Analyzing Outcome Information
GETTING THE MOST FROM DATA

Analyzing Outcome Information

GETTING THE MOST FROM DATA

The Urban
Institute

Contents

Exhibits

Preface

Raw data that nonprofit organizations obtain from their outcome monitoring procedures, no matter how good, need to be processed and analyzed before the information can be useful to managers and staff. This guide, the fifth in the Urban Institute's series on outcome management for nonprofit organizations, describes steps that nonprofit organizations can take in performing this analysis.

This guide is unique in offering suggestions to nonprofits for analyzing regularly collected outcome data. The guide focuses on those basic analysis activities that nearly all programs, whether large or small, can do themselves. It offers straightforward, common-sense suggestions.

Probably the major concern, for at least some small organizations, is the computer capacity needed to tabulate the numbers that would otherwise be done manually. Fortunately, most groups today have some basic computer capability, and easy-to-use software is readily available to make the calculations. Any problems with programming can likely readily be solved through outside technical assistance, which is easy to secure in any community.

This guide does not deal with the more complex analysis procedures that involve sophisticated statistical or mathematical knowledge. Those procedures are likely to be most feasible only when resources are available for in-depth studies.

The steps described here should be of major help in interpreting the collected outcome data and in making the data useful for decisionmaking—the ultimate purpose of any outcome measurement process.

Gordon W. Green
Vice President, Research
Independent Sector

Acknowledgments

Introduction

Any organization with an outcome measurement system will quickly accumulate a variety of data. Before those raw data can be used to help improve services, they need to be converted into useable information. This process is called analysis. Analysis is not a mysterious, mystical activity. Analyzing data is a normal part of life for almost everyone. The tasks involved are remarkably similar, no matter who is doing the analysis or what is being analyzed.

Analysis of quantitative data includes adding, subtracting, multiplying, dividing, or other calculations. However, it is also much more. Analysis requires human judgment. The combination of calculations and judgment often produces the best analysis. Analysis is as much about thinking as it is about calculating.

This guide suggests ways to extract information from outcome data—through analysis—with the goal of using the analysis to help improve services for clients and to ensure better outcomes in the future. A separate guide in this series provides suggestions on how to effectively use the analyzed outcome information.

The steps presented here assume that the nonprofit organization has already selected outcome indicators and is collecting information regularly, as discussed in other guides.[1] Once outcome data have been collected, they need to be turned into useful information.

These analysis procedures can be used each time the outcome data become available, whether monthly, quarterly, annually, or at whatever reporting interval the program uses. Done regularly, the analysis will provide the organization and its programs with a steady stream of key information about clients and results.

Little technical background and only a basic knowledge of mathematics or statistics are required. More complex analytical procedures can provide more sophisticated and in-depth analyses, but these are beyond the scope of this guide. Such advanced procedures are likely more useful for special studies, or to help with large-scale program decisions, than for regular outcome-monitoring purposes. (The fourth section of this report briefly discusses some such special analyses that a nonprofit organization might attempt, when and if appropriate.)

The set of analysis steps described in this guide is listed in exhibit 1. These steps are recommended for each reporting period. An organization with multiple programs

---

[1] See the series on outcome management for nonprofit organizations published by the Urban Institute in 2003. Another excellent source of information is *Measuring Program Outcomes: A Practical Approach,* published in 1996 by the United Way of America, Alexandria, VA.

EXHIBIT 1

Steps for Analyzing
Program Outcome Data

Begin with the Basics

Step 1. Calculate overall outcomes for all clients

Step 2. Compare the latest overall outcomes with outcomes from previous time periods

Step 3. Compare the latest overall outcomes with pre-established targets

Step 4. Compare the latest overall outcomes with clients in other, similar programs—and to any outside standards

Delve Deeper into Client and Service Characteristics

Step 5. Break out and compare client outcomes by demographic group

Step 6. Break out and compare outcomes by service characteristics

Step 7. Compare the latest outcomes for each breakout group with the outcomes from previous reporting periods and with targets

Step 8. Examine findings across outcome indicators

Make Sense of the Numbers

Step 9. Identify which numbers should be highlighted

Step 10. Seek explanations and communicate the findings

should establish separate analysis procedures for each program. A computer (rather than staff) can be set up to do most of the computations. A nonprofit organization new to outcome management and the analysis of outcome data might choose to focus initially on only a few of the procedures described here. As the organization becomes more familiar with outcome data, more questions will arise about the outcomes measured, and additional analysis will become increasingly useful.

This guide is more detailed and technical than others in the series. However, most of these techniques are really common sense.

Nonprofit organizations provide a wide range of diverse services. Inevitably, differences will exist in how analysis procedures are applied, from program to program and from organization to organization. The basics, however, are likely to be applicable to most programs and most outcome indicators. The steps can be used if the program has 25 clients, hundreds of clients, or thousands of clients.

Who Should Perform the Analysis

Who performs the analysis will depend on resources and preferences. Here are some guidelines:

- Direct service providers (such as caseworkers) should not be given the added burden of data analysis. If the direct service provider is interested, however, s/he should be encouraged to examine the relevant outcome data. As discussed later, direct service providers should be a major information source for explanations of key findings and for help in interpreting the data.

- If resources are available, someone, full-time or part-time, can be assigned responsibility for much of the analysis work.

- Program managers are likely the most important people in the analysis process. They should take time to examine the findings, identify highlights and issues, and seek explanations for unexpected or disappointing outcomes.

- In very small organizations, one manager may assume the responsibility for examining and interpreting the outcome data. If the manager needs assistance with particular tasks, such as performing some of specific procedures described in this guide, the manager should seek help, perhaps from a volunteer, a local college or university, or a consultant.

About This Guidebook

This guide presents 10 basic steps to turning outcome measurement data into useful findings, procedures that should be used on a regular basis. The steps are grouped into three sections. A fourth section covers more in-depth analyses for use in special situations, and a fifth covers a few other points about analysis.

Begin with the Basics describes the first steps in assembling outcome measurement data into a format for analysis and completing some initial comparisons.

Delve Deeper into Client and Service Characteristics covers procedures to break out the results by client demographic groupings and by program type, and to review groups of indicators.

Make Sense of the Numbers includes identifying and highlighting the key issues that the examination uncovered and then seeking explanations for unexpected results.

Special Analyses Using Outcome Information briefly identifies some special procedures that may be useful in further mining outcome information.

Final Points about Analysis lists some limitations that analysis users should be aware of and summarizes the benefits of analysis.

Begin with the Basics

At this point, don't try to look too deeply into the past or future. Simply compute overall client performance and start to make some basic comparisons.

Outcome information is used to describe the percent of clients who did, said, or achieved something. Outcome reports—presenting data on the outcome indicators—are prepared for each reporting period, whether monthly, quarterly, annually, or whatever. *Managers should usually be provided outcome reports displaying both the number and percentage of clients.*[2] If only percentages are reported, users will not know whether the percentages are based on very small or large numbers of clients. If only numbers of clients are presented, it will be difficult to know whether the numbers represent small or large percentages of clients.

