PLANNING AN EVALUATION AND ESTIMATING ITS COST

JOSEFINA J. CARD
CATHERINE GREENO
JAMES L. PETERSON
Social Research Applications

Scientific evaluations can provide funders and program administrators with useful input concerning where to allocate scarce service dollars. But evaluation itself is legitimately the subject of cost concerns. What will it cost to get the various potential benefits of a scientific evaluation? What are the evaluation options of a given program, given its budget and the size and expertise of its staff? Where can a program administrator go for consulting help? This article provides a helpful framework for answering these important questions. First, it describes two types of evaluations which vary in both the questions they can answer and in their consequent cost. Second, it delineates and briefly describes the technical elements or steps required by each evaluation type. Third, it describes the nature and potential variability of costs associated with each technical step. Finally, it steers the reader to available sources of expert information and help.

AUTHORS' NOTE: This article was produced with funds provided by the Charles Stewart Mott Foundation, the Northwest Area Foundation, the Henry J. Kaiser Family Foundation, and the William and Flora Hewlett Foundation. The authors wish to thank Marvin B. Eisen for his helpful review of earlier drafts of this article and Evelyn C. Peterson for production assistance.

EVALUATION & THE HEALTH PROFESSIONS, Vol. 15 No. 4, March 1992 75-89

As competition for scarce service dollars grows, the benefits of conducting a scientific evaluation of a program's effectiveness are becoming increasingly apparent. Both private and public funders have begun looking toward scientific evaluations to help them make judgments concerning what to fund, how to fund, and for how long to fund. Program administrators are similarly coming to realize that they have a lot to learn from the results of such scientific investigations. Interestingly, the very same information on the extent and quality of service delivery needed for a good evaluation can be used to inform program personnel where their dollars are going at present and where such resources might be shifted in the future.

A legitimate set of questions raised by many interested funders and program administrators centers on cost issues: What will these benefits of evaluation cost me? What are my options, given my budget, and the size and expertise of my staff? Where can I go for consulting help?

The purpose of this article is to provide a helpful framework for answering these important questions. First, we describe two types of evaluations which vary in both the questions they can answer and in their consequent cost. Second, we delineate and briefly describe the technical elements or steps required by each evaluation type. Third, we describe the nature and potential variability of costs associated with each technical step. Finally, we steer the reader to available sources of expert information and help.

Our own experience stems from providing technical assistance to dozens of teen pregnancy prevention and care programs around the country, and our examples will draw on these experiences. However, the technical and cost issues we discuss are generic to most other programs in the health arena.

THE IMPORTANCE OF ADVANCE PLANNING

Obviously, one article cannot answer all the questions a funder or program administrator may have concerning the costs and benefits of evaluation. However, our hope is that this article will encourage program personnel to consider evaluating their work, and — by providing a common framework and rationale for assessing the benefits and

costs involved—lead to greater dialogue between program personnel and expert evaluators.

Ideally, such dialogue should take place prior to setting up a program. Programs will almost always need to collect data on their clients—on the people they provide services to—be they pregnant teenagers, babies with AIDS, or any other clients. With appropriate advance planning, the very same data collection system that enables a program to monitor the provision of services to clients can be expanded to form an evaluation-related data collection system. The end result is a cost-effective enterprise.

It is generally costly to backtrack, or to discover late in the game that one cannot achieve or afford what one set out to do. A small investment up front, in the form of collaborative advance planning with an evaluation expert, will likely pay off in the future. A good rule to bear in mind is that it is always better to do a small job well than to do a big job poorly. The challenge is to design and implement the most powerful monitoring or evaluation system consistent with one's dollar and personnel constraints.

TWO LEVELS OF EVALUATION

In evaluation, a program's performance is measured against a performance standard. The standard can be a set of services to be delivered (a process evaluation), or a set of goals and objectives to be met (an impact evaluation). It is through evaluation that we can learn whether the program is being delivered as planned (process evaluation) and whether the program is having its intended effect of, for example, delaying sexual intercourse, encouraging use of contraception among the sexually active, or preventing teen pregnancy (impact evaluation).

