

# Program Operations Guidelines for STD Prevention



## Program Evaluation

# Table of Contents

---

<b>FOREWARD</b>	<b>iii</b>
<b>INTRODUCTION</b>	<b>iv</b>
Program Operations Guidelines Workgroup Members	vi
Program Evaluation Subgroup Members	vii
Program Evaluation Internal/External Reviewers	vii
<b>INTRODUCTION</b>	<b>E-1</b>
<b>PLANNING AN EVALUATION</b>	<b>E-2</b>
<b>STEPS IN DESIGNING AND CONDUCTING AN EVALUATION</b>	<b>E-3</b>
<b>TYPES OF EVALUATION</b>	<b>E-6</b>
Formative Evaluation	E-6
Evaluability Assessment	E-7
Process Evaluation	E-7
Outcome Evaluation	E-8
<i>Defining Program Effect</i>	<i>E-9</i>
<i>Designing the Study</i>	<i>E-9</i>
Randomized Trials	E-10
Quasi-experiments	E-10
Economic Evaluation	E-10
Types of Economic Evaluation	E-10
<i>Cost Analysis</i>	<i>E-10</i>
<i>Cost Effectiveness Analysis (CEA)</i>	<i>E-11</i>
<i>Cost-Utility Analysis (CUA)</i>	<i>E-11</i>
<i>Cost-Benefit Analysis (CBA)</i>	<i>E-11</i>
Measures	E-12
<b>PRACTICAL CONSIDERATIONS</b>	<b>E-13</b>
Overcoming Barriers	E-13
Resources	E-13
Ethics	E-14
Standards to Consider in a Practical Evaluation	E-14

<b>Appendix E-A</b>	<b>Logic Model</b>	<b>E-15</b>
<b>Appendix E-B</b>	<b>Examples of Objectives</b>	<b>E-16</b>
<b>Appendix E-C</b>	<b>Types and Uses of Evaluation</b>	<b>E-17</b>
<b>Appendix E-D</b>	<b>Evaluation Glossary</b>	<b>E-18</b>
<b>References</b>		<b>E-19</b>



# Foreword

---

The development of the Comprehensive STD Prevention Systems (CSPS) program announcement marked a major milestone in the efforts of CDC to implement the recommendations of the Institute of Medicine report, *The Hidden Epidemic, Confronting Sexually Transmitted Diseases, 1997*. With the publication of these STD Program Operations Guidelines, CDC is providing STD programs with the guidance to further develop the essential functions of the CSPS. Each chapter of the guidelines corresponds to an essential function of the CSPS announcement. This chapter on program evaluation is one of nine.

With many STDs, such as syphilis, on a downward trend, now is the time to employ new strategies and new ways of looking at STD control. Included in these guidelines are chapters that cover areas new to many STD programs, such as community and individual behavior change, and new initiatives, such as syphilis elimination. Each STD program should use these Program Operations Guidelines when deciding where to place priorities and resources. It is our hope that these guidelines will be widely distributed and used by STD programs across the country in the future planning and management of their prevention efforts.

Judith N. Wasserheit  
*Director*  
*Division of STD Prevention*

# Introduction

---

**T**hese guidelines for STD prevention program operations are based on the essential functions contained in the Comprehensive STD Prevention Systems (CSPS) program announcement. The guidelines are divided into chapters that follow the eight major CSPS sections: Leadership and Program Management, Evaluation, Training and Professional Development, Surveillance and Data Management, Partner Services, Medical and Laboratory Services, Community and Individual Behavior Change, Outbreak Response, and Areas of Special Emphasis. Areas of special emphasis include corrections, adolescents, managed care, STD/HIV interaction, syphilis elimination, and other high-risk populations.

The target audience for these guidelines is public health personnel and other persons involved in managing STD prevention programs. The purpose of these guidelines is to further STD prevention by providing a resource to assist in the design, implementation, and evaluation of STD prevention and control programs.

The guidelines were developed by a workgroup of 18 members from program operations, research, surveillance and data management, training, and evaluation. Members included CDC headquarters and field staff, as well as non-CDC employees in State STD Programs and university settings.

For each chapter, subgroups were formed and assigned the task of developing a chapter, using evidence-based information, when available. Each subgroup was comprised of members of the workgroup plus subject matter experts in a particular field. All subgroups used causal pathways to help determine key questions for literature searches. Literature searches were conducted on key questions for each chapter. Many of the searches found little evidence-based information on particular

topics. The chapter containing the most evidence-based guidance is on partner services. In future versions of this guidance, evidence-based information will be expanded. Recommendations are included in each chapter. Because programs are unique, diverse, and locally driven, recommendations are guidelines for operation rather than standards or options.

In developing these guidelines the workgroup followed the CDC publication “CDC Guidelines -- Improving the Quality”, published in September, 1996. The intent in writing the guidelines was to address appropriate issues such as the relevance of the health problem, the magnitude of the problem, the nature of the intervention, the guideline development methods, the strength of the evidence, the cost effectiveness, implementation issues, evaluation issues, and recommendations.

STD prevention programs exist in highly diverse, complex, and dynamic social and health service settings. There are significant differences in availability of resources and range and extent of services among different project areas. These differences include the level of various STDs and health conditions in communities, the level of preventive health services available, and the amount of financial resources available to provide STD services. Therefore, these guidelines should be adapted to local area needs. We have given broad, general recommendations that can be used by all program areas. However, each must be used in conjunction with local area needs and expectations. All STD programs should establish priorities, examine options, calculate resources, evaluate the demographic distribution of the diseases to be prevented and controlled, and adopt appropriate strategies. The success of the program will depend directly upon how well

---

program personnel carry out specific day to day responsibilities in implementing these strategies to interrupt disease transmission and minimize long term adverse health effects of STDs.

In this document we use a variety of terms familiar to STD readers. For purposes of simplification, we will use the word patient when referring to either patients or clients. Because some STD programs are combined with HIV programs and others are separate, we will use the term STD prevention program when referring to either STD programs or combined STD/HIV programs.

These guidelines, based on the CSPS program announcement, cover many topics new to program operations. Please note, however, that these guidelines replace all or parts of the following documents:

- Guidelines for STD Control Program Operations, 1985.
- Quality Assurance Guidelines for Managing the Performance of DIS in STD Control, 1985.
- Guidelines for STD Education, 1985.
- STD Clinical Practice Guidelines, Part 1, 1991.

The following websites may be useful:

- CDC [www.cdc.gov](http://www.cdc.gov)
- NCHSTP [www.cdc.gov/nchstp/od/nchstp.html](http://www.cdc.gov/nchstp/od/nchstp.html)
- DSTD [www.cdc.gov/nchstp/dstd/dstdp.html](http://www.cdc.gov/nchstp/dstd/dstdp.html)
- OSHA [www.osha.gov](http://www.osha.gov)
- Surveillance in a Suitcase [www.cdc.gov/epo/surveillancein/](http://www.cdc.gov/epo/surveillancein/)
- Test Complexity Database [www.phppo.cdc.gov/dls/clia/testcat.asp](http://www.phppo.cdc.gov/dls/clia/testcat.asp)
- Sample Purchasing Specifications [www.gwu.edu/~chsrp/](http://www.gwu.edu/~chsrp/)
- STD Memoranda of Understanding [www.gwumc.edu/chpr/mcph/moustd.pdf](http://www.gwumc.edu/chpr/mcph/moustd.pdf)
- National Plan to Eliminate Syphilis [www.cdc.gov/Stopsyphilis/](http://www.cdc.gov/Stopsyphilis/)
- Network Mapping [www.heinz.cmu.edu/project/INSNA/soft\\_inf.html](http://www.heinz.cmu.edu/project/INSNA/soft_inf.html)
- Domestic Violence [www.ojp.usdoj.gov/vawo/](http://www.ojp.usdoj.gov/vawo/)
- Prevention Training Centers [www.stdhivpreventiontraining.org](http://www.stdhivpreventiontraining.org)
- Regional Title X Training Centers [www.famplan.org](http://www.famplan.org)
- HEDIS [www.cicatelli.org](http://www.cicatelli.org)
- Put Prevention Into Practice [www.jba-cht.com](http://www.jba-cht.com)
- [www.cdc.gov/nchstp/dstd/hedis.htm](http://www.cdc.gov/nchstp/dstd/hedis.htm)
- [www.ahrq.gov/clinic/ppipix.htm](http://www.ahrq.gov/clinic/ppipix.htm)