For some programs, and some of their outcome indicators, only the number of clients may be needed (for example, when examining counts of the number of home-less clients given meals or beds in shelters).

Overall performance is fairly straightforward to calculate. Unfortunately, overall performance by itself provides very limited information for identifying where improvements are needed. Section II of this guide suggests ways to examine the out-comes data in more detail.

Before describing the steps listed in exhibit 1, some basic calculation issues need to be discussed:

---

[2] Throughout this guide, we use the word "client" as a general term for whomever or whatever a program attempts to improve. For some programs, such as environmental programs, "client" may not be appropriate.

■ determining the relevant clients;

■ using averages;

■ deciding whether outcome indicators should be calculated as the extent of change, only clients' final condition, or both; and

■ addressing issues with using samples rather than obtaining outcome information on all clients.

Determining the Relevant Clients. For most outcome indicators in a single reporting period, outcome calculations, and subsequent analysis, will only need to cover a portion of the program's clients. If "percentage of participants who show improvement three months after completing the service" is the outcome indicator, then only clients that fit that criterion within the reporting period should be counted.

In addition, when outcome indicators call for percentages, it is important to identify clearly the population that should be included in the denominator of the percentage. In the preceding example, the percentage would be calculated by dividing the number of participants showing improvement by the total number of clients who completed the service, not by *all* the program's clients. Nor should the denominator include clients who had been out of the program for six months but had dropped out before completing the program.

Programs will also likely find it useful to calculate and report the number and rate of *dropouts*. For example, the number of dropouts would include those who dropped out before completing the program during the reporting period. It may also be appropriate to calculate the cumulative number and rate of dropouts from the year to date as well as the number and rate of dropouts for the quarter. When calculating the dropout rate, the denominator needs to be carefully defined. In this case, it would probably be the number of clients in the program at the beginning of the period for which the outcome indicator is reported.

Using Averages. For some outcome indicators, a program can calculate an average "score." For example, adding the individual test scores for each student and then dividing by the number of students tested will produce average scores on school tests for a program's client. Averages can even be calculated from verbal outcome information, such as that shown in exhibit 2a. This can be done by assigning a number to each outcome level, multiplying those numbers by the number of clients in the respective cells, and then dividing by the total number of clients. For example, in exhibit 2a, if the values 1, 2, 3, and 4 were assigned, respectively, to responses of very satisfied, somewhat satisfied, somewhat dissatisfied, and very dissatisfied, the overall average value would be 2.3.

While this approach has the advantage of summarizing all the data into one number, the average score, a few very high or very low scores can dramatically change the average. A disadvantage is that no information is available on how scores were distributed. In general, for analysis purposes, we recommend using averages only to supplement and summarize the data on the distribution of outcome levels, not

as a substitute for examining the distribution. Examining the number and percentage of clients who achieve, or report, each outcome level will provide considerably more information on the intensity and extent of outcome success.

Deciding Whether Outcomes Should Be Calculated as the Extent of Change, Only Clients' Final Condition, or Both. Programs that seek to improve the condition of their clients—such as programs addressing substance abuse and other risk behaviors—have the following basic options for tabulating and analyzing the outcomes:

1. Calculate the difference between each client's condition (or knowledge, behavior, or attitude) at intake and at the time of follow-up. Then tabulate the "change score." The outcome indicator would be expressed as the number, and percentage, of clients who have met some predetermined improvement criterion. For example, the *percentage of clients who had reduced their weekly alcohol consumption by at least X drinks.*

2. Examine the client's condition only at the time of follow-up, after the service has been provided. This provides an outcome indicator like the *percentage of clients who had not smoked in the past 30 days.* This option is needed if the outcome indicator is expressed in terms of attainment of a particular outcome level.

3. Examine both amount of change and condition of the client after services have been completed.

The first two options provide different perspectives on the outcome. The first procedure provides change information. However, some clients whose condition had improved may still not have reached desired levels even though their risk behavior had been reduced. If only client condition at the time of follow-up is considered (the second procedure), this could encourage programs to focus their attention on clients whose starting conditions were better in order to be able to report better final outcomes. Preferably, the program would use the third option, of using both versions. The first option requires the program to record each client's condition at, or near, intake and to later calculate the difference between each client's intake and follow-up condition. Thus, using the third option adds some data work over only using the second option.

Addressing Issues in Using Samples of Clients. For programs with large numbers of clients, it may not be feasible to obtain outcome data on all clients. Instead, a program can seek outcome information from a representative sample of clients. If so, the program needs to employ random-sampling procedures so that the sample is representative of the full population of clients served. To estimate the total number of clients that achieved the specified outcome for that reporting period, a program would multiply the percent of clients in the sample that achieved the outcome by the total number of participants.

Step 1: Calculate Overall Outcomes for All Clients

Tabulate outcomes for all clients in the program whose outcomes were measured during the reporting period. This is the natural first step. It provides a basic summary

overview of a program's results. The usual way to summarize the findings is to use tables (sometimes called "frequency tables.") Exhibits 2 and 3 are examples of such tables.

Tables reporting these aggregated outcome data can be created manually if necessary, especially if the number of clients is small. However, standard word processing or spreadsheet software can reduce the time and resources needed to make the calculations for these tables.

Exhibits 2 and 3 provide examples of displays of outcomes by both percentage and number of clients. Exhibit 2a shows the numbers and percentages of client responses for each of four levels of satisfaction with recreation opportunities. To simplify or summarize large amounts of data, several outcome levels can be combined. Exhibit 2b illustrates combining the four levels of satisfaction into two levels. This is often helpful for readers of outcome reports. However, the combined version does not have the rich information of the original. For example, it may be important to know how many clients were "somewhat dissatisfied" versus "very dissatisfied." If one of the extreme outcome levels has a particularly high percentage—such as the percent of clients that rate the helpfulness of a service very low—this important finding might be lost. In such cases, those important outliers should also be presented.

Usually, managers should examine the full distribution of responses. However, for reporting outside the program it may be sufficient, and preferable to outside users, to provide summarized information. First, examine the more detailed information and then decide whether combining outcome levels is useful, at least for some reporting purposes.

Exhibit 3 presents the outcome data for a number of separate outcome indicators. It shows the levels of improvement for youth mentored by volunteers of one Big Brothers Big Sisters program on 21 outcome characteristics. This type of table permits a quick assessment of where the service is doing well and where not so well.

The table focuses on percentages but also identifies the *total* number of clients for whom data were obtained. To avoid overcrowding the table, the number of clients is not shown in each cell. However, a report user can readily calculate the number of clients with a particular outcome level for any behavior by multiplying the number of responses by the percentage reported for the change level.

Step 2:   Compare the Latest Overall Outcomes with Outcomes from Previous Time Periods

Comparisons are the name of the game. They help provide the context in which organization officials and program managers can begin to interpret outcome information. This step and many of the following steps offer different types of comparisons. Comparisons provide program and organization officials with ways to interpret the outcome data. Each comparison provides a different "benchmark" against which the latest values for any outcome indicator can be measured.

