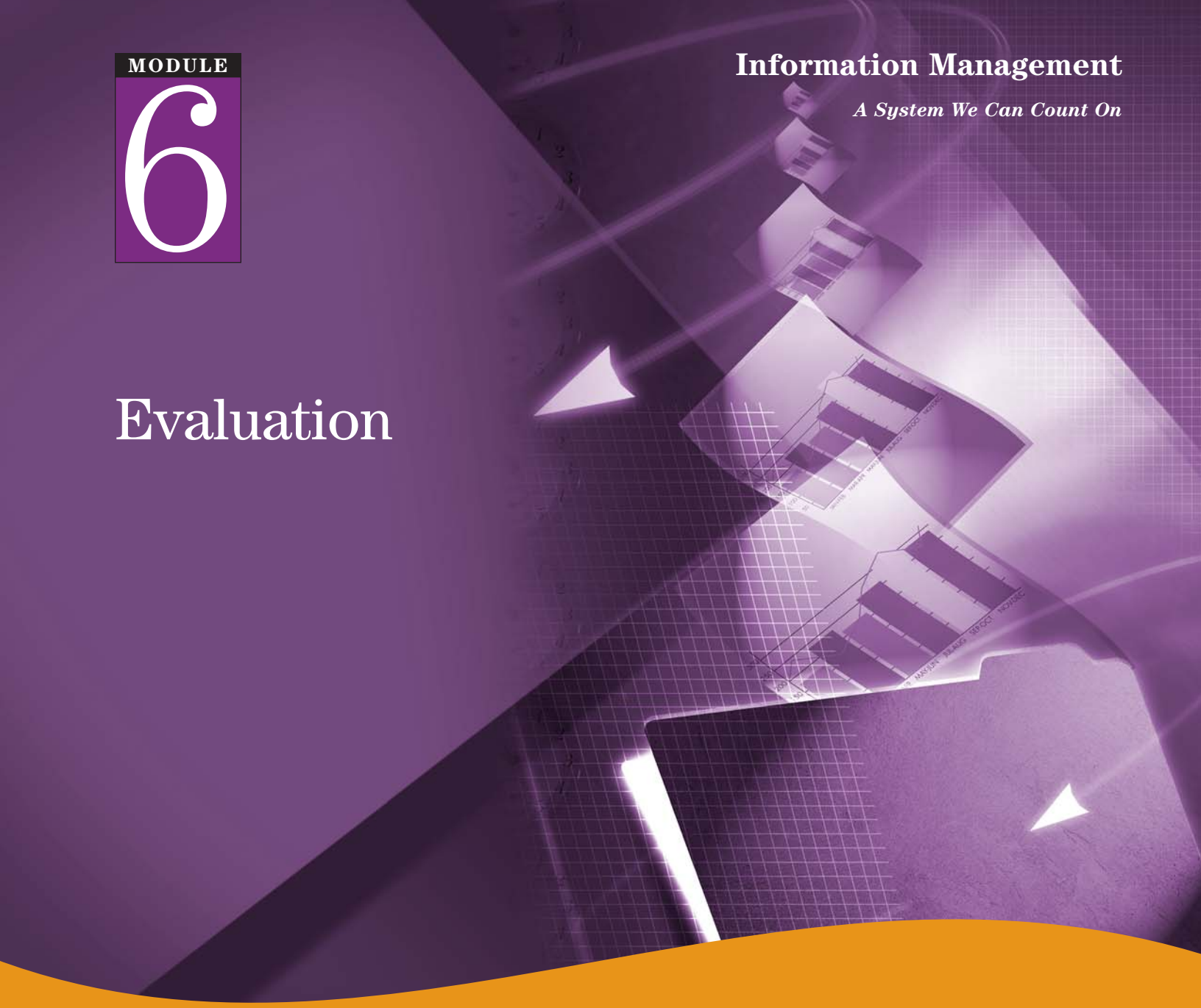
MODULE 6

Information Management

*A System We Can Count On*



# Evaluation

# The Health Planner's Toolkit

*Health System Intelligence Project – 2008*

Ontario

# Table of Contents

# About HSIP

## Health System Intelligence Project (HSIP)

The Health Planning Toolkit is produced by the Health System Intelligence Project. HSIP consists of a team of health system experts retained by the Ministry of Health and Long-Term Care's Health Results Team for Information Management (HRT-IM) to provide the Local Health Integration Networks (LHINs) with:

- sophisticated data analysis;
- interpretation of results;
- orientation of new staff to health system data analysis issues; and
- training on new techniques and technologies pertaining to health system analysis and planning.

The Health Results Team for Information Management created the Health System Intelligence Project to complement and augment the existing analytical and planning capacity within the Ministry of Health and Long-Term Care. The project team is working in concert with Ministry analysts to ensure that LHINs are provided with analytic supports they need for their local health system planning activities.

## Report Authors

**Sten Ardal**
**John Butler** (Module 6 Lead Author)
**Jane Hohenadel**
**Dawn Olsen**

## Acknowledgements

**Arnold Love**, Evaluation Consultant

# Introduction

Gauhar is an integration consultant on the staff of a Local Health Integration Network (LHIN) in Ontario. She has been asked to join a team to design an evaluation of an inpatient program for acutely ill elderly persons. The evaluation's sponsor (the local hospital) wants to know whether the program has achieved the outcomes proposed for it when it was established.

Gauhar has also been asked to join another team that will create an overall evaluation strategy for a community support program that has not yet been created – it is still in the design stage. She anticipates that the evaluation strategy for this program will include, but not be limited to, a strategy for evaluating the outcomes of this program. She believes it is also important to evaluate:

- the ingredients or resources that will go into the program;
- the way these resources are used; and
- the strengths and weaknesses of the program's activities.

Gauhar also recognizes that both evaluations face challenges because of a recent program evaluation in the community that went horribly wrong. It became an exercise in assigning blame rather than improving the program. She wants to be sure the two evaluations currently being designed will address stakeholder anxieties aroused by the recent failed evaluation. She also believes it will be necessary to explain to stakeholders that evaluation is not a mysterious and malevolent process.

Over coffee, Gauhar muses about these challenges with her colleague Gabriel, who brought a cake to work to share with the LHIN's staff. Gabriel points out that evaluation is not mysterious – *"I used basic evaluation processes to decide whether I successfully baked this cake – and these are the same processes that you and your team members will use to evaluate complex, expensive and crucial health care programs. It has its complexities – but it isn't rocket science."*

For readers who like both cakes and health services, Appendix A provides a comparison showing what can be evaluated, both in cake-baking and in the operation of a health service program. While a "master chef" or an evaluation expert is sometimes needed, the basics of a cake or an evaluation are easy to understand.

# The Module's Purpose

This module will not turn the reader into an evaluation expert. It will provide basic information about evaluation so the reader can grasp the essential concepts and activities that comprise evaluation. It will help the reader to get the most from evaluation and to know when and how to use it.

This module begins by identifying the three linked components of any activity or program – inputs, activities and outcomes (see Figure 1) – and by identifying and describing the types of evaluation appropriate for these components.

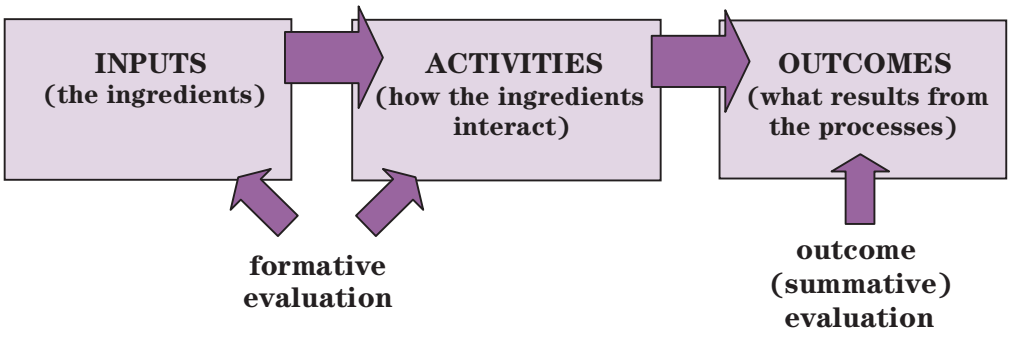
It then discusses planning an evaluation and outlines the steps involved in preparing and conducting an evaluation.

The module outlines evaluation challenges and provides tips on how to conduct an evaluation successfully.

This module does not identify and describe the many data-related tools and processes used in evaluation. However, since many of these tools and processes have to do with data as evidence (as clues about what happens and why), a reader wanting more background may want to read Module 3 (*Evidence-Based Planning*) in the Planner's Toolkit.

Evaluation can vary widely in scope. It may be limited to evaluating one activity within a program, or an entire program, or several programs comprising an agency, or several activities or programs scattered across a number of agencies. For the sake of simplicity this module assumes that the unit of analysis is a program.

**Figure 1: Program Components and Types of Evaluation**



Section 1

# What is Evaluation?

Program evaluation is:

**"The systematic gathering, analysis and reporting of data about a program to assist in decision-making."[1]**

Evaluation shows whether a program is accomplishing its goals. It also identifies program weaknesses and strengths, areas of the program that need revision, and areas of the program that meet or exceed expectations. To do this, analysis of any or all of a program's domains is required:[2]

- the need for the program;
- the design of the program;
- the program's implementation and service delivery;
- the program's outcomes; and
- program efficiency.

At its core, evaluation asks three broad questions:

- What should happen?;
- What actually happened?; and
- Why did it happen?

But what is a program? It is a group of related activities intended to achieve specific outcomes. *It is "the embodiment of ideas about means of achieving desired social objectives."*[3] Accordingly, "... how ideas get implemented and what is their impact are the dual concerns of program evaluation."[3]

## 1.1 Evaluation as a Key Planning Component

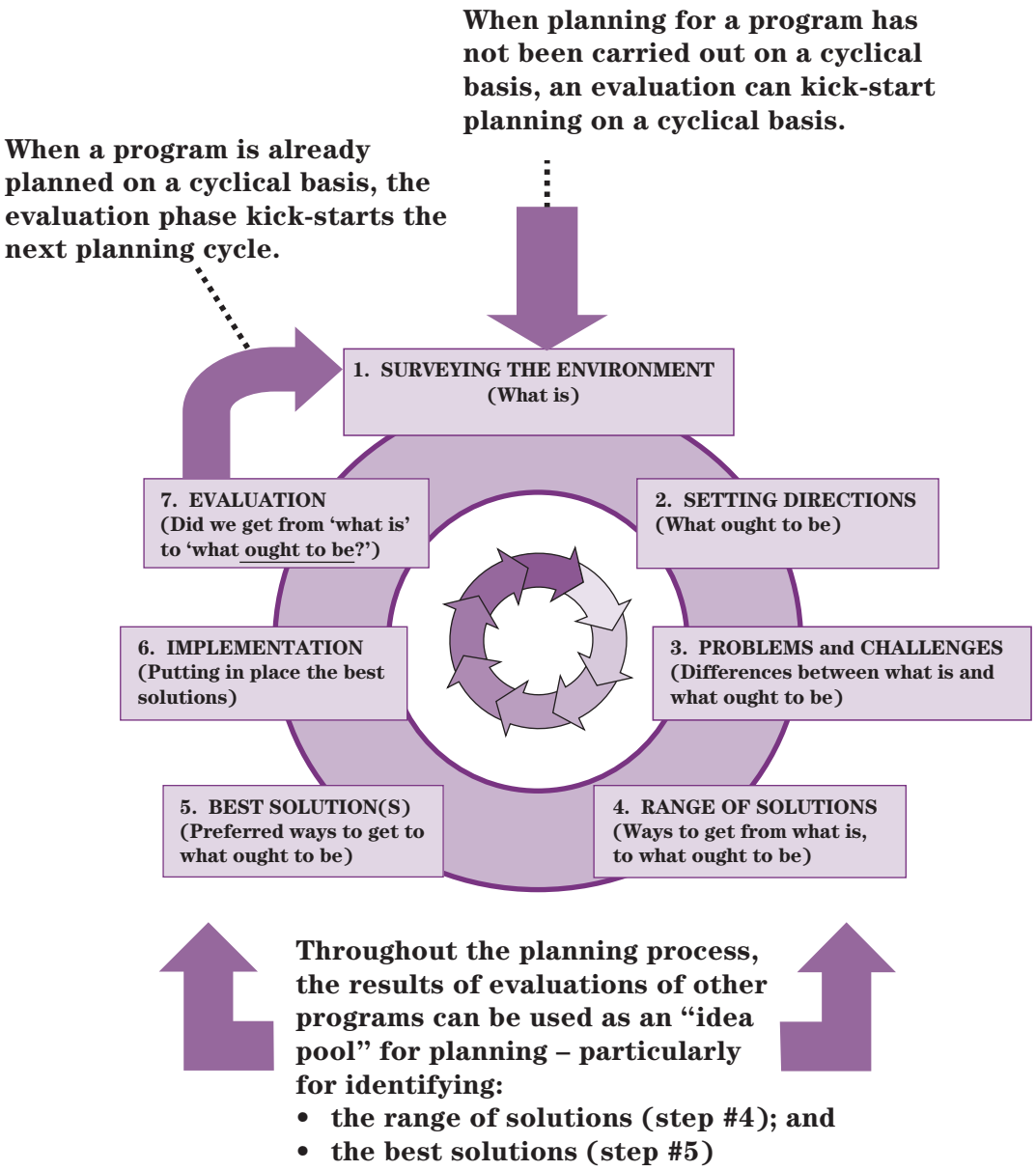
Evaluation is an essential component of planning. Module 1 (*The Planning Process*) in the Health Planner's Toolkit presents a cyclical planning model, and at several key points in the cycle, evaluation activities and other planning activities coincide or influence each other as illustrated in Figure 2.

"Evaluation, especially when it is focused on how well an organization or program is meeting its goals, can be quite turbulent… One of the best ways to prepare for and forestall this turbulence is to make the evaluation process part of an overall planning process."

– Randy Stoeker,
*Making Connections: Community Organizing, Empowerment Planning, and Participatory Research in Participatory Evaluation*[4]

Planning is a continuous cyclical process that takes into account both changed circumstances and the effects of implementing previous planning. However, one cycle of planning cannot learn from previous cycles unless monitoring and evaluation processes are put in place to determine the effects of earlier cycles. A key for success with any planning effort is agreement at the beginning on what will be tracked and evaluated.

**Figure 2: Evaluation's Role Within Planning**



## 1.2 When Is Program Evaluation Desirable?

Program evaluation is often used when programs have been functioning for some time. This is called **retrospective evaluation**. However, evaluation should also be conducted when a new program is being introduced or when a program from another jurisdiction is being introduced in a new environment. These are called **prospective evaluations**. A prospective evaluation identifies ways to increase the impact of a program on clients; it examines and describes a program's attributes; and, it identifies how to improve delivery mechanisms to be more efficient and less costly.

These benefits go beyond demonstrating the degree to which a program has succeeded or failed. Evaluations help program managers understand the reasons for program performance, which may lead to improvements or refinements to the program. Evaluations also help program funders to make informed judgments about the program's worth and help funders understand the reasons for a program's success or failure so it can be implemented successfully in other sites.

> "What gets measured, gets done. If you don't measure results, you can't tell success from failure. If you can't see success, you can't reward it. If you can't reward success, you're probably rewarding failure. If you can't recognize failure, you can't correct it. If you can demonstrate results, you can win public support."
>
> – D. Osborne and T. Gaebler, *Reinventing Government*[5]

The potential benefits of an evaluation are important considerations in making the decision to evaluate.

Evaluation often compares what **ought to happen** against what **actually happened** and attempts to account for any differences between the two. Put another way, it compares the optimal program (as its designers and managers envisioned it) with the actual program – and it can compare the "oughts" and the "actuals" for inputs, activities or outcomes. This is classic activity/outcome evaluation using the goal-based evaluation model of the 1970's and 1980's, but the use of evaluation has broadened considerably, exemplified by five key benefits identified in evaluation literature: [6]

1. accountability for program performance and spending;
2. improved decisions about program direction, allocation of resources, program design, implementation, management, efficiency and evaluation;
3. increased understanding of the program and of client needs, and increased capacity for program design, assessment, and improvement;
4. social change arising from the promotion of different programs, the shaping of public opinion, or the cultivation of pluralism and democracy; and
5. increased cohesion and collaboration among the program team and other stakeholders.

Though there may be a need for information to inform decisions, a formal evaluation may not always be the best choice. For example, when managing performance or tracking activity, monitoring rather than a formal evaluation might be a better choice (see section 1.4).

If the intent is to test the efficacy of a new intervention, an economic evaluation might make more sense. Economic evaluation involves a comparison between alternative courses of action, evaluating the options in terms of both their costs and their benefits.[7] Although economic evaluation can be complex, its scope is narrower than a full scale formal program evaluation (see Module 3, *Evidence-Based Planning*).

## 1.3 An Evaluation Matrix

Evaluation has been used in many disciplines and contexts, resulting in many different classifications of evaluation types. The broadest and most common classification of evaluation identifies two kinds of evaluation:

1. **formative evaluation.** This generally refers to evaluation of components of a program other than their outcomes. For instance, a formative evaluation may evaluate the degree of need for the program, or the activities used by the program to achieve its desirable outcomes, but without evaluating the degree of outcome.

2. **summative evaluation.** This generally refers to evaluation of the degree to which a program has achieved its desired outcomes, and the degree to which any other outcomes (positive or negative) have resulted from the program.

As the previous section of this module indicates, evaluation also has a timing dimension. It can be:

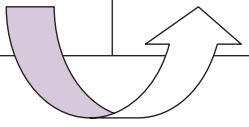
- **prospective**, meaning it determines what ought to happen (and why); or

- **retrospective**, meaning it determines what actually happened (and why).

Based on these two dimensions, a matrix describing the kinds of evaluation helps in understanding evaluation (see Table 1). This module's following sections provide descriptions of each kind of evaluation found in the matrix.

**Table 1: An Evaluation Matrix**

| THE COMPONENT DIMENSION | THE TIMING DIMENSION | |
|---|---|---|
| | **Prospective Evaluation** | **Retrospective Evaluation** |
| Input Evaluation / Activity Evaluation (combined and called Formative Evaluation) | What should the program's inputs be (and why)? | What were the program's inputs (and why)? |
| | What should the program's activities be (and why)? | What were the program's activities (and why)? |
| Outcome (Summative) Evaluation | What should the program's outcomes be (and why)? | What were the program's outcomes (and why)? |



**Prospective evaluations can produce monitoring strategies.**

**Retrospective evaluations can benefit from monitoring strategies.**

## 1.4 Monitoring

In addition to these kinds of evaluation, **monitoring** (sometimes called monitoring and assessment) should take place to support evaluation. Monitoring is the constant or recurring collection and examination of selected information on program activity over the life of the program. This information can be used for two purposes:

1. to alert the program to changes in program operation that might be signals of possible program failure; and

2. to provide a body of information that will be used when each kind of evaluation is carried out.

Monitoring can emerge from prospective evaluations, and can provide raw material for retrospective evaluations.

> Monitoring is the constant or recurring collection and examination of selected information on program activity over the life of the program.

Some evaluation analysts consider monitoring to be a variant of evaluation (a series of "mini-evaluations"). Other analysts consider it to be separate from evaluation but an important adjunct. In either case, developing an approach to evaluation should also include developing an approach to monitoring. Without monitoring, evaluators can find themselves scrambling to gather data that should have been gathered on an ongoing basis – and they may find that with the passage of time it is no longer possible to gather some of this information.

### An Example of Insufficient Monitoring

A community mental health agency operates a recovery program for people living with bipolar disorder. The program has three phases. Evidence from similar programs shows that positive client outcomes are much higher when clients participate in all three program phases before leaving the program.

This program was created based on a prospective evaluation that determined desirable inputs, activities and outcomes. As part of this evaluation the program's designers determined that the client drop-out rate prior to completion of the program should be no more than 10% of clients.

The program intends to conduct a retrospective process evaluation two years after the start of the program and a retrospective outcome evaluation four years after the start.

However, the program has not put in place an ongoing monitoring process. While each client record indicates the date on which the client leaves the program, the program does not track, on a monthly basis, the percentage of clients who leave the program before completing all three program phases (i.e., it does not track the drop-out rate).

When the program conducts a process evaluation two years after it started, it decides that it needs to know the drop-out rate as part of the evaluation. It must now go over two years of client records to calculate the drop out rates and whether they have increased or decreased over the two years. It finds that for the program's first year, the client drop-out rates per month averaged 10% per month, but for the next year, they averaged 25% per month – much higher than the anticipated drop-out rate.

**If the program had established a monitoring process to be used on an ongoing basis, it would have been able to identify, account for and develop corrective action on drop-out rates much earlier in the life of the program.**

The kinds of evaluation described are not necessarily separate from each other. Each is like a specialized radar set, scanning its own section of the sky so it can make a unique contribution toward a comprehensive evaluation. Taken together, the scans give a comprehensive picture of what is in the sky.

In an ideal comprehensive evaluation process, each kind of evaluation is carried out at the appropriate time in the life of the program and provides information that can enlighten the next kind of evaluation.

The evaluations described in this section are described in greater detail in the pages that follow.

