Published in final edited form as:

Sex Transm Dis. 2016 February ; 43(2 Suppl 1): S11–S17. doi:10.1097/OLQ.0000000000000281.

Program Evaluation for Sexually Transmitted Disease Programs: In Support of Effective Interventions

Marion W. Carter, PhD

Division of STD Prevention, Centers for Disease Control and Prevention, Atlanta, GA

Abstract

Program evaluation is a key tool for gathering evidence about the value and effectiveness of sexually transmitted disease (STD) prevention programs and interventions. Drawing from published literature, the Centers for Disease Control and Prevention evaluation framework, and program examples, this article lays out some of the key principles of program evaluation for STD program staff. The purpose is to offer STD program staff a stronger basis for talking about, planning, conducting, and advocating for evaluation within their respective program contexts.

Program evaluation has widespread support within public health and among sexually transmitted disease (STD) programs, but it also can raise misunderstandings and frustration. On the one hand, it is hard to argue with the idea that programs should engage in some evaluation, to better assess their work and know whether and how to change course. In its implementation, however, evaluation can become more problematic, when limited resources and capacity, and an ever changing funding and health care landscape run counter to the lofty goals of evaluation. Moreover, defining what does and does not constitute evaluation can be difficult, with many complementary concepts circulating in public health. This article is intended to lay out key considerations for conducting evaluation in state and local health department STD programs. It provides an overview of relevant concepts, principles, and steps, with the aim of offering STD program staff a stronger basis for talking about, planning, conducting, and advocating for evaluation within their respective program contexts.

WHAT IS EVALUATION?

The Centers for Disease Control and Prevention Framework for Program Evaluation in Public Health defines evaluation broadly as the “systematic investigation of the merit, worth, or significance” of a program, whether in part or in its entirety.¹ The specific uses of evaluation can vary: to gain insight into a program, to change practices, to assess effects, or even to catalyze self-reflection by program stakeholders about their program.¹ The methods used are just as varied and can involve a range of qualitative or quantitative methods and kinds of data. Methodology alone cannot be used to identify an activity as evaluation. A hallmark of evaluation is—or should be—its usefulness to the program.

Correspondence: Marion W. Carter, PhD, Division of STD Prevention, Centers for Disease Control and Prevention, 1600 Clifton Rd, MS-E-80, Atlanta, GA 30333. acq0@cdc.gov.

Conflict of interest: None declared.

Evaluation is burdened by jargon, and many terms are used loosely and differently, though not necessarily incorrectly. In public health today, “monitoring and evaluation” competes with “program evaluation,” “performance measurement,” and now “quality improvement,” among other concepts. In the absence of consensus definitions, it is important to define these terms clearly whenever discussing them. In this article, the term “evaluation” is also used broadly and includes a wide range of approaches. Table 1 offers a short glossary of some key terms, including brief examples.^{2,3}

Drawing from recent published literature, Table 2 demonstrates the wide variety of evaluation approaches and uses that have been applied to STD program contexts. Program evaluation can be highly technical, as in the case of the dynamic modeling done of a school-based screening program in Philadelphia,⁴ or relatively simple, as in the case of an evaluation of a change in the way partner notification was conducted using existing program data.⁵ They can be quantitative, qualitative, or use a mix of methods^{6–9}; some include cost analyses.¹⁰ The scope of evaluation can range widely as well, from small scale to large scale, involving multiple clinical settings or programs nationwide.¹¹ They can produce positive and negative findings, both of which are important evidence to bring to bear in assessing effectiveness.¹²

Of course many program evaluations do not reach publication. Most STD programs engage in evaluation of some kind or another on a regular basis, for example, when they systematically review their partner services data and indicators for any issues or when they pilot test a new way of conducting community education or outreach. Any time that staff use data or experience to make changes, decisions, or judgments about their program, they are undertaking evaluation in one sense. By drawing explicitly on the principles of evaluation when doing so, however, they may take those efforts to another level of rigor and usefulness.