EXHIBIT 2a

Sample Basic Outcome Table

Number and Percent of Youth Satisfied with the
Recreation Opportunities at Recreation Centers

Very Satisfied	Somewhat Satisfied	Somewhat Dissatisfied	Very Dissatisfied	Total
12% (*n* = 14)	53% (*n* = 60)	25% (*n* = 28)	10% (n = 12)	100% (n = 114)

Is action needed? A 35 percent level of dissatisfaction probably warrants attention. For example, further analysis could determine what types of clients were dissatisfied.

EXHIBIT 2b

Sample Outcome Table with Data Summarized across Rating Categories

Number and Percent of Youth Satisfied with the
Recreation Opportunities at Recreation Centers

Either Very or Somewhat Satisfied	Either Very or Somewhat Dissatisfied	Total
65% (*n* = 74)	35% (*n* = 40)	100% (*n* = 114)

EXHIBIT 3

Sample Detailed Outcome Report

Students That Reported Various Levels of Improvement on Outcomes of
Big Brother Big Sisters School-Based Mentoring Programs

	Number of Responses	Much Worse (%)	A Little Worse (%)	No Change (%)	A Little Better (%)	Much Better (%)
Confidence						
Self-confidence	6,929	1.1	3.8	21.0	51.1	23.0
Ability to express feelings	6,689	0.9	3.5	31.1	48.3	16.2
Decisionmaking ability	6,641	1.0	4.0	37.8	44.7	12.6
Has interests or hobbies	5,140	0.8	2.2	39.4	39.6	18.0
Personal hygiene, appearance	5,184	0.8	3.9	58.3	24.7	12.3
Sense of the future	5,054	1.0	3.0	43.4	36.9	15.7
Competence						
Use of community resources	3,318	0.6	1.7	53.1	31.4	13.1
Use of school resources	5,769	0.6	2.9	42.5	40.0	14.0
Academic performance	6,884	2.1	7.4	30.3	45.3	14.9
Attitude toward school	6,658	2.0	6.8	26.8	44.8	19.6
School preparedness	6,342	2.8	8.2	36.2	36.2	16.6
Class participation	6,313	1.5	4.8	34.7	41.8	17.2
Classroom behavior	5,914	2.3	8.8	35.0	37.5	16.4
Ability to avoid delinquency	2,869	2.2	5.8	45.0	32.7	14.3
Ability to avoid substance abuse	1,430	0.9	2.7	52.7	26.4	17.3
Ability to avoid early pregnancy	1,378	1.1	2.9	54.1	25.3	16.7
Caring						
Shows trust toward you	6,203	1.2	3.1	28.8	44.1	22.8
Respect for other cultures	4,592	1.0	3.8	46.1	33.0	16.0
Relationship with family	3,465	1.8	5.2	43.1	33.7	16.1
Relationship with peers	6,475	1.6	6.7	30.6	44.9	16.2
Relationship with other adults	5,779	1.7	4.5	31.5	44.8	17.5

Source: Adapted from a 2001 Big Brothers Big Sisters of America program-based outcome evaluation survey report on its school-based mentoring grant programs.

Comparing the latest outcome data with those for previous time periods is the most common comparison currently in use. This step can be taken once the program has collected more than one round of data for an outcome indicator. The comparison indicates the extent to which outcomes are getting better or worse. This comparison can also be used to indicate what effect past program changes have had.

Comparisons across time can have a number of variations, such as

■ the most recent year compared with the previous year

■ the last month or quarter compared with the previous month or quarter

■ the last month or quarter compared with the same period in the previous year

■ the cumulative performance thus far this year compared with the cumulative performance over the same period of time in the previous year

Exhibit 4 illustrates comparing the last quarter of one year with the same quarter in the previous year. The exhibit indicates that for one indicator, the outcome was worse than the previous year. For two other indicators, the outcomes were better than those of the previous year. The exhibit also identifies possible actions that the program might take based on the findings.

After the program has outcome data for more than one past year, it can examine trends—that is, data from more than one year on each outcome indicator. (Note, however, that any major changes to measurement indicators or data collection procedures may make comparisons over time misleading or inappropriate.)

Step 3: Compare the Latest Overall Outcomes with Pre-established Targets

Programs that regularly collect outcome data are also likely to establish annual targets for each indicator. This comparison will indicate the extent to which the program is meeting, exceeding, or falling behind expectations.

Programs generally should *not* set targets for their outcomes until at least one round of data has been collected. Without baseline data, a sound basis doesn't exist for establishing realistic targets. Once the targets are set, outcome data can then be used to compare actual performance against the targets.

Programs should probably also establish monthly or quarterly targets (or whatever time period coincides with the program's reporting intervals) against which results can be compared. This will provide the outcome information in a more timely way, so that mid-course corrections can be made. This will also enable the program to consider seasonal factors it believes will affect outcomes.

Exhibit 5 illustrates comparisons to targets for a number of outcome indicators. The report includes a column that shows how much the actual values achieved differed from the targets. This information is easy to calculate and will help users to more easily interpret how well the program is doing.

EXHIBIT 4

Sample Comparison over Time: Animal Shelter

	Oct.–Dec. 2002	Oct.–Dec. 2003	Difference
Percent of dogs adopted within 30 days of being taken into the shelter	75%	92%	+17%
Percent of cats adopted within 30 days of being taken into the shelter	66%	52%	−14%
Percent of animals kept after adoption	92%	99%	+7%
Percent of clients somewhat or very satisfied with adoption services	95%	95%	0
Total number of dogs and cats adopted during the time period	96	88	−8

The shelter has been successful at decreasing the length of time dogs wait to get adopted, and in reducing returned animals.

The length of time a cat stays in the facility before being adopted has increased. This should be examined to see if the shelter can improve these results. Were there actions used to reduce the time to adoption for dogs that would also work for cat adoptions?

EXHIBIT 5

Sample Comparison of Actuals to Targets: Community Arts Center Report

	Actual Oct.–Dec. 2003	Target Oct.–Dec. 2003	Difference between Actual and Target	Notes
New memberships during this period	100	250	−150	
Members up for renewal that renewed membership	86%	90%	−4%	200 members up for renewal during this period
Members satisfied with the **quality** of center performances	88%	85%	+3%	Based on 113 survey responses
Members satisfied with the **variety** of choices in types ofperformances	90%	85%	+5%	Based on 113 survey responses
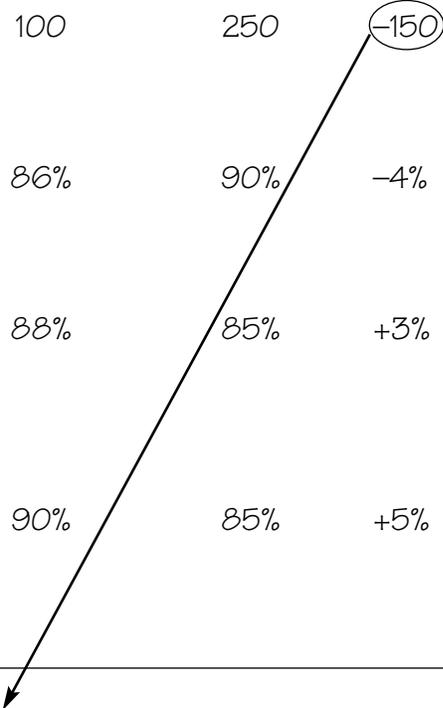
Though this program performed well keeping its members satisfied, it did not achieve its target of recruiting new members. This may suggest a need to investigate whether other types of performances might attract new members, or whether the marketing program is adequate. It might also be worth examining whether the target was, in fact, realistic.