## 1.5 Formative Evaluation

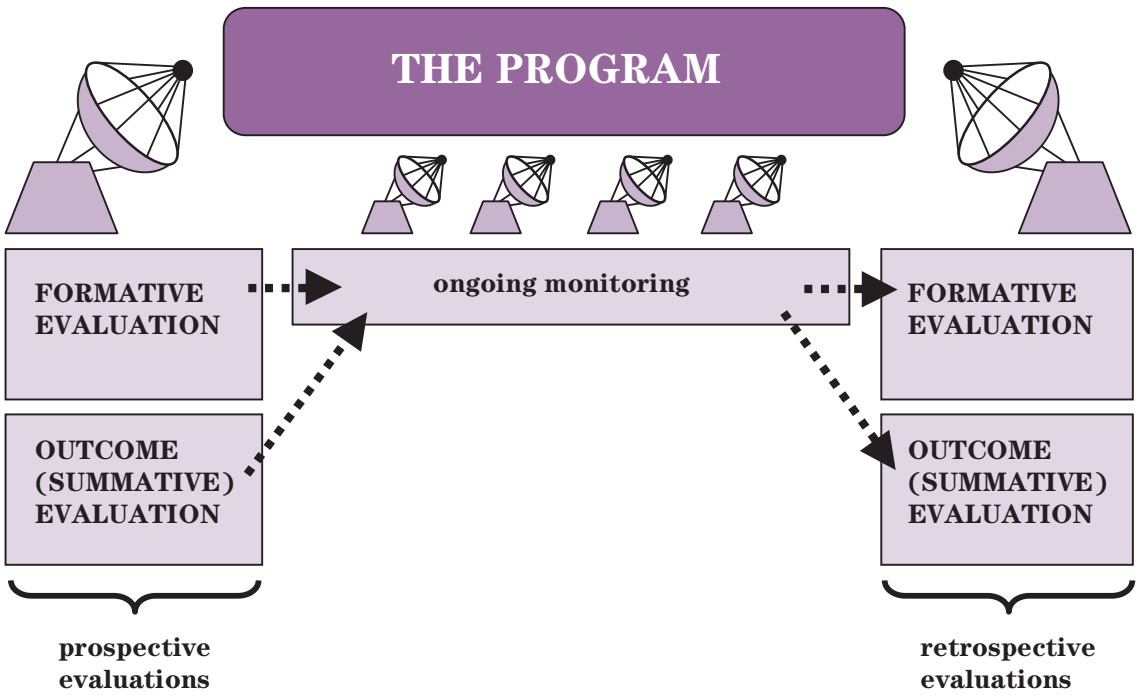
Formative evaluation is akin to Total Quality Management (TQM) and Continuous Quality Improvement (CQI) since all these approaches are a commitment to constantly improve operations, processes and activities to meet client requirements efficiently, consistently and cost-effectively.[8]

In formative evaluation the goals, objectives and criteria for judging success for a program are defined upfront.[9] Formative evaluation then evaluates whether the program will use, or does use, the right mix and volumes of human resources, materials and activities to carry out the program.[10]

**In a prospective formative evaluation**, the inputs deemed necessary to achievement of goals are specified. Inputs can include clients, staff, governance, volunteers, competence levels, money, practice protocols, service sites, operating supplies and levels of support from other programs and stakeholders – in short, any raw material deemed necessary to operate in concert with other raw materials to produce desirable outcomes. Prospective input evaluation should include the development of monitoring strategies so inputs and their effects can be measured constantly or recurrently.

A prospective formative evaluation also determines how the inputs/ingredients of the program should interact with each other as activities to produce outcomes.

**Figure 3: The Components of Comprehensive Evaluation**



A prospective formative evaluation should also include the development of a program logic model that can be used in creating and evaluating the program. A program logic model provides a framework for an evaluation – a flow chart that shows the program's components, the relationships between components and the sequencing of events. A logic model shows how the program's theory will be turned into practice and can be used in any type and size of program. Logic model development is described later in this module (see Appendix B).

**In a retrospective formative evaluation**, the evaluation determines if the inputs and activities of the program are the right ones to produce the desired outcomes. For instance, a retrospective input evaluation might determine that:

- some inputs/activities were missing altogether and should be added;
- some inputs/activities were valid but were not provided in sufficient quality or quantity;
- some inputs/activities were over-supplied, beyond the supply necessary to produce desired outcomes; and
- some inputs/activities were not necessary at all.

A retrospective formative evaluation also examines a program to understand how a program really works and how it produces its results. In other words, it evaluates how the inputs/ingredients of the program interact with each other as activities to produce outcomes. These evaluations are appropriate when programs have operated for some time, when there is evidence of inefficiencies in delivering program services or when staff, clients or other stakeholders express concerns about the program. Retrospective formative evaluations also help to accurately portray to external parties how a program operates (a help in replicating the program elsewhere).[7]

If monitoring strategies and tools were put in place prior to the retrospective evaluation, data from ongoing monitoring can provide information for use in the retrospective evaluation.

In a retrospective evaluation, a program logic model should be developed if such a model was not created earlier in the life of the program. If a logic model already exists it should be reviewed to determine if it is still considered valid and complete.

### An Example of a Prospective Formative Evaluation

Several agencies in Macklem Falls decide to collaborate in developing a navigation resource program for people living with diabetes.

After conducting a needs assessment, the program's organizers examine similar navigation programs in existence in several other communities. Since the needs assessment shows a higher incidence of more severe diabetes in Macklem Falls than in these other communities, the organizers of the Macklem Falls program determine that it should be staffed 20% higher than the programs in other communities. As well, Macklem Falls is a highly multicultural community and includes an extensive francophone population. The program's organizers therefore determine that:

- Staff of the program should be able to provide service in both English and French.
- A volunteer component should be part of the program, allowing for the provision of translation and cultural sensitivity services in Urdu, Vietnamese, Polish and Spanish (reflecting major ethnocultural groups in Macklem Falls).

In retrospective formative evaluations a discrepancy between the way the program was supposed to operate (as specified by the original program design) and how it actually operates (as shown by the formative evaluation) can lead to the question, *"Is there evidence that the current way of providing service is likely to reduce the expected positive client outcomes that were identified when the program was designed?*" The discrepancy also leads to "why" questions (which then lead to conclusions about whether to change program activities):

- Is the program delivered in a different way because the original way was inadequate?
- Is it delivered in a different way because staff were not properly trained in how to provide it in accordance with the original design?
- Is it delivered differently because the inputs (staff number and types, or their professional knowledge base for instance), are insufficient to allow it to be delivered in the planned way? If so, is this because of an original underestimate or undersupply of the resources needed – or is it because demand for service (and therefore number of clients served) has escalated sharply, making it necessary to provide less service to each client?
- Is it delivered in a different way because the needs of clients differ from the needs that were defined when the program was designed? If so, does this mean that the program is accepting the wrong clients – or that the needs of a different client group have legitimately superseded the priority client group defined in the program design?
- Is it delivered differently because of new insights into how to deliver the program – insights that were not available when the program was designed?
- Is it delivered in a different way because of a realization by managers that the current way of delivering it would work just as well as the original way – but at less cost?
- Is it delivered differently because there is a disincentive built into the program's reward system (e.g., something that discourages staff from providing service according to the original design)? If so, what is the disincentive?

Formative evaluations can include tracking the quantity and descriptors of people who are reached by a program, tracking the quantity and types of services provided, descriptions of how services are provided, descriptions of what actually occurs while providing services and descriptions of the quality of services provided.[11]

Formative evaluations should also examine structures that formalize, or should formalize, program activities. For example, collaboration with key external stakeholders at key client transition points is a **process** (a collection of activities) while an inter-program committee set up to maintain and improve this process is a **structure** (a formalized way of integrating and

### An Example of a Retrospective Formative Evaluation

Pine Point Memorial Hospital conducts an evaluation of activities involved in its cardiac catheterization program (a procedure used to diagnose heart disease) to see if the program works as planned.

The evaluation examines the referral system of local hospitals to track the activities involved in transferring patients for the procedure, including the length of time between referral and receipt of the procedure. It also examines who receives the procedure and whether wait times are longer for some patient groups than for others, why such discrepancies exist and whether the discrepancies are justified. As well, it looks at activities by which the results of cardiac catheterization are provided to referring clinicians and organizations.

focusing these activities). Module 4 (*Integration: A Range of Possibilities*) in the Health Planner's Toolkit provides a more extensive discussion of the difference between processes and structures.

Several other kinds of formative evaluation are commonly used, including:[11]

- **Needs assessment** to determine who needs the program, how great the need is, and what might work to meet the need – since knowledge of the nature and volume of need is an ingredient or input into a program. Measures may include service utilization, availability and accessibility of services and stakeholders' perceptions of their needs.[10] Module 2 (*Assessing Need*) in the Health Planner's Toolkit provides extensive information on needs assessment.

- **Evaluability assessment** to determine whether the evaluation is feasible and how stakeholders can help shape its usefulness. Such assessments should also be carried out as preliminaries to formative and summative evaluations. An evaluability assessment helps establish:
  - whether, and how, the program's inputs can be evaluated; and
  - whether, and how, evaluation questions can be asked and answered in ways that produce accurate results and allow and encourage decision-makers to use the results. If the questions and answers seem like a foreign language to decision-makers, they will not likely use the answers to shape their decisions.

When first introduced in the 1970's, evaluability assessment was an early step in summative evaluations. It later proved useful to conduct evaluability assessments as early in the life of a program as possible (preferably at the program design stage), since discussion of how to evaluate a program's elements helps to clarify elements of program design. It is now often a component of formative evaluation. Modern programs tend to be theory-based, and evaluability assessment helps clarify and diagram (via logic models) the program's theory of change.

A good introduction to evaluability assessment, including recommended assessment steps, is found in Trevisan and Huang's article *Evaluability Assessment: a Primer.*[12]

## 1.6 Summative Evaluation

Summative evaluations examine the changes that should or did occur as a result of the program. In short, they deal with outcomes. There are two types of summative evaluations:

1. A **prospective summative evaluation** determines what the outcomes of a program should be.

2. A **retrospective summative evaluation** examines a program that is already underway, to determine what outcomes (intended and unintended as well as positive and negative) it has produced and whether the program was the likely cause of the outcomes. If a prospective summative evaluation was done, a retrospective summative evaluation can compare actual outcomes with intended outcomes (as determined in the prospective evaluation) to determine the degree to which intended outcomes were achieved.

Retrospective summative evaluation includes several variations:[12]

- **Impact evaluation** compares program outcomes with an estimate of what would have happened in the absence of the program. This form of evaluation is often used when external factors are known to influence the program's outcomes, so the program's contribution to achievement of its objectives can be isolated.[13]

- **Cost-effectiveness analysis and cost-benefit analysis** address questions of efficiency by standardizing outcomes in terms of their dollar costs and values.

- **Meta-analysis** integrates the outcome estimates from multiple studies to arrive at an overall or summary judgment on an evaluation question.

## 1.7 The Difference between Outputs and Outcomes

The terms outputs and outcomes are often confused with each other. Both have their place in evaluations, but they are different.

- An **output** is a measurable result of activities within a program, reflecting the immediate result of the activities but not directly reflecting the effect on clients of the program. For instance, the activities of an in-home support program might produce 5,000 hours of service provided to 150 clients in the course of a year. The activities might also produce 10,000 promotional pamphlets for the program, 50 evening educational sessions for the families of clients and 400 hours of ongoing training for program staff in the year.

  Outputs are valuable because they represent the results of activities and act as vehicles through which positive outcomes for clients are produced. Outputs can be examined as part of evaluations but they are not the same as measuring outcomes such as the difference the in-home support program has made in the lives of its clients.

- An **outcome** is a measurable positive or negative change to clients of a program or to other stakeholders. For instance, positive client-focused outcomes of a residential long-term care program might be increased life span for residents and higher quality of life than they would experience without the care. However, not all outcomes are positive. Negative outcomes (outcomes that are a detriment to clients or to other stakeholders) are worth identifying and examining, both to determine whether positive outcomes outweigh negative ones and to find ways to reduce negative outcomes.

  Some negative outcomes are unintended. For instance, amputation of the leg of a client with bone cancer has the negative unintended effect of making walking more difficult for the client. However, this negative outcome is outweighed by saving the patient's life. Other negative outcomes may be unexpected. For instance, a weight reduction counseling program for obese clients may produce the positive outcomes of major sustained weight reduction and increased cardiac health – but evaluation may find that many clients experience depression after program completion. Identifying this negative outcome can set the stage for developing ways to reduce or address depression in post-counseling clients.

  Similarly, some positive outcomes may also be unintended or unexpected.

### An Example of a Retrospective Outcome Evaluation

A network of organizations was created to provide coordinated social services and health care services, to improve health care savings and to maintain more seniors in their homes. It is meant to provide a "one-stop shop" where seniors receive information, referrals, case management, care coordination and outcome monitoring from a single source.

An evaluation is conducted to compare program outcomes with the original program objectives, which stated that each client should have access to the programs and services appropriate to her continuum of needs. The evaluation identifies any duplication of services and administrative costs, both before program implementation and over time, to establish if benefits outweigh costs. The level of health care utilization is also measured before and after program implementation, with comparisons of inpatient acute care admission rates, inpatient length of stay, total costs of ambulatory care and whether additional system capacity is required, to ensure equitable access to the continuum of services.

Sometimes evaluators want to conduct an outcome evaluation but they have insufficient project resources to properly measure outcomes (since they are generally more complex to measure than outputs, particularly if the intention is to measure long-term outcomes). The evaluators may opt instead for an evaluation that looks thoroughly at outputs rather than conducting an inadequate outcome evaluation. Today's usual evaluation practice would concentrate on activities and short-term outcomes because they are easier to measure, require fewer resources and are valuable if they support or disprove the program theory that is intended to lead to longer-term outcomes.

## 1.8 Comparison as a Concept in Evaluation

It is possible to conduct a purely descriptive evaluation, describing what happened or should happen, without interpretation or comparison. However, most evaluations are based on comparisons. In the minds of many evaluators, comparisons are necessary for rendering evaluative judgments about the merits of a program. These comparisons may be made through the use of experimental and quasi-experimental designs or through other methods such as carefully designed case studies or rigorous qualitative and mixed-method designs. For instance:

- Prospective evaluations often compare the need for inputs, activities and outcomes with the inputs, activities and outcomes used in other programs. This is often done in search of a model for the program being prospectively evaluated. This comparison answers the question, "*Can we learn from other programs?*" The usual challenge today is implementing a theory-based or evidence-based program in a real-world setting, leading to the core evaluation questions, "*How can we make this program work well here in our program setting?*" and "*What were the lessons learned about the factors that make this program model successful or not in the real world?*"
- Retrospective evaluations often compare actual inputs, activities and outcomes against the desired inputs, activities and outcomes formulated earlier in the life of the program. This comparison answers the question, "*Did we accomplish what we planned to accomplish?*"

Several other comparisons can take place in evaluations:

- A retrospective summative evaluation can compare post-program client status with the status of a matched group of individuals who received no program service or who received a different service. This comparison answers the question, "*Would our clients have been just as well off if we had done nothing for them, or if they had received a lower cost alternative?*"
- An evaluation can compare a program with one or more programs with similar outcomes for similar clients (i.e., programs that are equally effective) to find out which program produces its outcomes most efficiently in terms of inputs and activities. This comparison answers the question, "*Can we get acceptable outcomes at less cost?*"

> "The original mission of program evaluation in the human services and education fields was to assist in improving the quality of social programs. However, for several reasons, program evaluation has come to focus (both implicitly and explicitly) much more on proving whether a program or initiative works, rather than on improving programs. In our opinion, this has created an imbalance in human service evaluation work – with a heavy emphasis on proving that programs work through the use of quantitative, impact designs, and not enough attention to more naturalistic, qualitative designs aimed at improving programs."
>
> – W.K. Kellogg Foundation Evaluation Handbook[11]

Section 2

# Planning the Evaluation

## 2.1 Who Conducts an Evaluation?

In terms of who-does-what, two broad kinds of evaluation can be conducted.

1. **Internal evaluation** (sometimes called self evaluation), in which people within a program sponsor, conduct and control the evaluation. Internal evaluation can more fully engage the insights of program personnel but runs the risk of overly subjective evaluation results.

2. **External evaluation**, in which someone from beyond the program acts as the sponsor and evaluator and controls the evaluation. External evaluation has the advantage of objectivity if done well, but it may lack buy-in from program stakeholders and may not be fully sensitive to their unique insights.

The two kinds of evaluation are not entirely separate. An internal evaluation may use external resources to help conduct the evaluation without surrendering control to the external resource – or an external evaluation may engage program personnel heavily in design of the evaluation without ceding control of the evaluation to program personnel.

In the past few years, variants of internal evaluation have emerged, known as collaborative, participatory, and empowerment evaluation. Yet as interest in internal evaluation has grown, criticisms have grown as well – most notably that "*self-evaluation is subject to the major bias of overrating oneself and one's own work*"[15] – countered by the statement: "*It may seem counter-intuitive, but we have found that most people are more self-critical of their efforts than traditional external evaluators, because it is one of the few opportunities they have to make things better (to improve their programs and address systemic organizational problems). In addition, empowerment evaluators are aware of bias and attempt to help people make their biases explicit.*"[16]

> "Everybody seems to hate external evaluation while nobody trusts internal evaluation."
>
> – David Nevo, cited in the Newsletter of the Standing International Conference of Central and General Inspectorates of Education (SICI), July, 2000[14]

> "The dilemma of whether to use external or internal evaluation is as false as that between qualitative and quantitative methods. The solution is always to use the best of both, not just one or the other."
>
> – M. Scriven, quoted in Foundations of Empowerment Evaluation[17]

A mix of internal and external evaluation sometimes brings the strengths of both to the evaluation process: "*Empowerment evaluation and external evaluation are not mutually exclusive… a second set of (external) eyes often helps the group avoid blind spots and provides another vantage point outside the internal vision of the program. Complementing an external evaluation's contributions, empowerment evaluation provides an extraordinarily rich source of information for external assessments. Empowerment evaluation and external evaluation thus can be mutually reinforcing efforts.*"[18]

Many experienced evaluators have found that internal evaluation is the best way to conduct formative evaluation and monitoring because bias is much less an issue and organizational learning is paramount. Internal evaluation provides the infrastructure needed by external evaluators for summative evaluation once the program has achieved maturity.

"In recent years, there has been growing debate between two broad approaches to program evaluation. In the more traditional model, an external evaluator is employed as an objective observer who collects and interprets quantitative and qualitative findings, and presents the information to management. A "scientific" paradigm is used which focuses on the quality of the data collected and an evaluation is considered valid to the extent that it meets specific standards of methodological rigor.

More recently… a participatory evaluation model has been used. The focus of this approach is to engage program staff, clients and other stakeholders in the evaluation process so that the information collected is used to improve the program. Because they rely on program staff both to formulate evaluation questions and collect data, these investigations may be less objective by the standards of the scientific paradigm. They are valued, however, because they improve the analytic capacity of program participants, and also increase the likelihood that evaluation results will be used to refine and improve programs."

– Allison H. Fine, Colette E. Thayer and Anne Coghlan, Program Evaluation Practice in the Nonprofit Sector[19]

## 2.2 The Steps in an Evaluation

This module divides the steps in an evaluation into two categories:

1. steps for preparing an evaluation; and
2. steps for conducting an evaluation.

**Table 2: Steps in Preparing and Conducting an Evaluation**

| **The eleven steps for *preparing* an evaluation:** |
|---|
| 1. identify and engage stakeholders; |
| 2. set the purpose of the evaluation; |
| 3. embed the program's objectives within a program logic model; |
| 4. conduct an evaluability assessment; |
| 5. address ethical issues; |
| 6. develop the evaluation project's terms of reference; |
| 7. develop the evaluation team; |
| 8. develop a project communications plan; |
| 9. confirm the evaluation design; |
| 10. design evaluation questions; and |
| 11. establish measurable indicators. |

| **The eleven steps for *conducting* an evaluation:** |
|---|
| 1. identify population and sampling; |
| 2. develop data collection tools and methods of administration; |
| 3. train personnel who will administer the tools; |
| 4. pilot test the tools and methods of administration; |
| 5. administer the tools and monitor the administration; |
| 6. prepare the data for analysis; |
| 7. analyze the results; |
| 8. interpret the results; |
| 9. develop recommendations for action; |
| 10. communicate the findings; and |
| 11. evaluate the evaluation. |

These steps are described later in this module.

Section 3

# Preparing the Evaluation

> "When evaluations are not well prepared, there is a danger that they can be carried out inefficiently. It is very easy to ignore important questions (is the programme at all evaluable? what is and what is not to be evaluated? for what purpose? how? by whom? for when? with what resources?) before evaluations are launched. These questions may seem obvious after the evaluation has taken place, but they need to be properly addressed beforehand."
>
> – European Commission, Evaluating EU Expenditure Programmes[20]

Preparing for an evaluation means setting up the preconditions for carrying out the work in ways that yield practical information for informed decision-making. The approach presented in this module was adapted from the evaluation literature, particularly from Porteus, Sheldrick and Stewart (1997) which presents a step-by-step guide to evaluating programs.[21] Note that the steps outlined below are arranged in sequence. However, the unique characteristics of a specific evaluation may require a different sequence of steps than the sequence this module presents and some steps may need to be concurrent rather than sequential.

The steps for preparing an evaluation are described in the next sections of this module.

## 3.1 Identify and Engage Stakeholders

### Identify Stakeholders

To support the development of a program evaluation, begin by identifying:

- **the people who will be affected by the evaluation's process or by its results.** These might include clients as well as program staff (including front-line, management and support staff);
- **the people who are the evaluation users.** These might include the program's managers, funders, board members of the agency hosting the program and community partner agencies; and
- **other people who can contribute to the success of the evaluation.** For example, skilled evaluators in the community may lend their expertise to the project or there may be beneficiaries of previous evaluation projects in the program, agency or sector who can help the current evaluation to understand the broad context within which it will be conducted.

