RESEARCH REPORT

Evaluation Guide for Public Service Program Managers

Harry P. Hatry
December 2020

URBAN INSTITUTE · E L E V A T E · T H E · D E B A T E

Contents

Acknowledgments

Evaluation Guide for Public Service Program Managers

This report[1] seeks to help program managers strengthen their decisionmaking using evaluation information. It describes how managers can get evaluation information and use it to help them continually improve the effectiveness and efficiency of their services. This report's suggestions focus on human services program managers; however, most suggestions apply to managers of other public services as well.

Introduction

Managers of state and local human services agencies as well as nonprofit organizations make dozens of program decisions throughout the year. Managers need varied information to make those decisions, especially evidence on how effective and efficient their services have been—answering the following questions: how are we doing? Are we getting the results we want? Are we making progress in improving the lives of our citizens? Obtaining such information is the main purpose of evaluation.

"Evaluation" has many interpretations. As used in this report, the word covers any approach that seeks to address questions of how a program is doing.

This report first discusses program managers' roles in evaluation and then suggests actions they can take to more effectively use evaluation. The suggestions focus on low-cost actions that provide timely information. This report only addresses procedures that public service agencies can use without requiring advanced statistical procedures or highly trained analytical staff.

We focus on three major forms of evaluation.[2] Each provides evidence on the extent to which a program works or not. Each accomplishes a different task:

- performance measurement systems, especially outcome measurement;

- in-depth program evaluations; and

- process (Implementation) evaluations.

The COVID-19 pandemic provides a major example of these differences. Daily, the media use *performance measurement systems* to get critical data on key outcomes such as the numbers of reported

cases and deaths. Data are also regularly calculated on important characteristics of these incidents, such as location (e.g., state or county), age group, and race/ethnicity. These data may not be as rigorous as desirable, but they provide approximations for tracking progress in slowing down the pandemic, as well as rough indications of the outcomes of treatment efforts. At the same time, numerous efforts are underway to develop vaccines. These efforts use *in-depth evaluations* that provide key information on the outcomes of each tested vaccine. Rigorous, scientific evaluation procedures are used to determine which procedures and vaccines are working and to what extent. Meanwhile, *process evaluations* assess how closely treatment procedures have been followed (e.g., whether health workers have been sticking to protocols for treating the disease). Such evaluations help identify needed corrections.

Many, if not most, human services agencies already participate in one or more form of evaluation. However, technology has advanced rapidly in recent years. Human services agencies can now do things with their evaluation systems, often at a low cost, that have not been practical in the past. This report focuses on these developments and how program managers can take advantage of new technology to improve their services for their participants.

Program Managers' Roles in Performance Measurement Systems

Performance measurement systems provide data that tell what outcomes and outputs have occurred and how these have changed over time. Performance measurement systems can help managers address many operational issues throughout the year. These systems primarily are known for providing performance data at regular intervals. Now these systems can provide the latest data related to performance issues *any time*.

New technology enables program managers at any level of the organization—from top-level management to frontline supervisors—to use performance measurement systems. It also enables an organization to add sets of outcome indicators as new issues arise. The pandemic provides an example: at first, the only indicators tracked and reported were numbers of cases and deaths. Now, performance measurement systems are adding "intermediate" outcomes, such as numbers of tests conducted and hospitalizations. This flexibility in performance measurement reporting offers program managers new opportunities to track outcomes. However, such flexibility comes at a cost—though hopefully low—for changing the system and, of course, collecting the needed data, if they are not already available.

Another major advance in information technology has enabled program managers to readily obtain data breaking out outcome indicators by key characteristics of the population served by the program, such as gender, age, race/ethnicity, household income, location, and disability status. Calculating outcomes broken out by demographic characteristics was seldom undertaken in the past century, in part because of the time and effort needed to compute these values. *Comparisons can now be made among such characteristics to help program managers tailor their activities to the participant groups served. And the calculations can be provided to program managers much more quickly.* The need for outcome breakouts has become increasingly important because of the urgent national issues of racial equity and COVID-19. These issues are just as relevant at the state and local levels and to most public services and programs provided by those governments and nonprofit organizations.

However, performance measurement systems are not designed to identify the extent to which outcomes can be *attributed* to the program or service. Nor do performance measurement systems tell *why* the outcomes were positive or negative. They only tell *which* outcomes were positive or negative and to what extent.

Program managers are increasingly responsible for maintaining performance measurement systems. Outside experts are often used to help design the measurement process. The cost of these systems is seldom explicitly budgeted, with the time and extra costs covered by overhead funds. The major added cost for maintaining the system is that of collecting data for performance indicators not already collected.

