Stage 4:

Planning Evaluation

Topics

- Developing an Evaluation Plan
- Establishing Evaluation Objectives
- Process (Formative) Evaluation Objectives
    - Accountability
    - Program improvement
    - Replication
- Summative Evaluation Objectives
    - Overall program effectiveness
    - Program effects-what else happens as a result of outreach?
- Evaluation Methods
    - Quantitative Method
    - Qualitative Method
- Selecting an Evaluation Design
    - Experimental design
    - Quasi-experimental design
    - Non-experimental design
- How Much Evaluation is Feasible?

Figures

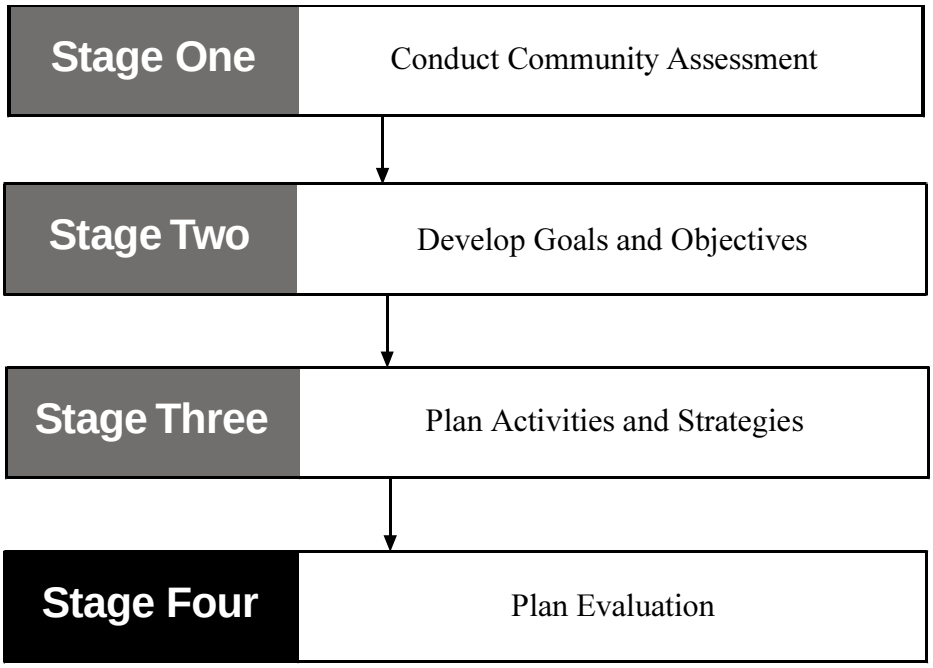
Figure 11: Program Evaluation Flow Chart

Figure 12: Evaluation Designs

Figure 13: Level of Resources for Various Evaluation Designs

Tool Kit

- References and Selected Readings
- Workform for Process Evaluation Objectives
- Gowan Library Example



What to Establish

Define Evaluation Objectives

Should process evaluation be conducted?
- to demonstrate accountability
- to monitor progress
- to make 'mid-project' adjustments
- to replicate a pilot project

Should summative evaluation be conducted?
- to document what was achieved
- to find out what else happened
- to research effectiveness of specific strategies

Consider for Best Results

What do stakeholders want to know?

How will results be used?

What decisions will be made?

Select Evaluation Design

Determine the independent and dependent variables of interest

Choose evaluation design that balances time, resources and method with desired level of validity

Consider for Best Results

Is it important to get valid, generalizable results?

Will quantitative data or qualitative data be used, or both?

If so, how can validity be maximized?

A typical model for program development includes the following phases:

1. Identifying a target audience and conducting a community needs assessment,
2. Developing written goals and objectives,
3. Implementing activities to accomplish those objectives, and
4. Evaluating the overall quality and success of those activities vis-à-vis the stated objectives.

In reality, planning and conducting a program and its evaluation is more complex than a four-step process. Different types of evaluation correspond to different phases of program development. Thus, as seen in Figure 11, the model should be at least a 6 step process that integrates various types of evaluation throughout.

The manual thus far has discussed ways to conduct evaluation for a community and audience assessment, as part of program development phases I-III in Figure 11. This chapter will describe an overview of evaluation planning to assess a program's implementation and outcomes.

For further information on evaluation planning, several sources are listed in the Tool Kits at the end of Stages 4 and 5. One outstanding and comprehensive source is the nine volume kit edited by Joan L. Herman called *Program Evaluation Kit,* Newbury Park, CA, Sage Publications, 1987.