There will likely be overlap among these three groups of stakeholders.

> "Because evaluation takes place within a political and organizational context, it requires group skills, management ability, political dexterity, sensitivity to multiple stakeholders and other skills that social research in general does not rely on as much."
>
> – Introduction to Evaluation, Web Center for Social Research Methods[22]

### Engage Stakeholders

Once stakeholders have been identified it is crucial to engage them in the evaluation. This engagement has four dimensions:

1. Engaging them in establishing the evaluation's purpose.
2. Engaging them so they can help shape the broad evaluation questions. This will help clarify the purposes of the evaluation, build commitment for it and fine-tune the questions the evaluation will address.[23]
3. Engaging them so they can ask anxiety/reassurance questions about the evaluation and so they can receive early frank answers to the questions. These questions reflect stakeholders' fears and worries about the evaluation. It is not always possible to allay all stakeholder fears, but much anxiety can be relieved by giving stakeholders a chance to describe their fears, phrased as answerable questions.

4. Engaging them in helping to identify how they will remain involved in the evaluation process and in the analysis and implementation of the results. Not all stakeholders will be involved in the evaluation to the same degree. Board members and managers, for instance, have important roles in authorizing and ensuring the implementation of recommendations arising from the evaluation – roles that might not be taken on by other stakeholders. It is also important to ensure that stakeholders do not make changes to evaluation methodologies and processes if those changes would result in unethical or sub-standard evaluation.

Different stakeholder groups will have different evaluation questions and anxiety/reassurance questions. It is never a good idea to assume what the questions of each stakeholder group will be – the only way is to ask them to generate questions. Stakeholders may pose questions such as those shown in Table 3.

The program manager must be involved and identified as either a key client or proponent for program evaluation in order to promote the involvement and cooperation of program staff, the relevance of the exercise and the use of findings for making program changes.[5] The program manager should communicate his or her commitment to program staff and clearly state the purposes it is expected to serve.

## 3.2 Set the Purpose of the Evaluation

Setting the purpose of the evaluation will help decide whether the evaluation will be a formative evaluation or a summative evaluation.

Examples of evaluation purposes are:

- to identify ways to improve the program;
- to determine if program benefits outweigh the cost of operating the program;
- to measure whether the program made a difference in the lives of participants/clients; and
- to help a funding body or administrator to understand the program and its results.

> "You've got to be careful if you don't know where you're going, 'cause you might not get there."
>
> – Yogi Berra, 1998

Even at this early stage in an evaluation project it is useful to review factors that affect the purpose of the evaluation and affect other steps in preparing the evaluation. These factors are described in Appendix C.

## 3.3 Embed the Program's Objectives within a Program Logic Model

In preparing to evaluate a program it is necessary to develop a program logic model to understand how the program is meant to be implemented (described below and in Appendix B). This step need not be expensive and time consuming. Its goal is to inform the evaluator about the program, not to draw conclusions about the nature and amount of its effects. This means identifying the activities that comprise the program components. A logic model communicates the underlying theory or set of assumptions or hypotheses about why the program will work or about why the program is a good solution to an identified problem.[24] It should be developed as part of a prospective formative evaluation, when a program is being planned. However, if no logic model was developed at the program's inception, a logic model should be developed as an early step in retrospective formative evaluation or in summative evaluation.

If a logic model was developed earlier in the life of the program it should be reviewed early in the evaluation to determine if it is still accurate and relevant. If no logic model was developed previously it should be developed at this point in the evaluation.

If program objectives do not exist, the evaluator must work with program staff and decision-makers to define them and embed them in the program logic model. Program objectives summarize the program's ultimate direction or desired achievement, and are usually expressed as short-term, intermediate-term or long-term objectives.[6] Some programs will have a single objective.

**Table 3: Examples of Stakeholder Questions**

| Stakeholder Perspective | Sample evaluation questions: What might they want to know? | Sample anxiety/reassurance questions |
|---|---|---|
| **Program Clients** | • Does this program provide us with high quality service?<br>• Are some clients provided with better services than other clients? If so, why? | • Is this evaluation being conducted because the program is doing a bad job, and are clients at risk?<br>• Will its results jeopardize my chances of receiving service to meet my needs?<br>• Will the evaluation results be available to clients and potential clients or will shortcomings be "swept under the rug"? |
| **Program Staff** | • Does this program provide our clients with high quality service?<br>• Are some clients provided with better services than other clients? If so, why?<br>• Should staff make any changes in how they perform their work, as individuals and as a team, to improve program processes and outcomes? | • Is this evaluation being conducted because there is a suspicion that staff is doing a bad job?<br>• Could staff be punished or blamed for any program shortcomings?<br>• Is this evaluation an excuse for expecting staff to do more, without resources to carry out the work?<br>• Will staff have full opportunity to see and comment on the results?<br>• Will the evaluation identify program excellence or will it only identify shortcomings? |
| **Program Managers** | • Does this program provide our clients with high quality service?<br>• Are there ways managers can improve or change their activities, to improve program processes and outcomes? | • Is this evaluation being conducted because there is a suspicion that managers are doing a bad job?<br>• Could managers be punished or blamed for program shortcomings? |
| **Board Members** | • Does this program provide our clients with high quality service?<br>• Does the program operate within broad parameters established by the board?<br>• How well is the board doing in terms of its broad oversight of the program? | • Could board members be blamed for program shortcomings? |
| **Funding Bodies** | • Does this program provide its clients with high quality service?<br>• Is the program cost-effective?<br>• Should we make changes in how we fund this program or in the level of funding to the program?<br>• Does the program meet requirements we established as conditions for the funding? | • Will this evaluation become just an excuse to ask for more money?<br>• Will it become just an excuse to ask for a relaxation of the conditions under which we provide funding? |
| **Partner Agencies/ Programs in the Community** | • Does the program provide its clients with high quality services?<br>• Should we continue to make referrals to, and receive referrals from, this program?<br>• Do we refer appropriate clients to the program?<br>• Can we help the program to deliver its services better? | • Will the evaluation results be available to partner agencies/programs or will shortcomings be "swept under the rug"? |

More complex programs may have several objectives. In complex programs it may be hard to specify objectives precisely. In other instances program administrators may have avoided specifying objectives for fear of setting performance standards that the program cannot meet[25] or because the program was considered so tentative or preliminary that objectives were not specified. In well defined programs, objectives are clearly stated in terms of a sequence of events or a *hierarchy of objectives*.

Health programs or services are designed to change or maintain something such as the health status, knowledge, beliefs, attitudes or behaviours of individuals, organizations, communities or other social groups.[6] An objective should tell how much of what should happen, to whom, by when. They provide a structure for designing evaluation questions. Program objectives must:

- identify the **source** of the change, i.e., the program and its components;
- define **who** will change after receiving the program;
- state **what** the program is going to change;
- identify by **how much**; and
- indicate **when** the change is expected.[6]

Objectives should include a direction (*increase*, *decrease* or *expand* for example) and be specific, measurable, realistic and based on a practical rationale drawn from sources such as a literature review, program documentation, experience and epidemiological data.[25]

There is no single right way to develop a logic model. No two models will look the same and the format will depend on the needs of planners, evaluators and other stakeholders. However, common steps to facilitate logic model development are found in Appendix B.

## 3.4 Conduct an Evaluability Assessment

An evaluability assessment (described in greater detail in Section 1.5 earlier in this module) helps determine if it is worth proceeding with an evaluation. At a minimum such an assessment should look at:

- **the program's circumstances**, including analysis of what infrastructure, data collection mechanisms and data bases are in place to support evaluation; and
- **the organizational climate**, including examination of the commitment and buy-in for evaluation, whether there are resources and capacity for evaluation, and what barriers to evaluation might exist.

The decision to proceed with an evaluation can be made even if circumstances are not ideal. What the evaluability assessment adds is an understanding of the challenges the evaluation will face. If the challenges are too daunting, the decision can be made to forego evaluation altogether or to postpone it in favour of strengthening the conditions that will make evaluation possible at a later date.

Occasionally an evaluability assessment will reveal such negative features in the program's circumstances or the organizational climate that an immediate evaluation is required to prevent harm to clients or staff, despite challenges the evaluation will face.

### An Example of the Results of an Evaluability Assessment

The Board of an in-home support program for post-stroke clients decided in June 2005 that it would conduct a program outcome evaluation starting in June 2007. However, in April 2007 it conducts an evaluability assessment that reveals barriers to evaluation:

- Data systems necessary to conduct outcome monitoring were not put in place.
- Most Board members are new and are nervous about proceeding with an evaluation until they have gained greater basic understanding of the program.
- A new Program Manager was appointed recently, and she is struggling to understand her role and to fully take on her operational responsibilities.
- A series of serious staff conflicts took place immediately before the appointment of the new Manager. She is working hard to resolve the clashes, but her staff is still under stress. They perceive an evaluation at this time as an attempt by management to put them in their places, and they will likely not cooperate with the evaluation.

**Based on the evaluability assessment the Board postpones the evaluation for six months.**

During that period it will:

- develop a greater understanding of the program;
- ensure that a data system for outcome monitoring is put in place;
- support the Program Manager as she fully takes on her operational responsibilities;
- support the efforts of the Manager and her staff to restore trust in the workforce; and
- encourage her to provide information to staff that will help them understand that both staff and clients will benefit from an evaluation in six months time.

## 3.5 Address Ethical Issues

Ethical issues, including confidentiality issues, must be addressed. Section 5.4 of this module discusses ethical issues in evaluation. Early discussion of ethics is important because ethics should drive all subsequent components of the evaluation. At the very least, evaluation projects should make an ethical commitment to cause no harm to participants and to avoid negative impact on beneficial services they receive.

Organizations that carry out research typically have a research and ethics committee to approve evaluation projects. If a steering committee has been created to oversee the evaluation, this committee might also act as the evaluation project's research and ethics committee. Because each organization has its own requirements and procedures for ethical reviews, it is prudent to check with the organization to understand its procedures.

By distributing a draft set of project ethics to the steering committee (based perhaps on the Canadian Evaluation Society's *Guidelines for Ethical Conduct* shown in Section 5.4 of this module), the evaluator can kick-start the discussion of ethics. If warranted, steering committee members and other stakeholders can then add or modify ethical components to fit the specifics of the current evaluation.

From time to time during the evaluation a discussion of whether the ethical guidelines are being followed is desirable, in part to maintain stakeholder commitment to the project by reassuring them that the project takes ethics seriously.

Depending on the evaluation design, consent forms may be required for participants to review and sign. For example, if the evaluation will report on personal information about clients participating in the evaluation, the consent of these clients is required. Clients and other participants must understand what their role will be in the evaluation and how information associated with them will be reported. The evaluator should clearly convey the terms of confidentiality regarding access to evaluation results and participants should have the option to participate or not. Appendix D provides a sample consent form that can be revised to reflect the nature of the evaluation. Participants review and sign such consent forms prior to participation.[26] Many evaluations are considered administrative and are covered by blanket consent obtained at intake, so it is wise to check with the organization to verify its policy on consent.

In some cases the act of agreeing to participate in a self-administered survey or a telephone interview is sufficient. Confidentiality of information needs to be guaranteed. This means that a participant could not be identified from any material resulting from the evaluation. This issue is usually explained in a cover letter for a mail-in survey or in an interviewer's script for a telephone interview. Again, it makes sense to check with the organization to find out its policy on notifying participants.

## 3.6 Develop the Evaluation Project's Terms of Reference

The project's terms of reference guide subsequent steps. The terms of reference may be a broad document that is subject to revision throughout the project. They should include:

- a statement of what is to be evaluated (the name of the program for instance);
- a statement of what kind of evaluation is envisaged (a formative evaluation, a summative evaluation or some combination of the two);
- a statement of the intended benefits of the evaluation (outcome improvement, for instance, or greater program efficiency);
- a statement of the authority under which the evaluation will be carried out (the sponsor of the evaluation for example);
- a statement of project timeline requirements or limitations (for instance, *"A final evaluation report must be provided to the evaluation's sponsor by May 1 2008"*);
- a statement of resource requirements or limitations for the project (for instance, *"The sum of $35,200 is available for completion of the project"*); and
- a statement of major project steps (including, for instance, the steps for preparing and carrying out an evaluation, described in this module).

## 3.7 Develop the Evaluation Team

Even in a modest evaluation, an evaluation team is needed. At a minimum an evaluation requires:

- one person responsible for carrying out evaluation activities ("the evaluator"); and
- one person responsible for managing relations with evaluation staff, solving organizational problems and enabling buy-in ("the client"). To reduce the potential for bias, this person usually does not direct or oversee the evaluator or have final say in evaluation matters.

In more complex evaluations the evaluator might be a team, with a team leader and other workers, and the client might be a committee bringing many insights to the project.

Whether the evaluation is simple or complex, the evaluation team must be designed and, if necessary, members must be recruited and trained.

Team development may have occurred earlier in the evaluation's planning process. If not, then developing the team – or at the very least, designing it – makes sense at this point in the project, before it further engages its stakeholders.

## 3.8 Develop a Project Communications Plan

A project communications plan serves three purposes:

1. It guides communications to reduce or eliminate anxiety, resistance and hostility.
2. It guides communication to maintain and increase support for the evaluation as well as support for the eventual uptake of the evaluation's findings.
3. It serves as a reminder to the evaluation's leaders that communication is essential.

It makes sense to develop the communications plan after, rather than before, the identification and engagement of stakeholders, because the kinds and numbers of stakeholders will influence the communications plan.

Like the initial project evaluation plan, the communications plan should be reviewed and revised during the course of the evaluation.

Module 5 (*Community Engagement and Communication*) in the Health Planner's Toolkit provides advice that will help in developing an evaluation communications plan.

## 3.9 Confirm the Evaluation Design

The evaluability assessment, the statement of the evaluation's purpose, the preliminary evaluation plan and stakeholder input have probably given the evaluation's sponsor a sense of what is doable and what is not. It is helpful at this point for the evaluation's sponsor and the evaluator to review and refine the evaluation's design. Most importantly the review should determine the degree to which the evaluation will be **descriptive** and/or **analytical**, as guidance in developing the evaluation questions.

> "I keep six honest serving-men
> (They taught me all I knew);
> Their names are What and Why and When
> And How and Where and Who.
> I send them over land and sea,
> I send them east and west;
> But after they have worked for me,
> I give them all a rest."
>
> – Rudyard Kipling, Just So Stories, 1902

**Descriptive** elements of the evaluation are meant to answer four of the questions that are the hallmark of good journalism, just as they are the hallmark of good descriptive evaluation:

**WHO WHAT WHEN WHERE**

Descriptive design primarily describes the characteristics of the population of interest or the characteristics of the program. It is relatively easy to implement, less expensive than analytical evaluations and can be used for all types of evaluations.[10]

Examples of descriptive designs include:

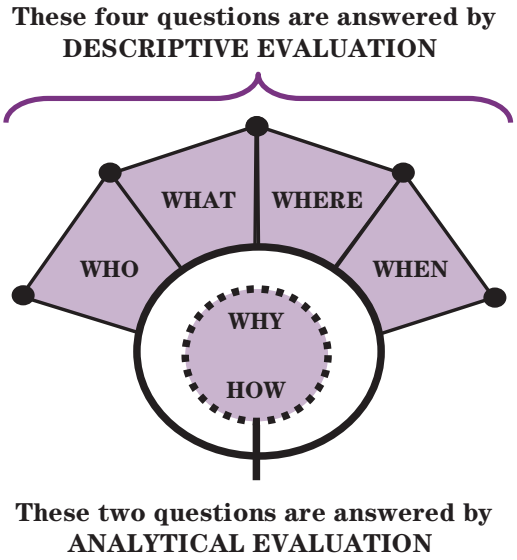
- cross-sectional surveys, in which a sample of the program population completes a questionnaire at one point in time; and
- pre-post designs that have measures taken both before a program is implemented and after it has been in place for a period of time.

Description alone does not answer two of the famous six journalistic questions:

**WHY** **HOW**

It is the **why** and **how** questions that are answered by analytical evaluation.

**Figure 4: Questions Answered by Descriptive and Analytical Evaluation**




**In general, the probability that A caused B is increased when:**

- A large data set supports a relationship between A (the potential cause) and X (the effect) – for instance, the relationship holds true for 90% of 1,000 client cases examined by evaluation, not just for 90% of 10 client cases examined in the evaluation;
- The data supporting the relationship between A and X have high levels of accuracy and reliability, based on valid measurement;
- Several data sets, rather than a single data set, support the relationship between A and X. For instance, the likelihood that a particular mental health counseling method delays relapses is stronger if:
  - clients say it does;
  - practitioners say it does;
  - other programs using this method show delayed relapses; and
  - programs that do not use this counseling method do not show significant delay in relapses; and
- Other potential causes are examined and do not show the association to the same degree. The more of these potential causes that are examined, and the larger and/or the more accurate the information set examined for each potential cause, the more confident one can be in excluding them as causes.

As well, hunting down the cause is more effective if it allows for potential causes to be examined together, to estimate their combined effect. For example, potential cause A alone, may not cause X (the effect), and potential cause B, alone, may not cause X, but causes A and B combined may be the cause of X.

**But…. increasing the probability of establishing causation through these methods takes time and money!**

As Section 5.2 of this module explains, evaluations do not prove causation. The most they can do is indicate, to a high degree of probability, what might have caused an outcome. Put another way, answering the "why" and "how" questions ends up with answers that indicate "probably why" and "probably how". Increasing the probability level for causation in an evaluation usually involves greater information as well as the time and resources necessary to gather the information.

Coming extremely close to proving cause or effect in an evaluation is difficult and expensive but it may be demanded in some situations. To do it well requires the elimination of other possible causes and it necessitates control over who receives and does not receive the program intervention. Most evaluations are descriptive and do not address the burden of proof, but others require analytical designs.[10]

An **analytical design** can involve a comparison of groups of target participants or programs to systematically identify whether or not the intervention has an effect or which program design works better by comparing groups receiving different programs. Two kinds of analytical designs draw their methods from experimental sciences:

- An **experimental design** controls the selection of participants in the study, who are randomly assigned to treatment and control groups. An example of an experimental design is the pre-test – post-test control group design in which the target group (older adults with 10 or more physician visits in the past six months for instance) are randomly allocated to the intervention group or the control group. Program effects would be estimated by calculating the average difference between the pre-test and post-test scores in the intervention group, and the average difference between the scores for the control group.

- A **quasi-experimental or observational design** does not randomize target groups to intervention and control groups. It is not always possible to randomize participants into intervention and control groups because of logistical constraints or ethical or legal issues. A quasi-experimental design might, for instance, use a comparison group whose members share the characteristics of the target group (but this is not a group to which members are randomly assigned). The comparison group would not receive the intervention – for example, clients of a program in a different district where the program is not offered. Multiple observations are collected for both groups before and after the program is launched.

It is wise to choose the evaluation design that best maximizes the validity of the evaluation within available resources. Module 3 (*Evidence-Based Planning*) in the Health Planner's Toolkit indicates that *"a measurement is valid if it measures what it was intended to measure"* and helps the reader to understand validity. To identify potential limitations of the evaluation approach, the following questions should be considered:[10]

- Did everyone in the program have equal chance of being measured?

- Were participants choosing (self selecting) to take part in the evaluation?

- Did participants drop out of the program before information needed for the evaluation was collected?

- Were standardized and valid methods of measurements used? If not, could results have been caused by how the measurements were taken?

- Were there other factors happening at the time of the evaluation that may have caused the outcome?

- Is it possible that the results were due to chance?

A resource by Campbell and Stanley (1966) titled *Experimental and Quasi-Experimental Designs for Research* can help identify and understand threats to validity.[27]

## 3.10 Design the Evaluation Questions

This step translates program objectives into answerable evaluation questions.[28]

Evaluation questions will most often focus on program implementation (via formative evaluation) and program outcomes (via summative evaluation).[6, 25]

> "Questions are the engines of intellect, the cerebral machines which convert energy to motion, and curiosity to controlled inquiry. There can be no thinking without questioning – no purposeful study of the past, nor any serious planning for the future."
>
> – David Hackett Fischer, Historians' Fallacies: Toward a Logic of Historical Thought[29]

Outcome questions usually ask whether a program achieved its objectives. Examples of outcome questions include:

- What do people do differently as a result of the program?
- Who benefits and how do they benefit?
- What do participants/clients learn, gain and accomplish?
- Are participants/clients satisfied with what they gain from the program?

While it is important to know about program outcomes (i.e., the descriptive component of evaluation), it is important to know how and why the outcomes were achieved or not achieved (i.e., the analytical component). The status of program outcomes alone offers little guidance about how to improve programs, how to identify and replicate successful program aspects in other settings or how to avoid unintended negative consequences of a program in the future. Accordingly a summative evaluation is often accompanied by a formative evaluation to help explain program outcomes.[6]

Section 5.2 of this module further explores the challenges of asking "why" and "how" questions in evaluations.

The evaluator should make a list of questions that she and stakeholders want to have addressed. Priority is usually given to the questions of the direct users of evaluation information. Otherwise the process becomes too unwieldy.