KEY PRINCIPLES TO APPLY IN EVALUATING STD PROGRAM STRATEGIES IN THE UNITED STATES

Given the programmatic and epidemiologic variety across STD programs in the United States, there is no single evaluation approach that will fit all contexts. No single set of evaluation questions is appropriate across programs, and no single set of outcomes and indicators has equal value to all programs. The following section describes a general approach that can be used by STD programs regardless of their level of morbidity or resources. Organized around the Centers for Disease Control and Prevention evaluation framework (Fig. 1) and using an STD program perspective, this section outlines some of these principles and potential pitfalls below. Table 3 summarizes key points.

Engage Stakeholders

Stakeholders for an evaluation are the people or agencies who are directly involved in implementing an evaluation or who are (or should be) invested in the evaluation findings. For an STD program evaluation related to partner services, for example, stakeholders may include disease intervention specialists (DIS) staff and supervisors, HIV program counterparts, and representatives from the top clinical agencies that work with the DIS

directly, among others. Engagement of them should entail genuine consultation through various available means, such as meetings, e-mails, conference calls, webinars, and so on. Program staff may wish to formally constitute an advisory group of stakeholders for an evaluation, or they may opt to engage stakeholders in a more informal, ad hoc way. Regardless of the approach to engagement, stakeholder engagement is often time consuming.

The core purpose of engaging stakeholders is to ensure the use of an evaluation. Although defined as the first step within the framework, engagement can benefit all other steps. Stakeholders can help define a program, prioritize evaluation questions and approaches, interpret findings, and map out the best ways to disseminate the findings. Stakeholders should also include those with some influence to make changes as a result of an evaluation. They not only guide the evaluation, offering technical and other advice, but they also should be positioned to advocate for it and its results, by virtue of their engagement. Token or minimalist engagement approaches (e.g., meeting and communicating rarely; not truly taking stakeholder feedback into account) can undermine both of these goals. Engaging the same set of individuals or organizations across evaluations also could mute the benefits of stakeholder engagement; stakeholders should be reviewed and identified for each evaluation effort. Different sets of people may be invested in or impacted by an evaluation of an STD partner services program than an evaluation of an STD surveillance system, for example.

Describe the Program

Clear program descriptions provide clear roadmaps for evaluation. When the intended outputs, outcomes, contextual influences, and relationships underlying a program or intervention are clarified, it is much easier to see what could be measured or documented and where additional study might most be needed. Logic models, theories of change, and systems diagrams are all tools for describing a program in ways that expose its intended logic. The choice of tool should depend on the preferences and needs of its users. Descriptions also can be developed at different levels. For example, an STD program may have a description for its entire program and could develop nested descriptions that go into greater detail about each component, such as its partner service work or policy strategies. For all their variations, STD programs tend to share short-term outcomes such as increased screening and treatment of STDs and long-term outcomes such as reduced STD prevalence and incidence.

The process of explicitly laying out the logic and context behind a program also can reveal strengths and weaknesses in a program's design, which have direct implications for the potential success of evaluation. Mismatches between program effort and intended outcomes may become apparent, as may important assumptions that might place a program at risk. For example, if an STD program planned to increase chlamydia (CT) screening rates in a city by holding 2 workshops with large groups of providers, the program description might help raise questions about how realistic it is to expect such change with relatively little intervention exposure. Measuring and tracking screening rates citywide as the primary marker of success of that particular intervention would be misguided. Also, are the STD program staff certain that lack of knowledge (as addressed through the planned workshops)

is the main reason CT screening rates are low? A program description can prompt these and other kinds of questions that may both strengthen a program and ensure that evaluation effort is well placed.