Step 4: Compare the Latest Overall Outcomes with Outcomes for Clients in Other, Similar Programs— and to Any Outside Standards

Unless a program is unique, another agency somewhere is working to help clients with similar problems, and with at least somewhat similar services. It can be useful to compare outcomes if the other program collects similar outcome data. Similarly, an outside organization, such as a federal agency or a national association, might identify a "standard" (target) desired value for particular outcome indicators.

Comparisons can be made if these outside outcome indicators are reasonably similar. The program needs to make sure there are no major differences in defining the outcome indictors that would lead to unfair comparisons. For example, such indicators as school dropout rates and success rates in reducing substance abuse are widely used. However, many differences can exist in how these indicators are defined.

Reporting comparisons appears to be increasingly done in the nonprofit world. Such comparisons may be particularly feasible where a national organization is identifying suggested outcome indicators for its constituents to use, such as the American Red Cross, Boys & Girls Clubs of America, and Volunteers of America.

The availability of comparable outcome data from other communities, or other programs in the same community, is likely to be quite limited, at least at present. As a consequence, this type of comparison is likely to be much less available than the others.

This type of comparison contains some other dangers as well. For instance, it can be difficult to ascertain accurately whether the data from another organization are of reasonable quality. Further, another program may have differences in definitions of the outcome indicator that are not easy to detect, rendering comparisons somewhat suspect.

Exhibit 6 from Northern Virginia Family Service in Fairfax (VA) compares its outcomes for each outcome indicator used by its "Healthy Families Program" with outcomes for other, nearby programs. The exhibit also identifies its target for each indicator. Exhibit 7 from MotherNet/Healthy Families of Loudoun (VA) illustrates comparisons to state and national data, as well as to the program's own targets.

Note, however, that the available national and state data in exhibit 7 were for prior years. This illustrates another problem with comparisons with other programs, especially in other communities. The available outcome data are likely to be for a different time period from those preferred for timely comparisons.

EXHIBIT 6

Sample Comparison of Outcomes across Organizations Measuring Similar Outcomes for Similar Services

Northern Virginia Family Service's Healthy Families and Early Head Start Programs
Outcomes FY 2003 (percent)

Outcome Objectives	Alexandria	Arlington Healthy Families	Arlington Early Head Start	Prince William Healthy Families and Early Head Start	Loudoun
80% of moms who delivered will keep ACOG schedule (adequate prenatal care)	90	68	43	70	50
90% of moms who delivered will have babies weighing at least 2,500 grams	95	91	100	90	100
85% of mothers will have an interval of at least 24 months on subsequent births (or not have a subsequent birth)	97	93	100	92	100
90% of target children will remain up-to-date on immunizations	93	98	100	91	100
95% of target children will be screened for developmental delays (ASQ) at least semiannually until 36 months old and at least annually thereafter	98	97	96	95	100
100% of target children identified with a possible developmental delay will be referred to early intervention services	100	100	100	100	100
90% of target children will be in developmental range (excludes congenital delayed)	95	90	88	91	73
85% of parents surveyed will report HF/EHS program has been helpful in raising their children	96	100	82	98	100
70% of total home visits due were done	70	71	66	73	68
95% of families will not have founded reports of child abuse or neglect on target children	NA	98	100	100	100

Source: Adapted from reports by Northern Virginia Family Service, Early Childhood Division. A few indicators were out because of space issues.
NA = not available

EXHIBIT 7

Sample Comparison of Outcomes to State and National Values

MotherNet's Healthy Families Program (MN/HFL), Loudoun, Virginia

Objective	MN/HFL Target	MN/HFL Actual	State of Virginia Actual	National Actual
Children will have a primary health care provider	95%	99%	87% (1995)	86% (1997)
Mothers enrolled prenatally will complete all recommended prenatal visits	80%	87% (*n* = 104)	85% 1st trimester care (1997)	87% 1st trimester care (1960)
Children will be immunized on schedule	90%	91% (*n* = 91)	73% (1997)	78% (1997)
Babies of prenatal enrollees will be at least 2,500 grams at birth	85%	98% (*n* = 43)	92% (1997)	93% (1997)
Teen mothers will not have an additional birth within two years of the target child's birth	85%	96% (*n* = 45)	81% (1997)	82% (1997)
Children will demonstrate normal child functioning	90%	98% (*n* = 81)	98%	98% (1995)
Enrolled families will not have a founded CPS report	95%	100%	55,253 reports; 5,916 confirmed 6.46 per 1,000 (1997)	3 million suspected cases per year; 8,000 per day

Source: Adapted from Donna D. Klagholz, Ph.D. and Associates, LLC, "Year Two Evaluation Report" (MotherNet/Healthy Families Loudon, November 2000).

Delve Deeper into Client and Service Characteristics

Outcome information generally becomes considerably more useful when analyzed by client and service characteristics. This helps pinpoint problem areas and identify needed actions.

An organization is likely to need some computer capability to generate the necessary tabulations. Making manual tabulations, as was possible for the steps described in the first section, can become a large and difficult chore if the program wants to examine outcomes for many subsets of clients. Organizations with computer capability can enter the relevant data into a basic software program capable of making the tabulations and translating them into tables showing the outcomes for each characteristic. Most computers have preloaded software such as a word processing or spreadsheet program that will be sufficient for these purposes. In addition, a wide range of tutorials and low-cost training courses are available in most localities for staff that do not feel comfortable using computers.

Steps 5–7 describe actions desirable for examining outcomes for subsets of clients. Step 8 suggests a quite different perspective—that of examining multiple outcome indicators to provide a broader perspective.

Step 5: Break Out and Compare Client Outcomes by Demographic Group

By examining outcome data broken out by important client characteristics, a program manager can determine what characteristics seem related to better out-

comes. The breakout indicates how effective the program has been for each client group. This can help guide the modification of services for specific client groups. A careful breakout of information can direct program staff toward more effective strategies for serving people who are not experiencing the level of benefits other client groups are displaying.

Exhibit 8 provides candidate client characteristics for which outcome data breakouts might be highly useful. (The exhibit also provides candidate service characteristics, which are discussed in step 6.) Each program should identify which client characteristics are likely to have a substantial effect on outcomes. The program should ask that the outcome data be broken out for each such characteristic.

Exhibit 9 illustrates breakouts for a number of client characteristics reported on one page. By examining such breakouts, managers can save time by focusing on the client groups most likely to warrant attention. (Selected data are highlighted using circles.)