> "Several years ago, I asked a former Deputy Minister of Social Services in Alberta to address my program evaluation class... Why did he think that program evaluations tended to be so ineffective?
>
> Most importantly, he cited the failure of most evaluators to ask the right questions, in other words, to ask questions that addressed the key problems in the program being evaluated. This failure is often related to the lack of knowledge that evaluators have about the programs they are evaluating. He said that good evaluators need to spend up to 80 per cent of their time checking and re-checking with the program sponsors to ensure that they are addressing the most important issues in an effective way."
>
> – Ian Greene, Lessons Learned from Two Decades of Program Evaluation in Canada[30]

It is useful to clarify the questions that the evaluation will answer by breaking larger questions into smaller components[23] as illustrated in Figure 5.

The list of evaluation questions can become lengthy and it may be necessary to prioritize the questions by considering two dimensions:

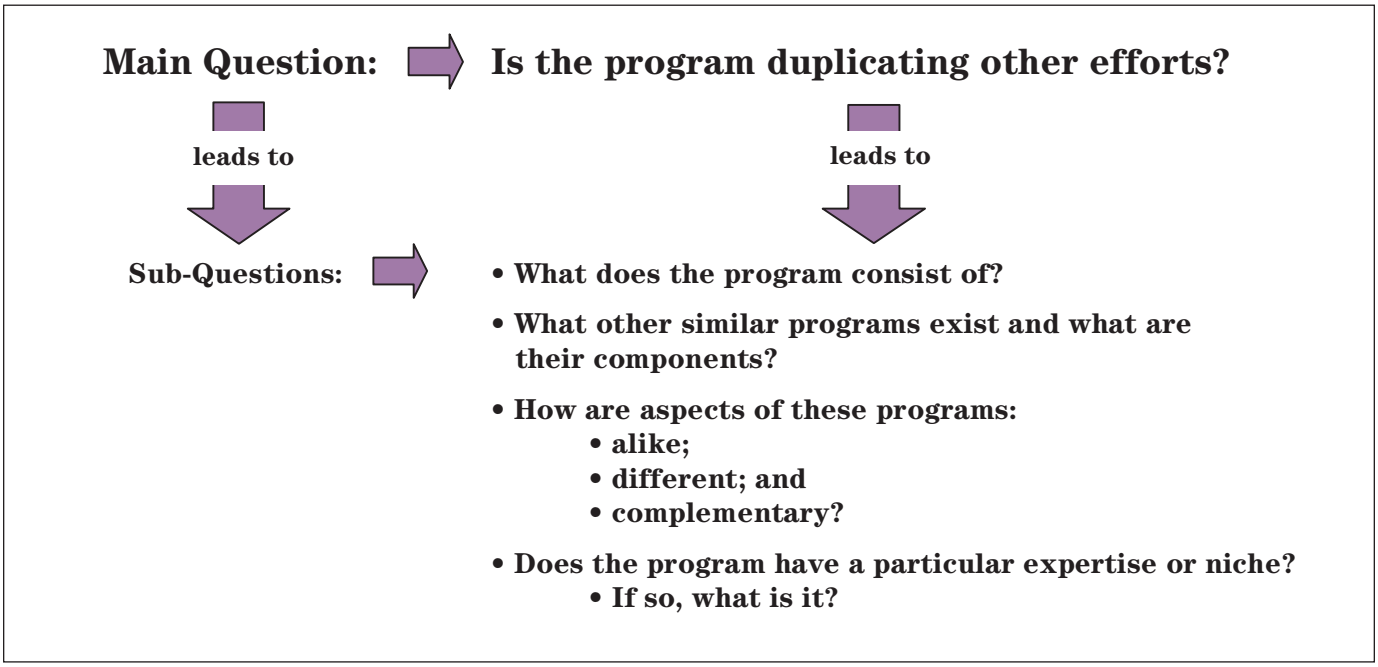
1. **The degree of importance of a question.** An attempt should be made to distinguish between what is needed and what might simply be nice to know.

2. **The feasibility of getting an answer to the question.** It may not be feasible to answer a question for either of two reasons:

   - **There is no known way to answer the question.** For instance, for many programs the question *"Will the program prevent the possibility of client relapse in future?"* is impossible to answer. When a question cannot be answered, it can sometimes be rephrased so it becomes feasible to answer it. For instance, the relapse question might be rephrased: *"What is the likelihood of relapse for the program's clients?"* This is answerable by using the indicator *"rate of re-entry to treatment for each of three age groups during the five years after discharge from the program"*. Whether a question is answerable, then, depends on whether there is at least one indicator that can be used to find the answer; and

   - **There are insufficient evaluation resources (time, money or personnel) to allow the question to be answered.** For instance, an evaluation of a five-year-old program might want to ask the question, *"Were program staff during the program's first year of operation optimistic or pessimistic about the program's chances of success?"* However, given the high staff turnover rate in the program, it may be prohibitively expensive to track down and interview staff who worked for the program during its first year of operation.

**Figure 5: Converting Large Questions into Sub-Questions**



The matrix shown in Figure 6 helps deal with various mixes of importance and feasibility regarding evaluation questions.

Cell #3 in this matrix (high importance/low feasibility) requires the most creative approach on the part of evaluators. Cell #2 (low importance/high feasibility) on the other hand poses a trap, since it may be tempting to answer unimportant questions simply because they can be easily answered.

Generally the goal of an evaluation is quality and not quantity. It is important to keep the evaluation manageable by addressing a few questions well.

Several other factors will help determine the importance and feasibility of questions:[25, 31]

- **Age of the program:** For new programs that are still unstable, questions that target the program's short-term and intermediate outcomes may provide useful and timely information to improve the management of the program. It would be premature to ask about longer-term outcomes of the program if not enough time has passed to allow these outcomes to be achieved.

- **Consensus:** It might make sense to choose questions that decision-makers, program staff and other groups all agree on. While an advantage to this approach is that a common set of expectations is generated about the kind of information to be disseminated about the program, a disadvantage is that consensus might not be possible. Another disadvantage is that groups may purposely avoid questions they see as a threat to the program's survival or to the reputations of people governing or working in the program. In short, the most popular questions are not necessarily the best ones.

  Alternatively, the evaluator may focus on questions raised by those stakeholders most committed to using the findings of the evaluation to improve the program.

- **Result scenarios:** Another strategy is to select questions whose answers will likely change the beliefs, attitudes, and behaviour of decision-makers, program managers and other stakeholders.[32] For example, if a program is found to have a beneficial impact, would that result lead to program expansion or other changes?

**Figure 6: An Importance/Feasibility Matrix**

| | Low importance | High importance |
|---|---|---|
| **Low feasibility** | **1.** **Don't bother!**  | **3.** **Get creative to increase feasibility.**  |
| **High feasibility** | **2.** **Don't do it if it takes resources from cells 3 & 4.**  | **4.** **A no-brainer – Do it!**  |

## 3.11 Establish Measurable Indicators

Indicators are the specific measures that answer evaluation questions. Establishing measurable indicators serves the important function of providing the criteria to judge the effectiveness of a program[6]. As described in Module 3 (*Evidence-Based Planning*) in the Health Planner's Toolkit, indicators are measures constructed to be comparable over time and across jurisdictions. More than one indicator may be needed to address an outcome accurately.

When using indicators it is often crucial to determine what level of achievement of the indicator is considered acceptable. For instance, in using the indicator *"percentage of clients who relapse within six months"* it may be decided that a relapse rate of 20% or less is acceptable for the program.

Indicators that are specific to a program may need to be developed or it may be possible to use indicators that have been developed and tested elsewhere.

The following questions can help determine measurable indicators:

- How will I know if an objective has been accomplished?
- What would be considered effective?
- What would be a success?
- What change is expected?[10]

Other criteria on which to base measurable indicators include consideration of the mandate of the program (for example, percentage of children immunized in a given year if a program's mandate is immunization of an entire child population). Advocated standards (for instance, standards set by professional organizations) can also be used. Another criterion might be the values and opinions expressed by recipients of a program (for example, the percentage rating 'excellent' for the quality of the service).

In summary, the identification and use of measurable indicators provide a systematic way to assess the extent to which a program has achieved its intended results. Table 4 provides examples of measurable indicators.

**Table 4: Examples of Measurable Indicators**

| Evaluation Area | Evaluation Question | Examples of Specific Measurable Indicators |
|---|---|---|
| **Formative Evaluation** | | |
| Staff Supply | Is staff supply sufficient? | • Staff-to-client ratios |
| Volunteer Supply | To determine volume of volunteer involvement | • Total volume of volunteer hours provided per year |
| Program Knowledge Base | Is the knowledge base of staff sufficient? | • Percentage of staff who meet or exceed recommended education/training levels for their positions |
| Service Utilization | What are the program's usage levels? | • Percentage of residents in the LHIN who used an emergency department in the past year |
| Accessibility of Services | How do members of the target population perceive service availability? | • Percentage of target population in the LHIN who are aware of the program in their area<br>• Percentage of the "aware" target population who know how to access the program |
| Staff Time | Has there been any decrease in the amount of time clinical staff spend on administrative duties? | • Proportion of clinical staff time spent on administrative duties before and after program intervention |
| Inquiries | Has there been any increase in volume of inquiries about the program? | • Number of phone inquiries to a crisis line in the calendar year<br>• Percentage change in number of phone inquiries from previous year |
| Resources Distributed | Do health professionals and community organizations have the opportunity to increase their knowledge about the program? | • Percentage of members of each target group who receive pamphlets about the program<br>• Percentage who read the pamphlets<br>• Percentage who say the pamphlet has substantially increased their knowledge |
| Client Satisfaction | How satisfied are clients? | • Percentage of clients who report being very satisfied or satisfied with the service received |
| **Summative Evaluation** | | |
| Changes in Behaviour | Have risk factors for cardiac disease been reduced? | • Compare proportion of respondents who reported increased physical activity |
| Morbidity/Mortality | Have hospital separations due to circulatory system diseases in 40-64 age group been reduced?<br>Has lung cancer mortality decreased by 10%?<br>Has there been a reduction in the rate of low-birth weight babies? | • Age-sex standardized hospitalization rate for circulatory system disorders for those age 40-64; compare year to year<br>• Age-standardized lung cancer mortality rates for males and females<br>• Compare annual rates of low-birth weight babies over five year period |
| Client Resilience | Has there been an increase in clients' self-confidence? | • Percentage of clients who feel their self-confidence has improved since involvement with the program (pre/post measurements) |

Section 4

# Conducting the Evaluation

Up to this point the evaluation process has engaged stakeholders, determined a purpose for the evaluation, developed a logic model, addressed ethical questions, determined if the program is evaluable, developed terms of reference and an evaluation team, confirmed the evaluation's design, determined the evaluation questions and selected measurable indicators.

Building on the 11 preparatory steps for the evaluation, it is now time to conduct the evaluation by carrying out 11 additional steps:

1. identifying population and sampling;
2. developing data collection tools and methods of administration;
3. training personnel who will administer the tools;
4. pilot testing the tools and methods of administration;
5. administering the tools and monitoring the administration;
6. preparing the data for analysis;
7. analyzing the results;
8. interpreting the results;
9. developing recommendations for action;
10. communicating the findings; and
11. evaluating the evaluation.

## 4.1 Identify Population and Sampling

In some situations all of the population of interest may be contacted for the program evaluation. For example, for a community mental health program for people with bipolar disorders that has operated for only a year, the number of clients served in the year may be less than one hundred. It may therefore be feasible to contact all clients.

However, it is not always feasible (due to time and resources) or necessary to contact all participants in a program. A sample of the population of interest may suffice. The responses from the sample will allow the evaluator to provide a reasonable estimate for the population with a level of precision that will depend on the sample size, the sampling design and the amount of variability within the population with respect to the measures of interest. If funds restrict the desired level of precision, a different sampling design or evaluation approach may be considered.

When a sample is selected, several factors should be considered:

- The sample could be representative of the target population so that the results can be generalizable. On the other hand, not all samples are representative – the design may call for purposive sampling or some other sampling approach.

- The sample needs to be large enough so that the data collected will provide reliable results.

- The sample must be accessible. For example, an ideal target population for emergency department evaluation might be all visitors to the emergency departments of all hospitals in a LHIN in the last six months. The questions to consider are:
  - can a list of these patients be obtained?
  - does it provide enough information to allow a generalizable sample to be derived?

This list would be the **sampling frame** and would contain contact information for the target clients.

It is beyond the scope of this module to provide technical detail for calculating sample size. An evaluation specialist or epidemiologist should be consulted to determine sample size. Several references for determining sample sizes are included in this module's reference list.[33, 34, 35]

There are a number of ways to select samples of the target population. The easiest in terms of selecting the sample and in analyzing the resulting data is **simple random sampling** in which everyone in the population has the same chance of being selected.[36] All statistical software packages will properly handle the analysis of data based on a simple random sample. That is not the case for other types of sampling designs.

The choice of sampling design should not just be based on ease of use. Other factors to consider are the homogeneity of the target population, questions to be answered by the evaluation and available resources and timeframe. For example, does the evaluation want to provide estimates of the level of client satisfaction for each sub-LHIN area – or just for the LHIN as a whole? To ensure a sufficient sample size in each sub-LHIN area, the evaluation may use a **stratified sampling design**. Each stratum would correspond to a sub-LHIN area. A stratified random sample is obtained by dividing the population into groups of individuals that are similar except for the stratifying variable, then selecting a simple random sample from each.

An evaluation that wants to survey clinical staff in hospitals within a LHIN area regarding a new professional development initiative may decide to select a sample of hospitals in the LHIN and then survey all staff in these hospitals. This is referred to as **cluster sampling**. A cluster sample is a simple random sample in which each sampling unit is a collection, or cluster, of elements.[36] The sampling unit here is a hospital and the staff within the hospital are the elements on which measurements are taken. An alternative sampling approach in this example could be a random sample of clinical staff across all hospitals in the LHIN area.

In **systematic sampling**, elements are selected from a sampling frame at regular intervals. A sampling interval and a random start are required. This would be a reasonable design for selecting patients to participate in a survey when client records are in paper form in cabinets. For example, if the evaluator determines (based on the sample size calculation) that every twentieth client record needs to be extracted, the first record extracted would be randomly selected from records 1 to 20, and then every 20th record thereafter.

Another sampling design is **convenience sampling** – the type of sampling encountered in a mall when an interviewer stops people and asks if they will participate in a survey. A survey of drivers wearing seatbelts, based on those who happen to stop at a certain street location at certain times of the day over a number of days is another example. Similarly, respondents of web-based surveys are self-selected and will not be representative of the general population. Convenience sampling is not recommended for providing reliable and generalizable estimates.

An evaluation project plan should state who is included in the target population (for instance, all emergency room clients during the month of July 2007 who presented with a pulmonary condition), and who is excluded (all emergency room clients presenting for any other reason). These are referred to as inclusion and exclusion criteria. Criteria on which to base eligibility may include age, gender, marital status, occupation, location of residence, literacy, health behaviours, health status and the presence or absence of one or more medical conditions.

## 4.2 Develop Data Collection Tools and Methods of Administration

Data collection tools are ways of gathering information that will answer evaluation questions and that allow answers to be expressed as:

- measurements; or
- classifications.

For instance, **measurements** can be expressed in terms such as *how often*, *how big*, *how long*, *how intensely* or *how broadly*, often with gradations similar to the degrees on a thermometer or amounts on a measuring cup. **Classifications**, on the other hand, allow answers to be placed into one or another of several mutually exclusive boxes. For example, a traditional agreement scale (*strongly agree*, *agree somewhat*, *don't know*, *disagree somewhat*, and *disagree strongly*) allows answers to be classified (and it also includes a rough degree of intensity measurement because it differentiates between "strongly" and "somewhat").

> "In Poland under communism, the performance of furniture factories was measured in the tonnes of furniture shipped. As a result, Poland now has the heaviest furniture on the planet."
>
> – attributed to the Report on Business, Globe & Mail (Toronto), circa 1996

Methods of administration are the ways the tools are actually applied. For instance, asking a series of satisfaction questions is a tool, but the questions can be administered through a questionnaire, an interview or a file review if the files contain evidence of client satisfaction or dissatisfaction.

Methods of administration may include:

- surveys/questionnaires;
- focus groups;
- face-to-face interviews;
- observation;
- case studies;
- activity logs;
- administrative records;
- patient/client charts;
- registration forms; and
- attendance sheets.

These are described, along with their advantages and disadvantages, in Appendix E.

A chicken-and-egg question often posed in evaluation design is, *"Should we develop the tools and then find the right way to administer them, or should we tailor the tools to fit our preferred method of administering them?"* The answer is not always simple, because tools influence the methods of administration and methods of administration influence tools. It generally makes sense to start with the tool and then determine how it can best be administered. If that method of administration is not feasible, an alternate method of administration will need to be chosen, leading to a revision of the tool if its original form does not fit with the method of administration.

It may be helpful to use multiple tools and methods for measurement and classification. To select the best methods, the evaluator should consider advantages and disadvantages of the type of information needed, resources available, cultural appropriateness and reliability and validity.[37]

A data collection plan can be developed by working through a methods worksheet to identify where the evaluator will get information, from whom, when the data should be collected and from how many people. An example of a worksheet is included in Appendix F, which also provides a list of questions to work through to assist in the completion of the methods worksheet. Evaluators in Canada often use an evaluation framework which is a spreadsheet with columns for evaluation questions, indicators, sources of data, data collection tools, who will collect data, when data will be collected and methods of analysis.

## Find or Develop the Tools

The first step for selecting tools is a literature search for published tools that can be used. Tools used in evaluations of similar health programs may be used or could be adapted. For instance, if a question from the Canadian Community Health Survey will provide the information needed by the evaluation, it is worth using. It will ensure that validation has taken place and allows comparisons between the results of the evaluation and the results of the Canadian Community Health Survey.

It may also be useful to contact colleagues or associates to see if similar evaluations have been undertaken. Using existing tools may increase the reliability and validity of measures because the tools may have already been tested:

- **Reliability** is about the extent to which a tool yields the same measurement or classification on repeated trials. For instance, a system for classifying material extracted from client files is not reliable if the same coder classifies the material differently during two attempts to classify the same material.

- A measure has **validity** if it measures or classifies what it purports to measure or classify.[38] A thermometer, for instance, is valid for measuring oven temperatures, but not valid for measuring quantities of baking soda.

If appropriate tools do not exist, new tools must be designed for the evaluation at hand. For instance, the education level of the target sample, as well as the estimated amount of time they will devote to responding to an instrument, must be considered when developing a tool.

It is important to consider how the data will be analyzed when the tool is developed, deciding in advance if measurements and classifications will be determined by

using software or by manual analysis. If responses will be entered into a data software template it is worth thinking carefully about the naming of questions and coding of question responses, to improve the efficiency and effectiveness of the analyst setting up the data entry mechanism and the person inputting the data.

Other items may need to be developed when designing the tools – cover letters, reminder cards, consent forms and interviewer scripts, for instance.

## Qualitative and Quantitative Information

There are two main categories of information – qualitative and quantitative. One or both can be gathered and used in an evaluation.

### Qualitative Information

As Module 3 (*Evidence-Based Planning*) in the Health Planner's Toolkit defines it, *"Qualitative information is narrative and reflects individual insights or observations. Qualitative information is usually non-numeric and is not analysed using statistical methods."*

There are five frequently used data collection processes in qualitative evaluation (more than one method can be used):

1. **unobtrusive seeing**, involving an observer who is not seen by those who are observed;

2. **participant observation**, involving an observer who does not take part in an activity but is seen by the activity's participants.

3. **interviewing**, involving a more active role for the evaluator because she poses questions to the respondent, usually on a one-on-one basis (although group interviews are possible);

4. **group-based data collection processes** such as focus groups; and

5. **content analysis**, which involves reviewing documents and transcripts to identify patterns within the material.[39]

For readers interested in fundamental assumptions underlying qualitative research and the different essential orientations or schools of thought in qualitative research, a citation is provided in the references section of this module.[40] For a "how-to" approach to conducting focus groups, several sources are cited in the references section.[36, 41, 42]

### An Example of Qualitative Data Collection

A LHIN would like to determine the interest in a mobile service for seniors for receiving check-ups. Each sub-area within the LHIN has been asked to conduct focus groups in its area. Planning consultants recruit residents aged 65 and over who live at home, to participate in focus group discussions. Focus groups of six to eight participants are held to determine seniors' interest and perceived need as well as to identify barriers to regular health care and the use of the proposed service. In addition, face-to-face interviews are conducted with senior managers who are responsible for mobile services for seniors in other LHIN areas to obtain lessons learned.

**Quantitative Information**

As Module 3 (*Evidence-Based Planning*) in the Health Planner's Toolkit defines it, *"Quantitative, or numeric information, is obtained from various databases and can be expressed using statistics."*

Forms for recording pieces of information (a client history form tracking people attending a flu clinic or the service record of a piece of hospital equipment for example) are quantitative data collection methods.

Quantitative data can also involve large administrative data sets such as hospital inpatient data and national survey data – for example, the Canadian Community Health Survey which measures factors such as health behaviours and access to services. These sources represent secondary or existing pre-collected data sets that provide historical data about a program.[43] If a secondary data set can provide the information needed to evaluate a program, it may be possible to avoid collecting new data.

### An Example of Quantitative Data Collection

The evaluation of an acquired brain injury rehabilitation program wants to understand the time-related experiences of clients of the program. The evaluator extracts the following information from client files to aid in this work:

- wait time, in days, from receipt of referral to the program to the time of admission, for a sample of 150 clients in 2006;
- time, in days, from day of admission to day of first rehabilitation procedure for clients in the sample;
- average hours per day of rehab therapy provided to clients in the sample, grouped in terms of severity of injury (mild, moderate and severe); and
- time, in days, from the time a client is notified of a discharge date to the actual date of discharge, for clients in the sample.