The manager's roles in performance measurement systems are summarized below:

a. Make sure the service to be evaluated has an outcome-oriented mission or vision statement that can be used as the starting point for selecting performance indicators.

b. Review the performance indicators periodically to assure they continue covering the program's mission.

c. Decide for which participant *demographic characteristics* outcome data should be broken out (e.g., age group, gender, race/ethnicity, location, disability status).

d. Decide for which *service characteristics* individual outcome indicators should be broken out (e.g., each service location and mode of delivery).

e. Oversee development of a data quality control process to assure the data produced will be as accurate as possible.

f. Assure the data collection and reporting processes do not violate participant privacy requirements.

g. Arrange for analysis and interpretation of the data, such as comparisons of the latest performance data with previous-year values, targets set by the program, and comparisons among participant demographic population groups.

h. Arrange for the clear and accurate display of findings for the manager and other intended users (i.e., take advantage of new data visualization capabilities, such as mapping, in making data displays more attractive and useful).

i. Identify to whom the findings of each performance report are to be provided and arrange for dissemination.

j. Use the information, as appropriate, throughout the year to improve service outcomes and peoples' lives.

k. Periodically review the performance measurement process, improving it as necessary.

Program Managers' Roles in In-Depth Program Evaluations

An in-depth program evaluation is an ad hoc procedure, undertaken when needed. These evaluations provide much more thorough and rigorous information on program outcomes than data from performance measurement systems. In-depth evaluations seek to identify the extent to which outcomes can be attributed to the program. (Such studies are often called "impact evaluations.") These studies are intended to provide rigorous evidence of the extent to which the program or policy is working. Some evaluations seek to identify best practices that can be adapted to multiple locations in the country.

However, such evaluations are costly; even large programs can only afford to conduct these evaluations infrequently. (The outcome data from the program's performance measurement system is likely to be useful for the in-depth program evaluation, reducing the evaluation's cost.)

Program managers' roles in in-depth evaluations are somewhat limited, especially for managers of small programs, and summarized below:

a. **Periodically, perhaps annually, review the need for such evaluations.** Box 1 lists criteria for selecting programs for in-depth evaluation.

b. **Seek funding for evaluations of one or more of the manager's own programs.** Many in-depth program evaluations have been funded by federal agencies. However, foundations and other philanthropic funders are becoming more interested in supporting such evaluations. If the program has funds to support a full in-depth evaluation, the program manager will likely need to have it done by external evaluators. It is good practice to have a formal evaluation done by an independent, usually external, researcher.

c. **Cooperate with the evaluation team in designing and implementing the evaluation**, such as by providing data or other information the team needs. This applies whether the in-depth evaluation is the program manager's choice or an outside requirement;

d. **Review the evaluation plan.** Assure the evaluation considers the program's needs. For example, does the evaluation plan consider appropriate outcome indicators, data collection procedures, participant groups for whom outcome data will be examined, and appropriate confidentiality requirements? This also applies whether the in-depth evaluation is the program manager's choice or an outside requirement. Also be sure the evaluation specifications require the final report to identify and discuss the implications of the finding's limitations.

e. **Specify whether the evaluation report should contain recommendations for service improvements** based on the findings of the program's impact. Evaluators are likely to get extensive information about potential program improvements, which can add considerable value to the evaluation.

f. **Ask staff to review external evaluations that may be relevant to the program.** Program managers should look out for findings from completed external evaluations that might provide information that could improve their own program.

g. **Consider less rigorous forms of evaluation that do not require specialized staff or high costs.** Unfortunately, as noted above, in-depth program evaluations are often expensive, require specially trained people to do them, and usually require considerable time before the evaluation findings are available to the program manager. In any year, only a small percentage of an agency's programs can be evaluated in depth. Few state and local government agencies or nonprofit organizations have staff to do these in-depth evaluations in-house.

h. Below are three low-cost options, listed from simplest to most complex. However, the simpler the process, the weaker the evidence will be about whether the program examined caused the outcomes. The program manager might choose one or more of these procedures:

» Compare the outcomes of new service delivery procedures from a period before changes were made with the outcomes for a period afterward, indicating whether any difference in outcomes is because of the program examined. Ask staff to look for other explanations.

» Compare the outcomes for participants who received the new service delivery procedures with the outcomes of participants who did not receive them—with little or no attempt to assure the groups are comparable. For example, the groups may differ substantially in the percentage of participants with difficult, hard-to-help problems.

» Use Mini-RCTs (randomized controlled trials). These are simplified versions of RCTs. This approach can be used if the service delivery is not complex; the outcomes sought are not expected for a long time, such as more than one year; and the outcomes are not difficult to measure. This new approach is discussed further in the last section of this report.

Program Managers' Roles in Process/Implementation Evaluations

Process evaluations are in-depth examinations of whether a program has been implemented as intended. Such information can be invaluable in indicating why program outcomes are not as expected and identifying needed changes. Process evaluations do not seek to identify program outcomes, and they are much less complex than in-depth program evaluations.

The need for a process evaluation might be triggered when data from either a performance measurement system or an in-depth evaluation shows disappointing outcomes from a program. The purpose is to help the program managers identify what went wrong and what worked, and to the extent possible what changes are needed to improve the outcomes.

BOX 1

Criteria for Selecting Programs for Evaluation

Potential impact of the program's service

▪ Expected number of people affected by the program

- Number of people whose lives are expected to be substantially improved or worsened by the program

- Expected effects of the program on improving diversity, equity, and inclusion (DEI)

- Size and cost of the program

Feasibility of evaluation for the program

- Expected cost of the evaluation

- Expected success of the evaluation, considering challenges such as difficulty in obtaining key data

- Political and citizen interest in the program

Suggestions for Program Managers

Recent advances in information technology—accompanied by state and local governments' and nonprofit organizations' greater understanding of how to use evaluation data—have led to human services programs more often using evaluation tools to strengthen the effectiveness and efficiency of their services. Below we provide more detailed suggestions for program managers to improve their collection and use of evaluation information. These suggestions are also listed together in the appendix.

