensuring that performance data would be sufficiently timely, complete, accurate, useful, and consistent.³ In our governmentwide review of agencies' 1997 strategic plans, we found that many did not discuss how they planned to use program evaluations in the future to assess progress toward achieving their goals.⁴ More recently, in anticipation of the required updating in 2000 of agencies' strategic plans, we noted our continued concern that many agencies lack the capacity to undertake the program evaluations that are often needed to assess a federal program's contributions to results where other influences may be at work.⁵

In the early stages of GPRA implementation, we reported that agencies' evaluation resources would be challenged to meet the increasing demand for program results under GPRA.⁶ Across the government, agencies reported devoting relatively small amounts of resources to evaluating program results in 1995 and making infrequent efforts to extend their resources by training others. However, some federal evaluation officials described efforts to leverage their evaluation resources through

- adapting existing information systems to yield data on program results,
- broadening the range of their work to include less rigorous and less expensive methods,

³Managing for Results: Opportunities for Continued Improvements in Agencies' Performance Plans (GAO/GGD/AIMD-99-215, July 20, 1999).

⁴Managing for Results: Agencies' Performance Plans Can Help Address Strategic Planning Challenges (GAO/GGD-98-44, Jan. 30, 1998).

⁵Managing for Results: Continuing Challenges to Effective GPRA Implementation (GAO/T-GGD-00-178, July 20, 2000).

⁶Program Evaluation: Agencies Challenged by New Demand for Information on Program Results (GAO/GGD-98-53, Apr. 24, 1998).

- devolving program evaluation to federal (or state and local) program managers, and
- developing partnerships with others to integrate the varied forms of performance information on their programs.

The agencies discussed in this report demonstrated evaluation capabilities of their own as well as the ability to leverage federal and nonfederal evaluation resources to improve understanding of program performance. All the agencies described in this report had prior experience and resources for conducting program evaluations. However, these agencies also provided examples of ways to leverage resources through

- drawing on the findings of a wide variety of evaluations and audits,
- putting the findings of complex evaluations to multiple uses,
- mining existing databases, and
- collaborating with state and local partners to develop mutually useful performance data.

Observations

The agencies whose evaluations we studied demonstrated creative ways of integrating the results of different forms of program assessment to deepen understanding of how well their programs were working. Program evaluations allowed these agencies to demonstrate broader impacts than were measured annually, as well as to explain the reasons for observed performance. In those agencies where outcome measurement was in the beginning stages, evaluations helped them to explore how best to measure program performance. These agencies' experiences provide examples of how program evaluations can contribute to more useful and informative performance reports through assisting program managers in developing valid and reliable performance reporting and filling gaps in needed program information, such as establishing program impact and reasons for observed performance and addressing policy questions that extend beyond or across program borders.

Several agencies have used GPRA's emphasis on reporting outcomes to initiate or energize their efforts to measure program outcomes, while others made no reference to evaluation in their performance reports. We continue to be concerned that some agencies may lack the capability to undertake program evaluations, and we believe it is important that the updated strategic plans contain fuller discussions of how agencies are using program evaluations. As agencies update their strategic and performance plans, the examples in this report might help them identify how evaluations can contribute to improving understanding of their programs' performance.

Agency Comments

The Departments of Health and Human Services and Veterans Affairs provided written comments that are reprinted in appendixes I and II. The other agencies either had no comments or provided technical comments that we incorporated where appropriate throughout the text. HHS said the report accurately reflects its approaches to link evaluation studies with performance measurement and believes that it will be helpful to agencies in coordinating their performance measurement and program evaluation activities. VA suggested that we note that the extensive resources involved in collecting primary data posed a challenge to collecting outcome data, and we have done so.

We are sending copies of this report to Senators Tom Harkin, Ernest F. Hollings, James M. Jeffords, Edward M. Kennedy, Joseph I. Lieberman, Richard G. Lugar, John McCain, John D. Rockefeller IV, and Arlen Specter; and to Representatives Thomas J. Bliley, Jr., William L. Clay, Larry Combest, John D. Dingell, Lane Evans, William F. Goodling, James L. Oberstar, Bud Shuster, Charles W. Stenholm, and Bob Stump in their capacity as Chairman or Ranking Minority Member of Senate and House authorizing or oversight committees.

We are also sending copies of this report to the Honorable Daniel R. Glickman, Secretary of Agriculture; the Honorable Hershel W. Gober, Acting Secretary of Veterans Affairs; the Honorable Alexis M. Herman, Secretary of Labor; the Honorable Donna E. Shalala, Secretary of Health and Human Services; the Honorable Rodney E. Slater, Secretary of Transportation; the Honorable Richard W. Riley, Secretary of Education; and the Honorable Jacob J. Lew, Director, Office of Management and Budget. We will also make copies available to others on request.

If you have any questions concerning this report, please call me or Stephanie Shipman at (202) 512-2700. Elaine Vaurio made key contributions to this report.