Once this **raw quantitative data** has been extracted the evaluator can perform a number of mathematical calculations using the data (arithmetic means and medians for example) to provide refined quantitative data.

## Using Surveys

Surveys are often administered in evaluations. They represent a quantitative data collection method that may also have qualitative components such as open-ended questions that allow respondents to answer in their own words.

Surveys (as well as forms collecting new data directly from people) are known as primary data sources.[43] Surveys can be conducted in a variety of ways, including face-to-face interviews, telephone interviews, self-administered questionnaires and Internet-based surveys (including interactive web sites and e-mail methods).

Factors determining which survey method is most appropriate include:

- the timeframe available for the evaluation;
- available funds;
- available human resources for administration and analysis;
- the type of information being collected (i.e. what is being measured); and
- the target population.

These five factors will also help to determine whether the evaluator wants to use qualitative or quantitative methods or both. In some situations, once a quantitative analysis is done a qualitative analysis is conducted to delve more deeply into issues that emerged from the quantitative analysis.

Given the extent to which surveys are used to gather evaluation data, writing good survey questions is very important:[44]

- Questions should be simple and tailored to the target group.
- Questions should be as clear as possible.
- Each question should not cover more than one issue.
- Questions should not lead the respondent to a specific answer.
- Questions should be crafted to avoid language that might offend respondents.

The layout of the questionnaire is also important in self-administered questionnaires and those administered over the Internet. The size of the print, the amount of white space on a page, the length of pages and overall questionnaire length should be considered.

Useful resources for survey methods and quantitative research are included in the references section of this module.[34, 44, 45, 46]

## 4.3 Train the Personnel Who Will Administer the Tools

Project personnel who will administer a measurement tool must receive training on the use of the tool. This can involve simulations of phone interviews or focus groups, to reduce the potential for variations of approach by interviewers that could impact on the responses.

During training, trainees may identify additional ways instruments and their supportive documents and processes can be adjusted to increase efficiency or effectiveness.

## 4.4 Pilot Test the Measurement Tools and Methods of Administration

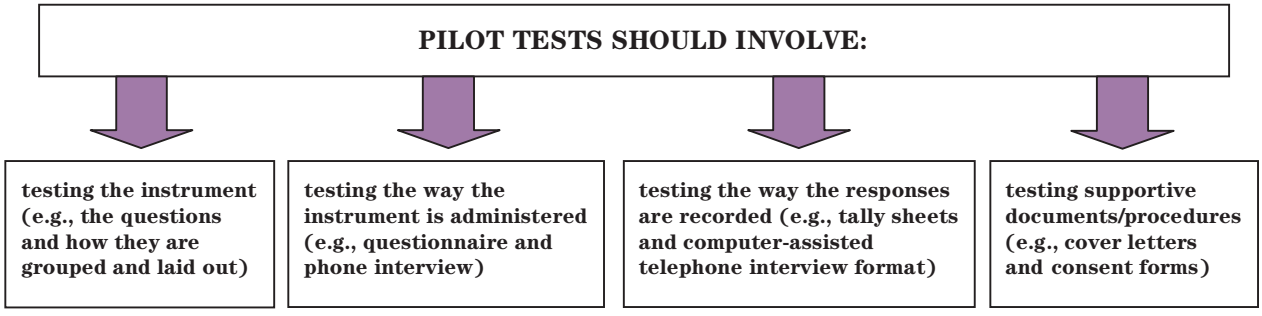
A pilot test can provide the evaluator with a sense of the reliability, validity and feasibility of tools. In some situations the validity of a measurement can be checked by comparing the results to another data set. For example, if respondents were asked to state whether they had used the emergency department within the LHIN area in the last year, a check of actual emergency department records for all respondents – or a sample of respondents – can indicate if clients' responses were valid.

As Figure 7 demonstrates, all aspects of the process involved in collecting the data should be tested.

Draft instruments should be tested with a few people with characteristics similar to those who will participate in the evaluation. Pilot testing an instrument determines how long it takes to complete, whether it is too long, whether its questions are understandable, whether questions are interpreted in a similar way by all respondents and whether the questions obtain responses for all the different response categories so that not everyone is responding the same way to each question.[45]

Following pilot testing, measurement tools and documentation are modified if necessary. The evaluation project's work plan, timelines and resources may need to be adjusted as a result of pilot testing.

**Figure 7: The Range of Pilot Test Activities**



## 4.5 Administer the Tools and Monitor the Administration

If preparatory work has been carried out well, the administration of the tools – through document review, interviews, focus groups, mail-outs, the Internet or other administration methods – should be relatively smooth. Nevertheless the project should allow for unforeseen glitches. "Glitches" must be identified before they become "crises". This suggests that close monitoring of the administration of tools is essential. For example:

- The first wave of responses should be closely examined to see if some of the items within the tool are not producing results (in other words, they may be unreliable or invalid despite earlier pre-testing of the tool and therefore need adjustment).

- The number of no-responses/refusals from respondents should be monitored to determine whether refusal rates are high enough to warrant adjustment to the conditions under which the tool is administered.

- Project staff responsible for administering the tools and collating the results should debrief with each other and with the evaluation project leader frequently to identify emerging issues that threaten the integrity or effectiveness of the data gathering or that present unforeseen opportunities to improve data gathering.

The evaluation team should remain open to adjustments to tools or their administration, even at mid-course. Stubborn adherence to a flawed process seldom acts as an antidote to the flaws.

## 4.6 Prepare the Data for Analysis

A quantitative data set must be "clean" before analysis starts. This means ensuring the data have been recorded accurately, the form and content of the data are consistent and the data are within acceptable ranges for what the measurement tool defined. For example, if data shows a participant aged 15 but the sample is meant to consist of persons aged 65 and over, an error has occurred.

> She said, 'You don't know it boy,
> but you just blew it.'
> And I said, 'Well that's my story
> and I'm stickin' to it.'
> That's my story.
> Oh, that's my story.
> Well, I ain't got a witness,
> and I can't prove it,
> but that's my story
> and I'm stickin' to it.
>
> I got that deer-in-the-headlight look."
>
> – Country singer/composer, Collin Raye,
> That's My Story

To facilitate data preparation a free software program called Epi Info is available from the Centers for Disease Control and Prevention (CDC) at http://www.cdc.gov/epiinfo.

Checking is also important for qualitative data. For instance, if a focus group was taped, transcribed notes can be compared to the audio tape to ensure they are complete and accurate.

## 4.7 Analyze the Results

Data analysis synthesizes information from all data sources to permit interpretation and to answer the evaluation questions. Different techniques are appropriate depending on the type of data produced during the evaluation. The analysis should be planned in advance of collecting and organizing the data – ideally when the evaluation plan is developed.

### Analysis of Qualitative Data

In qualitative evaluations the main goals are often to understand what has happened in the program and why, and to understand the program from the participants' perspective. [30, 31] Analysis will require identifying major themes in the field data which may be in the form of observation records, interview responses, focus group transcripts, tapes or other field notes.[47, 48] Qualitative

evaluations can use computer assisted qualitative data analysis to produce statistical and other analyses as aids to interpretation.

The results of focus group interviews or in-depth interviews should be interpreted carefully. In interpreting the findings from individual or group interviews, it is useful to include participants' interpretations.

The following should be considered when looking for trends and patterns in qualitative data:

- If different methods were used such as interviews, focus groups, observations and document reviews, is the evidence across the methods consistent or conflicting?
- Do the different sources of data yield similar results? For example, do interviews conducted with program managers yield similar findings to interviews conducted with staff? In how many interviews/groups did each theme appear?
- Are there common trends across multiple interviews/groups?[45]

Many analytical techniques can be used to examine qualitative data. These are described in detail in the qualitative methods literature.[40, 49, 50]

## Analysis of Quantitative Data

Quantitative data analysis begins by identifying all the numerical information to be used for answering each evaluation question. To avoid data overload it is useful to develop a question-oriented data analysis plan to help identify the information required to answer a question, to help analyze the information using one or more appropriate techniques, and to help formulate an answer to the questions based on the results.[43]

It is beyond the scope of this toolkit to fully discuss how to analyze variables and perform statistical analysis. The reader may find technical assistance through consultation with an epidemiologist, health analyst and/or statistical resource materials. The general protocol will be to determine how survey responses are to be organized or tabulated and then the statistical techniques to be used. Some general considerations have been included below.

A first step in analyzing the data is to review the types of data the evaluation has gathered. For example, is the data based on simple counts (e.g., 19 out of 20 participants) or was the information measured with the use of a scale (e.g., a 5-point scale where 1 means "least likely" and 5 means "most likely")? This will determine the type of analysis that can be carried out.

The use of descriptive statistics should be an early part of an analysis in order to get to know the data. Numerical descriptive methods include measures of central tendency (mean, median and mode for example), measures of spread (range, variance and standard deviation for example) and measures of relative standing (z-scores and percentiles for example). Graphical descriptive methods include bar graphs, histograms, line graphs, pie charts and box plots.[51] Most descriptive evaluations and many analytical ones use cross-tabulation to help answer evaluation questions (e.g., do clients who attend all of the sessions have better outcomes than clients who attend between 50% and 75% of the sessions?).

Some evaluations require more advanced statistical analysis. For example, as part of a summative evaluation, a multiple logistic regression may be appropriate to examine what variables appear to impact on clients who successfully completed a program. This module's references cite several resources on multiple logistic regression. [52, 53]

When analyzing the results it helps to start with the original evaluation objectives. For example, if one wants to improve a program by identifying its strengths and weaknesses, the evaluator can organize data into program strengths, weaknesses and recommendations to improve the program. If one wants to fully understand how a program works, the data could be organized in the order in which participants go through the program. Since the program's logic model identifies key performance indicators and critical questions for evaluation, the findings may be organized according to the elements in the logic model.

## 4.8 Interpret the Results

The next step is interpretation of the results so decisions can be made about the program and so an action plan can be designed.

Analysis means recording the facts as they were gathered and reporting what was found for each evaluation question in isolation.[45] Interpretation, on the other hand, is the process of attaching meaning to the analyzed data by viewing the findings as a whole. Numbers do not speak for themselves. They need to be interpreted based on careful judgements. To interpret the data or make sense of results, it helps to consider findings from other evaluations, baseline data and pre-defined standards of expected performance[54] as well as the original program goals.

The results of a statistical analysis can be supplemented with stakeholder interpretation. While the results may be statistically significant, the differences seen may not be very meaningful in terms of the decisions to be made (i.e., they may not be clinically significant). Discussion with stakeholders can provide possible explanations of the results. Greater understanding usually emerges when others are involved so the evaluator hears firsthand how different people interpret the same information. It may be useful to include program participants when discussing the meaning of the information.

It is also helpful to organize the results by the original evaluation questions and use the results to answer these questions. The evaluation questions provide useful categories around which to group information and develop themes. For example, questions that are asked as a way to explain program outcomes may be divided into three groups:

1. questions about why the program had no effects;

2. questions about why the program had beneficial effects; and

3. questions about why the program had positive and/or negative unintended consequences.

There may be several reasons why programs do not work as well as anticipated. The program may have been implemented but never reached expected levels of implementation, or was implemented in a different manner than intended.[55, 56] Another possibility is the intervention itself might not have been strong enough to make a difference, such as a healthy eating program that simply provides people with written information about the harmful consequences of poor diet. Another consideration is whether the program implementation followed established protocols. Health programs are often implemented by different people in different organizations in different geographic areas, thereby challenging the assumption that implementation is similar across settings.[43] Finally, it may be necessary to ask if the program's logic worked as expected. If a program's logic is flawed, no impacts or limited impacts might be found.

When a program has unintended impacts, detailed evaluation of program implementation may be required to uncover the likely causes. Especially when a program has harmful consequences, evaluators have a responsibility to try to identify the reasons, to avoid repetition in the future.

## 4.9 Develop Recommendations for Action

Some evaluations may be designed specifically to exclude a "recommendations" step. The decision may have been made, for instance, that the evaluation will only present analysis and interpretation of the results, without proposing what action to take based on the evaluation's findings. In these instances it may be expected that the sponsor or program management team rather than the evaluator will arrive at recommendations.

Some clients may expect the evaluator to make action-oriented recommendations. This is usually a good way to link evaluation with other planning components (as described in Section 1.1 of this module). However, problems can arise if the evaluator takes too active a role in developing recommendations. For instance, the evaluator may be excellent as evaluator but may not have detailed knowledge of the program's social and political context and may therefore develop naïve, simplistic or damaging recommendations. The evaluator's role may best be seen as encouraging and safeguarding the accurate use of the evaluation findings, but the recommendations should be a management responsibility.

The evaluator may also perform the important role of helping to set indicators that track whether recommendations are implemented.

Even if the evaluator does not develop recommendations, he should be available to the client after completion of the evaluation to provide insight and context about the evaluation when the client develops recommendations and an action plan.

To maximize the chance that the recommendations are implemented, an evaluator can stress the importance of the following characteristics of recommendations:[43, 57, 58, 59]

- **They should be defensible**. Recommendations should be linked to the evaluation findings and derived directly from the empirical evidence.
- **They should be timely**. Recommendations have little or no value if they are not ready when decision-makers want them or if decisions have already been made.
- **They should be realistic**. If implementation of the recommendation appears to be unfeasible it will likely be ignored by decision-makers.
- **They should be targeted**. Recommendations should indicate who has the authority to approve or disapprove them and who will be responsible for implementing them if they are approved.
- **They should be simple**. Recommendations are more easily understood when they are expressed in clear, simple language.
- **They should be specific**. Recommendations are more likely to be implemented when they address only one idea and are organized into specific tasks or actions.

## 4.10 Communicate the Findings

Evaluations are useful when their results are used by decision-makers, policy-makers or other groups.[47] An important step following an evaluation is to communicate the results and recommendations as a way to help decision-makers and stakeholders to interpret, understand and apply them.[43] To encourage the use of evaluation findings, evaluators must translate the answers to the evaluation questions into policy language or in ways that are understood by the audiences of the program evaluation. It is necessary to distil large amounts of data analysis and technical language into succinct sentences that can be understood by most people.

Current evaluation practice includes many alternatives to reports, such as presentations, oral briefings, one-page summaries, evaluation newsletters, attending team meetings to discuss evaluation, and web sites.

### Write the Evaluation Report

A report of the evaluation is critical but it is sometimes ignored because people are anxious to get on with the changes to the program. The report is a record of the evaluation that can be used by others, including other LHINs, the next evaluator of the program, and program stakeholders. The report should be produced within a reasonable time after data analysis. It should limit its content to what is needed and should be free of jargon. It should use simple examples and pictorial methods such as graphs and tables to describe and explain data in ways that improve the audiences' understanding of the results.

> "A theory of evaluation must be as much a theory of political interaction as it is a theory of how to determine facts."
>
> – L. J. Cronbach and Associates,
> Toward Reform of Program Evaluation[60]

The structure and emphasis of the evaluation report will vary depending on its intended audiences. For example:

- If the main stakeholder is the project funder, the report may focus more on the program's cost-effectiveness.
- Evaluators and researchers will be more likely interested in a comprehensive report that provides the details of the evaluation.
- A concise executive summary may be ideal for decision-makers who want to know only the bottom line results and recommendations.

A useful reporting strategy is offered by the Canadian Health Services Research Foundation (http://www.chsrf.ca/) based on the '1:3:25 rule': start with one page of main messages; followed by a three page executive summary; and present findings in no more than 25 pages. The Foundation's two page resource on how to write using the rule is available at http://www.chsrf.ca/knowledge_transfer/resources_e.php#commnotes

The report format should highlight key results; it is easy to become overwhelmed with too much information. It is important to focus on the evaluation questions and on information that answers those questions, since communicating findings to different stakeholders is essential so action can be taken on the results.[45] This module's references section cites a resource for information about disseminating results.[61]

The evaluator or client should balance the needs and interests of stakeholders when deciding how to communicate information to them. At a minimum a report should provide the program description (including its logic model), the evaluation questions and data collection methods and tools in addition to the principal findings. The findings should relate to the knowledge, experience and concerns of the target audience and should use language familiar to them. Even before the evaluation starts it is advantageous to

discuss and agree upon the distribution strategy with all stakeholders as part of the communications strategy described earlier in this module. When the evaluation is underway, interim updates help to maintain stakeholder interest and enthusiasm and establishes avenues for feedback on evaluation activities.

The evaluator or client may decide to share all or part of the results with participants in the program, possibly through a brief summary report. Cost, feasibility, ethical commitments and the interest of the participants should be considered before sharing takes place.

### Present and Share the Results

The evaluation report is the basis for further communications, which may include:

- meeting with the client to review the report and its findings; and
- presenting the report to key stakeholders in partnership with the client, focusing on stakeholders most likely to be influenced by the evaluation's results and most likely to influence implementation of the evaluation's recommendations.

## 4.11 Evaluate the Evaluation

Once an evaluation has been completed, it should be evaluated. The same approaches used to evaluate a program can be used to evaluate an evaluation:

- The evaluation can examine inputs to the program evaluation, the evaluation's activities and the evaluation's outcomes.
- It can look at short, medium and long-term outcomes of the evaluation.
- It can identify intended and unintended outcomes of the evaluation.
- It can be an internal evaluation (conducted by the same evaluators who conducted the original program evaluation) or it can be an external evaluation.
- Ongoing efforts to take action based on the evaluation can be monitored.

Even if the program evaluator does not play a major role in evaluation of the evaluation, she should leave a sufficient paper trail to allow the evaluation to be evaluated (for instance, a collection of tools used, as well as narrative material that allows a determination of who did what, and why, during the program evaluation).

Organizations that evaluate their evaluations often use the Evaluation Standards of the Joint Committee on Educational Evaluation (Sage, 1994) which is the de facto evaluation standard for most fields of practice (excluding personnel standards and student testing standards). This module references these standards in section 5.5 and in Appendix G.

Section 5

# The Evaluator's Challenges

While the world in which evaluators do their work may not be as gloomily adventurous as Patton describes, evaluations face limitations and challenges.

For example, evaluation in and of itself does not directly create much change – but it can influence change. Decision-makers may take the results of an evaluation and act on it by making decisions and changes that are supported by the evaluation's evidence.

An evaluation should only be conducted if there is evidence of commitment by stakeholders to act on the results. Evaluation is not warranted if there is no likelihood that it will lead to change or improvement. Evaluation requires a level of dedication that will not be sustainable or justified if there is no expectation of progress. As well, there may be negative consequences of evaluation if the evaluation does not lead to action. Failure to act may leave in place program features that, if improved on the basis of the evaluation, would have improved inputs, activities or outcomes. As well, failure to act may jeopardize the chances that stakeholders will support future evaluations.

"With each new evaluation, the evaluator sets out, like an ancient explorer, on a quest for useful knowledge, not sure whether seas will be gentle, tempestuous, or becalmed. Along the way the evaluator will often encounter any number of challenges: political intrigues wrapped in mantles of virtue; devious and flattering antagonists trying to co-opt the evaluation in service of their own narrow interests and agendas; unrealistic deadlines and absurdly limited resources; gross misconceptions about what can actually be measured with precision and definitiveness; deep-seated fears about the evils-incarnate of evaluation, and therefore, evaluators; incredible exaggerations of evaluators' power; and insinuations about defects in the evaluator's genetic heritage."

– Q. M. Patton, Utilization Focused Evaluation: The New Century Text (3rd Edition), 1997[47]

"The last word on how we may live or die
Rests today with such quiet
Men, working too hard in rooms that are too big,
Reducing to figures
What is the matter, what must be done... "

– W.H. Auden, The Managers, 1940

## 5.1 Evaluation Skepticism, Anxiety and Resistance

Skepticism about evaluation among those whose support is crucial to make evaluation successful is a challenge. Anxiety about evaluation – a more emotional reaction than skepticism – is an even greater challenge. If left untended, skepticism and anxiety can lead to damaging resistance to, or hostility towards, an evaluation. Negative consequences include:

- lack of access to important information and data;
- compliance and cooperation problems on the part of key stakeholders;
- false reporting; and
- reduced use of evaluation findings by decision-makers.

"Many evaluative situations cause people to fear that they will be found to be deficient or inadequate by others…"

– S. Donaldson, L. Gooler and M. Scriven, Strategies for managing evaluation anxiety: Toward a psychology of program evaluation[62]

A number of beliefs may lie at the heart of skepticism or anxiety among stakeholders:

- "The evaluators will reach whatever conclusions the evaluation's sponsor or funder wants them to conclude."
- "The evaluation is being done as a pretext for closing the program or reducing its funding."
- "The evaluation is being done as a way to get us to do more work, without the resources to do the work."
- "The evaluators already have their minds made up."
- "The evaluation is being done to find fault and to blame someone."
- "The evaluators will only look for deficiencies and won't identify successes."
- "The evaluation isn't necessary. It's just a bureaucratic exercise."
- "Our program is too unique to be evaluated."
- "The evaluators will never fully understand this program."
- "The evaluation is only about numbers, not about what people in the program think."
- "The evaluators are not qualified to perform the evaluation."
- "The evaluation process will take us away from our core duties."
- "The evaluators will find out that we've made mistakes."
- "The evaluation will frighten our clients."
- "The evaluation will invade our clients' privacy."
- "Other programs and stakeholders will believe we're being evaluated because we've done something wrong."
- "We will never get to see the results of the evaluation."
- "Nobody will pay attention to, or do anything about, the evaluation's results."
- "The results of the evaluation will generate additional unnecessary work for us."