Developing an Evaluation Plan

The three major components that should be addressed in an evaluation plan are:

1. Questions or issues you will address in the evaluation
2. What you will measure and how
3. Resources needed to accomplish the evaluation tasks

Figure 11: **Program Evaluation Flow Chart**

Program Phases	Question to Ask	Evaluation Phase
I. Identify Problem/Need	What is the targeted community? To what extent are information needs being met?	Community Assessment
II. Develop Goals and Measurable Objectives	What changes will address unmet needs?	
III. Select Activities and Strategies and Design Implementation Plan	What kinds of activities/strategies will produce changes desired? How will activities and strategies be tailored to the needs of the targeted group? How should the program be put into operation?	Audience Assessment
IV. Program Implementation	Is the program operating as planned? Are participants learning what is expected? Is the audience satisfied with results? Is the program reaching the intended audience?	Process Evaluation
V. Program Outcomes	Were objectives reached? Are there impacts regarding health information use? What other impacts have occurred?	Summative Evaluation
VI. Feedback	How realistic were initial goals? What programmatic changes need to be made?	

To be most effective, plans for evaluation should be in place before outreach activities begin. Thinking ahead will make it easier to plan whether and what baseline data to collect. Data collection instruments, such as surveys, may need to be developed and pilot tested in advance. If there are plans to compare a specific strategy with an alternative to see which is more effective, time is needed to work out the logistics about when and with whom the two strategies will be tested.

And, even though an evaluation report is completed at the end of the program, it is difficult, ineffective, and not very objective to begin thinking about evaluation after the program is over. Therefore, it is best to plan ahead, before activities begin, about what will be measured and how.

In developing the plan, the following issues require consideration:

1. Outreach goals and objectives
2. Plans for implementation, or what is currently happening if the program is already in place
3. Evaluation objectives – purpose of the evaluation and its role
4. Evaluation questions to be addressed
5. Methods and types of information that will be accepted as evidence of the effects of the program
6. Design – when and from whom data will be collected
7. Data collection – what and how data will be collected
8. Resources
9. Timeline for evaluation

The first two steps in evaluation planning involve clarifying the goals and objectives of your outreach program and plans for implementation. Both of these steps are described in detail in Stages 2 and 3. Equally important is establishing objectives for the *evaluation*, as described in the next section.

Evaluation objectives will help determine the specific issues or questions the evaluation will address. Decisions about how to gather measurements will include considering what types of information (qualitative or quantitative) will be most appropriate and accepted as evidence. Decisions about the research design – when and from whom data will be collected – will follow.

Each of these considerations are addressed in this chapter, with a brief discussion of how much evaluation is realistic for your program. Issues of data collection – what and how data will be collected – are discussed in Stage 5.

Establishing Evaluation Objectives

One of the most challenging aspects of evaluation is clarifying what it is you want to find out. A good first step is to identify the "stakeholders" who will have an interest in the evaluation results. They might include:

- Funding agency
- Targeted community
- Your boss
- Outreach staff

When planning what data to collect, think about what these stakeholders will look for in the evaluation report. For example, although information about the overall results of the program might be needed by the funding agency, key contacts of the targeted community may want to know the reactions and comments of outreach participants in order to make a decision about future outreach efforts. Other outreach programs with similar audiences may be interested in how you conducted your program and what worked best. Or, your outreach staff may be interested in determining whether one particular strategy is more effective than another.

Ask stakeholders about their criteria for success – what outcomes from the project are most important to them? Do they also want to know if it was successful compared to an alternative

(such as another type of outreach program, or no program at all)? Is the program being evaluated as a pilot study for possible replication?

One way of prioritizing the evaluation questions is to ask yourself *and* those interested in the evaluation how the information gained about a particular question will make a difference. What decisions will be made as a result of the data? Or, how will the information help improve the program?

It will be important to refine the broad purpose or objectives of an evaluation into specific questions. Questions addressed by evaluation during and after outreach can be categorized as *process* and *summative,* respectively. [Note: some evaluation textbooks differentiate process evaluation as part of formative evaluation and summative evaluation as another term for outcome/impact evaluation.]

Process (Formative) Evaluation Objectives

Process evaluation helps to keep track of an outreach program as it is happening so that modifications or improvements can be made on an ongoing basis.

Very generally, process evaluation questions address:

- Is outreach working as intended?
- How can it be improved (while it is going on)?

To focus the types of data you may want to address in a process evaluation, use the "Workform for Process Evaluation Objectives" in the Stage 4 Tool Kit. A sample filled-in workform is provided in Appendix K, "Sample Process Evaluation Objectives." Appendix L, "Sample Ways to Measure Program Process," provides selected measures for several of the evaluation objectives in Appendix K.

There are many possible questions for a process evaluation, and choosing which ones to ask will depend on how the data will be used. The following section provides examples, by purpose, for process evaluation data, based in part on a more thorough discussion by King, 1987 (1).