Even the smallest agencies can use simplified versions of these suggestions to track key performance indicators and use evaluation information to continuously improve their services for their participants.

We group these suggestions into two stages applicable to human services evaluation efforts:

- getting the right data

- analyzing, disseminating, and using the data

This report only addresses procedures that public service agencies can use without advanced statistical procedures or highly trained analytical staff. These suggestions do not cover all evaluation approaches. Rather, they are selected suggestions for program managers that reflect new data capabilities from recent information technology advances. Before these advances, implementing these suggestions was expensive and time-consuming, making obtaining some performance data infeasible for many program managers.

Suggestions for Getting the Right Data

1. Seek input from representatives of key stakeholder groups to help identify the relevant outcomes and outcome indicators to track regularly.

Do not rely solely on agency staff to identify what outcomes and outcome indicators are needed to evaluate progress in achieving program goals and objectives. What is important to participants and other people or organizations affected by a program should also matter to program managers. Seek input from representatives of key stakeholder groups, such as the following:

- your participants and other members of the public expected to be affected by the program, including people such as inmates and substance abusers who are also participants;

- interest groups with significant subject matter interests in the service or program, such as environmental advocacy groups and particular demographic groups (e.g., those representing older adults, children, Hispanic/Latinx groups, etc.);

- business representatives;

- choices of outcomes and outcome indicators made by other agencies providing similar program services; and

- staff, particularly program field staff who are likely aware of program participants' concerns.

The pandemic has made it difficult to use in-person sessions, such as focus groups and town halls, to get this information. However, virtual sessions and social networking apps can provide reasonable substitutes.

2. Get information about programs' effects on their participants. Survey participants using the latest information technology, such as inexpensive survey software, to obtain and tabulate outcome information from the people and organizations affected by your program.

Some information for evaluating progress in achieving desired outcomes can be obtained from agency records, particularly on the participant's progress while the participant is still receiving program services. For many—if not most—human services programs, extensive evaluation information needs to be obtained from program participants *af*ter they have exited. How can that be done at low cost? Today, those surveys are considerably more affordable, especially once the survey process has been installed.

For many in-depth program evaluations, surveying former participants has been and continues to be a major way to get this information. For some programs, agency records can provide the follow-up information; for example, state unemployment records track the employment status of former

participants in various employment program services. For many human services programs, such as substance abuse programs, evaluators have used surveys of former participants for their in-depth evaluations. Many key health outcome indicators are available to help track program outcomes.

Such follow-up has been expensive and can require great effort. Program managers understandably have not taken on these efforts to follow-up with former participants. However, we now have much better technology and increased pressure from funding sources (e.g., the federal government and foundations) to provide evidence that their programs work.

Program managers now have low-cost survey options to get outcome information from participants. Software is now readily available in many forms (e.g., Survey Monkey). Not only has the improved software made it easier to develop questionnaires and regularly survey participants on program outcomes, but also to tabulate survey responses. Electronic surveys are gaining widespread use. They can now reach most households with low incomes; many, if not most, households now have smart phones, including participants with low incomes. Participants can be surveyed by mail and electronically. Response rates will likely be as good as, and probably higher than, those currently achieved by national polls. For performance measurement systems, the survey questionnaire can be short, unlike questionnaires generally used for in-depth program evaluations.

The program manager will likely need help implementing this process. However, once the process has been implemented, the cost of regularly tracking former participants need not be a major cost. To install a regular, ongoing survey process, the program manager is likely to need help with (1) developing the survey questionnaire; and (2) programming the computer to enter, tabulate, and display the data. Another important element is making sure appropriate procedures are in place to assure respondent confidentiality. Once the system has been set up, the annual cost of maintaining the process should be low.

3. Break out outcome data by important participant demographic characteristics, such as gender, race/ethnicity, age group, household income, location, and disability status.

Program managers' ready access to such outcome data breakouts is a major breakthrough. Recent information technology software and hardware data-processing breakthroughs have made this procedure practical and accessible. Such tabulations can now be routine, quick, accurate, and low cost. Similarly, program managers can now more easily set up the desired calculations. A major example is programs' increased ability to regularly survey their participants for evaluation information on their services—and tabulate participant responses. However, programs with large numbers of participants may need to hire a contractor to administer the surveys.

Program managers can then compare outcomes among participant groups—vital information for addressing diversity, equity, and inclusion issues. Providing outcome data by participant group demographics is a first step toward identifying where corrective actions are needed for underserved and other populations. After program changes have been implemented, outcome data provides evidence on whether problems have been alleviated. (Remember, however, that outcome data by itself provides little evidence on whether the program changes *caused* the improvements.)

Program managers play a major role in identifying participant characteristics that will help their programs address decisions such as how to distribute scarce budgeted resources.