The sponsor of an evaluation, as well as the evaluators, should recognize that stakeholder skepticism, anxiety, resistance and hostility are not necessarily illogical. Stakeholders may have good reasons for their concerns. For instance:

- their previous experiences with evaluations may have been negative;
- the design of the evaluation may be flawed;
- they may have been left out of the processes of deciding whether to do an evaluation, what its purpose should be, and how it should be done;
- they may not have been fully and accurately informed about the evaluation;
- the evaluation may be taking place during a time of organizational discomfort or crisis – a time when the evaluation may be seen as adding to the crisis rather than resolving it;
- they may have reason to distrust their own leadership, or leadership in the broader environment, and they may conclude that even an excellent evaluation will not be acted upon honorably or competently by the powers-that-be; and
- the evaluation may pose the threat of unrealistic demands on their time.

Skepticism, anxiety and resistance occur not only at the beginning of an evaluation project. They can occur at any point. Evaluators need to ensure that a monitoring mechanism is in place to spot small concerns before they become big. A drop-off in attendance at meetings of stakeholders and evaluators, "no shows" for interviews, reluctance to supply available data, poorly completed questionnaires and verbal and non-verbal signs of anxiety or hostility during interviews may all be early warning signs.

To some degree, skepticism and evaluation anxiety can be reduced by clear, timely and accurate recurring communication with stakeholders from the very inception of an evaluation project, to allay concerns that are driven by lack of information. They can also be reduced through the engagement of stakeholders in the design, execution and ongoing monitoring of the

evaluation project, because engagement gives stakeholders a chance to test the validity of their concerns and to work with the evaluator to address concerns. Module 5 (*Community Engagement and Communication*) in the Health Planner's Toolkit provides useful ideas about engagement and communication in evaluation projects.

Despite best efforts at communication and engagement, it is unlikely that all skepticism and anxiety can be removed, particularly if the concerns of stakeholders are rooted in distrust of the evaluators. One useful technique in addressing trust issues involves discussion between evaluator, client and other stakeholders of the values and standards that guide evaluation. Sections 5.4 and 5.5 of this module address ethics and standards.

## 5.2 The Challenge of the "Why" And "How" Questions

In an evaluation it is easier to describe what happened than to explain why it happened or how it happened – yet answering "why" and "how" questions is often a crucial evaluation outcome. Sometimes an evaluation's stakeholders expect the evaluation to "prove" a particular outcome is caused by a particular factor, but the most one can expect is a statement of the probability of causation. Identifying causation is difficult for health outcomes. People and their problems are complex, and interventions can be multi-layered, particularly for people with long-term disorders. Clients may be served by several programs – and in a more integrated system, they are often served concurrently by several programs, adding to the number of factors that could cause outcomes.

> "Nothing is ever proven in science. There is always some uncertainty about the actual value of results obtained from some experiment or their interpretation… In the strictest sense, we never arrive at 'proof'; we simply arrive at a very high degree of probability that we understand something."
>
> – K. Prestwich, The Nature of Scientific Proof[63]

> "When I feed the poor, they call me a saint. When I ask why the poor have no food, they call me a communist."
>
> – Brazilian Archbishop Dom Helder Camara

In a more integrated system, closely aligned programs that serve many of the same clients may more often decide to carry out concurrent, linked evaluations as a way to address the "why" questions.

Asking "why" questions can also raise stakeholder anxiety levels, since these questions can be interpreted as attempts to fix blame rather than to improve a program. Evaluators should make it clear that the purpose of such questions is to explain, not to blame.

## 5.3 The Good versus The Perfect

Programs are seldom perfect – and evaluations are seldom perfect. There is usually not enough money and time to conduct an evaluation to the level of methodological rigor found in a good clinical trial for instance.

Faced with constraints, evaluators and their clients may need to cut corners in ways that leave the ultimate results open to the criticism that they are not perfect. They may, for instance, not address all the questions that one might want to ask about a program, or they may yield results that are better than educated guesses but far less conclusive than absolute proof.

To help ensure that a good (but not perfect) evaluation is a "good enough" evaluation, evaluators and their clients can do several things:

- ensure that the questions posed at the core of the evaluation are the essential questions, not just questions whose answers are "nice to know" but not essential;
- communicate with stakeholders throughout the evaluation process to make them aware that there are limits to what the evaluation will accomplish. This helps prevent the development of unrealistic expectations;
- prevent unnecessary scope creep during the project (scope creep is the tendency for projects to broaden their scope – sometimes almost imperceptibly – during a project); and
- beware of adopting logical fallacies to "explain" a program's inputs, processes or outcomes. An example of a common logical fallacy is shown below.

> "It doesn't really matter whether you can quantify your results. What matters is that you rigorously assemble evidence – quantitative or qualitative – to track your progress. If the evidence is primarily qualitative, think like a trial lawyer assembling the combined body of evidence. If the evidence is primarily quantitative, then think of yourself as a laboratory scientist assembling and addressing the data."
>
> – Jim Collins,
> Good to Great and the Social Sectors, 2005[64]

In some evaluations, qualitative data may need to be used instead of quantitative data as a way to live within the project's budget and its timeline. This is generally necessary when the quantitative data do not exist and would need to be generated from scratch. In these instances, qualitative data may not be as good as quantitative data but it may be good enough for the purposes of the evaluation.

However, in some evaluations qualitative data is more useful than quantitative data – and can be more expensive and time-consuming to gather.

**An Example of a Logical "Fallacy of Causation": post hoc propter hoc**
(the mistaken idea that if event **B** happened **after** event **A**, it happened **because of** event **A**)

An evaluation of a nutrition counseling program showed that in the second half of 2006, clients of the program reported higher satisfaction with the program than clients served in the first half of 2006.

The evaluators did not have sufficient resources to fully investigate the reasons for the improvement in satisfaction, but they noted that in mid-2006 a new Program Director was appointed. The evaluators therefore assumed that since the arrival of the new Program Director (event A) came before the improvement in satisfaction levels (event B), the arrival of the new Program Director must have caused the improvement in satisfaction levels – even though there is no corroborating evidence that event A caused event B.

## 5.4 What Ethics Govern Evaluation?

Evaluation is an endeavour for which ethical standards are essential, for the protection of all who are involved in or affected by evaluation.

Most associations representing evaluation specialists have codes of conduct. The Canadian Evaluation Society, for instance, has published *Guidelines for Ethical Conduct* comprising guidelines grouped into three categories (competence, integrity and accountability).[65]

Another good example of an ethical code for evaluation is *Guidelines For The Ethical Conduct of Evaluations*, published by the Australasian Evaluation Society. These guidelines can be accessed as a 16 page pdf document through the Society's web site: http://www.aes.asn.au/ (scroll down the home page and click on *"Guidelines For The Ethical Conduct of Evaluations"*).

**Canadian Evaluation Society Ethical Guidelines for Competence**

Evaluators are to be competent in their provision of service.

- Evaluators should apply systematic methods of inquiry appropriate to the evaluation.
- Evaluators should possess or provide content knowledge appropriate for the evaluation.
- Evaluators should continuously strive to improve their methodological and practice skills.

**Canadian Evaluation Society Ethical Guidelines for Integrity**

Evaluators are to act with integrity in their relationships with all stakeholders.

- Evaluators should accurately represent their level of skills and knowledge.
- Evaluators should declare any conflict of interest to clients before embarking on an evaluation project and at any point where such conflict occurs. This includes conflict of interest on the part of either evaluator or stakeholder.
- Evaluators should be sensitive to the cultural and social environment of all stakeholders and conduct themselves in a manner appropriate to this environment.
- Evaluators should confer with the client on contractual decisions such as: confidentiality; privacy; communication; and, ownership of findings and reports.

**Canadian Evaluation Society Ethical Guidelines for Accountability**

Evaluators are to be accountable for their performance and their product.

- Evaluators should be responsible for the provision of information to clients to facilitate their decision-making concerning the selection of appropriate evaluation strategies and methodologies. Such information should include the limitations of selected methodology.
- Evaluators should be responsible for the clear, accurate, and fair, written and/or oral presentation of study findings and limitations, and recommendations.
- Evaluators should be responsible in their fiscal decision-making so that expenditures are accounted for and clients receive good value for their dollars.
- Evaluators should be responsible for the completion of the evaluation within a reasonable time as agreed to with the clients. Such agreements should acknowledge unprecedented delays resulting from factors beyond the evaluator's control.

## 5.5 What Standards Govern Evaluation?

Standards can be defined as generally accepted principles, criteria and rules for the best or most appropriate way to carry out an activity. They may include ethical standards but are not limited to ethics. However, many professional ethical codes require adherence to both ethical and technical standards.

While there is no made-in-Canada set of standards for evaluation, The Canadian Evaluation Society has espoused the standards established by the Joint Committee for Educational Evaluation of the American Evaluation Association and is a member of this Joint Committee. While these standards were developed in an educational context they are considered applicable to evaluation in sectors such as healthcare.

These standards, found in Appendix G of this module, are grouped into four clusters:

1. **utility standards** to ensure that an evaluation will serve the information needs of intended users;

2. **feasibility standards** to ensure that an evaluation will be realistic, prudent, diplomatic, and frugal; and

3. **propriety standards** to ensure that an evaluation will be conducted legally, ethically and with due regard for the welfare of those involved in the evaluation as well as those affected by its results; and

4. **accuracy standards** to ensure that an evaluation will reveal and convey technically adequate information about the features that determine the worth or merit of the program being evaluated.

Section 6

# A Few Final Tips

The following tips may be of use to those venturing into evaluation.

1. **During evaluations, remain both charitable and realistic**. Evaluation is about fostering the pragmatic and possible rather than expressing impatience about failure to achieve the utopian and impossible.

2. **Treat evaluation as a group activity, not a solo activity**. Evaluation is not about someone wearing a green eye shade sitting alone in a room, making abstruse calculations that will seal the fate of a program. Evaluation requires the involvement of many stakeholders. Treat them as partners in the evaluation, not as its servants.

3. **Deal with evaluation skepticism, anxiety, resistance and hostility**. Stakeholders who are concerned about an evaluation may have good reasons for their concerns. Even if the concerns are not objectively valid, they can still turn an evaluation into an exercise in failure.

4. **Work toward comprehensive evaluations**. Such evaluations include formative and summative evaluation as well as ongoing monitoring. In the long haul, taking this approach will help make evaluation a smooth ongoing component of the planning process rather than a series of disjointed activities.

5. **Work toward creating a culture of evaluation**. Without this culture in organizations and systems, evaluation can be seen as an unwelcome outside intrusion rather than a necessary, desirable and normal way to do business. Appendix H in this module presents ten important factors for embedding evaluation in organizational and system cultures.

6. **Be sure evaluations are evaluated**. Evaluations can be evaluated in the same ways that programs can be evaluated. Evaluating evaluations is just as important as evaluating programs – and given the likelihood of continued scarce resources to conduct evaluations, evaluation of evaluations helps use available resources wisely. It also sends a positive message to stakeholders in program evaluations: *"Evaluation is so important that we, the evaluators, make sure that we are evaluated too."*

## Section 7

# Summary

Evaluation includes a wide variety of methods to evaluate many aspects of programs in different settings. This module describes the two main kinds of evaluation as well as the environment of ethics and standards within which they are conducted:

1. **formative evaluation** (which evaluates the inputs into a program and the activities meant to convert inputs into outcomes); and

2. **summative evaluation** (which evaluates the outcomes of a program).

The module also points out that evaluations can be prospective or retrospective.

After describing the planning stages of evaluation, the module describes the 22 steps in an evaluation:

**Table 5: Steps in Preparing and Conducting an Evaluation**

| **The eleven steps for *preparing* an evaluation:** |
|---|
| 1. identify and engage stakeholders; |
| 2. set the purpose of the evaluation; |
| 3. embed the program's objectives within a program logic model; |
| 4. conduct an evaluability assessment; |
| 5. address ethical issues; |
| 6. develop the evaluation project's terms of reference; |
| 7. develop the evaluation team; |
| 8. develop a project communications plan; |
| 9. confirm the evaluation design; |
| 10. design evaluation questions; and |
| 11. establish measurable indicators. |

| **The eleven steps for *conducting* an evaluation:** |
|---|
| 1. identify population and sampling; |
| 2. develop data collection tools and methods of administration; |
| 3. train personnel who will administer the tools; |
| 4. pilot test the tools and methods of administration; |
| 5. administer the tools and monitor the administration; |
| 6. prepare the data for analysis; |
| 7. analyze the results; |
| 8. interpret the results; |
| 9. develop recommendations for action; |
| 10. communicate the findings; and |
| 11. evaluate the evaluation. |

It also describes limitations and challenges commonly found in evaluation processes, including:

- the limited influence evaluation has when it is not accompanied by stakeholder commitment to act on the results;
- evaluation skepticism, anxiety, resistance and hostility sometimes exhibited by stakeholders; and
- dealing with the reality that evaluations are often limited by available resources.

Health and health services reflect an era of integration, community-based programming and partnerships in funding and service delivery. Given the prevalent focus on cost containment there is more competition for program funding and likely few resources for evaluation. With health system reform and a constantly changing program environment, evaluators may be challenged to maintain methodological rigor in evaluations.

There are many resources that provide detailed guidance for evaluation designs, techniques and analysis. Those resources (some of which are cited in Appendix I), coupled with this module itself, can assist LHINs and their community partners in planning for and conducting evaluations.

"Do you remember the play 'My Fair Lady'? Do you remember when Mr. Higgins, frustrated with the non-logical behavior of his lady friend says 'Why can't women be more like men?' Well Mr. Higgins might not get much bemused attention today when this kind of remark could get you sent to jail for sexist discrimination. But it helps to make this point. As a field, evaluation persistently faces a client base that to a large extent is not rational, analytical, empirical and so on... but we persistently, blissfully overlook this fact and then complain that no one listens to us. Of course, I am exaggerating here, to make a point.

What we really seem to feel, unconsciously, is that the world really should be more like us. We need to guard against giving the impression that we are superior. We need to guard against intimidating non-evaluators with imposing technical language. And we need to guard against projecting any sense that we 'know' the right way to do things, and the world should listen to us because we know how to think and act rationally – we've been trained to do it."

– Chris Wye,
Evaluation: The Path to the Future
(Canadian Evaluation Society Conference keynote speech, 2003)[66]

# References


1. Ontario Ministry of Health and Long-Term Care, Public Health Branch, 1996. In: The Health Communication Unit at the Centre for Health Promotion. Introduction to evaluation health promotion programs. November 23, 24, 2004. Accessed December 27, 2007,at http://www.thcu.ca/infoandresources/presentations/thcuevalslidesv2.0.Nov.2004.forweb.pdf

2. Rossi PH, Freeman HE, Lipsey MW. Evaluation: A systematic approach, 6th Ed. Thousand Oaks (California): Sage Publications; 1999.

3. Rutman L, Mowbray G. Understanding program evaluation. Newbury Park (California): Sage Publications, Inc.; Volume 31. 1983.

4. Stoeker R. Making Connections: Community Organizing, Empowerment Planning, and Participatory Research in Participatory Evaluation (draft Internet copy). Accessed December 27, 2007 at http://sasweb.utoledo.edu/drafts/evalppranon.htm#intro

5. Osborne D, Gabler T. Reinventing Government. Addison-Wesley Publishing. Company; 1992.

6. Zorzi R, Perrin B, McGuire M, Long B, Lee L. Defining the benefits, outputs, and knowledge elements of program evaluation. The Canadian Journal of Program Evaluation 2002, 17(3): 143-150.

7. Drummond M, Stoddard G, Torrance G. Methods for the economic evaluation of health programs. New York: Oxford University Press; 1997.

8. Industry Canada. Steps to competitiveness: Step 8: Quality assurance glossary. 2004. Accessed December 27, 2007 at http://strategis.ic.gc.ca/epic/internet/instco-levc.nsf/en/h_qw00037e.html

9. Scriven M. The methodology of evaluation, in Perspectives on curriculum evaluation. AERA Monograph Series on Curriculum Evaluation, Chicago (Illinois): Rand McNally. No.1, 1967.

10. The Health Communication Unit at the Centre for Health Promotion. Evaluating health promotion programs. Centre for Health Promotion, University of Toronto: Version 3.5, April 6, 2006. Accessed December 27, 2007 at http://www.thcu.ca/infoandresources/publications/EVALMasterWorkbookv3.6.03.06.06.pdf

11. W.K. Kellogg Foundation Evaluation Handbook. W.H. Kellogg Foundation (undated). Accessed December 27, 2007 at http://www.wkkf.org/pubs/Tools/Evaluation/Pub770.pdf

12. Trevisan M, Huang YM. Evaluability Assessment: a Primer, 2003.

13. United States Government Accountability Office. Performance Measurement and Evaluation: Definitions and Relationships. May 2005. Accessed December 27, 2007 at http://www.gao.gov/special.pubs/gg98026.pdf

14. Newsletter of the Standing International Conference of Central and General Inspectorates of Education (SICI), July 2000.

15. Patton M, quoted in Fetterman D. Empowerment Evaluation: Holding ourselves and our friends accountable (a view from our own lens). Accessed December 27, 2007 at http://homepage.mac.com/profdavidf/eeresponse.pdf

16. Fetterman D. Empowerment Evaluation: Holding ourselves and our friends accountable (a view from our own lens). Accessed December 27, 2007 at http://homepage.mac.com/profdavidf/eeresponse.pdf

17. Scriven M, quoted in Fetterman D (2001). Foundations of empowerment evaluation. Thousand Oaks, CA: Sage.

18. Fetterman DM. (2001). Foundations of empowerment evaluation. Thousand Oaks, CA: Sage.

19. Fine AH, Thayer CE, Coghlan A. Program Evaluation Practice in the Nonprofit Sector. Innovation Network Inc. 1998, Accessed December 27, 2007 at http://www.nonprofitresearch.org/usr_doc/Fine.pdf

20. European Commission, Evaluating EU Expenditure Programmes: A Guide to intermediate and ex post evaluation. Accessed December 27, 2007 at http://ec.europa.eu/budget/evaluation/guide/guide03_en.htm

21. Porteus N, Sheldrick B, Stewart P. Program Evaluation Toolkit. Public Health Agency of Canada. 1997. Accessed December 27, 2007 at http://www.phac-aspc.gc.ca/php-psp/toolkit.html

22. Introduction to Evaluation. Web Center for Social Research Methods. Accessed December 27, 2007 at: http://www.socialresearchmethods.net/kb/intreval.php

23. Taylor-Powell E, Steele S, Douglas M. Planning a program evaluation. University of Wisconsin, Cooperative Extension. February 1996. Accessed December 27, 2007 at http://www.uwex.edu/ces/pdande/evaluation/index.html

24. Kirkpatrick S. The program logic model: what, why and how? December 31, 2003. Accessed December 27, 2007 at http://www.charityvillage.com/cv/research/rstrat3.html

25. Shortell SM, Richardson WC. Health program evaluation. C.V. Mosby Company St. Louis. 1978.

26. McNamara C. Basic Guide to Program Evaluation. Accessed December 27, 2007 at http://www.managementhelp.org/evaluatn/consent.htm

27. Campbell D, Stanley J. Experimental and quasi-experimental designs for research. Rand McNally. 1966.

28. Fink A. Evaluation fundamentals: guiding health programs, research and policy. Newbury Park (CA): Sage Publications. 1993

29. Fischer DF. Historians' Fallacies: Toward a Logic of Historical Thought. New York: Harper. 1970.

30. Green I. Lessons Learned from Two Decades of Program Evaluation in Canada. Internet paper. Accessed December 27, 2007 at http://www.yorku.ca/igreene/progeval.html

31. Weiss CH. Evaluation: methods for studying programs and policies. Upper Saddle River (NJ): Prentice Hall. 1998.

32. Patton MQ. Utilization-focused evaluation: the new century text. Thousand Oaks (CA): Sage Publications. 1997.

33. Mendenhall W, Ott L, Scheaffer R. Elementary Survey Sampling. Belmont, California: Duxbury Press. 1971.

34. Salant P, Dillman DA. How to conduct your own survey. Toronto: John Wiley & Sons, Inc. 1994.

35. Centers for Disease Control and Prevention. Epi Info software program. Accessed December 27, 2007 at http://www.cdc.gov/epiinfo

36. Patton MQ. Qualitative evaluation checklist. Evaluations checklist project. September 2003. Accessed December 27, 2007 at http://www.wmich.edu/evalctr/checklists

37. National Resource Centre, Compassion Capital Fund. Measuring outcomes. Accessed December 27, 2007 at http://www.nascsp.org/documents/Outcomes.pdf

38. Carmines E, Zeller R. Reliability and Validity Assessment. Newbury Park, California: Sage Publications, Inc. 1990.

39. Peter RJ. Qualitative research: A practical guide. Toronto: RCI/PDE Publications. 1994.

40. Denzin NK, Lincoln YS (Eds.) Handbook of qualitative research. Thousand Oaks (California): Sage Publications Inc. 2000.

41. The Health Communication Unit, Centre for Health Promotion. Using focus groups. University of Toronto. Version 2.0; June 2002. Accessed December 27, 2007 at http://www.thcu.ca/infoandresources/publications/Focus_Groups_Master_Wkbk_Complete_v2_content_06.30.00_format_aug03.pdf

42. Morgan DL. Krueger RA. The focus group kit. Volumes 1-6. Sage Publications. Thousand Oaks (California); 2000.

43. Grembowski D. The practice of health program evaluation. Thousand Oaks, California: Sage Publications, Inc.; 2001.

44. Woodward C, Chambers L. Guide to questionnaire construction and question writing. Ottawa: Canadian Public Health Association. 1980.