Accountability: did you do what you said you would do? To provide accountability to stakeholders such as funders, partners, or directors, first decide what characteristics are important to the success of the program (do not forget the perspective of your targeted audience – what do they think is important)? Some might be:

- Costs (staff, materials, equipment, facilities)
- Relevance of equipment, resources (e.g. PubMed), and services (e.g. interlibrary loan) provided or promoted with respect to user need –e.g., are resources useful in terms of content, understandability, language, or cultural relevance?
- On-site administrative support
- Facilities (location, size, and number of computers allotted for training)
- Time allotted to activities
- Staff responsiveness to participants' needs

The above characteristics are just examples. Modify the list according to the characteristics most important to the success of your outreach program and decide how each will be monitored. Appendix K, under Accountability, provides an example list of characteristics important to one outreach program. Note that it is helpful to review the objectives, outcomes, and overall plan for implementing the program when selecting characteristics to monitor.

Program improvement: assessing progress toward objectives so adjustments can be made that are targeted and effective. Planners need to decide in advance what indicators to measure, which will depend on the outcomes identified in each objective (see Appendix D "Sample Outreach Objectives"). Some indicators could be:

- Numbers or percentage of target audience reached
- Evidence that promotional activities increase awareness of information resources

- Evidence that participants increase their level of self-efficacy (confidence) in search skills
- Evidence of quality (relevant or useful or efficient) search results
- An increase in ILL requests
- Evidence of intended or actual use of electronic resources (e.g. Website hits, if relevant, or survey responses about intentions to use electronic resources)

The data collected to measure these indicators will give valuable feedback about what might be working and what needs adjustment. This type of evaluation is measuring the effectiveness of specific strategies. You can look to the implementation plan you developed in Stage 3 to help clarify what assumptions you may want to test about causal links between strategies and outcomes.

Another way of thinking about what causal links to measure is by identifying the independent and dependent variables. An independent variable is what the planner has control over (e.g. the intervention). The dependent variable is the outcome or what changes (e.g. use of PubMed) as a result of the independent variable. For example, if assessing the effect of an outreach activity (e.g. skills training) on outcomes of interest such as attitudes, beliefs and behavior, the independent variable is the skills training and the dependent variables are changes in attitudes, beliefs and behavior. Thus, dependent variables are typically the outcomes identified in the outreach objectives.

If one is conducting a theoretically-based evaluation, it is important to track the variables identified in the theory to determine whether or not the intervention is operating effectively. For example, if a strategy based on Diffusion of Innovation theory is used to change information seeking behavior, you may want to test the assumption that the strategy actually *caused* the behavior change. By focusing your data collection on variables that are critical to the theories you use, your evaluation can identify those strategies that seem to make the most difference, so you can *explain* rather than just describe the outcome.

Say that the Extended Parallel Process Model was used to develop the intervention and evaluation. In a process evaluation, researchers would measure perceptions of threat (severity, susceptibility) and efficacy (response efficacy, self-efficacy) to determine whether the intervention was promoting danger control actions (i.e., adoption of the recommended response) or fear control actions (i.e., defensive avoidance, reactance against the recommended response). If the results of a survey indicated high threat and low efficacy, then according to this theory the intervention would be failing. However, if the survey indicated high threat and high efficacy, then one could be fairly confident that the intervention was producing the actions desired (2).

For a more detailed example of theory-based process evaluation see Appendix K, Program Improvement. Keep in mind that, ultimately, the outreach objectives themselves may need modification if they are not being reached. Meanwhile, monitoring progress *during* the outreach program will provide opportunities to make changes that might impact the overall level of success. Appendix M, Sample Exit Questionnaire, provides sample questions for an end of class survey to assess progress toward educational and behavioral objectives. Results from the exit questionnaire can be compared to the audience assessment (Appendix H), conducted prior to the training class that provided a baseline from which to compare.

Replication: If your outreach program is a pilot project, process evaluation will be important for effective replication of the program in other communities or locations. Here, the role of the process evaluation is to document the day to day operation of the program. If results of your outreach are successful and you can say – "It works!" – the descriptive information you gather here will answer the question – "What

works?" The description might be informal, such as a written outline generated from the implementation plan that is periodically updated to describe what actually happens. This serves as an historic record and a realistic picture of the time, staff, resources, problems, and successes involved. See the Stage 4 Tool Kit, "Workform for Process Evaluation Objectives," for sample evaluation questions regarding replication.

Summative Evaluation Objectives

While process evaluation questions help determine how well outreach is working while it is ongoing, summative evaluation helps determine what outreach accomplished.