## Program Operations Guidelines Workgroup Members

David Byrum	Program Development and Support Branch, DSTD
Janelle Dixon	Health Services Research and Evaluation Branch, DSTD
Bob Emerson	Training and Health Communications Branch, DSTD
Nick Farrell	Program Support Office, NCHSTP
Melinda Flock	Surveillance and Data Management Branch, DSTD
John Glover	Program Development and Support Branch, DSTD
Beth Macke	Behavioral Interventions and Research Branch, DSTD
Charlie Rabins	Illinois Department of Public Health
Anne Rompalo	Johns Hopkins School of Medicine
Steve Rubin	Program Development and Support Branch, DSTD, New York City
Lawrence Sanders	Southwest Hospital and Medical Center, Atlanta
Don Schwarz	Program Development and Support Branch, DSTD
Jane Schwebke	University of Alabama Birmingham
Kim Seechuk	Program Development and Support Branch, DSTD
Jerry Shirah	Training and Health Communications Branch, DSTD
Nancy Spencer	Colorado Department of Public Health
Kay Stone	Epidemiology and Surveillance Branch, DSTD
Roger Tulloch	Program Development and Support Branch, DSTD, Sacramento, California

## Program Evaluation Subgroup Members

Janelle Dixon	Health Services Research and Evaluation Branch, DSTD
Bob Emerson	Training and Health Communication Branch, DSTD
Nick Farrell	Program Support Office, NCHSTP
Mike Hennessy	Behavioral Interventions and Research Branch, DSTD
Guillermo Herrera	Health Services Research and Evaluation Branch, DSTD
Bill Kassler	Health Services Research and Evaluation Branch, DSTD
Bobby Milstein	National Center for Chronic Disease Prevention and Health Promotion
Nancy Spencer	Colorado Department of Public Health

## Program Evaluation Internal/External Reviewers

Gay Allen	Training and Health Communication Branch, DSTD
Tom Bertrand	Rhode Island Department of Public Health
Kitty Corbett	University of Colorado at Denver
Tom Gift	Health Services Research and Evaluation Branch, DSTD
Katy Irwin	Health Services Research and Evaluation Branch, DSTD
Mary McFarlane	Behavioral Interventions and Research Branch, DSTD
Dana Shelton	Office of the Director, DSTD



# Program Evaluation

---

## INTRODUCTION

This chapter gives a brief description of program evaluation and describes how evaluation can be used to help reach program goals and objectives. It does not include all methods, philosophies, or approaches to evaluation and touches on only a few aspects of STD prevention programs.

There are many reasons for program evaluation, some with emphasis on scientific methods to collect data, some with emphasis on the process of monitoring, and others with emphasis on the use of data to inform program managers and other key policy makers about how well a program is meeting its goals and objectives. CDC emphasizes evaluation as a way to improve and account for public health actions using methods that are useful, feasible, proper, and accurate. To accomplish this, CDC recommends that specific, systematic evaluations be carried out throughout the life span of a program, from program inception and planning to implementation, sustained delivery, and re-design (MMWR, 1999).

Ongoing evaluation in STD prevention programs is critical to developing and sustaining high quality, appropriately targeted STD prevention efforts. Evaluation offers the opportunity to review, analyze, and modify STD prevention efforts as necessary. It allows STD prevention programs to know where they have been, where they currently are, how they got there, and where they are headed. Good program managers use evaluation to improve program performance (See Leadership and Program Management chapter) and to monitor progress toward achievement of goals and objectives.

In addition to program self-evaluation, evaluation may be needed in other situations. Some of these situations are:

- To help prioritize activities and guide resource allocation
- To inform funders of the program whether their contributions are being used effectively
- To inform community members and stakeholders of the project's value (Rugh, 1996)
- To provide information that can be useful in the design or improvement of similar projects (Rossi, 1998)

Regardless of the reason for the evaluation, different strategies are called for in different situations and at various stages in programs. In the development stage, evaluations focus on assessing the extent and severity of the issues to be addressed and on designing effective interventions to address them (Wong-Reiger, 1993a). Once programs are initiated, it is important to examine various methods of operation to improve program effect or decrease costs in producing the desired effect (Wong-Reiger, 1993b). An example is a program improvement which increases the number of patients who voluntarily return for treatment while also decreasing the cost of follow-up.

To aid in decisions concerning continuing, expanding, or curtailing programs, evaluation should also consider costs in relation to benefits. It can compare an intervention's cost effectiveness with that of alternative strategies. For either new or ongoing programs, impact assessments estimate the effects of the intervention and the degree of effectiveness in providing the target populations with the resources, services, and benefits that are intended (Rossi, 1998).

Whether an evaluation is comprehensive or tries to answer only one question, “the aim is to provide the most valid and reliable findings possible within political and ethical constraints and the limitations imposed by time, money, and human resources” (Rossi, 1998).

---

### Recommendation

---

- Programs should conduct appropriate, regular and ongoing evaluation for self assessment and quality improvement.
- 

## PLANNING AN EVALUATION

Evaluation should be part of program planning from its inception. There should be a plan of evaluation for each essential program component, including how and when each will be evaluated and how the evaluation will be used to improve the program. While evaluations are conducted after the program has started, early planning for evaluation enables gathering the right data, at the right time, for the right purpose. This is especially important for determining if the program’s activities are having the desired outcomes, such as behavior change, and is essential in determining if the program was responsible for the desired impact.

While a single public health intervention is seldom shown to be the reason for achieving a particular end result, such as a reduction in disease morbidity, confidence and utility of most evaluations can be increased by designing the evaluation questions and methods when planning or changing program activities or interventions. For instance, if a comparison of indicators before and after the program is to be used, planning for this must be included in the beginning. In addition, managers and evaluators must be able to identify factors outside the program intervention which might confound the evaluation and affect the outcome. These should be taken into consideration in the design of the evaluation and the collection of data. (Wong-Reiger, 1993)

**EXAMPLE:** The STD program supports a risk reduction program which emphasizes delaying sexual intercourse for all teens, in a local community based or-

ganization. An adolescent female pregnancy prevention program in the same community based organization also has an effort to persuade teens to delay the onset of sexual activity. The effect on the STD program’s risk reduction program is confounded by a similar program for adolescent females which must be taken into consideration when designing and conducting the evaluation.

As managers plan an evaluation, they should begin with a clear purpose in mind. They must gather background information concerning what is to be evaluated and why, and determine the stakeholders of the program and the evaluation, how findings will be used, and the amount of fiscal and human resources available to design and conduct the evaluation. With this information in hand, the steps can be undertaken to begin the evaluation process. Throughout the evaluation process, input from stakeholders, staff, or evaluators may alter the extent of the evaluation or the resources available. However, at each juncture, keeping the purpose in mind will be useful in making decisions (Patton, 1997; Herman, 1987).

Designing an evaluation requires that choices be made between various ways of obtaining information; each choice is subject to trade-offs between accuracy, time, and resource constraints. Some of those choices are: type of information collected (e.g. descriptive or numeric), timing of measurements (e.g. pre and post), measurement techniques, (e.g. single versus multiple measures); and who and what is measured. The quantity and quality of information to be produced and the costs associated with each must be considered in the choice (NIDA, 1991).

---

### Recommendations

---

- Programs should plan evaluations early in the development of interventions.
  - Programs should have a plan of evaluation for all important program components, including how and when each will be evaluated.
  - Program evaluations should be designed and conducted with a clear purpose.
-

## STEPS IN DESIGNING AND CONDUCTING AN EVALUATION

There are six essential steps in designing and conducting an evaluation. These steps are to 1) engage stakeholders in the evaluation, 2) describe the program, 3) focus the evaluation design, 4) collect credible evidence, 5) justify conclusions, and 6) ensure use and share results (MMWR, 1999). Each step is described in greater detail below.

### 1.) Engage stakeholders in the evaluation.

In practice, evaluation is often an effort of only program managers and evaluators (external or internal). However, for evaluation to be successful, it is necessary that other stakeholders are included in the planning, implementation, and interpretation of the evaluation and its findings.

The range of stakeholders includes participants who expect services, funders who expect results for their support, other agencies or groups who serve the same or similar clients, the staff or volunteers who run the programs, and the administrators who are responsible for the delivery of services (Wong-Reiger, 1993). There are stakeholders of the program and stakeholders of the evaluation and some are both. The more involved stakeholders are, especially in the decision making process, the more cooperative they will be in providing information and being open to unexpected results. It is important to understand what various stakeholders want from the evaluation and how rigorous they expect evaluation methodology to be. It is also likely that these different motivations and expectations will cause conflict if not accounted for or resolved.

Stakeholder involvement will vary with the type of evaluation. The choice of which stakeholders to involve and at what level is a function of the purpose of the evaluation and who will use the results. Some evaluations may involve stakeholders only in decision making while others may be completely “participatory”. Participatory evaluations involve stakeholders in all aspects of the project including design, data collection, and analysis. The benefits of participatory evaluation are: 1) selecting appropriate evaluation methods,

2) developing questions that are grounded in the perceptions and experiences of clients, 3) facilitating the process of empowerment, 4) overcoming resistance to evaluation by participants and staff, and 5) fostering a greater understanding among stakeholders (Marris, 1998). Regardless of the level of involvement, it is important that responsibilities and roles of each person or group are clearly defined and agreed to at the beginning of the process.