Focus the Evaluation

Evaluations should be focused so that they are useful and feasible. Like all programs, STD programs are complex, and it is nearly impossible to evaluate all components of them equally well, across the variety of evaluation questions that programs might be interested in—even if resources were unlimited. Therefore, careful choices should be made about how to focus one's evaluation effort. The process of narrowing and defining the focus of an evaluation begins with identifying what part of a program to evaluate. Various considerations (budgetary, political, scientific, etc) come into play in this choice. Common criteria include how pressing and real the information needs are (i.e., whether there are high stakes decisions to make about that component), how many resources are or may be invested in a program component, and how important a program component is to a programs' overall logic of effectiveness. For example, if only 1% of an STD program budget goes toward maintaining its Web site, and the Web site is recognized as contributing relatively little to the program's broader impact on STD screening and treatment, then the program may opt to regularly review routine Web metrics about utilization of the Web site but not to undertake more in-depth evaluation (e.g., survey of users to see what they learned from it). If an STD program were piloting a new approach to reaching providers reporting use of incorrect gonorrhea (GC) treatment, this may be prioritized for evaluation, because of the importance of this activity to the program's primary outcomes and potential for rolling out that approach across all of its local health departments. The rationale for any evaluation should be explicit and strong; stakeholders often can assist significantly in such deliberations.

From this point, the process of focusing an evaluation should continue, toward identifying the specific evaluation questions to answer and the specific kinds of data and evidence that will be used to answer those questions. Brainstorming about what might be important or interesting to know about a program's new approach to GC treatment improvement may generate numerous questions, which could quickly overwhelm the available evaluation resources of a typical STD program and its partners. Final decisions must balance information needs with available resources. This does not mean that a program evaluation should only seek answers to the easiest questions that rely on existing, easily accessible data or information. Rather, the rationale for asking new questions and collecting new data should be well justified. That rationale is often compelling because it is usually where data are fewest that the information needs are most acute.

Sometimes, in the course of describing a program strategy clearly or determining the most pressing information needs about a particular program strategy, staff may determine that formative evaluation or assessment is most needed. For example, staff may determine that before evaluating whether a new approach to improving GC treatment practices works, they need to obtain more information about the target providers and what the barriers to improved GC treatment are, to develop a program model that has a higher chance of success. Further evaluation of that new (presumably improved) program may then follow.

Gather Credible Evidence; Justify Conclusions

Once an evaluation plan is made, the goal is then to gather the evidence. The emphasis in evaluation is on *credible* evidence. Credible evidence tries to strike a balance between the feasibility of data collection and analysis and the usefulness of that evidence to informing program decisions. Evaluation draws on scientific methods of data collection and data analysis to ensure credibility. Evidence can come in both quantitative and qualitative forms, involving a range of potential data sources and methods. For example, a set of focus groups with DIS staff may serve to more formally identify barriers and potential solutions to improving partner services for MSM with syphilis. This could be complemented by tracking key indicators from partner services data before and after efforts made to reduce those barriers. From a scientific point of view, the lack of a control group of DIS who did not change their approaches or lack of randomization of index patients over time diminishes the ability to determine with certainty whether the program changes were responsible for any changes seen in the key indicators. However, the collective evidence may be plausible and credible enough to inform whether to continue with that approach or try something else.

Much evaluation work revolves around identifying and using key indicators. By definition, strong indicators have high credibility and scientific merit, as well as clear meaning. However, for STD programs situated in health departments, there is little consensus about the best indicators and targets to use to assess their programs. How do they measure the strength of their partnerships with clinical partners? What is an appropriate target for a program's surveillance capacity? In a given epidemiologic context, what makes for adequate treatment indices for partner services programs? Given shifting health care contexts, what benchmarks for STD screening and treatment should be set for different clinical settings? Even in the arena of clinical care, only 4 measures related to STDs have been endorsed by National Quality Forum, and only one, related to CT screening among young women, has been taken up by the Centers for Medicare and Medicaid Services as a measure of quality clinical care to date.¹ As a result, STD programs must continue to use indicators that make sense to their own context, while collaborating with one another to identify measures that are strongest and have broad application.