Very generally, summative evaluation questions address:

- Did outreach meet its objectives?
- What differences (i.e. outcomes) resulted?
- Are the outcomes beneficial or deleterious? To whom?
- Are the outcomes those originally envisioned?

The purposes for a summative evaluation can range from making judgments about overall program effectiveness (were objectives reached?) to discovering program effects (whether or not predicted by objectives).

*Overall program effectiveness***:** Monitoring and compiling a final tally of whether goals and objectives have been achieved is one of the basic purposes of a summative evaluation. Note that monitoring progress toward objectives is also one purpose of process evaluation; however, in the process evaluation this progress need only be spot checked. For a summative evaluation, data should be collected from a representative sample of outreach sites or participants so that staff will have good information to describe what the program achieved, and documentation about whether it met its goals.

See Appendix N, "Sample Ways to Measure Outcomes," for an illustration of how objectives might be tracked. Appendix O, "Sample Measures of Behavior Outcomes," provides sample questionnaire items that will measure outcomes for objectives related to behavior.

*Program effects – what else happens as a result of outreach***:** Summative evaluation questions might also help determine the impact of outreach on variables not addressed by objectives, to provide a broader perspective.

For example, one objective might be: "at least 25% of participants will report that outreach training influenced the way they subsequently obtain information for patient care decisions." Note that this objective does not specify what *type* of patient care decision is influenced. Data about the type of decision might be collected in a summative evaluation and reported to a hospital administrator or other interested party.

Another example of variables not included in program objectives that could be assessed in a summative evaluation is impact on worklife, such as job productivity (see Anderson et al. 1993 for survey examples to measure impacts on worklife)(3).

The point is that summative evaluation can be designed to measure whatever outcomes are of interest. Planners may want to collect information about unintended outcomes, to provide a rich picture of the impact of outreach. For example, an open ended question might ask "what happened that was not expected (either positive or negative)?"

Evaluation Methods

Discussions of evaluation methods are typically characterized by the definition of two types of data: quantitative and qualitative. Each type of data is useful in both the extensive and intensive data collection approaches introduced in Stage 1 and reviewed here.

With *extensive data collection,* much is already known about the situation and the possible variables or factors involved. The purpose is to collect data about a community that can be considered truly representative of the entire user population. Data collected can be both qualitative and quantitative (described below). Statistical validity and reliability are key criteria, meaning that the research instrument measures exactly what was intended and, if repeated, results would be the same or very similar. Random sampling is also important, so that all people being researched have an equal chance of responding. (For more discussion of random sampling, see Appendix C).

In situations where little is known about the phenomena being studied, it may be helpful to use a more exploratory data gathering approach called *intensive data collection*. The purpose here is to understand patterns of behavior or identify particular impacts or problems impeding desired results. With intensive data collection, you want a practical understanding of what is happening, but not to make generalizations. You can get both qualitative and quantitative feedback that does not strive for statistical validity, but does provide data to help understand your audience.

Each approach can use a mix of quantitative and qualitative methods, described next.

Quantitative method

Quantitative methods produce numerically based data, such as counts, ratings, scores, or classifications. Examples of quantitative data would be numbers of outreach participants reached, percentage of users satisfied with class instruction, pretest scores about attitudes towards computers, or percentages of users who indicate increased use in a followup survey.

Quantitative methods provide systematic and standardized way of gathering data, through the use of predetermined categories into which all responses must fit. Surveys are typically used to gather quantitative data.

Extensive data collection approaches might use quantitative data in an experimental research design to compare results of the intervention group with those of other programs or groups. The components of an experimental research design are described in the next section. It provides a way to aggregate results statistically and make generalizations from a carefully selected research group to a larger population.

It is difficult to generalize results from one outreach evaluation to another program, however, unless the independent variable is consistent across programs. An independent variable is what the planner has control over (e.g. the intervention). The dependent variable is the outcome or what changes (e.g. use of PubMed) as a result of the independent variable. For example, if assessing the effect of class participation by opinion leaders (the independent variable) on behavior outcomes, a count of PubMed use in the following month is the dependent variable.

In programs that have standardized curriculum, such as curriculum for K-12 public schools, outcomes (such as standardized test results) can be measured with high validity and reliability using quantitative methods based on experimental design.

However, outreach programs tend to be tailored and customized to the unique and specific needs of the target audience and not based on standardized outreach curriculum. Therefore, what might be measured with high validity and reliability for one outreach program may not be important or indicative to all programs. (4).

Qualitative method

The qualitative approach is based on the need to discover rather than to test the impact of programs (5). The goal is to develop an understanding about what is happening during implementation of a program and how, as well

as why, results are or are not achieved.