### 2.) Describe the program, including the needs, expectations, activities, stage, and context.

Program managers will need to elicit information from a variety of sources including staff, data, and documents to fully describe the program. The description should include the mission and objectives and be detailed enough so that others may understand the program goals and strategies (MMWR, 1999).

In describing the program it is useful to have a logic model, a graphic presentation of the logical relationship among program components. A program logic model is ideally developed at the planning stage and assists in clarifying the relationships between activities, objectives, and goals of the program. The development of a logic model is similar to identifying goals and objectives. There are four main components in developing a logic model: 1) the activities (methods of operation), 2) the services delivered (process indicators), 3) the intermediate results (outcome indicators), and 4) the intended results (impact indicators), including targeted groups. The logic model is most useful if each element in it is linked to a quantified objective, so that process, outcome, and impact indicators are defined in terms of concrete numerical targets. To develop a logic model, managers must be able to clearly and accurately describe the program and who and what it intends to affect. Each of the program activities is measured by one or more service delivery results, which in turn measure the level at which that activity is provided. Each service delivery result is linked to one or more intermediate result, which is expected to occur as a result of participation in the program activity. For a program to propose that an intended result

can be achieved, it must show that there are one or more intermediate results linked to the intended result. Further, there should be evidence that each step will indeed bring about the next step in the process.

When program goals and objectives are appropriately written, that is, specific, measurable, realistic, and time-framed, the model is easier to develop. If not, flaws in the objectives (and program design) may also be easier to recognize. (See Appendix E-A for an example of a program logic model, and E-B for examples of good and poorly developed objectives.)

Adequately describing the program from beginning to end, both as it should be and as it is, will aid managers in determining whether the course the program is on is the correct one. For instance, if the services delivered are different from those which were planned, delivered in significantly fewer amounts, or to the wrong populations, it is necessary to rethink what changes in the program are indicated. If it is believed that a particular program is necessary, then it is difficult to attribute any results to the program if the activities were not delivered as planned (Wong-Reiger, 1993).

**EXAMPLE:** A manager implements a program to provide screening for syphilis in intake drug treatment facilities on the grounds that the exchange of drugs for sex is a part of the syphilis epidemic in the community. The expectation is that 98% of the clients will be tested and, if needed, treated for disease; 95% of those with the disease will be interviewed; and subsequently 80% of appropriate sex partners will also be examined and treated. However, for a variety of reasons, only 60% of the clients are treated for syphilis, 50% of them interviewed and only 40% of the named partners located. It would be unrealistic for this aspect of the program to be credited for a decrease in disease among drug users in the community.

### **3.) Focus the evaluation design.**

Before the design of the evaluation is decided, managers, evaluators, and stakeholders will need to determine the objectives of the evaluation. The objectives differ depending on what is being evalu-

ated and how the evaluation is intended to be used, but it is important that the objectives are realistic, focused on the need at hand, and designed to answer the right questions. Evaluation objectives help clarify what aspect of the overall program is being assessed (Schechter, 1993). Setting the objectives for the evaluation will help focus it and keep the process from becoming too cumbersome and all-inclusive. It is also important to understand the difference between evaluation objectives and program objectives.

**EXAMPLE:** A program objective might be “Ensure that 95% of females who test positive for chlamydia in the STD clinic are appropriately treated within 7 days”. An evaluation objective might be “Assess whether follow up systems for clients are ensuring an adequate response rate.”

There are a variety of evaluation designs and not all are equally suited to the type of evaluation needed or wanted. It is necessary for managers to understand the difference and plan the evaluation in accord with the most appropriate evaluation method. This will help ensure that the evaluation strategy has the greatest chance of being useful, feasible, ethical, and accurate (MMWR, 1999).

### **4.) Collect credible evidence.**

Protocols and instruments may need to be developed for use in data collection activities. These activities should be supervised closely by the evaluation director since these data will be used for analysis. If evaluation was not part of the planning process, some data may be very difficult or impossible to collect once the program has been initiated. There must be a plan for who can provide data and who can gather the data. For process evaluation, decisions should be made whether to collect all available data on an ongoing basis, sample on an ongoing basis, or sample at specific times. For outcome/impact evaluation there are many methodological issues to consider; in this case, it is best to seek help from program evaluation specialists (Program Evaluation Toolkit, 1997).

Not all evidence for program evaluations are quantitative data. Some issues in the evaluation

are best addressed through qualitative methods. Such methods include observations, semi-structured and unstructured interviews, and collection of vignettes and interpretations about program aspects and functioning. They are often more useful in evaluation of the early stages of program development, or assessment of the need for “mid-course corrections.” Qualitative methods may help uncover aspects of the program, such as diverse understandings of its goals, that lead to revision of the logic model or a new frame for understanding problems. These methods are less appropriate for examining program outcomes.

**EXAMPLE:** Data obtained from STD\*MIS can tell evaluators the length of time it takes to complete field work assignments, but a complete assessment of field work requires that supervisors observe how staff perform their activities. The quantitative data coupled with the information gleaned during observation is needed to determine how well that component is working and what changes may be necessary to improve field results.

Generally speaking, it is best to have trained evaluation staff who can assess the findings and objectively analyze the data. However, the person who analyzes the data will need to work closely with program managers to assist in the interpretation of findings. The evaluation report should not only document raw findings, but should also analyze and synthesize them (Schechter, 1993).

### **5.) Justify Conclusions**

Once the evidence has been analyzed and synthesized, conclusions can be made about program activities. These conclusions must be linked to the evidence. However, because there is an apparent linkage does not mean that the conclusions are correct or acceptable to the stakeholders. Understanding the results within program context is essential or the results are often meaningless. Identifying evidence regarding the program’s performance is not all that is needed to draw evaluation conclusions.

Conclusions made about the program lead to recommendations for some types of actions. Further, recommendations for continuing, expanding, redesigning, curtailing or terminating a program

are not the same as determining a program’s effectiveness. Recommendations about program activities should be aligned with areas that stakeholders can control or influence and be acceptable to them.

### **6.) Ensure use and share findings**

The practical use of evaluation results and recommendations is not automatic. Too frequently evaluations are performed and it is assumed that appropriate action will occur. Program managers also need to plan for and take deliberate action to ensure that findings are disseminated appropriately and used properly. Frequent feedback to and from all the stakeholders is essential for ensuring use. Managers may need to develop a system of follow-up to determine the who, how, and when of operationalizing the recommendations.

**EXAMPLE:** An evaluation of a STD prevention program in a major city showed that 60% of women were being screened for chlamydia. Subsequently, a recommendation was made that all three clinics should begin routine screening. Program managers need to develop a plan for ensuring that each clinician is aware of the new policy, given the opportunity to discuss and agree on the change, trained in testing procedures, and that a mechanism is developed to systematically track the number of women tested. In addition, mechanisms for corrective action should be anticipated.

### **7.) Disseminate findings broadly and in a timely fashion.**

The results of an evaluation should always be shared with stakeholders and, when possible, with other prevention and control programs. The results should be disseminated in a timely and unbiased fashion (MMWR, 1999). If the dissemination of the results is significantly delayed, either the situation may have changed or stakeholders may perceive that the evaluation is unimportant to them, management, and the evaluators. Results which are delivered in a biased fashion, such as punitive, will be ignored or possibly subverted.

National conferences are one possibility for widespread dissemination of evaluation findings. However, programs which discover significant



findings that could have important effects on the control of STD should seek other more immediate ways of getting the information to other programs. As electronic communications become more and more commonplace, there will be many opportunities for widespread, rapid dissemination of findings.

With the results of the evaluation, a new process should be undertaken to refine the program, cease activities which do not work, and/or develop new interventions in areas of need. Evaluations are opportunities to improve programs and plan for the future and should be conducted as such.

### Recommendations

- Program managers should develop a written description of the program, including the involvement of stakeholders.
- Programs are encouraged to develop logic models for goals, objectives, activities, and the targeted groups.
- Evaluation results should be shared with stakeholders.
- Evaluation results should be used for program improvement and further program planning.