To obtain this breakout information on client subgroups, of course, the program needs to make sure that this information is collected and recorded so it can later be linked to the outcome data for each client. Many programs probably already obtain much of this information as part of their client intake process. Some of this information can also be obtained from clients if the program surveys its clients to obtain outcome information.

For each client characteristic selected, clearly defined categories need to be created. For example, the program might need to select a number of age range and race/ethnicity categories into which each client would be placed for analysis. Exhibit 10 is an example of a tabulation for one demographic characteristic, in this case education level.

Exhibit 11a also displays data broken out for one client characteristic, gender. The data indicate that little difference exists on this outcome indicator between females and males. Thus, this characteristic does not appear to require further review by the program, at least for this reporting period. Some breakouts for a reporting period will indicate that no further examination of the outcome is needed. This is to be expected.

Step 6: Break Out and Compare Outcomes by Service Characteristics

In this step, subsets of clients are grouped by service characteristics. These characteristics might include breakouts by each particular facility or office (if the program has more than one); the type of service procedure used (such as the mode of delivery, type of staff, or venue—if more than one); the amount of service provided the client; and the specific service provider (such as individual caseworkers).

EXHIBIT 8

Possible Client and Service Characteristic Breakouts

Client Characteristics (at intake)

Gender	Examine outcomes for men and women separately.
Age	Examine outcomes for different age ranges. Depending on the program, the age groups might span a large range of ages (such as examining clients under 21, between 21 and 59, and 60 and older), or the program might focus on a much smaller age range (such as youth programs wanting to compare outcomes for youth under 12, 13–14, 15–16, and 17 or older)
Race/Ethnicity	Examine outcomes for clients based on race/ethnicity.
Disability	Examine outcomes based on client disability. For example, some programs might want to determine whether clients with disabilities rate services differently than those without disabilities, as well as the outcomes for clients with various types of disabilities.
Educational level	Examine outcomes for each client based on the educational level achieved before starting service.
Income	Examine outcomes for clients grouped into specific income ranges based on the latest annual household income at the time clients began service.
Household	Examine outcomes for households of various sizes, generations, and numbers of children.
Difficulty of problem at intake	Examine outcomes by incoming status based on expected difficulty in being able to help the client. Inevitably, some clients are more difficult to help than others. For example, an employment program might want to consider the literacy level of its new clients. An adoption program might want to relate outcomes to the age and health of the children.

Service Characteristics

Facility/Office	Examine outcomes for individual facilities or offices.
Service provider	Examine outcomes for clients of individual service providers, such as caseworkers.
Type of procedure	Examine outcomes for clients who were served using each distinct procedure. For example, a youth program might have used workshops, field trips, classes, and so on.
Amount of service	Examine outcomes for clients who received varying amounts of service. This might be expressed as number of "sessions" a client attended, the number of hours of service provided each client, or whatever level of service measurement the program uses.

EXHIBIT 9

Sample Comparison of All Breakout Characteristics

Clients That Reported Improved Functioning after Completing Group Therapy

Characteristic	Number of Clients	Considerable Improvement (%)	Some Improvement (%)	Little Improvement (%)	No Improvement (%)
Gender					
Female	31	10	19	(55)	(16)
Male	43	30	40	21	7
Age Group					
21–30	13	23	31	31	15
31–39	28	21	32	36	11
40–49	24	21	29	38	13
50–59	9	22	33	33	11
Race/Ethnicity					
African-American	25	32	20	32	16
Asian	5	0	60	20	20
Hispanic	20	15	40	40	5
White/Caucasian	24	21	29	38	13
Sessions Attended					
1–2	13	15	8	(54)	(23)
3–4	21	24	33	33	10
5+	40	23	38	30	10
Facility					
Facility A	49	24	27	35	14
Facility B	25	16	40	36	8
Caseworker					
Therapist A	19	26	26	42	5
Therapist B	18	11	39	33	17
Therapist C	18	6	17	(56)	(22)
Therapist D	19	42	42	11	5
All Clients	**74**	**22**	**31**	**35**	**12**

EXHIBIT 10

Sample Client Breakouts: An Employment Program, by Entering Education Level

Clients Employed Three Months after Completing Service

Education Level	Total Clients	Percent Employed
Completed high school	80	63%
Did not complete high school	180	50%
Total	**260**	**54%**

By examining outcome data broken out by important service characteristics, a program manager can determine which service characteristics seem to relate to better outcomes. This can help identify service characteristics associated with successful outcomes and those associated with less successful outcomes. This step is applicable to programs with different service characteristics for particular clients—whether the differences are intentional or unintentional. For example, some service locations might be providing the same services to the same types of clients. Outcomes should be calculated for each location. The outcome comparisons will identify to what extent differences in outcomes have occurred.

Another example: A program may believe that the number of service hours that individual clients receive has important effects on outcomes. If the program keeps track of the number of hours of service provided to individual clients, it can later calculate the success rate for clients receiving different amounts of service. (The program would need to keep a record of the number of hours, days, sessions, or whatever unit of service it uses.)

Programs with outcome measurement systems can also use outcome data to *experiment with new or modified service delivery procedures*. The program might apply the newer procedure to a sample of clients and later calculate the success rates for both the new and original procedures to determine which procedure was more successful. (Such experiments are discussed further in the last section of this report.)

Exhibit 11b shows the outcomes from Exhibit 11a broken out by one service characteristic, the particular unit providing the service. The exhibit indicates that a substantially higher percentage of unit 1 clients reduced their drug use (reporting either "considerable" or "some" reduction) than clients of unit 2 (80 versus 52 percent). Based on these findings, the program is likely to want to examine how each unit delivers its services, in hopes of identifying better practices at unit 1 that unit 2 might use, or at least motivating unit 2 to improve its outcomes.

EXHIBIT 11a

Breakout by Client Demographic Characteristics

Client Reduction in Drug Use by Gender, Current Quarter

	N	Considerable Reduction (%)	Some Reduction (%)	A Small Reduction (%)	No Reduction (%)
Females	42	21	45	24	10
Males	31	26	42	26	6
Total	**73**	**23**	**44**	**25**	**8**

EXHIBIT 11b

Breakout by Service Characteristics

Client Reduction in Drug Use by Service Unit, Current Quarter

	N	Considerable Reduction (%)	Some Reduction (%)	A Small Reduction (%)	No Reduction (%)
Unit 1	42	25	55	15	5
Unit 2	31	21	31	36	12
Total	**73**	**23**	**44**	**25**	**8**

Percent of Clients with Somewhat or Considerably Reduced Drug Use

| Unit	Percent
Unit 1	80%
Unit 2	52%
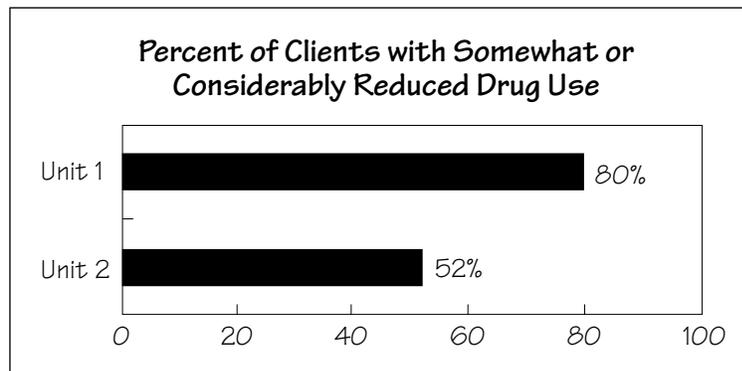
Exhibit 11b also contains a bar-chart version of the summary outcome data. This is a good display device, particularly for external reports.