Qualitative methods consist of at least three kinds of data collection:

1. In-depth, open-ended interviews or focus groups
2. Direct observation
3. Written documents, such as open-ended survey questions, personal diaries, and outreach records

The descriptive information collected is then organized into major themes, categories, and case examples through content analysis and other methods.

Qualitative research is a good method to use for understanding the meaning of a program and its outcomes based on the participants' own words instead of predefined responses. Using qualitative methods will help gain a better and perhaps more genuine understanding about participants' opinions or behaviors.

The credibility of qualitative methods depends on the methodological skill, sensitivity, and training of the evaluator. As with quantitative methods, achieving valid and reliable measures involves systematic and rigorous techniques. For a thorough and easy-to-use discussion about qualitative methods, see "How to Use Qualitative Methods in Evaluation" by Michael Quinn Patton (6).

Combining quantitative methods with a qualitative approach, described next, can provide information in greater depth than use of either method alone.

In a 1989 evaluation by the National Library of Medicine (NLM), researchers used qualitative data as the primary descriptive information, with quantitative data as a supplement. NLM used the Critical Incident Technique (CIT), in which 552 users of MEDLINE responded to a highly structured set of open-ended questions via telephone interviews. The purpose of the study was to develop a detailed understanding of the impact of MEDLINE-derived information – in what ways it is used, and with what effect. The interview technique provided a detailed understanding of user motivation and behavior, which can be determined only very generally if using traditional survey methodology with quantitative techniques (pre-defined response categories).

Quantitative techniques in the CIT study included pre-coded responses to characterize interviewees on such variables as specialty, work setting, community size, and the nature and extent of MEDLINE searching experience (7). Thus, the CIT study shows how qualitative methods can be usefully combined with quantitative techniques, offering ways to better understand the needs, opinions, or experiences of study participants.

Selecting an Evaluation Design

A consideration in planning an evaluation will be whether you want to base your analysis of the data on a particular design. An *evaluation design* structures how one will assess or measure the effect of an independent variable on a dependent variable(s); it dictates when and from whom measurements will be gathered during the course of an evaluation (8). In the health sciences, randomly controlled clinical trials use the experimental design that is quite rigorous (as explained below). Recognizing the difficulties of this approach in studying human behavior, the field of social science research offers several alternative designs that are considered by many to be preferable.

One consideration when determining design is *when* measurements are conducted. Options usually include a pretest/posttest, posttest only, or a time series where measurements are taken at multiple times before and after the intervention.

The advantage of a pretest/posttest or time series design is that one can determine how

much change there was from before to after the intervention, especially if results are compared between the intervention group and a control or comparison group. However, some prefer to use a posttest only design because they are afraid a pretest will sensitize individuals to respond in a certain way and may result in socially desirable responses where people indicate change because "they're supposed to" (2).

Decisions about *from whom* data is gathered will dictate whether the design is non-experimental, quasi-experimental, or purely experimental as seen in Figure 12. Some of these designs focus exclusively on outreach participants, while others compare participants (called the intervention group) with similar persons or groups (called the comparison or the control group, depending on whether random assignment is used). A common and practical approach is to focus only on the intervention group—collecting data after the intervention, or both before and after (the "nonexperimental design"). A more rigorous way to determine the

Figure 12: **Evaluation Designs**

I. Experimental design								
1. Pretest-posttest design								
-Intervention group	®	O	X	O				
-Control Group	®	O		O				
2. Posttest-only design								
-Intervention group	®		X	O				
-Control group	®			O				
3. Time series design								
-Intervention group	®	O	O	O	X	O	O	O
-Control group	®	O	O	O		O	O	O
II. Quasi-experimental design								
1. Pretest-posttest design								
-Intervention group		O	X	O				
-Comparison group		O		O				
2. Time series design								
-Intervention group		O	O	O	X	O	O	O
-Control group		O	O	O		O	O	O
III. Nonexperimental design								
1. Pretest-posttest design								
-Intervention group		O	X	O				
2. Time series design								
-Intervention group		O	O	O	X	O	O	O

Key: ® = Random assignment
O = Measurement

effects of a treatment is to compare results of those who receive outreach with similar persons who do not receive it (the "quasi-experimental design"). The experimental design requires that participant and non-participant groups are comparable by assigning people randomly to the intervention group and the comparison (or "control" group).

Experimental design

The most rigorous design is the powerful comparison between individuals or groups randomly assigned to intervention and control conditions. The advantage of this design is that random assignment ensures valid and accurate comparison of results. The disadvantage of this design are the difficulties, practically speaking, of achieving random assignment.