## TYPES OF EVALUATION

The evolution of evaluation research has led to a proliferation of evaluation methods and approaches, each of which has its role depending on what is being evaluated. Evaluations reflect many different scientific philosophies (Alkin, 1990). Because all programs have a set of guiding principles concerning how they should operate and how they achieve their outcomes, all interventions have a “program theory”. However, what is often lacking is an explicit recognition of the program theory (Weiss, 1997; McClintock, 1990). The explicit recognition of the theory behind STD prevention and control activities is often overlooked in daily operations because many have been used routinely for many years. However, it may be necessary

for managers to think about and identify the scientific and rational reasons why activities are conducted. STD prevention programs should be based on scientific evidence and evaluation should examine how the program in practice differs from the program in theory.

Evaluation activities should also follow the program’s developmental stages. In general, there is a natural developmental sequence that intervention programs follow, and the evaluation activities should match the development level of the intervention appropriately. The program stage will determine the level of effort and the methods to be used. See Appendix E-C for uses and types of evaluation.

### Formative Evaluation

When new programs, new interventions, new procedures, or new elements of existing programs are proposed, formative evaluation is indicated. Formative evaluations in the pre-implementation and design phase of a project emphasize needs assessment, and their data gathering may involve extensive community analysis or community identification procedures in addition to inquiry into a program setting and existing clientele. Formative evaluations are designed to help identify needs or gaps in service which the new program should address or to answer other questions that need to be answered (e.g., What is the most efficient way to recruit participants?, What types of program activities are desired?, What are consumer preferences for different STD test procedures?) (Wylle, 1992; Tessmer, 1994).

**EXAMPLE:** It is assumed that female clients would prefer urine based testing over those involving pelvic examinations; however, until formative information is obtained, program planners may not be initiating an intervention most acceptable/desired by the clients.

Thus, formative evaluations are conducted to collect data which provide information about the intervention that is being delivered. It is not just process information (how many tests will be done), but also how the clients react or respond to the intervention.

**EXAMPLE:** Whether female clients prefer a urine test because it is quick and does not require being undressed and undergoing an invasive procedure, or

whether they prefer to feel that their “test” is more complete because a pelvic examination or pap smear was done.

### Evaluability Assessment

When the evaluation of existing programs is desired, an evaluability assessment should be conducted. An evaluability assessment will determine to what extent an evaluation is possible (Smith, 1989, 1990; Smith 1981; Fisher, 1982). In conducting an evaluability assessment, the evaluator must be able to clarify program goals and objectives, determine the extent to which the goals and objectives can be achieved, determine what data are available or could be collected to assess program activities, determine the program performance measures and if they can be gathered at a reasonable cost, and explain how the results will be used. In addition, they should be able to identify the programmatic activities responsible for bringing about the intended results (Wholey, 1994). If the program cannot be adequately described in this way, program managers should focus on gathering the appropriate information and clarifying goals and objectives before any other evaluation tasks are undertaken.

**EXAMPLE:** The STD prevention program has obtained the assistance of outside experts to evaluate its efforts to increase screening in adolescent females in managed care settings. However, on examination, the evaluators learn that the program’s stated objective was to “educate providers in all managed care settings on the need for screening.” Further, they learn that the program did not specify the number of providers targeted, the number who received training, the type and extent of managed care settings, the number of adolescent females being screened before the intervention and had no way to ascertain the number now screened. Thus, this aspect of the program cannot be evaluated unless additional data are gathered.

### Recommendations

- A formative evaluation should be conducted when a new intervention or program is undertaken or when a different way of conducting an intervention is developed.
- An evaluability assessment should be conducted when planning an evaluation of any portion of an existing program.

### Process Evaluation

As programs develop there is a need to assess how well the implementation of the program is going and, if needed, to make corrections. In these stages, there are many evaluation questions that could be asked, all having to do with program monitoring and evaluation activities related to this problem. Answering these questions involves process evaluation. Process evaluations include documenting actual program functioning (Dehar, 1993; Finnegan, 1989), measuring exposure to and diffusion of the interventions (Fortmann, 1982; Hausman, 1992; Steckler, 1992), and identifying barriers to implementation (Demers, 1992). Process evaluation includes the identification of the target population, a description of the services delivered, the use of resources, and the qualifications and experiences of the personnel participating in them (NIDA, 1991). It involves determining what services were actually delivered, to whom, and with what level of resources.

**EXAMPLE:** Process evaluation of the effort to increase screening in adolescent females would include, at a minimum, the number of adolescent females in the population and the number screened before and as a part of the intervention, the tests used, and a description of the providers.

Documenting program functioning is important for two reasons. If the program is working well, there will be interest in replicating the program in other locations that serve similar or other populations. If the program is not working well, it is of tremendous use to know exactly how the program failed, in which component, and in what population (Chen, 1990).

**EXAMPLE:** Program A conducted a formative evaluation and determined that female clients really

do prefer urine based screening for chlamydia and based on prevalence data, a plan was developed to test 90% of the target population. However, the process information showed that half way through the intervention period, only 10% of the women had been tested. Instead of concluding that the effort was a failure, additional qualitative information was gathered which showed the drop off point for urine specimens was too public and women felt embarrassed at leaving urine specimens where everyone could see.

There are program monitoring tasks which must also be conducted before an outcome or impact evaluation can take place. Program monitoring tasks are concerned with documenting actual program functioning. Several major questions posed in this evaluation component are:

- **Which elements of the program actually have been implemented?**

Usually the practical problem here is that there are no data readily available to answer the question. When that occurs, the “answer” may be a guess rather than supported by evidence.

Example: One of the program’s surveillance objectives is for all laboratories in the area to report all positive syphilis serologies within a specified time. Unless the program staff can document how many laboratories there are, how many do serologies, and how often results are reported, that aspect of the program can only be estimated.

- **What are the types and volume of treatments or services actually provided to clients?**

This question is important to answer both for accountability purposes and also to assist in the development of an outcome evaluation subsequent to program implementation.

Example: If the program is concerned with preventing congenital syphilis, it is not only necessary to have laboratory data on syphilis serologies, but is also necessary to know how many pregnant women there are, how many receive testing for syphilis and at what stage of pregnancy.

- **What are the characteristics of program participants?**

It is important to determine if the recipients of program services resemble the intended “target group”

as identified in the program design and development stage. An effective intervention administered to a non-target group may be just as useless as an ineffective intervention administered to a targeted group.

EXAMPLE: If the STD prevention program has determined that most congenital syphilis cases have occurred in newborns of adolescents, but syphilis testing occurs mostly in adult women with private insurance, then the target population is not being reached.

Program monitoring can function as quality assurance of activities. Managers and staff should develop tools to ensure that the daily operations are functioning as they should. Corrections are more easily made when detected early and are less likely to create long-term, large scale damage to program progress.

EXAMPLES: Program monitoring may include chart reviews, direct observation of interviews and counseling sessions, routine analysis of laboratory reporting, and analysis of screening procedures and results.

An increased focus on accountability by funding sources has also increased requirements for evidence that a program is delivering what was paid for. Regular feedback from monitoring can be one of the most powerful tools a program manager has for documenting the operational effectiveness of a project, justifying staff, defending the continued existence of the program and even requesting additional support.

Finally, the information gained through program monitoring is necessary to determine which (if any) aspect of the program is appropriate for impact evaluation. The reason for this should be obvious, but it's often overlooked in the rush to evaluate program impact: programs (or components of programs) that don't exist or don't exist as intended should not be evaluated for impact (Rossi, 1998).

## Outcome Evaluation

When process evaluation shows that the program was implemented properly, there is often interest in measuring the effectiveness of the actual program (Mohr, 1995). Outcome evaluation is concerned with the end

result of STD prevention and interventions that have an effect on the health of populations. Criteria for using outcomes for evaluation include: (1) being objective, in that outcomes can be observed; (2) being measurable in ways that are reliable and valid; (3) being attributable to the intervention delivered; and (4) being sensitive to the degree of change expected by the intervention. For STD prevention programs there are a number of different outcomes that can be measured: biological, behavioral, cognitive, economic, and health status. The ultimate outcome is a change in morbidity or mortality. Because the expertise and time commitments to conduct outcome evaluations are often not available to STD prevention programs within health departments, such evaluations may be done by outside evaluators. Outcome evaluation typically requires some understanding of research design. [Key points are discussed below so that managers can work effectively with evaluators.]

In some cases, it may be relevant to consider outcomes that are not directly measurable (for example, some of the sequelae of PID typically occur years after the initial chlamydial or gonorrheal infection). Such outcomes may still be worthwhile to consider, especially for purposes of economic evaluations. In such cases, it may be advisable to use estimates from published literature of the rates at which outcomes occur and vary the rates over a reasonable range (as an example, PID is estimated to occur in 10% - 40% of untreated gonorrheal and chlamydial infections; the effectiveness of the program in preventing PID could be assessed at each end of that range, plus some figure in the middle, such as 20% or 25%). This is a technique known as sensitivity analysis, and can also be used with figures that are known and measurable to determine how program performance may be affected if circumstances change. For example, an on-site syphilis screening program may not be justifiable given the current rate of positive tests, but might be worthwhile to conduct if syphilis incidence increased from current levels (Haddix, 1996; Gold, 1996).