Table 4 provides a *starting point* for thinking about key outputs and outcomes for various kinds of STD interventions, as well as some example indicators for each. The specifics for all are highly dependent on program and epidemiologic context and on the specific program description and focus of an evaluation.

Ensure Use and Share Lessons Learned

In some ways, the measure of success for an evaluation is whether the results were used in a meaningful way. The prior steps should lead toward this. If relevant stakeholders understand the program and the evaluation, if the focus of the evaluation is important, and if the evidence gathered and analysis conducted are credible, then the results have a high chance of being actionable and used to inform the program. Use of results could lead to expansion of a

¹To search for these measures, go to the following Web sites for National Quality Forum and AHRQ and search for CT and for HPV (last accessed March 3, 2015): <http://www.qualityforum.org/QPS/QPSTool.aspx> and <http://www.qualitymeasures.ahrq.gov/content.aspx?f=rss&id=48812&osrc=12>

strategy, adjustments to a strategy, or discontinuation of a strategy. For example, a program may opt to conduct a simple cost analysis of their outreach screening work and find that in each of the last 6 months, they used approximately \$50,000 in staff time and \$10,000 in materials and transportation to identify 7 new cases of syphilis. Given their other budget pressures, they may use these results to discontinue their outreach screening and divert resources toward other strategies, while identifying whether and how to continue outreach work. This would be a laudable evaluation result.

The reason that evaluation results go unused usually ties back to weaknesses in one of the prior steps in evaluation. They also may get shelved because they are disseminated poorly, through modes that key stakeholders do not understand readily (e.g., highly technical, dense reports) or that do not facilitate discussion of what the implications of the results might be (e.g., a one-way presentation that reports results). The frequent absence of cost and other resources needed for an intervention or strategy also is an important barrier to evaluation results being used by other program contexts.

Applying the Framework to Funding Agencies

Sexually transmitted disease programs often receive outside funding and distribute funds to other agencies. The principles described earlier generally apply to funders as well. Funders should view funded agencies as key stakeholders who need to be engaged throughout the development and application of any evaluation requirements attached to funding. Funders' own program description, including the intended outputs and outcomes of the grant or contract, should be clear and realistic. Such descriptions should help set realistic expectations of funded agencies and guard against setting the program and funded partners up for failure. For example, funding a local health department a small amount, while asking them to increase CT screening countywide by 50% within 2 years, could be unrealistic and counterproductive. Similarly, funders also need to be focused about their own evaluation needs and questions. Asking for a great deal of data or evidence without clear rationale for the need for those data also can weaken those funded partnerships. The information obtained from funded agencies must be used with scientific integrity and credibility and clearly used for appropriate evaluation purposes. Not doing so can weaken the collective culture of evaluation and people's attitudes toward evaluation in general.

CONCLUSIONS

This article provided a brief primer on program evaluation in the context of STD programs. The aim was not to serve as a how-to for evaluation, but rather to make some of the high-level principles and concepts more accessible to STD program staff. Various resources exist for more in-depth explanations, examples, and tools to use to implement various aspects of evaluation for STD programs.² It is worth recognizing that some STD program staff and their partners have negative experiences or perceptions about evaluation, including performance measures—experiences that usually can be traced back to some of the pitfalls outlined above. Working to overcome these barriers is essential because evaluation is one of the tools for identifying more effective and efficient ways forward for STD programs.

Evaluation of all STD programs, large and small, is within reach. Having research expertise is not a prerequisite for planning or conducting evaluation. Sexually transmitted disease program staff often have relevant expertise among themselves or within their larger organizational units. Obtaining technical assistance from external evaluators or scientists on evaluation design and analysis can be worthwhile and does not necessarily need to be costly. Although obtaining expertise can be helpful, evaluation is not only for evaluators; rather, it is a function and responsibility that should engage staff from across an STD program.