Nancy Kingsbury
Assistant Comptroller General
General Government Division

Contents

Letter	1
Appendix I Comments From the Department of Health and Human Services	24
Appendix II Comments From the Department of Veterans Affairs	26
Related GAO Products	28

Abbreviations

APHIS	Animal and Plant Health Inspection Service
C/MHC	Community and Migrant Health Center
DOL	Department of Labor
DOT	Department of Transportation
GPRA	Government Performance and Results Act of 1993
HHS	Department of Health and Human Services
HRSA	Health Resources and Services Administration
OIG	Office of the Inspector General
OSHA	Occupational Safety and Health Administration
RSPA	Research and Special Programs Administration
SAMHSA	Substance Abuse and Mental Health Services Administration
SAPT	Substance Abuse Prevention and Treatment
TANF	Temporary Assistance for Needy Families
TOPPS II	Treatment Outcomes and Performance Pilot Studies Enhancement
USDA	United States Department of Agriculture
VA	Department of Veterans Affairs

Comments From the Department of Health and Human Services



DEPARTMENT OF HEALTH & HUMAN SERVICES

Office of Inspector General

Washington, D.C. 20201

SEP 18 2000

Ms. Nancy Kingsbury
Assistant Comptroller General
United States General
Accounting Office
Washington, D.C. 20548

Dear Ms. Kingsbury:

The Department of Health and Human Services appreciates the opportunity to comment on the General Accounting Office's (GAO) draft report, "Program Evaluation: Studies Helped Agencies Measure or Explain Program Performance" before its publication.

We believe that GAO's draft report accurately reflects discussions with Department staff as well as the approaches being used to link evaluation studies with performance measurement. We believe GAO's report will be helpful to agencies in explaining the relationship between performance measurement and program evaluation, how to leverage evaluation resources, and the need to develop a clear linkage between strategic plans and evaluation planning.

These comments represent the tentative position of the Department and are subject to reevaluation when the final version of this report is received.

Appendix I
Comments From the Department of Health and Human Services

Page 2 - Ms. Nancy Kingsbury

In addition, we provided some technical comments directly to your staff.

Sincerely,



for June Gibbs Brown
Inspector General

The Office of Inspector General (OIG) is transmitting the Department's response to this draft report in our capacity as the Department's designated focal point and coordinator for General Accounting Office reports. The OIG has not conducted an independent assessment of these comments and therefore expresses no opinion on them.

Comments From the Department of Veterans Affairs



THE SECRETARY OF VETERANS AFFAIRS

WASHINGTON

SEP 12 2000

Ms. Nancy Kingsbury
Assistant Comptroller General
General Government Division
U. S. General Accounting Office
441 G Street, NW
Washington, DC 20420

Dear Ms. Kingsbury:

We have reviewed your draft report, ***PROGRAM EVALUATION: Studies Helped Agencies Measure or Explain Program Performance*** (GAO/GGD-00-XX) and agree with your observation that program evaluations can offer a variety of benefits. We suggest that, in the "Challenges to Collecting Outcome Data" section, GAO make reference to the resource factors involved in collecting primary data. It is both expensive and time consuming to track, locate, and interview eligible program participants. The paperwork reduction process (Federal Register notices/OMB approval) takes a minimum of 90 days. The sampling, locating, and interviewing of eligible program participants requires several months.

We have passed separately several technical corrections to your office. I appreciate the opportunity to comment on your draft report.

Sincerely,

A handwritten signature in cursive script, appearing to read "Hershel W. Gober".

Hershel W. Gober
Acting

Related GAO Products

Hazardous Materials Training: DOT and Private Sector Initiatives Generally Complement Each Other (GAO/RCED-00-190, July 31, 2000).

Managing for Results: Continuing Challenges to Effective GPRA Implementation (GAO/T-GGD-00-178, July 20, 2000).

Community Health Centers: Adapting to Changing Health Care Environment Key to Continued Success (GAO/HEHS-00-39, Mar. 10, 2000).

Drug Abuse Treatment: Efforts Underway to Determine Effectiveness of State Programs (GAO/HEHS-00-50, Feb. 15, 2000).

Performance Plans: Selected Approaches for Verification and Validation of Agency Performance Information (GAO/GGD-99-139, July 30, 1999).

Managing for Results: Opportunities for Continued Improvements in Agencies' Performance Plans (GAO/GGD/AIMD-99-215, July 20, 1999).

Managing for Results: Measuring Program Results That Are Under Limited Federal Control (GAO/GGD-99-16, Dec. 11, 1998).

Managing for Results: An Agenda to Improve the Usefulness of Agencies' Annual Performance Plans (GAO/GGD/AIMD-98-228, Sept. 8, 1998).

Grant Programs: Design Features Shape Flexibility, Accountability, and Performance Information (GAO/GGD-98-137, June 22, 1998).

Program Evaluation: Agencies Challenged by New Demand for Information on Program Results (GAO/GGD-98-53, Apr. 24, 1998).

Performance Measurement and Evaluation: Definitions and Relationships (GAO/GGD-98-26, April 1998).

Managing for Results: Agencies' Annual Performance Plans Can Help Address Strategic Planning Challenges (GAO/GGD-98-44, Jan. 30, 1998).

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

Ordering Copies of GAO Reports

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.

Order 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 touch-tone phone. A recorded menu will provide information on how to obtain these lists.

Viewing GAO Reports on the Internet

For information on how to access GAO reports on the INTERNET, send 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

Reporting Fraud, Waste, and Abuse in Federal Programs

To contact GAO FraudNET use:

Web site: <http://www.gao.gov/fraudnet/fraudnet.htm>

E-Mail: fraudnet@gao.gov

Telephone: 1-800-424-5454 (automated answering system)

**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