The use of the terms outcome and impact have been used in conflicting ways in the past. However, one useful description of the definition of each is: outcome evaluation is the measuring of the effectiveness of an intervention on the target population, whereas impact

evaluation attempts to measure the total effect of a prevention program on the community as a whole (NIDA, 1991). In this document we will use the term "outcome".

"Outcome" implies measures of effectiveness of an actual program. To assess outcomes, it is first essential to define in specific quantitative terms what the intended program effect is. To carry out a credible assessment of outcomes, it is then essential to design a scientific study, as rigorous and systematic as resources allow.

### ***Defining Program Effect***

To define the effect of the program it is necessary to define measurable goals. This is often difficult or impossible because theoretical goals of the program must be connected to empirical, measurable indicators in the real world. Programs without measurable goals cannot be rigorously evaluated.

### ***Designing the Study***

Designing an appropriate outcome or impact study is complicated; evaluators must overcome the challenges of building into the evaluation plan the ability to unambiguously infer that, if there is a change recorded in outcomes measurements, the change is due to the actions of the program and not to other external or internal influences.

External influences, often called confounders, are a potential explanation for program outcome. If the design is not well developed, it is often easy to jump to inaccurate conclusions (i.e., the intervention had an effect, when in reality there is little or no correlation.)

EXAMPLE: Reduced STD morbidity might actually be due to the effects of increased screening and treatment programs or education. Internal influences also need to be considered. For instance, peer counseling programs may purposely or inadvertently recruit adolescents who are already motivated to change behaviors. When testing for STDs in this group shows a lower prevalence than in similar adolescents, the results probably show an unrealistically high estimate of program efficacy compared to the adolescents who did not volunteer for the program.

## Randomized Trials

The evaluation design that is considered to produce the strongest evidence that a program intervention or activity contributed to change is the randomized control trial (RCT). The rationale for this design is well established. In brief, the essence of a randomized trial lies in the random assignment of subjects to be exposed to the intervention or to be a control (not exposed to the intervention). By using the rule of chance, intervention and control groups are, on average, comparable before exposure. Because of this initial equivalence, if outcome differences between those who do and do not receive the intervention are statistically detected, they are highly likely to be due to the operation or processes of the intervention.

## Quasi-experiments

Although randomized trials provide the strongest evidence about a program's effectiveness they may not be feasible to implement. RCTs are costly, time consuming, can be subject to methodological flaws, and may not be considered ethical to conduct if withholding an intervention from one group may adversely affect opportunities for improved health status. Thus, evaluators turn to the analysis of quasi-experiments, defined generally as any research design that does not utilize random assignment to deliberately construct an initial equivalence between groups. Quasi-experimental designs use a control group which is separate from the experimental group and not randomized. When randomized trials are not possible and quasi-experiments are substituted in their place, specific design features usually have to be instituted to rule out or eliminate each alternate explanation to the hypothesis of treatment effects.

## Economic Evaluation

Economic evaluation considers both the outcomes of a program and the cost of producing those outcomes. In some cases, the most effective program may also have the lowest cost, but it is not necessarily true that the lowest-cost option is the most cost effective. It is also possible that the program that produces the most units of a given outcome may be impractical to implement because it is so costly that it diverts too many

resources from other uses, or requires more resources than are available. An example is provided at the end of this subsection.

To conduct an economic evaluation, it is necessary to know what resources are used in a program, and what these resources cost. In some cases, the costs are not direct (i.e., they don't have to be paid), but indirect (such as an opportunity cost, which is the cost of using a resource in a given program that could be used elsewhere). This process involves measuring or estimating the value of facilities, equipment, personnel, and other resources used. Sometimes patient time commitments and travel costs are relevant, as well (Drummond, 1987). Adequately determining appropriate costs can be difficult, and should not be undertaken without the help of someone familiar with economic analyses (Rossi, 1998).

What costs are included in the analysis will depend upon the perspective chosen. The perspective of an analysis determines which costs are considered. The broadest perspective is societal, which includes all costs borne by all parts of society, including local programs, the health care system as a whole, and patients. More limited perspectives are also often used which do not consider the costs borne by some groups in the economy.

**EXAMPLE:** Client travel costs and time costs for clinic visits would not be relevant from a health care system perspective because the health care system does not pay for them, but would be for a societal perspective analysis which includes all costs. The perspective should be appropriate for the particular issue being analyzed (Haddix, 1996).

## Types of Economic Evaluation

### *Cost Analysis*

The simplest form of economic evaluation is a cost analysis. Because it considers only the costs, however, it is a partial economic evaluation (Drummond, 1987). To conduct a cost analysis the costs of a program must be determined, making sure to collect all relevant costs for the perspective being used (Haddix, 1996).

**EXAMPLE:** The STD prevention program might determine the cost of screening for chlamydia in family planning clinics, or the cost to follow women who tested positive for chlamydia in a private medical facility to get them treated.



It is important to conduct cost analyses when appropriate. However, at a minimum, the state/local health department should calculate the cost per service unit for each of its major prevention programs (the 'service unit' will depend on the program; for example, in an STD clinic, costs could be expressed as dollars per patient visit; dollars per gonorrhea, syphilis, or chlamydia test; or dollars per infection identified and treated).

Once costs are determined, there are three common methods used for comparing the costs and consequences of different interventions: cost effectiveness, cost-utility, and cost-benefit analysis.

### ***Cost Effectiveness Analysis (CEA)***

CEA divides the net cost of a program by the outcomes produced by the program. The outcomes chosen are generally the health effects targeted by the program, such as cases of disease prevented or lives saved. The result will be expressed as the net cost per unit of outcome.

EXAMPLE: In comparing programs that promote the detection of chlamydia, the unit of measure for the CEA might be "cases of PID averted".

This differs from the per-unit cost analysis presented in the previous section in that the cost savings associated with the adverse outcomes averted or with the desirable outcomes produced are incorporated into the net cost. This is the most commonly used type of economic analysis in the health field, and is especially well-suited to comparing different interventions or programs that share the same outcome (Haddix, 1996). The interventions can be ranked in order of increasing effectiveness, and the cost effects of moving from one intervention to the next most effective one can be easily determined. It is less effective in comparing interventions that produce different outcomes, because it does not provide a common outcome measure.

EXAMPLE: CEA would be more helpful in comparing two chlamydia screening programs than in comparing a chlamydia screening program with a cancer prevention program.

### ***Cost-Utility Analysis (CUA)***

CUA is similar to CEA, except that the program outcomes are measured in common terms across interventions, most commonly quality-adjusted life years

(QALY) (Haddix, 1996; Farnham, 1996). With this approach, interventions that produce different outcomes (such as chlamydia prevention and cancer prevention) can be compared -- the different outcomes are translated into QALYs; it is then theoretically possible to determine the most efficient use of resources to produce the maximum amount of health. However, actually determining the QALYs gained by preventing a case of infection is not a straightforward task. QALY measures for STD outcomes are not well-developed. In other programs, CUA is most commonly used in programs with significant non-health benefits and is often used to determine whether to fund the program or not (Farnham, 1996).

### ***Cost-Benefit Analysis (CBA)***

CBA is also similar to CEA, except that it places a monetary value on the outcomes of programs. In the above example of CEA, the monetary value per case of PID averted would be determined and factored into the net cost. In theory, this is the broadest form of analysis because it can be determined whether the benefits of a program justify its costs. However, in practice it is also limited to a comparison of those specific costs and benefits that can easily be expressed in terms of money (Drummond, 1987). Cost benefit analysis often presents controversial questions, such as, "What is the value of saving a life?" or, "Is the life of an older person worth as much as the life of a younger person?" Determining the answers to these questions is not straightforward, and no clear consensus methodology has emerged. Because of these difficulties, CEA and CUA are more often used in health programs (Farnham, 1996).

EXAMPLE: The following example provides hypothetical data to illustrate the concepts presented above regarding economic analyses.

A program manager wishes to evaluate the gonorrhea screening program at one of the program's clinics. All women under 25 years of age are routinely tested when they present to the clinic. After collecting the costs for staffing, supplies, testing equipment, and clinic overhead, it is determined that the program currently costs \$50,000 per year. This is the cost analysis of the program. It is further determined that 2,500 tests are performed each year, for a cost of \$20 per test. The screening program leads to the detection and

treatment of 50 cases of gonorrhea, and is estimated to prevent 10 cases of PID per year.