In random assignment, it is presumed that any pre-existing differences among subjects (skill level, intelligence, race, etc.) will be evenly distributed between the intervention and control groups. Random assignment avoids "selection bias" that may be an issue when, for example, individuals self-select into one or another group based on pre-existing characteristics such as familiarity with computers.

Random assignment also controls "threats" to the validity or accuracy of results. For example, how do you know that your intervention alone caused increased usage of PubMed? Perhaps a new promotion by America Online featuring free Internet access caused the increase in usage and not your persuasive message.

How random assignment is achieved

Random assignment can occur at the individual level (i.e., each person may or may not receive the intervention) or at the group level (i.e., different groups may or may not receive an intervention). If there is concern that members of a group will talk to each other about an intervention, then it is best to randomly assign by the group instead of by the individual. Otherwise, if those in the control group were exposed to the intervention through friends or colleagues, you will not get a clear picture of how the intervention worked.

Typically, each subject or group is given a number from one on up and then a random numbers table (which may be found in the back of any basic statistics text) is consulted to place subjects in either intervention or control group. An arbitrary decision is made beforehand, which numbers in the table will be the control group and which will be the intervention group (e.g., odd entries = intervention, even entries = control).

Alternatively, one can simply place each person or group's name on a piece of paper, throw the names into a hat, and designate the first 20 draws as the intervention and the next 20 draws as the control group, and so on.

Quasi-experimental design

Random assignment is the key feature of an experimental design, distinguishing it from a *quasi-experimental* design in which a *comparison* group is included but participants, though they are as similar as possible to the intervention group, are not randomly assigned.

In most outreach situations, it may not be possible or ethical to randomly assign participants to a control group, so the quasi-experimental design is a good option. For example, one can create comparison groups by dividing potential participants into several groups and staggering the intervention. Individuals or groups should still be matched on various characteristics (like demographics) and then compared for results.

A quasi-experimental design results in interpretable and supportive evidence of outreach effectiveness, but usually cannot control for all factors that affect the validity of results. For example, if variations exist between the groups, it may be because of the intervention (you hope) or it may be because of other unique, idiosyn-

cratic factors (e.g., one group has unrestricted access to the Internet, the other does not). There are ways to statistically control for known covariates (influences on outcomes), but it is best to randomly assign groups or individuals to either the intervention or control group.

For either the experimental or quasi-experimental design, the size of the intervention and control or comparison groups is determined according to "power" estimates. Specifically, you want enough people per group to detect significant differences between the groups, if in fact significant differences exist. Usually a minimum of 20 per group can provide an adequate degree of power for attitudes toward an intervention; however, it is best to consult power tables when determining how many individuals or groups you need per condition, given a specific outcome (2).

Non-experimental design

If it is impossible to assign a control or comparison group for your research, you can use the one-group pretest/posttest approach. This design is relatively inexpensive and easy to administer. However, it is a weak design if trying to answer questions such as:

1. How good are the results? Could they have been better? Would they have been the same if the outreach had not been carried out?
2. Was it the outreach that brought about these results or was it something else?

Time series measurements of a single intervention group can provide better information than a simple pretest/posttest. For example, surveys may be administered to a sample of randomly selected individuals of an intervention group at multiple times before and after an intervention.

How Much Evaluation is Feasible?

A number of factors may affect the feasibility of an evaluation, including:

- Costs
- Staffing
- Timing
- Political or ethical considerations

A good baseline rule is that five percent or more of a program's budget should be allotted to program evaluation activities (9). Different evaluation designs require different levels of resources, as seen in Figure 13.

Reisman describes key implementation factors that influence the amount of resources required, including:

- Number of participants
- Frequency of data collection
- Length of time for which data will be collected
- Number of data collection instruments involved
- Availability of existing sources of data
- Availability of staff with data analysis skills or access to computers and statistical consultants
- Ease of administering data collection instruments
- Willingness of outreach participants to contribute to the evaluation.

Decisions related to selecting an evaluation design should consider implementation factors as well as timing and staffing requirements. Political or cultural considerations of your targeted audience are also important (see page 62 for further discussion of cultural factors in data collection).