PROCESS EVALUATIONS

Because the goal of a process evaluation is to discover whether or not service delivery goals are being met, it examines accurate and complete records monitoring service delivery and compares the quan-

tity and quality of services delivered against the program's service delivery objectives. Many process evaluations also determine who is receiving services and clients' level of satisfaction with various aspects of the program. Process evaluations are useful both as a program development and a management information tool.

IMPACT EVALUATIONS

Impact evaluations are designed to test whether or not the program is effective: Does the program work? For whom? Under what circumstances? Only impact evaluations can provide information about whether a program is actually benefiting clients in the manner prescribed by the program's goals and objectives.[1] Impact evaluations are more costly and complex than process evaluations, but they also provide more information. For an evaluation to be considered an impact evaluation it must have several features including: (a) inclusion of measures of the program or intervention as delivered (similar to process evaluation), (b) inclusion of objective measures of each of the program's short-term and long-term objectives, and (c) inclusion of data from a comparison group.

TECHNICAL AND COST ELEMENTS OF PROCESS AND OUTCOME EVALUATIONS

Figure 1 lists the technical components or elements required by process and impact evaluations. It shows that the following technical elements are common to both types: program model development, research design planning, instrument development, instrument pretest, interviewer training, data collection at program intake and exit, data management, data analysis, and report writing. However, impact evaluations require four elements not required by process evaluations: sampling and subject identification and selection, collection of data from a comparison group, a minimum sample size, and postprogram follow-up data collection from program dropouts and graduates. Because it costs money to implement each of these extra elements, impact evaluations will generally cost more than process evaluations. But, as we have seen, impact evaluations can also answer a greater number

Technical Elements	*Required For Process Evaluation?*	*Required For Impact Evaluation?*	*Cost Elements, Personnel*	*Cost Elements, Nonpersonnel*[a]	*Variability of Cost*
Program Model Development	yes	yes	Entire program & evaluation staff	none	low
Research Design Planning	yes	yes	Evaluation expert; program director	none	low
Instrument Selection/ Development	yes	yes	Evaluation expert; clerical staff	Library; acquisition	high
Instrument Pretest	yes	yes	Evaluation or program staff	Photocopy; travel; mail; phone; subject payment	moderate
Interviewer Training	yes	yes	Evaluation expert	Photocopy; travel	moderate
Sampling and Subject Identification	no	yes	Evaluation expert	Possible phone; mail; travel	moderate
Data Collection			Evaluation or program staff	Photocopy; travel; mail; phone; subject payment	
Subject Recruitment/ Permissions					high
Comparison Group	no	yes			high
Minimum Sample Size	no	yes			moderate
Time Frame Intake Data	yes	yes			high
Postprogram Follow-up of Dropouts and Grads	no	yes			high
Length of Instrument	yes	yes			high
Mode of Data Collection	yes	yes			moderate
Data Management	yes	yes	Evaluation or program staff	Microcomputer; word processing or data base management software	high
Data Analysis	yes	yes	Evaluation staff	Microcomputer; statistical analysis software	high
Report Writing	yes	yes	Evaluation expert; program director	Microcomputer; word processing software	high

Figure 1: Technical and Cost Elements in Designing and Conducting an Evaluation

a. Other than standard office equipment and consumable supplies.

Evaluation Steps	*Estimate of Required Staff Time*
Program Model Development	1-5 days per year per staff member
Research Design Planning	1-10 days for the program director and evaluation expert
Instrument Selection/Development	Varies widely: From one week to several months
Instrument Pretest	One day to one month
Interviewer Training	half day-14 days
Sampling & Subject Identification	Varies widely: From one week to several months
Data Collection	Varies widely: From one week to several months
Data Management	Varies widely: From one week to several months
Data Analysis	Varies widely: From one week to several months
Report Writing	Varies widely: From 2 days to several months

Figure 2: Estimates of Staff Time Required for Program Evaluation

of questions than process evaluations. In order to determine if a process or impact evaluation best meets your needs, the technical requirements and associated costs must be considered. The following sections describe, and Figures 1 and 2 summarize, these requirements and costs.