4. Similarly, break out the outcome data by important service characteristics.

This enables the program manager to compare outcomes by these service characteristics. The program manager can ask staff to suggest service characteristics that vary and are likely to have different outcomes that should be compared. For example, some programs provide similar services at more than one location, and data could be calculated showing the outcomes for each site. Another example: when a program uses different service delivery approaches (such as different modes of communicating with participants), the outcomes for each approach can be measured and compared. An underused breakout is to disaggregate *outcome data by level of case difficulty or complexity*. (For health patients, the term risk level is sometimes used.) Calculating and reporting success rates grouped into a small number of difficulty levels (perhaps three or four levels) provides more meaningful and fair information. Examining outcomes by complexity also alleviates the temptation to focus on the easiest participants to claim higher success rates. The outcome indicator might look something like this: "Percent of young people *with multiple problems at entry* who improved after receiving the service." This indicator reports the findings for one participant difficulty level. Similar calculations would be made for each other difficulty level, as well as for the overall counts.

The program manager needs to (1) make sure each difficulty level is well defined; (2) oversee development of the procedure for assigning a level to each participant when the participant enters the service; and (3) arrange for the computer to link the difficulty level to the outcomes achieved for that participant.

Outcomes can also be calculated for *each service employee or each team of employees*, such as each case worker. However, judgments comparing employee productivity should not be made based on outcome data without also considering the difficulty levels of each employee's mix of participants.

These outcome indicator breakouts, both those for demographic and service characteristics, can provide powerful information for program managers. Recent computing technology makes such calculations quick and low cost.

5. Include post-service outcome indicators in your performance measurement system.

The biggest weakness of most government and nonprofit organization performance measurement systems is the lack of major indicators of post-service outcomes. This gap is present in many local, state, and federal performance measurement systems. A major share of human services programs seeks to improve citizens' behavior and quality of life after they have left the service.

Measuring outcomes solely when the participant exits from the service neglects the usually more important goal of improved conditions sustained after exit. Classic examples are smoking, drinking, and substance abuse cessation programs. Have former participants ceased their unhealthy behaviors, say at 12 months after exiting the service?

Again, recent advances in information technology make getting such information more practical. Two major approaches are used to get follow-up outcome data: (1) review existing data in agency records, such as records that enable recidivism, adoption rates, health status, or post-service employment status to be calculated (perhaps the available data is not exactly what is sought but is at least a reasonable proxy); and (2) survey former participants (see suggestion two above for more on such surveys).

For many human services programs, reasonable post-service outcomes cannot be found in agency records. Surveys of former participants have become a major way to get post-service outcome data. This approach is commonly used in in-depth program evaluations. New information technology now enables program managers to use affordable surveys more easily. In the smoking cessation program example, the program would establish the time for each former participant to be surveyed, such as 12 months after the participant's service exit. At this time, the program would send out an email with a short survey questionnaire to each participant, asking about their smoking since they left the program. The office could also give participants the choice of a mailed or emailed survey.

However, the program office needs to assure respondents that confidentiality will be preserved and the data will not be used in ways that identify the respondents without their permission.

This process of following up with former participants has additional possible benefits. It can be used to ask the respondent if they needed further help. And, as noted in the next suggestion below, a few

questions can be included in the questionnaire asking respondents to assess specific characteristics of the quality of the program's services.

6. Include in your performance measurement system indicators that capture the quality of care received from participants' perspectives.

One major concern gaining attention in performance measurement systems in recent years is the quality of services participants have received. Quality could include characteristics such as the service's

- timeliness,
- accessibility,
- friendliness,
- accuracy of information provided, and
- helpfulness.

In some cases, service quality data might be captured by existing agency data, such as recorded waiting times for assistance and tabulations from a government's complaint records. As noted above, participants can also be surveyed for their ratings on these characteristics. A few questions could be added to the questionnaire used to get evaluation data on former participants, as discussed above.

Even more useful, the survey questionnaire could end by asking respondents for their suggestions for program improvement, which could be highly useful to program managers.

7. Develop performance data collaborations across programs, agencies, and even sectors.

Program managers will, of course, focus on the data needed for their own program. However, many state and local programs face intersecting issues. Evaluation-related data may be needed from multiple programs and agencies. For example, reducing juvenile delinquency in a community is likely to involve the school system (e.g., absenteeism, disciplinary problems, and progress in school); health agency (e.g., health problem-incidence history); social services agencies (e.g., records of attempts to help and their outcomes); and the criminal justice system (e.g., number of interactions with the police and type of crime). Program managers may want to participate in such efforts, even taking initiative to form performance partnerships.

Information technology advances have only recently made such performance partnerships practical, especially those that rely on tracking individual participants through the whole system. This requires linking datasets across partners. For example, this procedure enables tracking outcome

indicators such as "Percentage of juveniles who had participated in a Department of Recreation summer camp program and with a remedial school education program, who had a reduced number of interactions with the police department over the following school year."

However, although the other suggestions in this report are likely low cost, developing this procedure of tracking individual participants across programs can be expensive. Tracking participants across agencies and/or programs can also raise major anonymity concerns, and the partner programs need to work out security and privacy arrangements. This procedure of linking datasets is used for some in-depth evaluations. Privacy issues in linking the datasets can be a major problem for both setting up an ongoing performance measurement process or undertaking a one-time in-depth program evaluation.