As the stewards of public resources, local, state, and federal STD programs need to ensure that their resources are being used in the most effective and efficient ways possible. Application of evaluation approaches can help identify programmatic activities that may be of limited value, whether they are old or new strategies. Similarly, evaluation can help establish the evidence base for strategies that may be underresourced, yet have a large potential for effectiveness. Evaluation and quality improvement approaches in particular also facilitate systematic tinkering with program components to make gains in efficiency and effectiveness. Increasingly, STD programs need to see when and how they can track outcomes, in addition to outputs, and identify opportunities to include cost analysis in their evaluations, as part of a broader effort to assess effectiveness and value of different approaches. Optimally, this evidence builds from the ground up, with observations and small evaluations contributing toward an evidence base that all STD program staff can draw from. Large-scale evaluations and formal research and demonstration projects can contribute as well. Collectively, these efforts help broaden the evidence base for STD programs more generally and help us all move more swiftly toward the ultimate goal of decreased STD incidence and disease burden.

Acknowledgments:

The author would like to thank Brandy Maddox, Dayne Collins, Tom Peterman, Kyle Bernstein, and Tammy Foskey for comments on earlier drafts.

The content of this manuscript are those of the author and do not represent the official position of the Centers for Disease Control and Prevention.

REFERENCES

1. Centers for Disease Control and Prevention. Framework for program evaluation in public health. MMWR Recomm Rep. 1999; 48:1–40.
2. Centers for Disease Control and Prevention. Practical Use of Program Evaluation among Sexually Transmitted Disease Programs. 2014 Available at: <http://www.cdc.gov/std/program/pupestd.htm>. Accessed November 10, 2014.
3. Centers for Disease Control and Prevention (CDC). Program Operations Guidelines for STD Prevention: Program Evaluation. 2001 Available at: <http://www.cdc.gov/std/program/progevaluation.pdf>. Accessed November 20, 2014.
4. Fisman DN, Spain CV, Salmon ME, et al. The Philadelphia High-School STD Screening Program: Key insights from dynamic transmission modeling. Sex Transm Dis. 2008; 35(11 suppl):S61–S65. [PubMed: 18607306]
5. Osterlund A Improved partner notification for genital chlamydia can be achieved by centralisation of the duty to a specially trained team. Int J STD AIDS. 2014; 25:1009–1012. [PubMed: 24616118]
6. Udeagu CC, Shah D, Shepard CW, et al. Impact of a New York City Health Department initiative to expand HIV partner services outside STD clinics. Public Health Rep. 2012; 127:107–114.

7. Huppert JS, Reed JL, Munafo JK, et al. Improving notification of sexually transmitted infections: A quality improvement project and planned experiment. *Pediatrics*. 2012; 130:e415–e422. [PubMed: 22753557]
8. Introcaso CE, Rogers ME, Abbott SA, et al. Expedited partner therapy in federally qualified health centers—New York City, 2012. *Sex Transm Dis*. 2013; 40:881–885. [PubMed: 24113414]
9. Hutchinson J, Evans D, Sutcliffe LJ, et al. STIFCompetency: Development and evaluation of a new clinical training and assessment programme in sexual health for primary care health professionals. *Int J STD AIDS*. 2012; 23:589–592. [PubMed: 22930298]
10. Rukh S, Khurana R, Mickey T, et al. Chlamydia and gonorrhea diagnosis, treatment, personnel cost savings, and service delivery improvements after the implementation of express sexually transmitted disease testing in Maricopa County, Arizona. *Sex Transm Dis*. 2014; 41:74–78. [PubMed: 24326585]
11. Chesson H, Owusu-Edusei K Jr. Examining the impact of federally-funded syphilis elimination activities in the USA. *Soc Sci Med*. 2008; 67:2059–2062. [PubMed: 18952341]
12. Rietmeijer CA, Westergaard B, Mickiewicz TA, et al. Evaluation of an online partner notification program. *Sex Transm Dis*. 2011; 38:359–364. [PubMed: 21343844]
13. Riley W, Lownik B, Halverson P, et al. Developing a taxonomy for the science of improvement in public health. *J Public Health Manag Pract*. 2012; 18:506–514. [PubMed: 23023274]

Figure 1.
A framework for program evaluation in public health (1999).¹

TABLE 1.