PROGRAM MODEL DEVELOPMENT

Technical requirement. A program model is a diagram that contains a succinct portrayal of: (a) the intervention, (b) the population that will get the intervention, and (c) intermediate and long-term objectives of the intervention. Often, mediating variables — other variables (such as age, race, sex) expected to affect the outcomes — are also included. Without a program model, no evaluation is possible, because the model defines what is to be evaluated; what independent, intervening, and dependent variables to measure. Even if program effectiveness is not subsequently measured, program model development is an essential first step that is likely to lead to better administration of a program.

An illustrative program model is given in Figure 3. Program Lookout is a hypothetical curriculum that consists of 10 two-hour sessions and includes education on reproduction and skills training to help teens communicate with their parents and say no to sexual advances. A program model can be expected to evolve as the program develops. For example, as experience with Program Lookout is gained, sessions may be dropped, added, or modified in response to client interest or

Interventions	*Population*	*Mediating Factors*	*Short-Term Goals*	*Long-Term Goals*
LOOKOUT:	13- to 15-year-olds	Age	Increase knowledge	Delay first intercourse
A 10-session curriculum stressing reproductive knowledge, decision skills, and parent-child communication[a]		Grade School	Increase decision skills Increase parent-child communication	Reduce teen pregnancy

Figure 3: Sample Program Model
a. A booklet describing each session should accompany the program model.

early clear indications of effectiveness. When that decision is made, the change can be recorded on a revised program model. Regular staff meetings during which everyone is brought up-to-date on progress and program changes can help ensure proper implementation of the program.

An examination of the program model presented in Figure 3 shows that much important program information can be contained in a simple figure. It is not difficult to create a program model; yet, going from the program description that exists in other program documents, such as a grant proposal, to a program model sometimes leads to surprising results.

The most challenging aspect of transforming the program description into a program model usually comes in specifying the goals, especially in specifying the short-term goals. The short-term goals of a program are a crucial statement of how the intervention is expected to cause the desired long-term changes. Specifying these goals can be difficult, and many programs never even realize that such goals have never been specified at all. The process of reaching consensus with program personnel about the intervention and its goals – and regularly updating this consensus – is a prerequisite of effective program implementation as well as competent evaluation.

Associated costs. The cost of program model development is low and the payoff high. All programs – those doing process evaluations, impact evaluations, or no evaluation at all – should undertake this step. The entire program and evaluation staff should be involved in developing the program model at the start of the program, as well as in

updating the model at periodic (semiannual, for example) intervals, if required. Variability of cost is low: Regardless of the size of a program, this step will take approximately the same level of effort – about 1-5 days per year per staff member.

RESEARCH DESIGN PLANNING

Technical requirement. This step includes making the basic decisions regarding program evaluation: Will I evaluate my program? What type of evaluation – process or impact – should I undertake? How will I develop my research questionnaires or instruments? How will I recruit my subjects? Will I have a comparison group? Will I attempt an experimental design? How far into the future do I want to follow up my subjects? In short, this step encompasses a program's plans for all the subsequent steps to be described below. In keeping with our advice to do advance planning whenever possible, this step should be undertaken at the start of the program, immediately after consensus has been reached on a program model. Many programs will have been funded with an evaluation research design already in place (as part of the funded proposal). Such programs only have to review the evaluation plan in the funded proposal, to see if there are any changes staff members would like to make, now that the funds to proceed have been obtained.

Associated costs. Research design planning is best undertaken jointly by the program director and the program's evaluation expert (the latter could be an outside consultant or the head of an in-house evaluation staff). Cost variability across programs is low: in the range of 1-10 days of planning each for the program director and evaluation expert. As was the case with the program model component, the planning component is frequently iterative, with intermittent adjustments made as the plan faces implementation realities.