Considered together, Exhibit 11a and 11b indicate that outcomes appear related to the unit but not to gender.

Program managers and their staffs may find it quite informative to compare outcomes across all the client and service breakouts for which the program collects outcome data, including both the client demographic and service characteristics. Exhibit 9 illustrates such a comparison, presenting outcome information for three demographic and three service characteristics together.

If such tables are prepared, reviewers can quickly go down the tables to identify those client and service characteristics that appear most related to better, or worse, outcomes. This saves time for managers, and others, reviewing the outcome data. They can then focus on areas most likely to need attention. Users are likely to concentrate on the poorer outcomes, which provide the most room for improvement, but should not neglect to give plaudits to personnel who are associated with high-outcome levels. In general, such tables can make it considerably easier for data users to spot key issues.

For example, exhibit 9 on page 16 shows that very large percentages of female clients, clients attending only one or two sessions, and therapist C's clients, reported that the therapy sessions had not helped them to function better. The program would likely want to explore why this is so and identify implications for service adjustments.

Computers greatly ease the calculation of such breakouts. Many computer software programs can be used to make cross-tabulations. Breaking out data manually consumes considerably more time.

Simultaneous Analysis of Multiple Client and Service Characteristics

A somewhat more advanced analysis procedure is to examine simultaneously the effects of multiple client and service characteristics on a particular outcome. This procedure enables the program staff to bore in considerably deeper on the relation of these characteristics to the outcomes. It is a more complex level of analysis.

Exhibit 12 illustrates this analysis. The exhibit simultaneously breaks out an outcome indicator by one service characteristic (length of the program that the client attended) and one client characteristic (entering education level). The data indicate that overall substantially better employment outcomes occurred among clients with the higher education level when they entered the program (63 percent versus 48 percent). Program length appears not to have affected the outcomes for clients who had completed high school. On the other hand, outcomes differed substantially by program length for clients who had not completed high school at the time of entry, 73 percent for those completing the long program versus 26 percent for those completing the short program. The program might want to encourage clients who have not completed high school to take the longer program, but not those who have.

EXHIBIT 12

Sample Two-Characteristic Breakout

Percent of Clients Employed Three Months after Completing Service

Education Level at Entry	N	Short Program	Long Program	Total
Completed high school	100	62% employed (of 55 clients)	64% employed (of 45 clients)	63% (of 100 clients)
Did not complete high school	180	26% employed (of 95 clients)	73% employed (of 85 clients)	48% (of 180 clients)
Total	280	39% (of 150 clients)	70% (of 130 clients)	54% (of 280 clients)

Is action needed? Encourage clients who had not—rather than had—completed high school to attend the long program. Use these figures to help convince clients of the longer program's success with helping clients secure employment.

Without this two-way breakout, users of the outcome data would not know that better employment outcomes were related to attendance at the longer program, and particularly significant for clients who had not completed high school at the time of entry.

Such two-way outcome indicator breakouts could be prepared on a regular basis, or could be used for special analyses to be done only occasionally.

Step 7:   Compare the Latest Outcomes for Each Breakout Group with Outcomes from Previous Reporting Periods and with Targets

In this procedure, outcomes for the individual client breakout categories are compared either with previous outcomes or with targets for the particular breakout category. This is similar to steps 2 and 3, but the comparisons focus on the progress of particular client groups. This analysis shows how successful the program has been compared with previous years and with its targets.

Exhibit 13 illustrates comparisons against targets for a number of breakout groups. (The exhibit is a version of the data previously shown in exhibit 9.) As noted earlier, it is a good practice to provide a column that shows the amount of difference between actual values and targets (as shown in the last column of exhibit 13). In this example, the target is the same for all demographic and service characteristics. However, a program will often want to set somewhat different targets for different subgroups of clients.

Exhibit 13 indicates that outcomes for a number of breakout categories have fallen far short of targets, such as for female clients, for clients attending only 1–2 sessions, and for therapist C's clients. The exhibit also indicates that for this reporting period, the outcomes for therapist D's clients were considerably better than the target.

With computer help, generating such tables should be quite feasible, and the tables can be provided shortly after the data have been obtained for each reporting period.

Without such detail, program personnel would only know that the overall target had not been met. They would not have recognized that missed targets were associated with a particular client group and particular service characteristics, enabling program personnel to focus their improvement efforts on these particular characteristics.

EXHIBIT 13

Comparisons of Outcomes with Targets for Subsets of Clients

Clients That Reported Improved Functioning after
Completing Group Therapy Sessions

Characteristic	N	Percent Improved	Target	Difference (percentage points)
Gender				
Female	31	29%	60%	(−31)
Male	43	70%	60%	+10
Age Group				
21–30	13	54%	60%	−6
31–39	28	54%	60%	−6
40–49	24	50%	60%	−10
50–59	9	56%	60%	−4
Race/Ethnicity				
African-American	25	52%	60%	−8
Asian	5	60%	60%	0
Hispanic	20	55%	60%	−5
White/Caucasian	24	50%	60%	−10
Sessions Attended				
1–2	13	23%	60%	(−37)
3–4	21	57%	60%	−3
5+	40	60%	60%	0
Facility				
Facility A	49	51%	60%	−9
Facility B	25	56%	60%	−4
Caseworker				
Therapist A	19	53%	60%	−7
Therapist B	18	50%	60%	−10
Therapist C	18	22%	60%	(−38)
Therapist D	19	84%	60%	(+24)
All Clients	**74**	**53%**	**60%**	**−7**

Step 8:   Examine Findings across Outcome Indicators

The previous steps have focused on comparisons that can be made when examining each individual outcome indicator. But most programs will collect and report data on more than one outcome indicator. (The outcome reports illustrated in exhibits 3–7 present data on multiple outcome indicators.) Additional perspective can be gained by comparing results across multiple outcome indicators to identify which outcomes warrant attention by the program.

This comparison can be primarily subjective, with program staff reviewing the values for each outcome indicator for the reporting period.

For example, in exhibit 3, by examining the results across all 21 individual outcomes, program staff can identify for which outcomes students of a mentoring program reported substantial, or little if any, improvements, relative to the other outcomes. In this case, the program might choose to adjust its curriculum to place more emphasis on personal hygiene and appearance and other outcomes that had relatively poor results.