Glossary of Key Evaluation-Related Terms and STD Program Examples

An indicator is a specific, observable, and measurable accomplishment or change that shows whether progress has been made toward achieving a specific program output or outcome.²

- % of syphilis patients interviewed within 2 wk; % of sexually active young women tested annually for CT

Outputs are the direct products of your program activities or services delivered.² These are more likely to be in the program's "zone of control."²*

- Provider education packets distributed; index patients interviewed

Outcomes are the intended effects or changes in the target populations/organizations that result from your program.² These are more likely to be in the program's "zone of influence."²*

- Increased STD screening by HIV care providers; new cases identified through screening; new cases identified through partner services

Process evaluation is a type of evaluation that determines whether your program and its activities are implemented as intended and why/why not. Information gathered is used for refining or modifying these activities and related procedures. The inputs, activities, and outputs of a logic model (the left side) are used to plan a process evaluation.²

- A program wanted to improve their partner services interview rate. So they conducted interviews with DJS and focus groups with a sample of former patients to identify how the current protocol was working and areas for improvement. Then they developed a new protocol, and then tracked interview-related indicators over time to decide whether to finalize the new approach.

Program monitoring is an element of process evaluation in which program activities are observed and recorded to ensure the quality and fidelity of daily program operations.³

Performance measurement is like program monitoring but emphasizes the use and tracking of process or outcome indicators against benchmarks, to assess whether programs are heading in the right direction.

- A program had 8 indicators (6 process and 2 outcome measures) that it tracked and reported to its Division head. The division and program leadership used the information as a basis for identifying and discussing areas that appeared to be on track and areas that may have had problems.

Outcome evaluation is a type of evaluation that determines the effects of your program activities in the target population or organization (e.g., changes in knowledge, attitudes, skills, and practices). The outcome components of a logic model (the right side) are used to plan an outcome evaluation.²

- Program wanted to see whether their efforts to increase screening rates among 3 high-priority HIV care providers worked. They obtained data from all providers and tracked screening rates and new cases identified, preintervention, and postintervention with providers.

Formative evaluation is an evaluation designed to make sure that program plans, procedures, activities, materials, and modifications will work as planned.³

- A series of interviews conducted with primary care providers about barriers to screening

Quality improvement is the use of a specific method to improve a public health process with an identifiable beginning and end, using QI tools based on a defined aim statement and measurable metrics.¹³ QI complements evaluation.

- An STD clinic wanted to reduce clinic wait times, so they used QI tools to identify likely causes of the longer wait times, instituted some clinic changes, and tracked key indicators over time to see if the changes improved the wait time.

QI indicates quality improvement.

* Training slides delivered by Tom Chapel, Chief Evaluation Officer, Centers for Disease Control and Prevention, April 2012.

TABLE 2.
Examples of the Breadth of Published Program Evaluations Conducted in STD Program Contexts