INSTRUMENT SELECTION AND DEVELOPMENT

Technical requirement. In this step measures are selected or developed for the independent, intervening, and outcome variables specified in the program model. The measures could take many forms: for

example, questionnaires to be filled out by subjects themselves; questionnaires to be filled out by subjects' parents, teachers, doctors, counselors, or other knowledgeable individuals; abstracts of subjects' medical records or school records transferred to standard forms by a research assistant; in-person or telephone interviews of subjects. Some of the measures could be copied or adapted from similar work by previous investigators, others devised especially for a given program evaluation.

Associated costs. Instrument selection or development is generally done by an evaluation expert with the assistance of clerical staff. The cost of this step can vary widely as a function of both the number of variables to be measured and the extent of original development required. Small, simple evaluations are frequently well-advised to select from existing measures in the public domain. For example, in the teen pregnancy program area, recommended minimum data sets exist for the evaluation of both pregnancy prevention and pregnancy care programs (Eisen & Brindis, 1989; McBride, 1988). In contrast, evaluations of larger programs with more complex evaluation research designs would likely encompass both selection of existing measures for some variables and development of original measures for others. The cost of this step could thus vary widely across programs, from about a week to several months. A good rule to follow is: don't reinvent the wheel. If a good measure exists in the public domain for a variable in your program model, use it as is, or—if possible—adapt it to suit your population.

INSTRUMENT PRETEST

Technical requirement. Whether measures are adapted, original, or a hybrid, the resulting instrument should always be pretested (tried out) on one's own subjects or on individuals drawn from the same population pool as one's subjects. This is the only way to find out whether the instrument is too long, whether certain questions are unclear, whether subjects find some questions too intrusive, whether the reading level required of subjects is appropriate, whether subjects can actually be reached by telephone (if a telephone interview), for example.

Associated costs. Pretesting of an evaluation instrument will likely be done by the evaluation or program staff. This has two benefits. First, staff will become more familiar with the instrument and with subjects' reactions to it. Second, staff will get firsthand experience with the nonpersonnel costs of the data collection effort, such as instrument photocopying costs, mail, telephone, or travel costs, and subject payment costs (if required). The cost of the instrument pretest step will vary as a function of the number of pretest subjects, the mode of data collection, and the length of the instrument: Obviously an in-person one-on-one hour-long interview is more expensive to administer than a 10-minute paper-and-pencil questionnaire filled out by an entire classroom of subjects in one sitting. Because instrument pretesting is generally done on only a small number of subjects, the variability of such costs at the pretest stage is moderate, generally about one person-day to one person-month. An exception is the very large, flagship evaluation where several cycles of pretesting and instrument revision might involve several staff members over a period of several months.

INTERVIEWER TRAINING

Technical requirement. In this step the evaluation expert, or whoever was primarily responsible for instrument development, trains the data collection staff on how to administer the instrument properly.

Associated costs. Trainer and trainee compensation are the primary costs of this step. There are also minor nonpersonnel costs such as photocopying and travel costs (if training is to be done in subjects' homes). Such costs can vary in moderate fashion across various evaluations. There could be significant variance in training time: It clearly takes less time to train staff to administer a short, simple instrument (as little as a half day perhaps) than a long, complex instrument (as much as two weeks). On the other hand, regardless of the length and complexity of the instrument, this step is done infrequently: up-front, and whenever there are new interviewers, or a new data collection round. If the study is of any duration, interviewers' performance should be periodically checked.