A less powerful, but easier, approach is for each partner to commit to at least providing grouped data relevant to the evaluation. The partner programs would together select the overall outcomes relevant to the issue they are addressing and identify the performance data needed from each partner. This strategy avoids needing data that can be linked to individual participants, when that is an issue. The findings would be limited to those that can be attributed to each partner.

An example is a recent effort in Cobb County, Georgia, headed by its Office of the Attorney General, aimed at strengthening the prosecution of sexual assault cases. It has developed a partnership with its police department and a nonprofit organization that provides support to victims of sexual assault. The county, as with the rest of the country, has a low rate of reported cases going to trial with a potential prison sentence (nationwide, only 5 out of every 1,000 perpetrators end up in prison)3 and an unknown number of cases that the victims did not report to the criminal justice system. The county has developed a management information system (MIS) to which each partner provides data links, enabling case tracking through the criminal justice system. The nonprofit organization guarantees anonymity to victims who do not want to report the case to the police department or prosecutor. However, the nonprofit organization provides data on the number of such cases and tallies of the reasons for nonreporting obtained by the nonprofit organization's caseworkers.[4]

8. Track efficiency but do not settle for output efficiency only. Also, track outcome efficiency.

Program managers are concerned not only with the outcomes of their services but also with their service delivery efficiency. A common way to measure efficiency has been to use performance indicators in the form of "number of output units per dollar spent." A truer measure of efficiency should consider the outcome achieved by the output. Defective outputs should not be counted in tabulations of outputs.

Similarly, program managers should consider including with their efficiency indicators not only an output-focused efficiency indicator (e.g., cost per participant served), but also an outcome-focused performance measurement (e.g., "cost per young person served who graduated on time"). The latter is a much truer measure of efficiency. Efficiency measurements should only include outputs that have at least achieved a minimum level of quality and outcomes. This will reduce the temptation to produce quantity at the expense of service quality and outcomes.

Suggestions for Analyzing, Displaying, and Using the Data

9. Assign staff to examine, summarize, and highlight the performance data.

Currently, few state and local program managers, and fewer nongovernmental organizations, have analytical staff who, as part of their regular job, examine, summarize, and highlight findings from performance reports. Large public service organizations are beginning to hire such employees. This gap appears to be narrowing as the information technology evolves. The need for such help, even in small agencies, has grown as the amount of data available to program managers has skyrocketed.

This analyst would review the regular performance reports (i.e., monthly or quarterly). Just as important, if not more so, they would examine the latest performance data any time during the year as issues arise that call for the program to examine performance.

Some program managers may want to do this examination themselves. Others (probably most program managers) will prefer to have other staff identify the highlights and warning signs indicated by the data.

The amount of performance data available to the program manager can be considerable, especially if, as suggested in this report, the performance data are presented not only as totals but also broken out by demographic and service characteristics. The analyst would be asked to examine and compare breakouts to identify such issues as: (1) inequities and (2) outliers (i.e., outcomes that are much better or much worse than others) that indicate the need for corrective action or might be used to identify best practices that could help elsewhere.

We note that increasing numbers of young people are exposed to data and data analytics. Identifying staff with the ability and interest to undertake the basic analytic work should be getting easier, especially as the next generation enters the workforce. For special in-depth analytic work, such as national complex in-depth evaluations, more specialized employees will still be needed.

10. Seek Information that can help explain unexpected, unusual performance levels.

After examining the data, a big question arises as to <u>why</u> undesirable performance levels occurred so corrections can be made. What can program managers do to get at least rough explanations for unexpected performance? It is usually better to be roughly right than precisely ignorant. Here are some ways to get explanatory information:

- Ask program staff, especially field staff, to share their views about why the unexpected outcomes occurred.

- Hold focus groups or similar discussion groups with participants or other stakeholders.

- Establish a working group to consider the causes.

- When surveying participants for ratings of various program service qualities (see suggestion six), ask respondents to share their reasons for any poor ratings.

- At the end of stakeholder survey questionnaires, ask respondents to share suggestions for improving the program services (such information can provide important leads for how to improve the service). If applicable, use a mini-RCT, as discussed in the next suggestion.

- If the program's service has become a major concern and funds are available, fund an external in-depth evaluation of that service.

11. Use mini-RCTs (randomized controlled trials) to test alternative service delivery procedures, providing more rigorous evidence on their outcomes.

The key limitation of performance measurement systems is that although they can provide data on outcomes, they offer little evidence about whether the program contributed to those outcomes.

To obtain stronger evidence, randomized controlled trials (RCTs) have been considered the gold standard approach. A key characteristic of RCTs is that participants are randomly placed into service recipient groups, with each group receiving a different service delivery approach. Generally, RCTs have required specialized professional help and can be quite expensive. Other advanced statistical tools are available for obtaining stronger evidence on causality, known as quasi-experimental designs. As noted earlier, this report only addresses procedures that public service agencies can use without advanced statistical procedures or highly trained analytical staff.