The manager wants to compare the effects of two possible alternatives to the screening program which routinely tests all women under 25 years of age: selective screening based upon a risk assessment, and expanded universal screening for all women under 35 years old. This is a cost effectiveness analysis (CEA). After adding up the costs and subtracting the savings from the cases of PID averted, the costs and outcomes of the three alternatives are:

Testing Approach	Net Cost	Cases of PID Prevented	Cost per Case Prevented
Risk Assessment	\$30,000	9	\$3,333
Test All < 25 Years	40,000	10	4,000
Test All < 35 Years	66,000	12	5,500

Which program is “best” will be partially determined by the resources that the health department can devote to screening. The risk assessment approach is not necessarily the most cost effective, despite having the lowest net cost. It also prevents the fewest cases of PID. Similarly, while testing all women less than 35 years old prevents the most cases of PID, the cost per case prevented is highest and may require a level of funding that is unavailable. If the net cost of testing all women under 35 years old had the lowest net cost, it would unequivocally be the most cost effective, because it would represent the lowest cost program that also prevented the largest number of cases of PID. Even when CEA does not provide a clear-cut best choice it gives policy makers information that can help them make resource allocation decisions.

### Recommendation

- At a minimum, programs should calculate the cost per service unit for each of its major prevention programs.

### Measures

Issues pertaining to data collection and measurement are relevant for all levels of evaluation. The selection of indicators, instruments, measures, and data sources depends largely on the purpose and state of the program. One of the most basic considerations in selecting indicators is that they reflect the central goals and objectives of the program (NIDA, 1991). There are alternate methods that can be used to gather data for specific indicators, which vary in reliability, validity, depth, and cost. The method should reflect the priority being given to the indicator and the resources available for the evaluation. For example, if the process by which adolescents are believed to have access to health care needs to be thoroughly analyzed, the question may be approached by means of focus groups or in-depth individual interviews, a high cost approach because of the use of highly trained moderators or interviewers. However if only a cursory picture is required, then a few questions in a process evaluation might be sufficient (NIDA, 1991).

Evaluation activities in recent years have gone beyond basic budget and staff monitoring to count program outputs, such as services delivered to clients. Measurement of some outputs, such as counting the number of women screened and treated for chlamydia, captures the intended result of the program (Newcomer, 1997). Assessment of service delivery at the local level is not new, but linking the measures or indicators to program mission, setting performance targets, and regularly reporting on the achievement of target levels are relatively new features in performance measurement (Newcomer, 1997). The website [www.cdc.gov/nchstp/dstd/hedis.htm](http://www.cdc.gov/nchstp/dstd/hedis.htm) includes information on the HEDIS chlamydia measure and software developed to evaluate resource allocations.

Performance measurement is an inclusive term that may refer to the routine measurement of program inputs, outputs, intermediate results, or eventual outcomes (Newcomer, 1997). Performance measurement “consists of the systematic description and judgement of programs and, to the extent feasible, systematic assessment of the extent to which they have the intended results” (Wholey, 1994; Newcomer, 1997).

In-depth program evaluations are usually done by organizations such as contractors and universities, while performance measurement is often done by the programs. The ability to truly measure outcomes is limited. For instance, it is very difficult to measure the prevention of congenital syphilis in a population; therefore, many agencies and programs measure non-prevention as a substitute, e.g. number of pregnant women screened for syphilis (Hatry, 1997).

Beginning October 1, 1997, Federal departments and agencies were required to prepare strategic plans which were forwarded to the Office of Management and Budget, to the President, and on to Congress. These plans are a part of the Government Performance and Results Act (GPRA) of 1993. GPRA compares actual performance with the goal levels that were set by the agency's annual performance plans. The goal levels set by CDC, and the achievement thereof, are in part dependent on the achievement of state and local STD prevention program goals and objectives.

## PRACTICAL CONSIDERATIONS

### Overcoming Barriers

The different motivations and goals of stakeholders can bring conflicts to the evaluation process. Likewise, there can also be tensions between the program staff and evaluators, especially if the evaluators are from "outside" and were not a part of the project from the beginning. There may be mutual misunderstanding about the purposes and conduct of the program evaluation. Plans for conflict resolution should be determined at the beginning of any evaluation effort (Short, 1996). Conflicts in any of these areas can result in one or more of the stakeholders or evaluators erecting barriers to the process and almost guaranteeing that the results, however well founded, will not be used.

### Resources

It is very important to make sure that appropriate resources are available for doing the evaluations throughout the various steps and stages of the program. If the organization has no monetary resources dedicated for evaluation and is unwilling to free up staff time, the

feasibility of conducting an adequate evaluation must be questioned. Therefore, it is important that programs budget for evaluation in the planning stage. It can also be helpful to look for other options for doing evaluations, such as volunteers or students from local universities. Volunteers can help train staff and mobilize a process of self-evaluation useful to the organization. Students may be interested in conducting evaluations to gain work experience, develop professional relationships with health departments, or work in collaboration to obtain data to assist in educational requirements, such as completing a thesis.

Another issue of real importance is that of "overkill" in evaluation efforts. Some evaluations can cost as much as the program to be evaluated. While there are no hard and fast rules about the proportion of a program's budget to be spent on evaluations, it is useful to tailor the size of the evaluation to the importance of the decisions which need to be made, or the importance of the intervention to the overall public health objective.

The amount of time an evaluator can devote to the project is dependent on the available budget. Available time, combined with other resources, significantly influences the choice of methods. Site visits, for example, incur costs in terms of staff time as well as travel. Special outcome measures require substantial staff time for development, pilot testing, and analysis. Assessing more rather than fewer program participants, as another example, has significant cost implications for the program. Abundant resources available for evaluation are generally not found in STD and other public health programs, so there must be an effort to maintain a reasonable balance between the demand of scientific rigor and credibility and those of the budget (Herman, 1987).

At a minimum, the number and type of people involved in conducting an evaluation are the program manager, who understands the program and evaluation and knows of the resources available, the program staff who have experience with the program's activities and their target groups, and an evaluator, who knows the evaluation process and has expertise with specific methods and technical issues (Porteous, 1997). Another option is to collaborate with the evaluation team of an equivalent or sister program, or to join forces in a multi-site evaluation effort.

As the scope and limits of activities are defined, financial and staff resources, as well as expectations of the evaluation must be kept in mind. Organizations with an in-house evaluation staff of experienced epidemiologists, behavioral scientists, health planners, and evaluation specialties can be expected to conduct a more comprehensive, sophisticated, resource intensive, and wider array of evaluation activities than organizations with smaller budgets and fewer in-house staff resources (Schechter, 1993). However, an outside source for the evaluation should be considered when objectivity is needed, the necessary skills are not available in the organization, there is a lack of time or interest among staff, and if the budget is available (Marris, 1998).

Conducting an evaluation takes time and resources that are easily forgotten or overlooked. It is essential, when designing programs, that detailed evaluation steps and costs are included as part of the project action plan and budget. Consider a budget and time line for each of the following steps:

- Designing the evaluation
- Developing the measurement instruments
- Pilot testing measurement instruments
- Revising measurement instruments
- Collecting the data
- Processing the data
- Analyzing the data
- Writing the report
- Disseminating the results
- Critiquing the whole process.

## Ethics

Many evaluations have the potential for putting individuals or groups and their work into a bad light or a good one. In addition, some evaluations have the potential for doing real or perceived harm to individuals or groups, both in the target population and in the staffs who are part of the program being evaluated. Evaluations which are conducted without regard to the people involved can also cause harm to the relationship between the communities and the health department or within parts of the agency itself. Those collecting information should be vigilant in keeping

the confidentiality of participants and staff and be respectful of cultural values, language differences, and be as least disruptive as possible (Sanders, 1994). Data collection should be conducted in light of ethical considerations (e.g. it may not be ethical to randomize a subgroup not to receive an intervention which has proven to be effective), resources (e.g., data collection should not compromise delivery of critical program services), and sensitivities of program personnel (e.g., staff may feel concerned that evaluation results may shed a negative light on the program or their own work).

## Standards to Consider in a Practical Evaluation

CDC has accepted the set of standards developed by the Joint Committee on Educational Evaluation (1994) which have been endorsed by a number of other professional organizations. As the steps in evaluation are undertaken, the standards are used to help determine the soundness of public health evaluation efforts.