SAMPLING AND SUBJECT IDENTIFICATION

Technical requirement. All the preceding steps—program model development, research design planning, instrument development, instrument pretest, and interviewer training—are required whether one is doing a process or an impact evaluation. In contrast, sampling and subject identification need to be done only if one is attempting an impact evaluation. Process evaluations are interested only in information gathered from current clients, and therefore sampling and subject identification are not required. Impact evaluations, in contrast, require that data be collected from dropouts and program graduates, as well as members of a nonclient comparison group. Impact evaluations require that such data be collected through the end of the evaluation time frame, whether this be six months, a year, two years, or ten years. Moreover, impact evaluations require both low attrition (subject loss to the data collection) and sufficient power (minimum sample sizes) during all these data collection periods: a sufficient number of subjects in both the treatment and comparison groups to (a) adequately represent the group and (b) detect group differences at statistically significant levels, when such differences exist. Thus, impact evaluations require that decisions be made on whom to study and for how long they should be studied, and these are precisely the tasks encompassed by the sampling and subject identification step.

Associated costs. An evaluation expert is generally the person to direct this step. The costs involved can vary from a week (for a small program with a handy comparison group) to several weeks (a medium-sized program that has to search for a suitable comparison group) or months (a large-scale evaluation that involves location of several comparison groups, or screening by mail, telephone, or house-to-house visits to find eligible subjects).

DATA COLLECTION

Technical requirement. This step assumes that the preceding steps have all been accomplished, that a program has a pretested instrument measuring all relevant variables in its program model; that interviewers have been trained to administer the instrument; and that a subject pool has been identified.

Associated costs. Data collection is generally done by evaluation and/or program staff members. Occasionally, graduate students are trained and hired on a temporary basis to collect data. Costs associated with data collection can vary widely from one program to the next. Figure 1 lists the main determiners of cost of this step: the difficulty of recruiting subjects, or of obtaining their or their parents' permission to participate in the study; the size of the treatment and comparison group samples from whom data are to be collected; the time frame of the study and the number of follow-up data collection rounds; the length of the instrument; the mode of data collection (paper-and-pencil questionnaires administered to a group in a single setting are the least expensive, followed by mail questionnaires, telephone interviews, and in-person interviews in a school or clinic setting; in-person interviews conducted in the subject's home are the most expensive). Because process evaluations require only that data be collected from current clients, marginal costs for a process evaluation (those beyond a program's normal service- and client-related data collection system) are minimal. In contrast, marginal data collection costs for an impact evaluation can vary widely from a few hundred dollars to tens of thousands of dollars.

DATA MANAGEMENT

Technical requirement. Data management refers to the steps involved in transforming the collected data—whether the information is in the form of completed questionnaires, test scores, abstracts of medical or school records, telephone or in-person interview notes—into a clean and documented machine-readable data file, suitable for analysis with microcomputer statistical analysis software programs. The steps involved include coding the data (transforming the subjects' answers to each question into a consistent set of numbers, e.g., 1 for male and 2 for female), data entry (entering all such number codes from each subject into a computer), data cleaning (making sure that only valid numbers are included, e.g., that no subject has a code of 3 for the variable *gender*), and creation of composite variables (adding related answer codes to come up with a composite test or scale score, such as a Math Achievement Score or a Self-Esteem Score).

Associated costs. Data coding, entry, and cleaning are generally done by evaluation and/or program staff members; associated non-personnel costs include a microcomputer and word processing or data base management software. The total cost associated with data management will vary as a function of both the length of the instrument and the number of subjects, from under a person-week per data collection round for a small and simple evaluation to several person-months for a large one.

DATA ANALYSIS

Technical requirement. The clean data file produced under the previous step can be analyzed with the help of appropriate microcomputer hardware and software to answer various evaluation questions of interest. For a process evaluation, the data file can be analyzed to determine whether the program is being, or has been, implemented as planned, and whether service delivery objectives were or are being met. If the evaluation effort has met the previously described technical requirements for an impact evaluation, the data file can be analyzed to determine if the program was effective in meeting its short- and long-term objectives; which parts of the program appear to be most effective; and which subgroups of clients appear to benefit most from the program.