Even a small government agency or nonprofit organization can themselves use the principles of RCTs to get better information on what works without high added costs for the evaluation. These mini-RCTs, however, are not likely appropriate for: (1) complex issues; (2) issues whose outcomes are not

expected to show up for a long period; and (3) issues whose important outcomes are difficult to measure.

It is, however, appropriate for human services agencies and their program managers to use this type of analysis to evaluate many basic service delivery issues they regularly face, by incorporating randomization strategies in their operations. For example, a program manager might want to decide whether casework for a service would be more effectively delivered using group or one-on-one sessions, or if assistance could be provided electronically rather than in-person. The agency can randomly assign incoming participants to each approach. The outcomes for each group of participants are then calculated and compared to identify which procedure produced better outcomes.

Random assignment can be done simply now, using many free internet-based random assignment tools, which simulate older methods (e.g., flipping a coin for each incoming participant; drawing numbers out of a hat; using a random numbers table on a computer or handheld device; or assigning every other person or every third person to a different approach).

For example, several years ago, the Minnesota Department of Revenue tested a potential new procedure to collect owed taxes. It randomly selected 100 taxpayers with unpaid liability to be sent a letter presenting an opportunity to resolve their liability before being billed. It then randomly selected another 100 taxpayers with liability to which a letter was not sent. The department then compared the percentage of taxpayers in each group that paid their liabilities. Approximately 75 percent of those sent a letter paid their liability; only 25 percent of those not sent a letter paid (Liner et al. 2001).

Such a procedure provides particularly strong evidence on the relative effectiveness of each approach. This procedure is likely especially appealing to managers who like to innovate and try new service delivery approaches.

Box 2 suggests the basic steps necessary in undertaking these small-scale experiments. These mini-experiments are primarily applicable to alternative service delivery practices for which (1) the data on the important outcomes are available; (2) major outcome improvements can be expected to occur in a relatively short time frame, such as within a year; (3) the alternative service delivery approaches are not complex; and (4) ethical and confidentiality concerns are not violated. In other words, we suggest a simplified version of the in-depth evaluations needed for more complex program issues.

Testing Alternative Service Delivery Approaches

1. Identify the service approaches to compare. Typically, one would be the existing approach and the other a new one.

2. Choose a randomization method for deciding which incoming participants will be served by the new approach. The method should be one that selects a representative sample of participants for each approach. Randomization helps assure the comparisons will be valid and greatly increases the strength of the evidence. (For example, with random assignment, approximately the same share of difficult-to-help participants will likely be included in each group.)

3. As each participant enters the program, assign them to one of the two groups using the randomization procedure the program has identified.

4. Record which participants are assigned to which service approach.

5. Track the outcomes for each participant in each approach over whatever period the program believes is necessary to identify the outcomes of these approaches.

6. Tabulate the values on each outcome indicator for each approach.

7. Compare the findings and adjust program service practices, as appropriate. The program may want to drop the approach that shows the poorer outcome. Alternatively, it might decide that it has not yet obtained a clear enough picture from the outcome data to decide, in which case the program might continue the comparison for a longer time.

Source: Harry P. Hatry, Jake Cowan, and Michael Hendricks, *Analyzing Outcome Information: Getting the Most from Data* (Washington, DC: Urban Institute, 2004).

The mini-RCT option uses attributes of RCTs, providing program managers with much more and better information for program decisions than they had before recent technology advances.

12. Use the major advances in displaying data findings to communicate more effectively with program stakeholders.

Major developments in information technology have not only made it possible to get new performance data but also provide new and attractive ways to display the data. For example, mapping performance data has become widespread. The technology has spawned a new occupation: "data visualization specialist."

User-interaction software with data displays enables users to quickly change how the data are visualized. For example, now program managers and other interested parties can use the computers to

quickly display information such as previous years' performance levels and breakouts of outcomes for different population groups.

The available data presentation options enable program managers to get a better picture of important comparisons and better grasp the comparative data. And we now have vast options for color displays and various charting formats, such as the traditional bar charts, scatter diagrams, time trend displays, and flow charts. For program managers, graphic data displays are likely much more useful than presenting data in spreadsheets or even tables.

But don't overdo it. Resist the temptation to include too much data or too many colors in a data visualization.

Suggestions to Program Managers on Using Evaluation Information

13. Hold regular data-driven performance reviews with staff, using data from the latest performance measurement system report and relevant completed program evaluations as a starting point for such meetings.