Figure 13: **Level of Resources for Various Evaluation Designs**

Type of Design	Description	Disadvantages	Advantages	Resource Intensity
Post-Outreach Measures	Use of evaluation tools to describe outcomes (e.g., behavior, attitudes, or knowledge) *following* outreach	No comparison with people not exposed to outreach No certainty that outcome has changed (may have been the same prior to outreach)	Simple to administer Inexpensive	Low
Post-Outreach Measures with a Control Group	Same as described above, with the addition of collecting similar scores for a *control group*	Using a control group requires additional research participants Additional participants will not receive the outreach (unless it is offered to them at a later point) It is difficult to randomly assign outreach participants	Avoids pre-test sensitization Strong basis for comparison, so if there are differences in outcomes between the groups, can have confidence that outreach had some effect	Moderate
Pre- and Post-Outreach Measures	Describes participants' "scores" on expected outcome variables (e.g. behavior, attitudes, or knowledge) both *prior* to and *following* outreach	Changes in scores could be due to some other source (e.g. media promotion of health resources) No comparison with people not exposed to outreach	There is some basis for comparison (before and after) Every participant receives outreach	Moderate
Pre- and Post-Program Measures With a Control Group or Comparison Group	Same as described above, but with the addition of collecting similar scores for a *control group* or a *comparison group*	Using a control or comparison group requires additional research participants Additional participants will not receive the outreach (unless it is offered to them at a later point) It is difficult to randomly assign outreach participants to a control group If comparison group used (not randomly assigned), cannot control all factors affecting validity	Strong basis for comparison, so if there are differences in outcomes between the groups, can have confidence that outreach had some effect	High
Multiple Pre- and Post-Outreach Measures (Time Series)	Same as pre- and post-outreach measure approach, with additional scores *obtained several times before and several times after the intervention*	Additional measures must be obtained If obtaining behavioral measures, need to allow sufficient time to measure behaviors before intervention can occur	Helps to validate whether changes in outcomes sustain over time Helps to obtain a more complete picture of dependent variables before intervention occurs.	High

References

1. King JA, Morris LL, Fitz-Gibbon CT. How to assess program implementation. (Second ed.) Newbury Park: Sage Publications, 1987. (Herman JL, ed. The Program Evaluation Kit; vol 3).
2. Witte K. Theory-based interventions and evaluation of outreach efforts. Seattle, WA: National Network of Libraries of Medicine, Pacific Northwest Region Web site. http://www.nnlm.nlm.nih.gov/pnr/eval/witte.html, 1998.
3. Anderson JG, Aydin CE, Jay SJ. Evaluating health care information systems: methods and applications. Thousand Oaks: Sage Publications, 1994.
4. Dignan MB, Carr PA. Program planning for health education and promotion. Philadelphia: Lea & Febiger, 1992:164.
5. Glitz B. Focus groups for libraries and librarians. New York: Forbes, 1998.
6. Patton MQ. How to use qualitative methods in evaluation. Newbury Park, CA: Sage Publications, Inc., 1987. Program Evaluation Kit; vol 4).
7. Siegel E, Rapp B, Lindberg D. Evaluating the impact of MEDLINE using the Critical Incident Technique. Proceedings of the annual symposium on computer applications in medical care 1991:83-87.
8. Fitz-Gibbon CT, Morris LL. How to design a program evaluation. Newbury Park: Sage Publications, 1987. (Fitz-Gibbon CT, Morris LL, eds. Program Evaluation Kit; vol 3).
9. Reisman J. A field guide to outcome-based program evaluation. Seattle: Organizational Research Services, Inc., 1994.

Selected Readings

Berg BL. Qualitative research methods for the social sciences. Boston: Allyn and Bacon, 1995.

Herman JL. Evaluator's handbook. (2nd ed.) Los Angeles: Center for the Study of Evaluation, 1987. (Program Evaluation Kit; vol 1).

Herman JL. Program evaluation kit, vol 1-9. (2nd ed.) Newbury Park, CA: Sage Publications, 1987.

Hernon P, McClure CR. Evaluation and library decision making. Norwood, NJ: Ablex Publishing Corporation, 1990.

Isaac S, Michael WB. Handbook in research and evaluation : a collection of principles, methods, and strategies useful in the planning, design, and evaluation of studies in education and the behavioral sciences. (3rd ed.) San Diego, Ca: EdITS Publishers, 1995.

Patton MQ. Utilization-focused evaluation. Beverly Hills: Sage Publications, 1978.

See Appendix K for a filled-in example

ACCOUNTABILITY

Will I be accountable for documenting what occurred as the program happened? If so, what is most important to document?

a. Briefly describe the program's goals and objectives (*Ask evaluation stakeholders to verify or modify)*

b. What do you see as the most important results or outcomes of the program? (*Ask evaluation stakeholders to verify or modify)*

c. How will the program be implemented? Describe the resources, activities, services, and administrative arrangements that constitute the program.

Accountability measures: Obtain periodic updates on characteristics of the program (context, activities, and best practices) that will most determine its success. (*Determine in advance what the report questions will include. Ask evaluation stakeholders to verify or modify)*

Context: tangible features of the outreach program and its site

- _____
- _____
- _____
- _____

Activities: how the program is being implemented

- _____
- _____
- _____
- _____

Best practices: what is being done to leverage success?