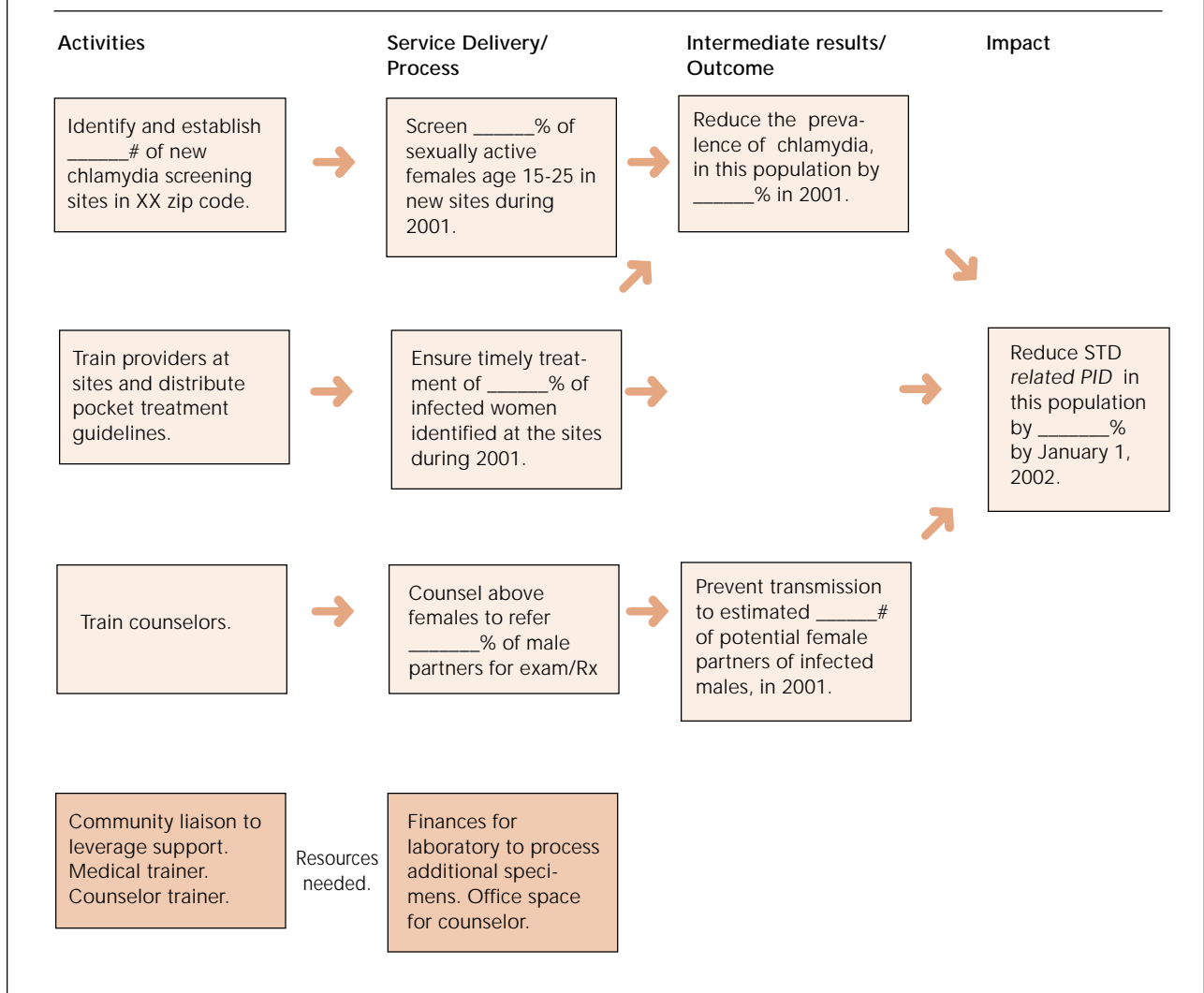
These standards are intended to ensure that:

- an evaluation will serve the information needs of its intended users. (Utility standards)
- an evaluation will be realistic, prudent, diplomatic, and frugal. (Feasibility standards)
- an evaluation will be conducted legally, ethically and with regard for the welfare of those involved in the evaluation, as well as those affected by the results. (Propriety standards)
- an evaluation will reveal and convey technically adequate information about the features that determine worth or merit of the program being evaluated. (Accuracy standards) (MMWR, 1999)

Recommendations
<ul style="list-style-type: none"> <li>• Programs should include funds for evaluations in their budgets.</li> <li>• Programs should consider utility standards, feasibility standards, propriety standards, and accuracy standards when performing evaluations.</li> </ul>

## Appendix E-A

### LOGIC MODEL





## Appendix E-B

### EXAMPLES OF OBJECTIVES

---

#### Good Objectives

1. Reduce the number of cases of STD-associated PID by 10% by January 1, 2002, in women age 15-25 who receive medical care in the ABC Managed Care Group (medicaid managed care contractor). The following objectives all pertain only to women in this target population and are to be achieved in the calendar year 2001.
  - A. Ensure annual chlamydia screening of 95% of sexually active women between the ages of 15-25.
  - B. Ensure appropriate treatment for chlamydia, in 95% of this population, within 7 days of receipt of test results.
  - C. Ensure appropriate treatment for chlamydia in an average of at least one male sex partner per infected woman, for 80% of women who are treated for chlamydia.

#### Poorly Designed Objectives

1. Reduce the number of cases of chlamydia-associated PID by 50%, in women in the ABC Managed Care Group. (50% is probably set too high, all women in the group would not be at risk for chlamydia, and the cost for screening would be prohibitive.)
  - A. Ensure chlamydia screening of all women. (How often should the screening be done? All women are not at risk for chlamydia - some are young girls, some are old, and a few are not sexually active. This is not specific, realistic or time-framed.)
  - B. Ensure treatment for chlamydia for all women. (Only appropriate, recommended therapy should be acceptable. A target level is needed as 100% is usually not attainable, and a reasonable time period for treatment needs to be set; otherwise, a lapse of months would be acceptable. This is not specific, measurable, realistic, or time-framed.)
  - C. Ensure appropriate treatment for male sex partners of women with chlamydia. (How many male sex partners? For what percent of women will the program strive to provide this intervention? This is not specific, realistic, or time-framed.)

#### Good Objective

Provide “enhanced risk-reduction counseling” (according to Counseling Guidelines) to 80% of males 15-18, who attend the Local Community Clinic, during the summer months, June-August 2000. (This process objective tells specifically what will be offered, to whom, and when.)

#### Poorly Designed Objective

Reach 500 youths through outreach. (All the key words here are vague - reach, youths, outreach. This objective could be measured in any way and so is not really measurable, it is not specific, not time-framed.)

## Appendix E–C

### TYPES AND USES OF EVALUATION

Types of Evaluations	When to use	What it shows	Why it is useful
Formative Evaluation Evaluability Assessment Needs Assessment	<ul style="list-style-type: none"> <li>• During the development of a new program.</li> <li>• When an existing program is being modified or is being used in a new setting or with a new population.</li> </ul>	<ul style="list-style-type: none"> <li>• Whether the proposed program elements are likely to be needed, understood, and accepted by the population you want to reach.</li> <li>• The extent to which an evaluation is possible, based on the goals and objectives.</li> </ul>	<ul style="list-style-type: none"> <li>• It allows for modifications to be made to the plan before full implementation begins.</li> <li>• Maximizes the likelihood that the program will succeed.</li> </ul>
Process Evaluation Program Monitoring	<ul style="list-style-type: none"> <li>• As soon as program implementation begins.</li> <li>• During operation of an existing program.</li> </ul>	<ul style="list-style-type: none"> <li>• How well the program is working.</li> <li>• The extent to which the program is being implemented as designed.</li> <li>• Whether the program is accessible and acceptable to its target population.</li> </ul>	<ul style="list-style-type: none"> <li>• Provides an early warning for any problems that may occur.</li> <li>• Allows programs to monitor how well their program plans and activities are working.</li> </ul>
Outcome Evaluation Objectives-Based Evaluation	<ul style="list-style-type: none"> <li>• After the program has made contact with at least one person or group in the target population.</li> </ul>	<ul style="list-style-type: none"> <li>• The degree to which the program is having an effect on the target population's behaviors.</li> </ul>	<ul style="list-style-type: none"> <li>• Tells whether the program is being effective in meeting its objectives.</li> </ul>
Economic Evaluation: Cost Analysis, Cost-Effectiveness Evaluation, Cost-Benefit Analysis, Cost-Utility Analysis	<ul style="list-style-type: none"> <li>• At the beginning of a program.</li> <li>• During the operation of an existing program.</li> </ul>	<ul style="list-style-type: none"> <li>• What resources are being used in a program and their costs (direct and indirect) compared to outcomes.</li> </ul>	<ul style="list-style-type: none"> <li>• Provides program managers and funders a way to assess cost relative to effects. "How much bang for your buck."</li> </ul>
Impact Evaluation	<ul style="list-style-type: none"> <li>• During the operation of an existing program at appropriate intervals.</li> <li>• At the end of a program.</li> </ul>	<ul style="list-style-type: none"> <li>• The degree to which the program meets its ultimate goal on the overall rate of STD transmission (how much has program X decreased the morbidity of an STD beyond the study population).</li> </ul>	<ul style="list-style-type: none"> <li>• Provides evidence for use in policy and funding decisions.</li> </ul>

## Appendix E–D

### EVALUATION GLOSSARY

---

**Evaluation:** The process of determining whether programs—or certain aspects of programs—are appropriate, adequate, effective, and efficient.