A relatively common example of the potential usefulness of examining findings across outcome indicators is if a program surveyed clients about their overall satisfaction with the service they received and about their satisfaction with specific service characteristics, such as the timeliness, hours of service availability, and accessibility of the service. Analysis across these indicators would indicate the relative importance of individual service characteristics to overall client satisfaction.

Sometimes a program might find it useful to combine the data from more than one indicator into one number, producing an "index." The index provides a single number intended to summarize the program's outcomes for the reporting period. Being able to look at only one number that represents the overall program results is attractive. However, such indices if used alone, without also reviewing the data on individual outcomes, will very likely hide many important findings. Indices, if used, should be used in conjunction with the examination of each outcome indicator.[3] As with the values on individual outcome indicators, index values for any reporting period can be compared with the index values from previous reporting periods or with targets.

---

[3] Creating an index requires that the program assign a "weight" to each outcome indicator. The weight represents each indicator's importance relative to the other indicators. The simplest way is to assume that each indicator has equal importance. This makes calculation of the index easier. However, this simplified procedure should not be used if the program believes that some outcome indicators are considerably more important than others.

Exhibit 3 illustrates a situation where an index might be useful. The program has obtained data on a considerable number of related outcomes, each of which represents a different dimension of client improvement. The program might find useful an index that combines the findings across all 21 outcome dimensions, or at least indices for each of the three sections—confidence, competence, and caring. These indices would indicate how much each client has improved—whether overall or in each of those three sections. The procedure here would be to assign weights to each of the 21 characteristics and also to assign values to each of the five levels of client improvement (such as one to five). Finally, the average percentage for each section would be calculated.

Make Sense of the Numbers

The previous analysis steps should provide the program with considerable information on how well the outcomes from the program's services are being achieved. However, even for small programs these steps will likely result in large amounts of data. Now it is necessary to sort out what particular information needs attention. That is the purpose of step 9.

Deciding what outcome information warrants attention is an important step. Attempting to determine *why* the outcomes occurred is the next step. This can be more difficult than the mathematical calculations. It is vital that program managers and organization officials explore why there are unexpected good or bad findings—to take the next steps of improving services. This is the purpose of step 10. It is up to program managers and staff, and to other policymakers, to attempt to find out what these outcomes mean. This requires going beyond calculating outcomes to interpreting them. Data do not speak for themselves.

Step 9: Identify Which Numbers Should Be Highlighted

Once the data have been tabulated and analyzed, as in steps 1 through 8, the program manager and staff need to identify those findings that appear unusual, unexpected, overly low, or overly high. This applies whether the actual outcome figures are compared with outcomes for previous reporting periods, with targets, with other similar programs, or whether outcomes are compared among client demographic groups, or among clients served in different ways.

This step is primarily judgmental. For example, how much worse than the previous reporting period, or worse than the target, does the latest outcome value have to be before it warrants significant attention?

If the program has sampled its clients to obtain outcome information, statistical analysis can help. The analysts can compute the likelihood that observed differences in outcome values occurred by chance. For example, suppose the target for clients achieving successful outcomes is 56 percent and the actual percentage for the reporting period is 49 percent. How likely is it that the 7 percentage point difference occurred by chance? It is beyond the scope of this guide to discuss the statistics involved.

More important than the statistical likelihood that the difference occurred by chance is *the practical importance* to the managers of the amount of the difference. The main consideration is how important the difference is. Practical importance is primarily a judgment call by those who would act on the information. Here we offer some "rules of thumb" for outcome indicators expressed as percentages:

- If the difference in percentage points is 10 or more, the difference warrants attention ("red light" treatment).
- If the difference is between 5 and 10 percentage points, this warrants modest attention, or at least flagging, to examine what happens in the next reporting period ("yellow light" treatment).
- If the difference is 5 percentage points or less, the difference is likely too small to be of concern ("green light" treatment).

Highlight those outcome findings that appear unusual or unexpected, whether overly low or overly high. This will signal to users the need for attention.

How should such findings be highlighted? The most basic approach is merely to call attention verbally to trouble spots or very good outcomes. However it is likely to be more useful to highlight important findings on outcome reports by such simple approaches as underlining or circling the outcome numbers that appear to warrant further attention. This will make sure that report users are alerted to key information on the reports. Using color usually grabs attention. Several earlier exhibits illustrate highlighting by circling key outcome values. For example, exhibit 13 calls attention to the values substantially below targets for female clients, for clients that attended less than three sessions, and for therapist C's clients. The exhibit also highlights the high percent of therapist D's clients that had improved.

For breakouts grouped by client location within the community, mapping, with different outcome levels represented by different levels or type of shading, can be useful. Inexpensive GIS software is now available and is increasingly used for displaying a variety of data on citizens.

As discussed under step 1 and illustrated in exhibit 2b, sometimes combining outcome levels into a smaller number of data points can provide a useful perspec-

tive, as shown in exhibit 13. It combines the percents of "considerable improvement" and "some improvement" shown in exhibit 9. These sums were used in exhibit 13 for comparing actual outcomes to targets. (However, as discussed in step 1, if one of the extreme outcome levels has particularly high percentages, this important finding might be lost. In such cases, those important extreme conditions should also be highlighted.)

Step 10: Seek Explanations and Communicate the Findings

Data by themselves say little, if anything, about why they are high or low. Before actions are taken on outcomes that appear poor (or very good), program managers should seek explanations for the findings. They should look for the specific factors that affected the highlighted outcome findings for the reporting period.

Internal and external factors, or both, may have affected the results. For example, increased employment in the community, not the services provided by the program, might have lowered domestic violence among the program's clients. Or, a reduction in federally subsidized food programs may have worsened the health of women in a prenatal program, despite the program's strong efforts.

Factors likely to contribute significantly to unexpectedly poor or good outcomes include the following:

External Factors (factors over which the organization or program usually has limited or no control)

■ Changes in external conditions in the community, such as unexpected local or national economic conditions (perhaps affecting client employment opportunities).

■ Changes in the mix of clients that came in for service, such as an unexpectedly large proportion of clients with major difficulties. In Olympic diving competitions, each dive is assigned a degree of difficulty, so judges can rate all the different divers, no matter if they attempt easy, moderate, or very difficult dives. To adjust for this factor, programs can define levels of difficulty—say, three or four—and then assign a level to each entering client. (Degree of difficulty is one of the candidate client characteristics listed in exhibit 8.) Then the program can directly take into consideration such differences by calculating separately the outcomes for clients with each degree of difficulty.

■ Unexpected reductions or gains in program funding and staff.

■ Various social and political influences in the community.

Internal Factors (factors over which the organization or program usually has a reasonable amount of control)

- Problems with the design of the program.
- Problems in implementing the program, such as inadequate staff training or poor communications with clients.

- Unexpected staff turnover or absenteeism.

- Changes in direction by agency management or the organization's board.

Programs will usually need to look for both kinds of factors.

Exhibit 14 lists a number of procedures to obtain explanations for unexpectedly poor or good outcomes. It may be useful to use a number of these procedures to obtain reasonable, satisfactory explanations.