Reference	Program Evaluation Purpose	Data Sources	Select Findings
Osterlund ⁵	Determine whether the centralization of partner notification (PN) duties to a specialty trained team improved their PN outcomes	Existing program data on PN outcomes such as number of PNs conducted	Centralization improved the number of partners notified per index case
Introcaso et al. ⁸	Assess the status of EPT policy knowledge and evaluate uptake among FQHCs and barriers to using	Survey sent to approximately 100 entities and sites	EPT practices do not align with guidelines; various barriers identified
Hutchinson et al. ⁹	Evaluate the acceptability, reach, costs, and outcomes associated with a new sexual health training program	Phone interviews, teaching logs, questionnaire, look at pre-post-CT screening rates	Program was rated highly, but had weaknesses; CT screening rose but not clearly due to intervention
Udeagu et al. ⁶	Determine whether the creation of an HIV field services unit to outpost STD DIS to do partner services in HIV clinical sites in select neighborhoods improved P outcomes	Partner services program data	Outcomes of partner services improved at outposted sites; providers and staff satisfied. Led to expansion of program
Rietmeijer et al. ¹²	Determine whether implementation of Colorado "inSPOT" (online PN program) was effective	Clinic-based surveys of clients, pre- and post-; use of Web site data	"inSPOT" recognition was low; strong preference for in-person notification; evidence of usage of system for jokes
Rukh et al. ¹⁰	Determine how the implementation of an express STD clinic affected clinic operations and clients seen	Program data on, e.g., number of clients, client profile, turn-aways, costs	New system routed the right people, led to some cost savings and reduction in clinic turn-aways
Huppert et al. ⁷	Identify ways to increase STD testing and STD result notification among adolescents seen in the ED	QI methods; QI database created, using EMR	Improvements were observed on both fronts, particularly thanks to dedicated cell phone
Chesson and Owusu-Edusei 2008 ¹¹	Determine whether federal funding for syphilis elimination affected syphilis rates	Dataset created from existing data, such as funding, SE cases, etc. by jurisdiction	Elimination funding seems to have had an effect at decreasing syphilis
Fisman et al. 2008 ⁴	Determine whether, and for whom, the high school STD program screening program was effective	Program data, published literature; dynamic modeling program	School based screening was cost-effective and enhanced by including males

ED indicates emergency department; EMR, emergency medical responder; QI, quality improvement; EPT, expedited partner therapy; FQHC, federally-qualified health centers; PS, partner services.

TABLE 3.

Summary of Key Aspects of Conducting Evaluation in an STD Program

	Key Points	Potential Pitfalls
Engage stakeholders	<ul style="list-style-type: none"> Purpose is to lay groundwork to ensure that evaluation is useful (i.e., understood, supported, and positioned to make a difference) Stakeholders may vary by evaluation focus; various means of engagement possible Engagement should occur throughout the evaluation 	<ul style="list-style-type: none"> Token engagement Wrong/Same people at the table
Describe the program	<ul style="list-style-type: none"> Purpose is to clarify a program or project's goals and use that as a roadmap for what to evaluate Logic model is one or various tools for doing that effectively (other tools may suit situation better) Process of describing a program this way can reveal program strengths/weaknesses or key assumptions 	<ul style="list-style-type: none"> Having unrealistic outcomes for a given program effort/setting oneself up for failure Rushing this process, assuming everyone has same understanding of a program Getting caught up in the format of the program description (e.g., what is an output vs. outcome)
Focus the evaluation	<ul style="list-style-type: none"> Purpose is to focus evaluation effort toward what the program really needs and can do, within its constraints Various criteria for prioritization are relevant such as how pressing the questions are, how relate to program resources, and how important to potential program impact Focus occurs at various levels: the evaluation topic, evaluation questions, and data and information sought 	<ul style="list-style-type: none"> Planning to evaluate too much or what is easy as opposed to needed Asking important evaluation questions too late; evaluations are retrospective or after-the-fact Focusing on what's easy and known, to avoid exposing weaknesses (risk averse)
Gather credible evidence; justify conclusions	<ul style="list-style-type: none"> Purpose is to use scientific methods to create actionable and credible results Obtaining scientific expertise is often beneficial but not a prerequisite for participation in evaluation Strong, universal metrics of a good STD program are relatively few; more progress and program collaboration are needed 	<ul style="list-style-type: none"> Data collection efforts that are burdensome but not well reasoned or useful Not asking for technical assistance in evaluation design and analysis, when warranted Misinterpretation or misuse of data or results
Use results and share lessons learned	<ul style="list-style-type: none"> Usefulness of an evaluation is a key measure of its success or failure Links back to the stakeholder engagement 	<ul style="list-style-type: none"> Not taken seriously or given sufficient time to do thoroughly or in ways that promote use of results Lack of a culture of shared success and failure, which allows all results to be really used as much as they could be

TABLE 4.