#### *Evaluation Types/Stages*

**Cost Analysis:** The simplest form of economic evaluation which considers only the cost of the program.

**Cost-Benefit Analysis:** An evaluation of the relationship between program costs and outcomes expressed in monetary terms.

**Cost-Effectiveness Analysis:** An evaluation of the relationship between program costs and outcomes. Can be used to compare different interventions with the same outcomes to determine effectiveness.

**Cost-Utility Analysis:** An evaluation of the relationship between program costs and outcomes usually measured in quality-adjusted life years (QALY). Can be used to compare interventions with different outcomes.

**Economic Evaluation:** Evaluation that considers both the outcomes of a program and the cost of producing those outcomes.

**Evaluability Assessment:** A way of determining to what extent an evaluation is possible by examining program goals and objectives, available and collectable program data, and program activities.

**Formative Evaluation:** An evaluation designed to make sure that program plans, procedures, activities, materials, and modifications will work as planned.

**Impact Evaluation:** Examines the ultimate impact of a program on some type of community problem.

**Needs Assessment:** Used in program planning to plan for decisions about program implementation. It is a process by which information is collected from the target population or community in an effort to match the needs and wants of the target audience, the program organization, and the community.

**Outcome Evaluation:** An evaluation that measures changes the program has made in participants/clients based upon program objective.

**Process Evaluation:** An evaluation designed to assess how well the implementation of the program is going, such as the extent to which the program is operating consistently with objectives and procedures originally defined for them.

**Program Monitoring:** An element of process evaluation in which program activities are observed and recorded to ensure the quality and fidelity of daily program operations.

#### *Evaluation Methods*

**Qualitative Methods:** Ways of gathering data that are open-ended and descriptive and used to collect in-depth information from the target population. Examples: interviews, focus groups, observations, etc.

**Quantitative Methods:** Ways of gathering data (example surveys) that can be expressed numerically to collect information and draw conclusions about the changes the program had on participants.

**Quasi-Experiments:** An evaluation design/method where participants are not randomly assigned into groups.

**Participatory:** A more inclusive approach to evaluation that seeks out different levels of stakeholders to help shape evaluation design.

**Randomized Trials:** An evaluation design where participants are randomly/by chance assigned to groups that will either receive the program or not. This method is considered to produce the strongest evidence that the program contributed to changes in participants.

#### *Other Evaluation Terminology*

**Activities:** What the program does with the inputs to fulfill its goals.

**Effectiveness:** The extent to which a program was successful in producing a change among participants.

**Inputs:** Resources dedicated for program delivery. Examples are: money, staff, time, facilities, equipment, etc.

## References

- Alkin M. Databases on Evaluation. Newbury Park: Sage, 1990.
- CDC, Practical Evaluation of Public Health Programs. Prepared by the University of Texas Health Science Center School of Public Health and the Texas Department of Health. 1998.
- Chen HT. Theory Driven Evaluations. Newbury Park: Sage, 1990.
- Dehar M, Casswell S, Duignan P. Formative and Process Evaluation of Health Promotion and Disease Prevention Programs. *Evaluation Review* 17 #2: 204-220, 1993.
- Demers A, Renaud L. Formative Evaluation of a Nutritional Marketing Project in City-Center Restaurants. *Evaluation Review* 16 #6: 634-649, 1992.
- Drummond MF, et al. Methods for the Economic Evaluation of Health Care Programs. Oxford University Press, 1987.
- Farnham PG, Ackerman SP, Haddix AC. Study Design. Prevention Effectiveness: A Guide to Decision Analysis and Economic Evaluation. New York: Oxford University Press, 1996.
- Finnegan J, Murray D, Kurth C, McCarthy P. Measuring and Tracking Education Program Implementation: the Minnesota Heart Health Program. *Health Education Quarterly* 16 #1: 77-90, 1989.
- Fisher RJ, Peters L. The Role of Evaluability Assessment in Mental Health Program Evaluation. *Canadian Journal of Community Mental Health* 1:34. 1982.
- Fortmann SP, William PT, Hulley SB, Maccoby N, Farquhar JW. Does Dietary Health Education Reach Only the Privileged? *Circulation* 66 1: 77-82, 1982.
- Haddix AC, Ackerman SP, Shaffer PA. Cost-Effectiveness Analysis. Prevention Effectiveness: A Guide to Decision Analysis and Economic Evaluation. New York: Oxford University Press, 1996.
- Hatry HP. Where the Rubber Meets the Road: Performance Measurement for State and Local Public Health Agencies. Using Performance Measurement to Improve Public and Nonprofit Programs. American Evaluation Association: 31-44, 1997.
- Hausman A, Spivak H, Prothrow-Stith D, Roeber J. Patterns of Teen Exposure to a Community-based Violence Prevention Project. *Journal of Adolescent Health* 13: 668-675, 1992.
- Herman JL, Morris LL, Fitz-Gibbon CT. Evaluator's Handbook. Newbury Park: Sage, 1987.
- Marris BV, King B. Evaluation of Health Promotion Programs. Toronto: Center for Health Promotion, University of Toronto, 1998.
- McClintock C. Evaluators as Applied Theorists. *Evaluation Practice* 11 #1: 1-12, 1990.
- Mohr LB. Impact Analysis for Program Evaluation. Thousand Oaks: Sage, 1995.
- MMWR, Framework for Program Evaluation in Public Health Practice. 48 (RR11), CDC, September, 1999.
- National Institute on Drug Abuse. Handbook for Prevention Evaluation. Department of Health and Human Services. 1991.
- Newcomer KE. Using Performance Measurement to Improve Programs. Using Performance Measurement to Improve Public and Nonprofit Programs. American Evaluation Association: 5-14, 1997.
- Patton MQ. Utilization-Focused Evaluation, 3rd ed. Sage, 1997.
- Porteous NL, Sheldrick BJ, Stewart PJ. Program Evaluation Toolkit: A Blueprint for Public Health Management. Public Health Research, Education and Development Program, Ottawa-Carleton Health Department. 1997.
- Rossi PH, Freeman HE. Programs, Policies, and Evaluations. Evaluation: A Systematic Approach. Newbury Park: Sage, 1998.
- Rugg DL, O'Reilly KR, Galavotti C. AIDS Prevention Evaluation: Conceptual and Methodological Issues. *Evaluation and Program Planning* 13; 79-89, 1990.
- Rugh J. CARE Design, Monitoring, and Evaluation Resource Packet. Version 1.1, 1-118, 1996.
- Sanders JR. The Program Evaluation Standards, 2nd Edition. Sage, 1994.
- Schechter C, et al. Planning and Evaluating HIV/AIDS Prevention Programs in State and Local Health Departments, A Companion of Program Announcement #300. Department of Health and Human Services, PHS, CDC. 1993.
- Short L, et al. Family Violence: Building a Coordinated Community Response. A Guide for Communities. American Medical Association, 1996.
- Smith M. Evaluability Assessment: Reflections on the Process. *Evaluation and Program Planning* 13: 359-364, 1990.
- Smith M. Evaluability Assessment: A Practical Approach. Boston: Kluwer Academic, 1989.
- Smith N. Evaluability Assessment: A retrospective Illustration and Review. *Educational Evaluation and Policy Analysis* 3 #1: 72-82, 1981.
- Steckler A, Goodman RM, McLeroy KR, Davis S, Koch G. Measuring the Diffusion of Innovative Health Promotion Programs. *American Journal of Health Promotion* 6 #3: 214-224, 1992.

- Tessmer M. Formative Evaluation Alternatives. Beverly Hills: Sage, 1984.
- Weiss CH. Theory-based Evaluation: Past, Present, and Future. *New Directions for Program Evaluation* #76: 41-55, 1997.
- Wholey JS. Assessing the Feasibility and Likely Usefulness of Evaluation. *Handbook of Practical Program Evaluation*: 15-39, 1994.
- Wong-Reiger D, David L. Step-By-Step Guide to Program Design. *A Hands-On Guide to Planning and Evaluation*. Canadian Hemophilia Society: 33-60, 1993.
- Wong-Reiger D. Introduction to the Program Logic Model. *A Hands-On Guide to Planning and Evaluation*. Canadian Hemophilia Society: 23-32, 1993.
- Wyllie A, Casswell S. Formative Evaluation of a Policy Oriented Print Media Campaign. *Health Promotion International* 7 #3: 155-161, 1992.