Such "How Are We Doing?" sessions should encourage group discussion to answer the following questions: where do the data indicate problems exist in service delivery? Why are they occurring? What can we do to alleviate them? Can these learning events be adapted to our programs and services? (This suggestion is the same as the PerformanceSTAT process started by the New York City Police Department and picked up by many local and state governments, especially in the early 21st century.)
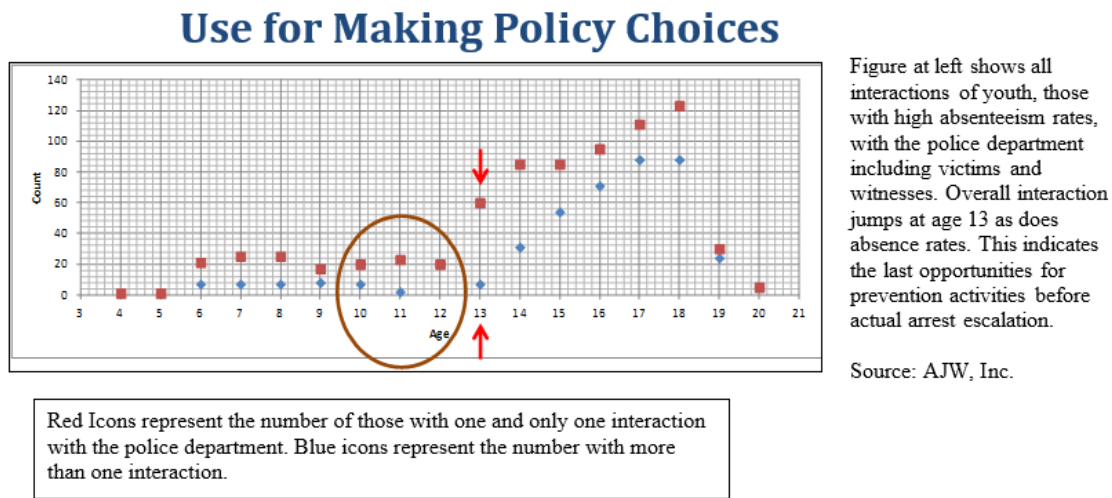
Avoid making these blaming sessions. Include discussion of the findings from any relevant in-depth program evaluations available.

14. Use the performance measurement system outcome data and information from relevant in-depth evaluations as a major basis for developing and justifying budget and policy program choices.

Figure 1 illustrates performance measurement data use, as part of a special, more in-depth, examination. In this case, it is used to inform program and policy issues related to youth delinquency and how to reduce it. The number of interactions with the police department (the outcome indicator) of young people who had high absenteeism levels was calculated for each age level. The calculations, as shown in the graph, make a strong case for intervention programs for young people at high risk before age 13. Data from both the school and criminal justice systems were needed. Social services programs also have a substantial interest because they are likely involved in intervention strategies.

FIGURE 1

Example of Performance Measurement Data Used to Inform Policy around Youth Delinquency

Use for Making Policy Choices

Figure at left shows all interactions of youth, those with high absenteeism rates, with the police department including victims and witnesses. Overall interaction jumps at age 13 as does absence rates. This indicates the last opportunities for prevention activities before actual arrest escalation.

Source: AJW, Inc.

Red Icons represent the number of those with one and only one interaction with the police department. Blue icons represent the number with more than one interaction.

Source: Cited in Harry P. Hatry, *Transforming Performance Measurement for the 21st Century* (Washington, DC: Urban Institute, 2014).
Notes: This figure was not developed by Urban researchers. The figure caption is not part of the original figure created by AJW, Inc., but contextualizes the image within this report.

15. Use performance information to motivate employees and contractors—but with caution.

Provide relevant performance information regularly to frontline staff—not only supervisors. Set program targets for each performance indicator and track and regularly report to staff on progress. Get input on targets from employees who contribute to them. However, emphasize that most performance indicators result from team, not individual, efforts.

Avoid the unintended effect of motivating employees to push for output quantity at the expense of quality, especially if target accomplishment is used for employee awards. For example, make sure that meeting service quality targets is also considered.

For contractors, consider including targets for performance indicators over which the contractor has a substantial role. Financial penalties, and bonuses, need to be set with care to avoid the unintended effect of motivating the contractor to push for output quantity at the expense of service quality. Much performance data would likely need to come from the program's performance measurement system.

16. Provide training, technical assistance, and/or mentorship to employees in accessing, interpreting, and using performance information. Bring in information Technology (IT) support to help.

Executing the ideas suggested in this report requires people who can implement them properly. This work is challenging but achievable. Details need to be worked out and training and technical assistance provided.

Knowing when you need help is important these days, especially because technology is rapidly changing. IT support will likely be necessary for the initial software programming, as will be training and the technical assistance to program staff for tasks such as data entry.

What These Three Evaluation Tools Do Not Cover

Evaluations are essential but have important limitations:

1. **They look backward.** They examine what happened in the past, but program managers make decisions about what should be done in the future. Information from these three tools provides starting points for estimating future levels of effectiveness and efficiency. These estimates will be valid to the extent that conditions are similar to those in the past.

    Estimation of future outcomes is much more complex, especially for programs for which some major outcomes occur far into the future.

2. **They do not examine costs of the programs evaluated.** They seldom provide program managers with data to address questions such as, "Are the outcomes worth the monetary costs?"

3. **These evaluation tools typically do not assess *alternative* ways to deliver a program's service.** Other analytic tools are available for comparing the costs and outcomes of alternative ways to deliver a program's services, such as cost-effectiveness and cost-benefit analyses. They have been around for decades but have not yet seen much use outside the federal government. Both tools can be considered natural extensions of in-depth evaluation. However, they expand considerably on in-depth program evaluation. They alleviate all three limitations identified here. How they can be adapted for program managers' use is not within the scope of this report.