45. The Health Communication Unit, Centre for Health Promotion. Conducting survey research workbook. University of Toronto. Version 2.0; March 1999. Accessed December 27, 2007 at http://www.thcu.ca/infoandresources/publications/Surveys_Master_Wkbk_V2_Formating%2008.09.03_Content%2003.31.99.pdf

46. Streiner DL, Norman GR. PDQ Epidemiology. Second Edition. St. Louis; Mosby-Year Book, Inc. 1996.

47. Patton Q M. Utilization Focused Evaluation: The New Century Text (Third Edition). 1997.

48. Caudle SL. Using qualitative approaches. In: Wholey JS, Hatry HP, Newcomer KE (Eds.). Handbook of practical program evaluation. San Francisco; Jossey-Bass; 1994: 69-95.

49. Rothe JP. Qualitative research: A practical guide. Toronto: RCI/PDE Publications, 1994.

50. Lofland J, Lofland L. Analyzing social settings: a guide to qualitative observation and analysis. Belmont (California): Wadsworth Publishing. 1984.

51. Mendenhall W, Beaver RJ, Beaver BM. Introduction to probability and statistics, 9th ed. Wadsworth, Inc. Belmont, California. 1994.

52. Hosmer D, Lemeshow S. Applied logistic regression. 2nd ed. Wiley & Sons, 2000.

53. Rosner B. Fundamentals of biostatistics. 6th ed. Duxbury Press. 2005.

54. Taylor-Powell E. The logic model: a program performance framework. University of Wisconsin: Cooperative Extension, 1989. Accessed December 27, 2007 at www.uwex.edu/ces/pdande/evaluation/pdf/LogicETP.pdf

55. Gottfredson D, Gottfredson G, Skroban S. Can prevention work where it is needed most? Evaluation Review, 1998. 22(3): 315-340.

56. Scheirer MA. Designing and using process evaluation. In: Wholey JS, Hatry HP, Newcomer K.E. (Eds.) Handbook of practical program evaluation. San Francisco; Jossey-Bass. 1994: 40-68.

57. Hendricks M. Making a splash: reporting evaluation results effectively. In: Wholey JS, Hatry HP, Newcomer KE (Eds.) Handbook of practical program evaluation. San Francisco; Jossey-Bass; 1994: 549-575.

58. Hendricks M, Papagiannis M. Do's and don'ts for offering effective recommendations. Evaluation Practice. 1990: 11(2): 121-125.

59. Sonnichsen RC. Evaluators as change agents. In Wholey JS, Hatry HP, Newcomer KE (Eds.) Handbooks of practical program evaluation. San Francisco; Jossey-Bass; 1994: 534-548.

60. Cronbach LJ and Associates. Toward Reform of Program Evaluation. San Francisco; Jossey-Bass; 1980

61. Morris LL, Fitz-Gibbon CT, Freeman ME. How to communicate evaluation Findings. Sage Publications, Inc; 2nd ed. December 1, 1987.

62. Donaldson SI, Gooler LE, Scriven M. (2002). Strategies for managing evaluation anxiety: Toward a psychology of program evaluation [Electronic version]. American Journal of Evaluation. 23(3), p. 261-272.

63. Prestwich K. The Nature of Scientific Proof. Biology Department, College of Holy Cross, (undated). Accessed December 27, 2007 at http://www.holycross.edu/departments/biology/kprestwi/behavior/e&be_notes/E&BE_04_Sci_Meth&Philo.pdf

64. Collins J. Good to Great and the Social Sectors: A Monograph to Accompany Good to Great. New York; Harper Business. 2005.

65. CES Guidelines For Ethical Conduct. Canadian Evaluation Society. Accessed April 12, 2007, at: http://www.evaluationcanada.ca/site.cgi?s=5&ss=4&_lang=EN

66. Wye C. Evaluation: The Path to the Future. Canadian Evaluation Society Conference keynote speech, 2003. Accessed December 27, 2007 at http://www.evaluationcanada.ca/site.cgi?s=1&ss=1&_lang=en&num=251

67. The Health Communication Unit at the Centre for Health Promotion. Evaluating comprehensive workplace health promotion. University of Toronto, Version 1.0, March 15, 2005.

68. Porteous N, Sheldrick B, Stewart P. Introducing program teams to logic models: facilitating the learning process. Canadian Journal of Program Evaluation. Special Issue 2002; 17(3): 113-141.

69. Schmitz CC, Parson BA. Everything you wanted to know about logic models but were afraid to ask. W.K. Kellogg Foundation. Insites: Boulder, Colorado. 1999. Accessed December 27, 2007 at http://www.insites.org/documents/logmod.htm

70. McNamara C. Some myths about program evaluation. Last revision February 16, 1998. Accessed December 27, 2007 at http://www.managementhelp.org/evaluatn/fnl_eval.htm

71. Fink A. Evaluation fundamentals: guiding health programs, research and policy. Newbury Park (CA); Sage Publications. 1993

72. Ferguson L. Developing an evaluative culture. Presentation to the 2003 Australasian Evaluation Society International Conference in Auckland, September 17, 2003. Accessed December 27, 2007 at http://www.evaluationcanada.ca/distribution/20030917_ferguson_linda.pdf

# Appendix A

# Let Them Eat Cake

**A Comparison of Making a Cake and Operating a Program – Everything on this page can be evaluated.**

| THE CAKE | THE PROGRAM |
|---|---|
| **ingredients (inputs)** | |
| people who need to eat, and who like to eat cake | a population with a defined unmet health need |
| flour, butter, milk, eggs, salt, baking powder, walnuts | clients drawn from the population in need, funding, staff, volunteers |
| a basic cooking skill set | a professional knowledge base sufficient to allow service to be delivered |
| a kitchen countertop, a mixing bowl, a spatula, a cake pan and an oven | a program site, furnishings, promotional material, other basic program supports |
| a recipe, and instructions on how to use the oven | a program logic model and a detailed description of how the services will be provided (what does what and when) |
| a thermometer and a smoke detector | an ongoing process for monitoring/assessing the program |
| **input constraints:** The oven is only big enough to bake a small cake, and since there is no butter in the fridge, lard will be used instead. | **input constraints:** Several health professions needed for this program will remain in short supply, and there is only enough money to run a program serving the immediate community, not the whole region. |
| **activities** | |
| mixing the right ingredients in the right amounts and in the right order (based on the recipe) and baking the cake at the right temperature for the right length of time (based on the oven instructions) | providing the right resources at the right time in the right combination and the right order, to deliver the program (based on the detailed description of how the services will be provided) |
| tasting the batter to be sure it tastes good, and checking from time to time to see if the cake is baking properly | monitoring/assessing program activities |
| **outcomes** | |
| **short term intended outcomes:** A cake thoroughly baked but not burnt, with the right texture and an appealing appearance | **short term intended outcomes:** Clients complete the program, they believe they have benefited from the program, and there is evidence that on leaving the program their health status has improved |
| **medium term intended outcomes:** Family and friends eat pieces of the cake, they enjoy it, and they benefit nutritionally from eating the cake | **medium term intended outcomes:** Six months after program completion, clients maintain their health gains attributable to the program |
| **long-term intended outcomes:** Family and friends ask the cook to bake more cakes, and continue to enjoy and benefit from eating the cakes | **long-term intended outcomes:** One year after completion of the program, clients maintain their health gains attributable to the program |
| **positive unintended outcomes:** The cook has developed skills that allow him to bake cookies, and his family appreciates his talents. | **positive unintended outcomes:** Given their improved health status, clients are able to perform better in their workplaces |
| **negative unintended outcomes:** The kitchen is a mess and it takes two days and a fire hose to clean it up. | **negative unintended outcomes:** Some clients overestimate the degree to which their health has improved, and take health risks they would not otherwise take. |

# Appendix B

# Developing a Logic Model

## What is a Program Logic Model?

A program logic model provides a framework for an evaluation. It is a flow chart that shows the program's components, the relationships between components and the sequencing of events. It shows what a program is intended to do, who the program is for and why the program exists.

A logic model:

- shows the relationship between what is invested (inputs), what is done (processes), who is reached (outputs), and what results (outcomes);
- is comprised of a sequence of "IF-THEN" relationships; and
- represents the core of program planning and evaluation.[67]

Evaluation can determine whether the program is working as shown in the logic model. The logic model also sets the stage for determining if an evaluation is feasible (i.e., if the program can be evaluated).

## What Does a Logic Model Look Like?

Figure B.1 shows the overall framework of a logic model, describing components of a program in system terms and identifying program dependencies. Outcomes depend on outputs, which depend on inputs or activities.

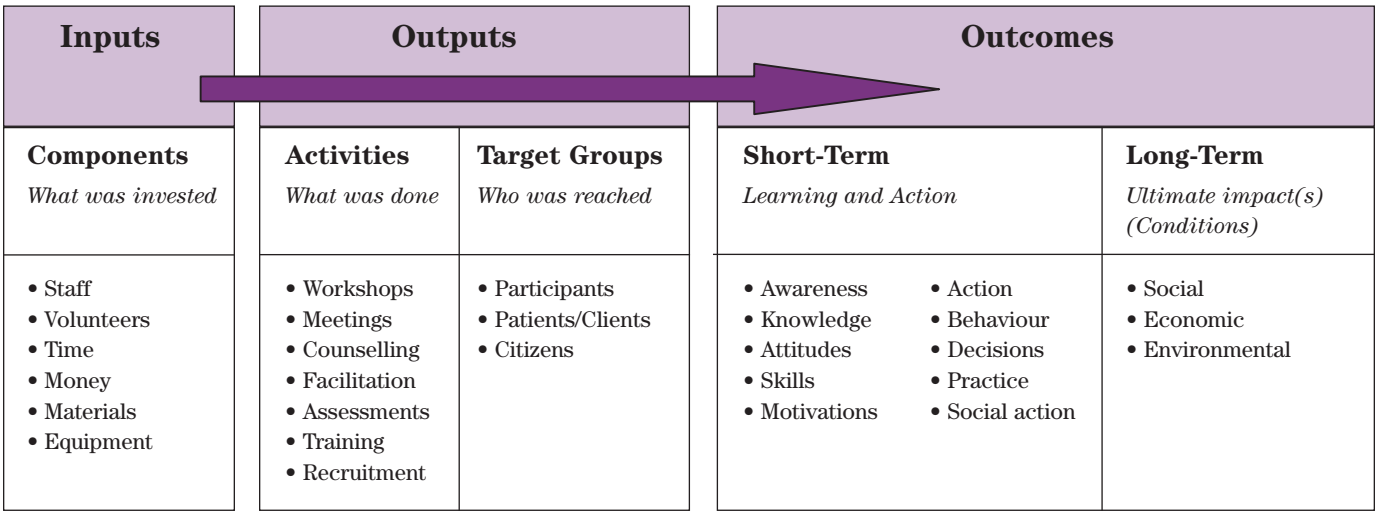
## Why Use a Program Logic Model?

**Table B.1: Usefulness of the Logic Model[68]**

| Benefits of Developing a Logic Model: |
|---|
| • builds a link between strategic and operational planning; |
| • provides the opportunity for stakeholders to discuss the program and agree upon its description; |
| • identifies different understandings or perceptions of the program; |
| • clarifies the difference between the activities and the intended outcomes of the program; and |
| • helps identify critical questions for evaluation. |

| Benefits of a Completed Logic Model: |
|---|
| • summarizes key elements of a program; |
| • makes explicit the assumptions underlying the program including the theory behind program activities; |
| • shows cause-and-effect relationships (i.e., which activities are expected to lead to which outcomes); |
| • helps in negotiating who is accountable for which outcomes over what time period; and |
| • helps develop performance measures for ongoing monitoring and assessment. |

**Figure B.1: High Level Overview of a Logic Model[54]**



## Use of IF-THEN Logic Model Statements

To support logic model development, a set of "IF-THEN" statements helps determine if the rationale linking program inputs, outputs and objectives/outcomes is plausible, filling in links in the chain of reasoning.[17] As Figure B.2 shows, the rationale flow is, *"if such and such can be achieved or is allowed to happen…then such and such will follow. And if such and such follows, then we should see some decrease in the problem we are addressing, or increase in the type of outcome we're looking for."*[69]

The use of "IF-THEN" statements will be illustrated later in this appendix when the steps for developing a logic model are detailed.

## How Do Logic Models Differ from Action Plans?

Logic models are often confused with action plans:

- An **action plan** contains program objectives, both a timeline and task outline, and will specify exactly what the staff or personnel need to do to implement a project (launching training sessions for instance).
- A **logic model**, on the other hand, illustrates the *presumed effects* of launching training sessions to increase awareness of services, and thereby increase the number of people accessing services.

## What Are the Steps in Developing a Program Logic Model?

This section recommends nine steps for developing a logic model.[24, 68] Before introducing these steps, it is important to set expectations about the resources required for preparing a logic model. The length of time it will take to develop a logic model will depend on the size and complexity of the program, the degree of consensus on the objectives of the program and the amount of experience in working with logic models.

A logic model is a flow chart showing components of a program, relationships between components, and the sequencing of events. There are different ways this flow can be described, but they all define three key themes – what, who and why.

Table B.2 illustrates a model that breaks this down further, yielding the mnemonic "CAT SOLO".[68]

**Figure B.2: An Example of the Use of IF-THEN Model Statements**

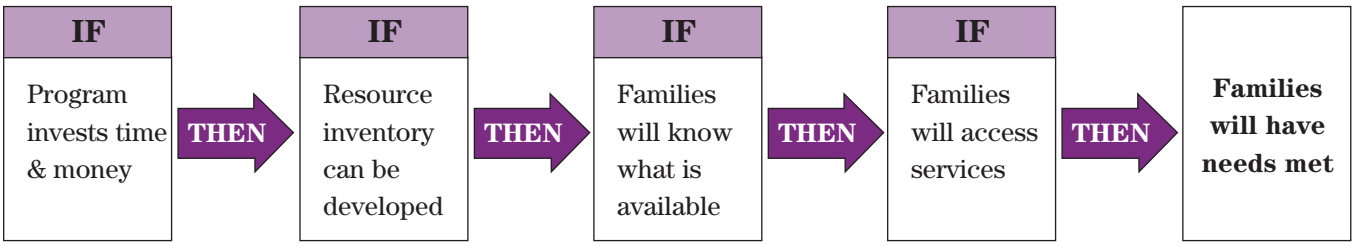


**Table B.2: CAT SOLO Mnemonic**

| Themes | Logic Model (CAT SOLO) | Descriptions |
|---|---|---|
| **WHAT** | **C**omponents | Groups of closely related activities in a program, such as educating, social marketing, etc. |
| | **A**ctivities | Action steps or those things that the program does to attain outcomes |
| **WHO** | **T**arget Groups | Individuals, groups, and/or communities to whom the program is directed, defined on the basis of age, sex, income, health characteristics, area of residence, ethnicity, etc. |
| **WHY** | **S**hort-Term **O**utcomes | Changes or benefits expected to occur in relatively short time frames |
| | **L**ong-Term **O**utcomes | Changes that will take longer to be realized |

There is no right or wrong place to start developing a logic model. The decision about where to start may hinge on the developmental stage of a program. Beginning with activities might be easier for existing programs while starting with outcomes may be more appropriate for new programs.

Two worksheets can be used to assist in the development of a logic model:

1. a CAT Worksheet; and
2. a SOLO Worksheet.

Before completing these, consider Step 1 below.

**Step 1**. Form a small workgroup of program planners, staff, evaluators and other stakeholders who offer expertise needed to describe the program and its intended results. This group may meet several times to develop and revise the logic model.

## The CAT Elements of a Logic Model

Next, the CAT Elements (**C**omponents, **A**ctivities and **T**arget Groups) of a logic model can be examined.

- Examples of **C**omponents include groups of related program activities such as coordination, community development, counseling, crisis intervention, fundraising, outreach, public education and training.
- Examples of **A**ctivities include the program's action steps to attain outcomes, phrased using action verbs like conduct, develop, distribute, educate/teach/train, provide, offer, identify, refer, set up or support.
- **T**arget groups comprise groups or communities to whom the program is directed. Examples of target groups include aboriginal women, low-income families living in rural areas, new immigrants, residents with end stage renal disease who receive in-hospital dialysis treatment, seniors living in long-term care facilities, and urban children between the ages of birth to six years.

Steps 2 to 6 will assist in completing the CAT worksheet.

**Step 2**. Review program reports, mission statements, strategic planning documents and relevant literature to identify components or information that will go into the boxes in the logic model. Most of the content in the logic model should be found in program documents with the exception of the cause and effect relationships. Step 9 will address making inferences for cause and effect relationships.

**Step 3**. Define the target group(s), considering socio-demographic variables and health characteristics.

**Step 4**. List the program activities in terms of what the program intends to do to achieve its objectives. Objectives should drive activities. This makes more sense than trying to determine the objectives based on already-planned activities.

**Step 5**. Group program activities into components such as counseling, training and advocacy.

**Step 6**. Working with a program team or work group, work through the CAT worksheet.

Next, the SOLO elements of a logic model are examined.

## The SOLO Elements of a Logic Model

Outcomes refer to the reasons why a program is delivered – the results or changes to be achieved with each target group, or changes that did occur. Outcomes answer the questions *"What difference has the program made in people's lives? Whose lives?"* The focus is on what the program makes happen, and not on the process of achieving them. Outcomes fall along a continuum from short-term to long-term effects or results.

- **S**hort-term **O**utcomes are the direct benefits of the activities delivered to program participants. Examples include increased awareness or knowledge, improved skills or a change in attitudes.

- **L**ong-term **O**utcomes reflect the social and economic consequences of a program in the broader community. They refer to the ultimate goals of the program. Long-term outcomes may be expressed as a change in practice or behaviour, or a change in status or condition such as reduced number of ALC days or improved health status.

The distinction between short- and long-term outcomes is about sequence and does not necessarily refer to specific timeframes. It can be useful to focus on the

**Table B.3: Sample CAT Worksheet**

| Components<br>**What are the main sets of activities?** | Activities<br>**What things are done? What services are delivered?** | Target Groups<br>**For whom are activities designed?** |
|---|---|---|
| Health Education | • Organize series<br>• Facilitate sessions | • Parents of children 2 to 4 years, especially parents with high school education or less |
| Recruitment | • Advertise in stores, libraries, community resource centres<br>• Write articles for community newspapers<br>• Send letters | • General public<br>• Parents of children 2 to 4 years, especially parents with high school education or less<br>• Physicians<br>• Community resource centres<br>• Other community organizations |

"IF-THEN" sequence and think of it as an outcomes hierarchy or an outcome path. For example, "If outcome A occurs, then outcome B should occur next, which should lead to outcome C."

For short- and long-term outcomes it is useful to include the direction of change and precisely what the program is trying to change. Expressions of outcomes may include use of the following words:

| | | | |
|---|---|---|---|
| Decreased | Expanded | Increased | Reduced |
| Diminished | Extended | Lowered | Raised |
| Eliminated | Improved | Prevented | |

**Step 7.** Work through the SOLO worksheets by answering each of the questions.

The reader is now equipped to prepare the logic model.

**Step 8**. Draft the logic model. Place elements outlined in the CAT and SOLO worksheets into a logic model diagram and add directional arrows to demonstrate causal relationships, depicted vertically or horizontally. Ideally a logic model is contained within a single page with enough detail to be explained easily and understood by other people. A logic model may be divided into key parts or phases with each part or phase on a separate page with additional detail if required.

**Step 9**. Check the logic to ensure each element outlined in steps 7 and 8 are causally linked to the next. Are objectives clear and measurable? Are causal linkages realistic? Then verify the accuracy and readability of the logic model and modify accordingly. It helps to verify the logic model by interviewing program managers and program staff because the way the program is portrayed in the logic model may differ from how the manager and key staff managers view it. Questions to ask might include:[3]

- Are any program components missing from the model?
- How does each component operate?
- Are these the activities that actually happen?
- Do you think that the program's activities are carried out in a uniform and systematic manner?
- Are any objectives missing from the model?
- Do you consider the objectives realistic?
- Is each objective and output precise enough to permit measurement?

Figure B.3 is a practical example – a logic model for the Sudbury and District Public Health Unit Healthy Sexuality and Risk Reduction Outreach Program, adapted from an example provided by the Health Unit (October 2006).

The Healthy Sexuality and Risk Reduction Outreach Program is a new program aimed at populations engaging in high risk behaviours as well as health

**Table B.4: Sample SOLO Worksheet**

| What is the direction of change? | What does the program intend to change? | Is it short-term or long-term? | What components contribute to this outcome? |
|---|---|---|---|
| Increased | awareness of the program | short-term | recruitment |
| Increased | participation in the program | long-term | recruitment |
| Increased | number of participants adopting healthy behaviours | long-term | health education |
| Improved | caregiver skills | short-term | health education |

professionals providing services to this population. It is intended to increase the availability of practitioners who can provide sexual health services, and thereby increase availability of screening services. Early stages of the program involve negotiations for locations of outreach clinics, developing policies and procedures for the clinics, training of practitioners, promoting awareness of the program and integration with other service providers.

Once practitioners are in place, the program will offer early detection services, resources, supplies, information for, and referrals to, other community services.