A Starting Point for Thinking About Key Domains for Outputs, Outcomes, and Associated Measures for STD Prevention and Control Programs

Outputs		Outcomes	
Clinic-based, community-based, and school-based screening or treatment	• Partnerships: e.g., #/% of targeted medical providers, schools, or organizations where an agreement (written/verbal) to implement, increase, or otherwise improve screening, or treatment practices exists	• Screening: e.g., CT screening among young sexually active women	
	• Reach: e.g., #/% of those above who made changes; #/% of those who received training or education; % of eligible community or school population that was offered services; #/% of community or school population that used services	• Positivity: e.g., associated % positive, % treated or (for HIV) linked to care	
	• Intensity and mode of delivery: e.g., # of minutes of education/training/technical assistance provided per provider or per community member; type/degree of any structural changes made at a provider site	• Treatment: e.g., correct GC TX provided, % of positive cases identified treated; #/% of eligible CT patients receiving partner treatment (Rx or meds)	
	• Cost: e.g., program costs (to health department; to provider system) to implement structural change made; program costs (staff, transportation, materials, laboratory) to implement screening program	• Cost: e.g., cost per new positive CT case treated	
Partner services	• Priority cases interviewed: e.g., % of cases with early syphilis interviewed within Y days of initiation	• New disease identification: e.g., % of partners to GC that are newly diagnosed as having GC	
	• Partners claimed: e.g., total # of partners claimed per GC case	• New disease identification: e.g., % of partners to early syphilis cases that are newly diagnosed as having HIV	
	• Partners initiated: e.g., total # of partners initiated per HIV case	• Timely disease intervention: e.g., % of GC cases where at least 1 partner was brought to treatment or epi treated for GC within Y days of interview of the index patient	
	• Partners examined/tested: e.g., total # of partners examined per early syphilis case or % of partners to GC cases that are tested for HIV		
Linkage to care	• Reach, referral: e.g., #/% of those clients who were eligible for linkage to care services and who were given a written referral for linkage to care	• Timely linkage to care: e.g., #/% of those clients linked to HIV care within 90 d of HIV dx	
		• Maintenance in care: e.g., #/% of those clients retained in care for a year, 2+ visits in 12 mo, at least 90 d apart	
		• Reengagement in care: e.g., #/% of clients who had fallen out of care that were linked to HIV care within 90 d of being identified as out of care	
Social marketing	• Strategies: e.g., # and type of products disseminated, # of channels used	• Knowledge or attitudes: e.g., % of population with correct knowledge	
	• Reach: e.g., % of surveyed population aware of campaign	• Campaign-specific actions: e.g., %/# that called in to the hotline; % who discussed STDs with a friend	
	• Intensity: e.g., average # of times audiences saw a campaign message	• Health care seeking behaviors: e.g., % of population that was tested for STDs in last year	
	• Cost: e.g., cost of staffing, placement and incentives, evaluation, etc	• Sexual behaviors: e.g., % of population that used condoms at last sex	
Behavioral counseling	• Approach: e.g., type(s) of counseling models used; % type of provider group (STD clinic or primary care practice)	• Knowledge or attitudes: e.g., % of clients receiving behavioral counseling with an increase in STD prevention skills; % of clients receiving	

Outputs		Outcomes	
•	Reach: e.g., #/% of clients that received behavioral counseling when indicated via a sexual risk assessment	•	behavioral counseling with an increase in sex communication or sex negotiation skills
•	Intensity: e.g., total # of sessions of behavioral counseling provided per eligible client	•	Sexual behaviors: e.g., % of clients receiving behavioral counseling who had an increase in condom use consistency
•	Cost: e.g., staff and materials cost/client that received behavioral counseling; % delivered with a health care worker reimbursable by Medicaid	•	Incident infections: e.g., % of clients receiving behavioral counseling who tested positive for STDs within 3 and 6 mo of behavioral counseling