Final Observation

Implementing the suggestions above is now feasible for most human services programs—whether their programs are small or large—thanks to advances in information technology. Program managers can utilize opportunities to adapt some attributes and capabilities of in-depth program evaluations to performance measurement systems. For example, linking participant outcomes to participant demographic and service characteristics as part of their performance measurement systems enables program managers to compare, when needed, service outcomes across different participant groups. Program managers can now regularly survey former participants to get basic outcome and service quality data at a low cost.

Program managers should consider seeking funds for in-depth evaluations when a service provided meets such criteria as those suggested in box 1. At this time, both foundations and the federal government are amenable to, and often require, in-depth evaluations, especially if the program manager can establish performance partnerships with other organizations.

Program managers should also consider using simplified versions of RCTs—Mini-RCTs—to test the outcomes of new or modified service delivery approaches.

Human services agencies still need in-depth program evaluations to address complex and expensive program and policy issues. New information technologies enable the processing and linking of large datasets to provide much stronger information on what is and is not working.

Program managers in state and local governments continually face evolving challenges with service delivery and need to know how well their procedures are working. These managers need timely, accurate information on their program outcomes. Regular performance measurement systems enable programs managers to obtain at least basic information on the outcomes levels of their programs but need added information about why the outcomes are at those levels. In-depth program evaluations can, when available, provide some of this information, and process evaluations can provide additional information about causes. However, perfect information will likely never be available. We need to accept this and hope it is usually better to be roughly right than precisely ignorant.

Appendix. Summary of Suggestions for Program Managers

Suggestions for Getting the Right Data

1. Seek input from representatives of key stakeholder groups to help identify the relevant outcomes and outcome indicators to track regularly.

2. Get information about programs' effects on their participants. Survey participants using the latest information technology, such as inexpensive survey software, to obtain and tabulate outcome information from the people and organizations affected by your program.

3. Break out outcome data by important participant demographic characteristics, such as gender, race/ethnicity, age group, household income, location, and disability status.

4. Similarly, break out the outcome data by important service characteristics.

5. Include *post-service outcome indicators* in your performance measurement system.

6. Include in your performance measurement system indicators that capture the *quality of care* received from participants' perspectives.

7. Develop performance data collaborations across programs, agencies, and even sectors.

8. Track efficiency but do not settle for output efficiency only. Also, track outcome efficiency.

Suggestions for Analyzing, Displaying, and Using the Data

9. Assign staff to examine, summarize, and highlight the performance data.

10. Seek information that can help explain unexpected, unusual performance levels.

11. Use mini-RCTs (randomized controlled trials) to test alternative service delivery procedures, providing more rigorous evidence on their outcomes.

12. Use the major advances in displaying data findings to communicate more effectively with program stakeholders.

Suggestions to Program Managers on Using Evaluation Information

13. Hold regular data-driven performance reviews with staff, using data from the latest performance measurement system report and relevant completed program evaluations as a starting point for such meetings.

14. Use the performance measurement system outcome data and information from relevant in-depth evaluations as a major basis for developing and justifying budget and policy program choices.

15. Use performance information to motivate employees and contractors—but with caution.

16. Provide training, technical assistance, and/or mentorship to employees in accessing, interpreting, and using performance information. Bring in information Technology (IT) support to help.

Notes

1 This report draws from Hatry 2013 and 2014.

2 This report does not address performance evaluations used for staff performance reviews, such as for salary reviews.

3 RAINN (Rape, Abuse & Incest National Network), accessed September 29, 2020, https://www.rainn.org/about-national-sexual-assault-telephone-hotline.

4 See Long et al. (2020) for the volume describing the full performance management process proposed for the local office of attorney general and its performance partnership with the police and nonprofit victim support organizations.

References

Hatry, Harry P. 2013. "Sorting the Relationships among Performance Measurement, Program Evaluation, and Performance Management." In *Performance Management and Evaluation: New Directions for Evaluation* 137, edited by Steffen B. Nielsen and David E. K. Hunter, 19–32. Washington, DC: American Evaluation Association.

———. 2014. *Transforming Performance Measurement for the 21st Century*. Washington, DC: Urban Institute.

Liner, E. Blaine, Harry P. Hatry, Elisa Vinson, Ryan Allen, Pat Dusenbury, Scott Bryant, and Ron Snell. 2001. *Making Results-Based State Government Work*. Washington, DC: Urban Institute.

Long, Jennifer, Holly Fuhrman, Harry Hatry, Janine Zweig, Elaine Borakove and Rey Cheatham Banks. 2020. *Model Response to Sexual Violence for Prosecutors (RSVP Model): Measuring the Impact—Volume II: Performance Management*. Washington, DC: AEquitas.

About the Author

Harry P. Hatry is a distinguished fellow at Urban Institute. He has been a leader in developing performance management/measurement and evaluation procedures for public agencies since the 1970s. He has worked with a wide range of local, state, and federal agencies—internationally and nationally—to develop outcome measurement procedures for such services as public safety, health, transportation, education, parks and recreation, social services, environmental protection, and economic development. He is a co-editor of the *Handbook of Practical Program Evaluation*, now in its fourth edition, and *Performance Measurement: Getting Results*, now in its second edition.

URBAN INSTITUTE · ELEVATE · THE · DEBATE

500 L'Enfant Plaza SW
Washington, DC 20024

www.urban.org