- _____
- _____
- _____
- _____

PROGRAM IMPROVEMENT

Will there be an opportunity to make adjustments to the activities and strategies targeted at program objectives? If so, how can progress toward objectives be tracked? *Ask yourself and your staff*:

a. What are the outcomes listed in each objective?

b. What indicators will provide measurable evidence of those outcomes?

c. How can those indicators be tracked?

d. What variables can be measured to show whether the theory-based strategies are working? (Review objectives and strategies identified in the implementation plan outline developed in Stage 3 - see Appendix I for an example).

REPLICATION

Is the outreach program considered a pilot project, or is it likely to be replicated at another site? If so, what types of information would be most useful to track for eventual documentation? Check off the types of information to track from the following list, and ask relevant stakeholders to add other data you may want to collect:

- ❑ Where exactly has the outreach program been implemented and what was done?
- ❑ How many and what sorts of people participated in the outreach? (e.g. age, sex, health profession)
- ❑ What are the characteristics of their information needs? (e.g. type of practice, types and purposes of information needed, frequency of information needed, sources used)
- ❑ What are the socioeconomic characteristics of the setting?
- ❑ What does (do) the outreach site(s) look like?
- ❑ What are the program's greatest successes? What facilitated each one?
- ❑ What are the program's biggest challenges (frustrations, barriers, or disappointments)?
- ❑ What sociopolitical factors may have impacted the outreach?
- ❑ What were the outreach costs in staff time, materials, equipment, and facilities?
- ❑ Other questions?

In Stage 3, your library staff at Gowan Library thought about their strategies and activities for reaching the objectives of the outreach program. At this point, you are on the way to beginning the program. However, you know this is the best time to begin thinking about the project evaluation. Careful consideration at this early stage will help make sure that the right data will be collected. For example, it is soon time to conduct the audience assessment discussed in Stage 3 that will help to tailor the educational activities planned. Staff already have some ideas about what they want to find out in the audience assessment. But, before conducting the assessment, think through the questions to be asked for the project evaluation. Is the audience assessment an opportunity to collect baseline data *before* the outreach training that can then be compared to results or outcomes at the end?

To begin considering what your project evaluation will assess, you list who would be interested in evaluation results, including:

- Geneva Health administrator
- State chapter of the primary care association
- Regional rural health association
- Funding agency
- Gowan Library outreach staff
- Gowan Library director
- Health librarian community

With this list in mind, you consider what these individuals might want from an outreach evaluation. For example, the evaluation question—were objectives reached?—may be of interest to several people, such as the funding agency and you, the director. This phase of evaluation is called the summative evaluation — asking questions about what happened in the overall picture, such as did outreach meet its objectives and what were the outcomes? The types of data collected for this phase of evaluation might include a comparison of pre- and post-measures of attitudes, awareness, skills, and behaviors, measured both during the audience assessment and in a followup after outreach training is completed. Other outcomes are tallied throughout the program (such as number of classes conducted). These measures also contribute to an overall summative assessment.

In addition to evaluating results, much is learned by tracking ongoing progress, so that you can identify what works well, what does not, and what can be improved as the project is ongoing. This phase of evaluation is called the process evaluation.

You find that the task of figuring out what evaluation questions to ask takes careful consideration before you can specifically define what you will measure. General questions, such as “were we successful?” is not meaningful until you define your criteria for success very specifically. Fortunately, you can look at the objectives you constructed in Stage 2 that include measurable indicators. But, you also want to evaluate other interesting data that will help you improve another similar outreach program in the future. You think about how you designed this outreach program—there were several assumptions you made in thinking through the whole process. For example, your plan to develop onsite expertise for information services support is a worthy objective. But, what if it doesn’t work? How will you know what went wrong? You realize you must think about what data might be helpful to collect along the way to help examine reasons for whatever results transpire.

You also realize that data collection requires effort and it is important to avoid asking evaluation questions if the answers will not be useful to you for making decisions or improvements. Too many measures might dilute your evaluation resources, and you will avoid asking questions just because they are "interesting." You have decided that you do not plan to use the results to make generalizations about any outreach program targeted to primary care clinics. You want practical results that will help you understand what would appear to be happening in your project only. Going any farther than that means using highly structured techniques or methods designed for statistical validity, such as control or comparison groups. At this exploratory level of research, you do not want to extend the evaluation resources necessary to conduct that type of rigorous research.

Finally, after figuring out what you really want to know from an evaluation and what you will do with the answers, your next step is deciding the types of data you need to collect and how you will do that. Stage 4 provides a discussion of various evaluation methods, some more rigorous than others. There are a range of possibilities and the planning tools in Tool Kits for Stage 4 and 5 and Appendices K through O help to think through what will be measured.