Finally, how *clearly and cogently the findings are presented* from these steps can also be critical to the usefulness of the analysis. A number of exhibits throughout this report have illustrated ways to present the outcome findings. These include using tables (with clear and complete labels and avoiding too much clutter); bar charts; highlighting techniques, such as circling key findings; and adding suggestions on likely necessary actions to key findings.

EXHIBIT 14

Obtaining Explanations for Unexpected Outcome Values

1. Examine findings from the previous steps to identify which client and service characteristics (such as those listed in exhibit 8) appear particularly related to the unexpected outcome values. While this will not explain why the outcomes occurred, it will enable the program to narrow in on where to search for explanations. For example, the group therapy program whose data are presented in exhibit 9 might focus its search for explanations on problems with females and those that attended less than two sessions, particularly the clients of caseworker C. In this example, the problem may be that clients of caseworker C only attended one or two sessions.

2. Discuss the reasons for shortfalls with supervisors and their service providers. Such discussions might be with individual supervisors and providers, or might be in groups. One very useful approach is to hold "How are we doing?" sessions shortly after each outcome report becomes available and is disseminated. At this session ask staff to provide their reasons for unusually good or poor outcomes identified by the report.

3. Examine client responses to "open-ended" questions included in surveys, particularly responses to questions asking respondents to provide their reasons for low ratings to particular service characteristics or their suggestions for improving services.

   Programs that survey their clients for outcome information should include such questions and examine the responses. The responses should be categorized and grouped by common themes to identify specific problems and client suggestions. For example, a number of clients might have stated (though probably in somewhat different words) that they had difficulties getting through to program staff by telephone, or getting to the program's facilities. Such information can suggest corrective actions that the program can take to ease such problems.

4. Hold focus groups with clients (former and/or current) to elicit their views about the problems. (These are 90- to 120-minute meetings with perhaps 8–12 clients in which a facilitator solicits comments and thoughts about the program's service.)

5. Form a working group of staff, and perhaps volunteers, to examine the problem and why it occurred (and ways to correct it).

6. For major problems, seek an independent in-depth program evaluation. Perhaps recruit an outside organization, such as a local university or community college, to examine the reasons for the problem.

Special Analyses Using Outcome Information

The analysis procedures described in the previous sections are suggested for use by programs after outcome data become available for each reporting period. This section identifies a few special analyses that programs might undertake when they believe these are warranted and have the resources.

1. For those programs that seek outcome information both at the time of exit and at a prescribed post-service follow-up time (such as six or 12 months after exit), *tabulate the difference between the client's condition at exit and follow-up.* This will identify whether, and to what extent, any substantial drop-off of client benefits occurred after exit. This analysis will enable the program to identify the extent to which outcomes are sustained after clients exit the service. If the drop-off appears major, this is likely to trigger an examination to identify whether the program needs to be modified.

2. *Assess whether the program's outcomes appear to be on an improving or declining trend.* If outcomes have been trending downward over a few years, a full-scale program review may be needed.

3. If a program has made a significant change to its service procedures or policies, *compare the outcomes for a period before the change was made to the outcomes for a similar period after the change.* The comparison should provide evidence of whether the change led to the desired improved outcomes.

4. *Use data from the outcome monitoring process to test new procedures before they fully replace existing procedures.* This procedure has thus far rarely been used by private nonprofit organizations. However, it has considerable

potential to encourage program personnel to try new ideas, to innovate. The basic procedure is to randomly assign clients to the new and old procedures—even if only by flipping a coin. (Randomization is often the most ethical way to assign clients, especially when trying out a new procedure whose improvement over the current procedure is unknown.) After sufficient time has elapsed for the new service procedures to be tested, tabulate and compare the outcomes for each group of clients. This will provide strong evidence on which procedure led to better outcomes, and to what extent.

For example, an agency providing a 12-week group therapy intervention on anger management could test a new curriculum and compare it with the old curriculum. If 40 new clients enrolled for group therapy sessions, clients would be randomly assigned to two groups. Group A, consisting of about 20 clients, would be given the existing curriculum, and group B, consisting of the other clients, would be given the new curriculum. Six (or 12) months after the end of therapy, the outcome improvement for each client could be assessed. If group B clients had significantly better outcomes than clients in group A, then this new procedure might be adopted for all future clients.

Exhibit 15 outlines the procedure for conducting such tests of new service delivery approaches.

5. Programs with internal capacity, or access to specialized statistical external help, can *undertake more advanced statistical analyses of the data.* Such specialized statistical procedures as use of regression analysis, chi square tests, and the like can sometimes add considerable insight into the outcome data. For example, multiple regression analysis could be used to better examine the relation of individual outcomes to client demographic or service characteristics. The analysis would indicate which characteristics had the most significant statistical relationship to the observed outcomes. Such analyses are beyond the scope of this guide.

EXHIBIT 15

Testing Alternative Service Delivery Approaches

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

2. Choose a method for deciding which incoming clients will be served by the new approach. The method should be one that selects a representative sample of clients for each approach. Some form of random assignment is necessary. Randomization helps assure that the comparisons will be valid and greatly increases the strength of the evidence. (For example, with random assignment, approximately the same proportion of difficult-to-help clients will likely be included in each group.) Methods of random assignment include flipping a coin and using a table of random numbers. Another method is to assign incoming participants alternatively to each practice.

3. As each client enters the program, assign the client to one of the two groups by the procedure the program has identified.

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

5. Track the outcomes for each client in each approach over whatever period of time 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 make adjustments to program 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 make a decision, in which case the program might continue the comparison for a longer time.

Final Points about Analysis

Periodic review of the analysis procedures. The organization should seek periodic feedback from staff involved in the outcome measurement process. This should include both those participating in the data collection and those using the analysis findings. Users of the analysis should be asked if the analysis has been useful and for suggestions on how the analysis can be improved.

Important analysis limitation. Outcome data by themselves can never prove—as a scientist would want to prove—what causes the outcomes. This is a key difference between outcome measurement and program impact research or evaluation, which can use random assignment and sophisticated statistical methods to separate a program's influence from other factors. But those advance steps are expensive and time-consuming, and not feasible for nonprofit organizations to use very often, if at all.

In addition, of course, those making decisions on outcomes need other information, not just information on outcomes. Additional information, such as resource requirements, costs, "political" issues, and staff morale, needs to be considered.

Benefits of analysis. Should the above limitation be used as an excuse for not performing outcome measurement and analysis? No, of course not. Ultimately, program and organization officials have to make frequent decisions on their programs—regardless of what information is available or how good it is. Outcome measurement and the analysis of the resulting outcome data, if done with reasonable care, should provide considerably improved information to help with decisions.

Analysis can greatly enhance the information content, add considerable perspective to the raw numbers coming from data collection procedures, and provide a much better basis for determining what actions are needed in the future.

Decisions about how program services will be delivered are ultimately judgment calls by program officials. Outcome data enriched through these analysis steps can much better inform those decisionmakers—to help make their services better, producing better future outcomes for clients.