Specific program outcomes include:

- increased access to prevention and early detection services and resources;
- increased availability of practitioners who can provide a variety of sexual health services;
- increased number of referrals to existing community supports and services; and
- increased awareness and coordination of health services and the establishment of partnerships to address primary health care needs and gaps.

In the long-term it is hoped that this program will decrease the incidence of key diseases including cervical cancer, and decrease risk behaviours in the target population.

The logic in Figure B.3 flows as follows: **IF** there is marketing via community partners and training for professionals via preceptorships and certification, **THEN** there will be increased awareness of the program, increased availability of practitioners via provision of services in strategic locations and increased screening. **IF** there is increased screening and increased availability of practitioners in the program, **THEN** the result is an increased number of individuals accessing appropriate services, a decrease in overall risk behaviours and subsequently, a decrease in incidence rates of key diseases.

To summarize, the steps for building a logic model are:

**Step 1**. Form a small workgroup of program planners, staff, evaluators and other stakeholders who have knowledge of the program and can offer the expertise needed to describe the program and its intended results accurately.

**Step 2**. Review program reports, mission statements, strategic planning documents and relevant literature to identify the main components or the information that will go into the boxes in the logic model.

**Step 3**. Define the target group(s). Consider socio-demographic variables, health characteristics, etc.

**Step 4**. List the program activities. That is, list what the program is intended to do in order to achieve its objectives.

**Step 5**. Group program activities into components such as counseling, training and advocacy.

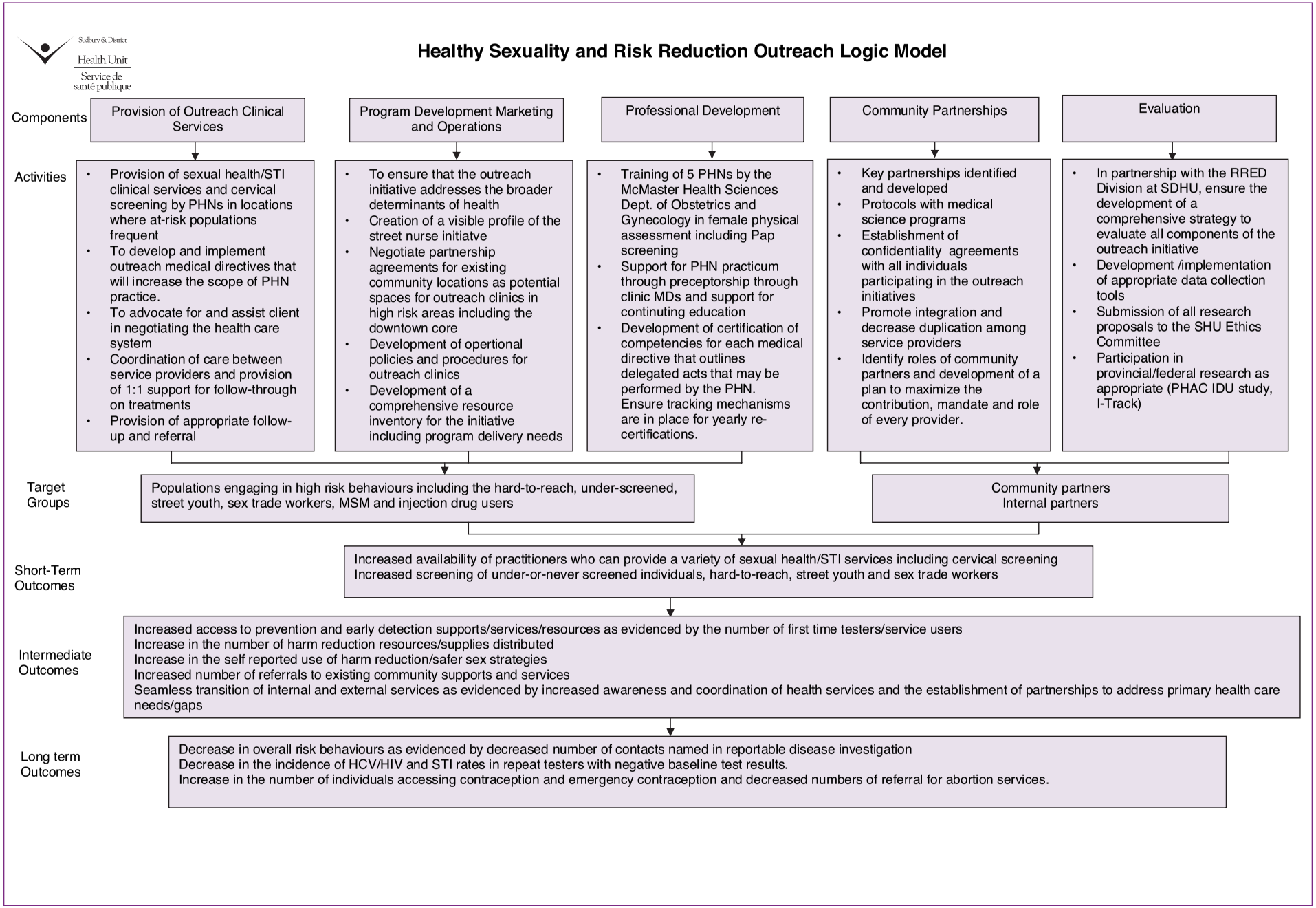
**Step 6**. Work through the CAT worksheet.

**Step 7**. Work through the SOLO worksheet.

**Step 8**. Draft the logic model. Place elements outlined in the CAT and SOLO worksheets into a logic model diagram and add directional arrows to show causal relationships.

**Step 9**. Check the logic to ensure each element outlined in steps 7 and 8 are causally linked to the next. Verify the logic model by interviewing program managers and program staff to ensure accuracy and completeness.

**Figure B.3: Sample Logic Model**



# Appendix C

# Factors to Consider in Planning for an Evaluation

| | |
|---|---|
| **Stage of program development at the time when evaluation takes place:** | |
| • Evaluation before the program is in operation<br>• Evaluation during the program's operation but before outcomes can be reliably determined<br>• Evaluation during the program's operation, when outcomes can be reliably determined | **Important because** the stage of program development will determine whether you can achieve the purpose(s). |
| **Previous history of evaluating this program:** | |
| • A formative evaluation has been conducted<br>• A process evaluation has been conducted<br>• An outcome/summative evaluation has been conducted | **Important because** it helps determine if previous evaluation tools and outcomes for the program can shape the current evaluation. |
| **Previous history of evaluating similar programs:** | |
| • Similar programs have been evaluated<br>• Similar programs have not been evaluated | **Important because** it lets you know if you can borrow tools from similar evaluations. |
| **Internal diversity of the program being evaluated:** | |
| • A program without diverse sub-programs<br>• A program with diverse sub-programs | **Important because** it helps determine if different approaches should be used for each sub-program. |
| **Uniqueness of the program:** | |
| • The program is unique (it has not operated elsewhere)<br>• The program is not unique | **Important because** it helps determine whether input and processes from other programs can be used for the intended new program. |
| **Outcome horizons:** | |
| • Short-term outcomes are of interest<br>• Medium-term outcomes are of interest<br>• Long-term outcomes are of interest | **Important because** it helps determine what tools, information and other resources you need. |
| **Range of program outcomes to be examined:** | |
| • Direct and/or indirect outcomes<br>• Intended and/or unintended outcomes | **Important because** it helps determine what tools, information and other resources you need. |
| **Availability of existing useful data that can be used in the evaluation:** | |
| • No such data, or little data, exists<br>• A moderate amount of such data exists<br>• A great deal of such data exists | **Important because** it helps determine whether to use existing data or generate new data. |
| **Stakeholder readiness:** | |
| • Stakeholders are largely ready<br>• Stakeholders are largely not ready | **Important because** it helps determine the effort required to increase stakeholder readiness, and how soon stakeholders can be fully engaged. |
| **Resources available for conducting the evaluation:** | |
| • Substantial resources are available<br>• Few resources are available | **Important because** it helps determine the evaluation's scope and complexity. |
| **Degree of urgency of the evaluation:** | |
| • The evaluation is urgent<br>• The evaluation is not urgent | **Important because** it helps determine the evaluation's scope, complexity and time line. |

## Appendix D

# Sample Informed Consent Form

I voluntarily agree to participate in the evaluation of [ABC Health program]. I understand that this evaluation is being conducted by [INSERT Name, Title, Organization], to improve the program.

I understand that the evaluation methods which may involve me are:

1. [The evaluator's] recorded observations of my participation with the program and its process; and/or

2. my completion of an evaluation questionnaire(s); and/or

3. my participation in a 30-60 minute interview.

I grant permission for the interview to be tape recorded and transcribed, and to be used only by [evaluator] for analysis of interview data. I grant permission for the evaluation data generated from the above methods to be published in an evaluation report to the funder, [insert name].

I understand that any identifiable information with regard to my name or personal information will not be listed in the report or any future publication(s).

If I have any questions about this evaluation or my role in it, I can contact [INSERT Name, position and contact phone number, address and/or e-mail address]

_____

Research Participant

_____

Date

*Adapted from Reference [26]*

# Appendix E

# Common Types of Data Collection Methods Used in Evaluations

Sources[68, 70]

| Description | Advantages | Disadvantages |
|---|---|---|
| **Survey, Questionnaire** | Refer to Module 5 (*Community Engagement and Communication*) in the Health Planner's Toolkit for description of advantages and disadvantages. | |
| **Focus Groups** | Refer to Module 5 (*Community Engagement and Communication*) in the Health Planner's Toolkit for description of advantages and disadvantages. | |
| **Face-to-Face Interviews** | Refer to Module 5 (*Community Engagement and Communication*) in the Health Planner's Toolkit for description of advantages and disadvantages. | |
| **Observation**<br>• Evaluator directly observes skills or behaviour. The purpose is to gather accurate information about how a program actually operates. | • view operations/processes of a program as they are actually occurring; and<br>• can adapt to events as they occur. | • can be difficult to interpret observed behaviors;<br>• can be complex to categorize observations;<br>• can influence behaviors of program participants; and<br>• can be expensive. |
| **Case Studies**<br>• Create a narrative to describe an activity or participant. The goal is to fully understand or depict clients' experiences in a program, and conduct comprehensive examination through cross comparison of cases. | • fully portrays clients' experience in program input, process and results; and<br>• a compelling means to portray program to outsiders. | • very time consuming to collect, organize and describe; and<br>• represents depth of information rather than breadth. |
| **Activity Logs**<br>• Staff record of day-to-day activities in program, e.g., topics covered, materials distributed, session format (lecture, discussion group, drop-in for example). | • low cost;<br>• can be developed or modified to meet evaluation needs; and<br>• easy for staff to complete. | • reporting detail and consistency of completing log data may vary among staff;<br>• requires analyzing written information in diaries, which may be cumbersome;<br>• changes in definition and kinds or types of data may make it difficult to compare data from different time periods; and<br>• some data may be confidential and may require special consent. |
| **Administrative Records**<br>• Refers to the data associated with the program's operations. Examples include financial (cost of materials, rentals, staffing; facility/equipment utilisation (location, use); personnel (assigned staff in terms of numbers, time); may include use of a Computerised Activity Reporting System (activities and staff time). | • low cost;<br>• easiest data to understand; and<br>• data usually exists. | • may be incomplete, inaccurate or inappropriately organized;<br>• not usually comparable to other organizations or programs; and<br>• limited to data currently being collected. |

*continued on next page...*

*continued from previous page...*

| Description | Advantages | Disadvantages |
|---|---|---|
| **Charts**<br>• Charts and records on individual participants. | • low cost; and<br>• easily available. | • some data may be confidential and may require consent;<br>• data may not be recorded consistently from chart to chart;<br>• analysing written information in charts may prove challenging; and<br>• data abstraction prone to errors; need to ensure people abstract data from chart in the same way. |
| **Registration Forms**<br>• Record of detailed participant personal data and other information such as how referred to program.<br>**Attendance Sheets**<br>• Sign-in sheets or staff-recorded. | • low cost;<br>• easily available; and<br>• can develop or modify to meet evaluation needs. | • some data may be confidential and may require consent; and<br>• changes in definition of terms and kinds/types of data may make it difficult to compare data from different time periods. |

# Appendix F

# Methods Worksheet

| Evaluation Questions | Data Collection Plan | | | | | |
|---|---|---|---|---|---|---|
| | **Does data exist?** ***Yes No*** | **Type of tool** | **Who can provide the data?** ***(Source)*** | **Design** | **How Many?** | **Timeframe** |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |

*Adapted from Reference* [68]

## Questions and Answers to Assist with Completion of Methods Worksheet

### Are all the data already available?

- When data are not readily available, it will be necessary to consider the tools that might be used to collect the data you need.

### What type of data collection tool would provide the data?

- There may be more than one tool suitable to use. Think about the quality of the data that a tool will produce. The tool should provide data which are as close to the truth as possible (validity). The tool should also give consistent answers if you ask the same person the same questions at different times (reliability).

### Who could provide the data, if asked?

- Identify the source(s) of the information such as program participants, program staff, stakeholders, others.

### What is the best design?

- The design will depend on whether all of the target group will provide data on an ongoing basis, a sample of the target group will provide data on an ongoing basis, or all or a sample of the target group will provide data at only one specific time or at several specific times.

### From how many people or things should data be collected?

- If the design uses a sample, a program evaluation specialist or epidemiologist can help determine how many participants should be included in order to ensure an accurate picture.

### What is the required timeframe for data collection?

- The timeframe will be the same as the program duration when the intention is to include all program participants involved and when data collection is ongoing. If the data is to be collected at a specific point in time, the timing will be based on the evaluation questions.

# Appendix G

# Evaluation Standards

**Summary of Program Evaluation Standards**
**of the Joint Committee for Educational Evaluation:**
**American Evaluation Association**

http://www.eval.org/EvaluationDocuments/progeval.html

## A. Utility Standards

*The utility standards are intended to ensure that an evaluation will serve the information needs of intended users.*

**U1 Stakeholder Identification** – Persons involved in or affected by the evaluation should be identified, so that their needs can be addressed.

**U2 Evaluator Credibility** – The persons conducting the evaluation should be both trustworthy and competent to perform the evaluation, so that the evaluation findings achieve maximum credibility and acceptance.

**U3 Information Scope and Selection** – Information collected should be broadly selected to address pertinent questions about the program and be responsive to the needs and interests of clients and other specified stakeholders.

**U4 Values Identification**–The perspectives, procedures, and rationale used to interpret the findings should be carefully described, so that the bases for value judgments are clear.

**U5 Report Clarity**–Evaluation reports should clearly describe the program being evaluated, including its context, and the purposes, procedures, and findings of the evaluation, so that essential information is provided and easily understood.

**U6 Report Timeliness and Dissemination** – Significant interim findings and evaluation reports should be disseminated to intended users, so that they can be used in a timely fashion.

**U7 Evaluation Impact** – Evaluations should be planned, conducted, and reported in ways that encourage follow-through by stakeholders, so that the likelihood that the evaluation will be used is increased.

## B. Feasibility Standards

*The feasibility standards are intended to ensure that an evaluation will be realistic, prudent, diplomatic, and frugal.*

**F1 Practical Procedures** – The evaluation procedures should be practical, to keep disruption to a minimum while needed information is obtained.

**F2 Political Viability** – The evaluation should be planned and conducted with anticipation of the different positions of various interest groups, so that their cooperation may be obtained, and so that possible attempts by any of these groups to curtail evaluation operations or to bias or misapply the results can be averted or counteracted.

**F3 Cost Effectiveness** – The evaluation should be efficient and produce information of sufficient value, so that the resources expended can be justified.

## C. Propriety Standards

*The propriety standards are intended to ensure that an evaluation will be conducted legally, ethically, and with due regard for the welfare of those involved in the evaluation, as well as those affected by its results.*

**P1 Service Orientation** – Evaluations should be designed to assist organizations to address and effectively serve the needs of the full range of targeted participants.

**P2 Formal Agreements** – Obligations of the formal parties to an evaluation (what is to be done, how, by whom, when) should be agreed to in writing, so that these parties are obligated to adhere to all conditions of the agreement or formally to renegotiate it.

**P3 Rights of Human Subjects** – Evaluations should be designed and conducted to respect and protect the rights and welfare of human subjects.

**P4 Human Interactions** – Evaluators should respect human dignity and worth in their interactions with other persons associated with an evaluation, so that participants are not threatened or harmed.

**P5 Complete and Fair Assessment** – The evaluation should be complete and fair in its examination and recording of strengths and weaknesses of the program being evaluated, so that strengths can be built upon and problem areas addressed.

**P6 Disclosure of Findings** – The formal parties to an evaluation should ensure that the full set of evaluation findings along with pertinent limitations are made accessible to the persons affected by the evaluation, and any others with expressed legal rights to receive the results.

**P7 Conflict of Interest** – Conflict of interest should be dealt with openly and honestly, so that it does not compromise the evaluation processes and results.

**P8 Fiscal Responsibility** – The evaluator's allocation and expenditure of resources should reflect sound accountability procedures and otherwise be prudent and ethically responsible, so that expenditures are accounted for and appropriate.

## D. Accuracy Standards

*The accuracy standards are intended to ensure that an evaluation will reveal and convey technically adequate information about the features that determine worth or merit of the program being evaluated.*

**A1 Program Documentation** – The program being evaluated should be described and documented clearly and accurately, so that the program is clearly identified.

**A2 Context Analysis** – The context in which the program exists should be examined in enough detail, so that its likely influences on the program can be identified.

**A3 Described Purposes and Procedures** – The purposes and procedures of the evaluation should be monitored and described in enough detail, so that they can be identified and assessed.

**A4 Defensible Information Sources** – The sources of information used in a program evaluation should be described in enough detail, so that the adequacy of the information can be assessed.

**A5 Valid Information** – The information gathering procedures should be chosen or developed and then implemented so that they will assure that the interpretation arrived at is valid for the intended use.

**A6 Reliable Information** – The information gathering procedures should be chosen or developed and then implemented so that they will assure that the information obtained is sufficiently reliable for the intended use.

**A7 Systematic Information** – The information collected, processed, and reported in an evaluation should be systematically reviewed and any errors found should be corrected.

**A8 Analysis of Quantitative Information** – Quantitative information in an evaluation should be appropriately and systematically analyzed so that evaluation questions are effectively answered.

**A9 Analysis of Qualitative Information** – Qualitative information in an evaluation should be appropriately and systematically analyzed so that evaluation questions are effectively answered.

**A10 Justified Conclusions** – The conclusions reached in an evaluation should be explicitly justified, so that stakeholders can assess them.

**A11 Impartial Reporting** – Reporting procedures should guard against distortion caused by personal feelings and biases of any party to the evaluation, so that evaluation reports fairly reflect the evaluation findings.

**A12 Meta-evaluation** – The evaluation itself should be formatively and summatively evaluated against these and other pertinent standards, so that its conduct is appropriately guided and, on completion, stakeholders can closely examine its strengths and weaknesses.

# Appendix H

# Factors in Building an Evaluation Culture

### Structural factors:

- the appointment of an internal evaluator; and
- the location of evaluation within the executive area.

### Procedural factors:

- formal requirement to undertake and use evaluation in projects;
- reflection on key learnings at the end of major projects;
- inclusion of evaluation in corporate and business planning processes;
- the use of steering committees to oversee major evaluation projects; and
- the use of monitoring and performance indicators by senior executive staff.

### Philosophical and attitudinal factors:

- an agreed vision that underpins the importance of evaluation;
- evaluation of programs, processes, products – not people; and
- interactive evaluations as a way of increasing ownership of findings and building capacity.[72]

# Appendix I

# Other Sources of Information on Evaluation

## 1. The Canadian Evaluation Society

The Canadian Evaluation Society (CES) is a membership-based organization representing evaluators across Canada. Twice yearly it publishes the *Canadian Journal of Program Evaluation*, sent to members of CES. Archived editions are also available to members. CES also provides access to unpublished documents (also referred to as Grey Literature) which may be of interest to evaluators. This grey literature is available to CES members and non-members at the CES web site. As well, the website has copies of the CES newsletter (available to members and non-members) that contains links to interesting evaluation articles. The CES web site (accessed December 27, 2007) is at: http://www.evaluationcanada.ca/site.cgi?s=1 (accessed on December 27, 2007).

## 2. The Evaluation Exchange, Harvard Family Research Project, Harvard Graduate School of Education

*The Evaluation Exchange* is a free newsletter on evaluation issues. Its primary focus is evaluation in educational settings but its articles are often relevant to health service evaluation. To subscribe, please go to http://www.gse.harvard.edu/hfrp/eval.html (accessed on December 27, 2007).

## 3. Program Evaluation Methods, Treasury Board of Canada

In 1998 the Treasury Board of Canada produced *Measurement and Attribution of Program Results (Third Edition)*, a comprehensive guide of particular use to those interested in summative (outcome) evaluation. The document is found at: http://www.tbs-sct.gc.ca/eval/pubs/meth/pem-mep_e.asp (accessed on December 27, 2007).

## 4. Encyclopedia of Evaluation, Sandra Mathison, Sage Publications Inc., 2005

Organized alphabetically in traditional encyclopedia format, this 520 page book is particularly helpful in providing definitions of specialized terms used in evaluation.

## 5. Health Planner's Toolkit, Ontario Ministry of Health and Long-Term Care

This toolkit comprises seven modules, each containing information relevant to evaluation and its context. These modules are available at

http://www.health.gov.on.ca/transformation/providers/information/im_resources.html#health
(accessed on December 27, 2007).

# Notes

Ontario