Associated costs. Data analysis requires previous scientific training. This step is thus generally done by the evaluation expert or a member of the evaluation staff. The ready availability of microcomputers and statistical analysis software has caused the costs associated with this step to diminish dramatically in recent years. Nevertheless the variability of data analysis costs across programs remains great. First, data files vary in size (number of included questions and subjects), as we have seen in the previous section. Beyond these intrinsic differences in the amount of data available for analysis, there are idiosyncratic preferences for how much effort to devote to "mining the data." Some evaluators opt to do only simple and high-level analyses of their data base, such as basic experimental-versus-control-group comparisons on a few key variables. Other evaluators opt to do more numerous and complex statistical analyses. Thus, marginal costs

for the data analysis step can range from a few hundred to many thousands of dollars.

REPORT WRITING

Technical requirement. Written reports on the findings of one's evaluation can range from simple five-page executive summaries to the program's funder to multiple books and journal articles written for scholarly and/or lay audiences.

Associated costs. The brief summary report can be written by a program director, but the books and journal articles generally fall within the purview of the evaluation expert. The same microcomputer hardware and software that served the data management and analysis phases of the evaluation can be used for report writing. Cost of this last step can vary widely from a couple of days for a simple five-page report to a year or more for a scholarly book or journal article series.

RECOMMENDATIONS

Evaluation options range from the simple and inexpensive to lengthy, complex, and costly enterprises. The following recommendations may help you determine which evaluation option will best serve your needs.

1. *Familiarize yourself with evaluation options and costs.* This article was intended only as a helpful beginning. Other sources of information about how to get help with evaluation are available (Card, Peterson, & Greeno, 1992). You should get consulting advice from an outside evaluator if you do not have anyone on your staff familiar with evaluation research. A local university would be a useful first place to look.

2. *Do as much advance planning as possible, both with your staff and with an outside consultant (if available), so that you get the most mileage for your dollars.* Compare the requirements for a process or impact evaluation, as described in this article, to your personnel and budget constraints, and make decisions accordingly. For example, if you do not have access to, nor funds for, an evaluator, do not attempt

to conduct an impact evaluation. If you do not have the money to establish a comparison group, or the staff to collect information over time from members of the comparison group as well as from program dropouts and graduates, do not attempt an impact evaluation. But, you may want to consider doing a process evaluation, especially if you will be collecting service delivery information anyway as part of your management information system.

3. *Always remember that it is better to do a small job well than to do a big job poorly.* Strive for the most ambitious undertaking that, in your considered opinion, you have a good chance of doing well. Remember that we all can and should learn from our experiences. A process evaluation of today's program could very well give you the confidence and experience to attempt an impact evaluation of tomorrow's replication, or tomorrow's new program.

NOTES

1. The literature occasionally distinguishes between an *outcome* evaluation, which focuses on short-term objectives (such as increasing knowledge among teens about contraception) and an *impact* evaluation, which focuses on longer-term objectives (such as the prevention of pregnancies among unmarried high school students). Because the technical requirements of these two types of evaluation are similar, with costs differing only because of the different time frames involved, we use the term impact evaluation throughout this article to stand for both types.

REFERENCES

Card, J. J., Peterson, J. L., & Greeno, C. (1992). Designing, monitoring, and evaluating teen pregnancy prevention programs. In B. Miller, J. J. Card, R. Paikoff, & J. L. Peterson (Eds.), *Preventing adolescent pregnancy: Model programs and evaluations*. Newbury Park, CA: Sage.

Bisen, M., & Brindis, C. (1989). Compilation of a prevention minimum evaluation data set (PMEDS). In J. J. Card (Ed.), *Evaluating programs aimed at preventing teenage pregnancies* (pp. 69-125). Los Altos, CA: Sociometrics.

McBride, D. (1988). The Minimum Evaluation Data Set for care programs. In J. J. Card (Ed.), *Evaluating and monitoring programs for pregnant and parenting teens* (pp. 135-151). Los Altos, CA: Sociometrics.