Quality Improvement & Innovation Partnership

Advancing Improvement in Primary Healthcare in Ontario

PROGRAM EVALUATION
RESOURCE GUIDE

May 2010

Amended December 2010

Ontario

Acknowledgements

Table of Contents

Quality Improvement & Innovation Partnership

A Brief Guide to Program Evaluation for Primary Healthcare Teams

Purpose

Program evaluation involves a complex set of activities that, if carried out well, can provide information helpful to improving the health care programs that primary healthcare teams deliver[1]. Many programs require only targeted evaluation to answer a question about the program and much of that information may exist in administrative records used to monitor the program. Less frequently, more comprehensive program evaluations are needed that attempt to examine multiple, complex, interrelated questions: the extent to which the program is reaching the people it targets, if the program is being delivered as was intended when it was planned, the extent to which outcomes intended are being achieved and unintended ones avoided and whether the costs of the program are reasonable given the benefits achieved. This guide differentiates between program monitoring, brief simple evaluations that attempt to answer one or two questions and comprehensive program evaluation. It attempts to help you decide whether you need a comprehensive evaluation and, if you do, how to ensure that the results obtained are seen as credible and will lead to action.

What is Program Evaluation?

Program evaluation is a process that answers important questions about how a program is functioning by systematically gathering pertinent new information and combining it with existing information. It uses a wide variety of research tools to address questions raised, including surveys, focus groups, and collection of information from medical charts or laboratory reports. It may also adopt a formal research design (randomized controlled trial, cohort study, etc.). When done well, program evaluation can improve the effectiveness and efficiency of the program. Program evaluation can be limited, by focusing on one or two important questions, or comprehensive, by exploring many different questions about the program such as how well it targets its clients and recruits them into the program, how efficiently and effectively it delivers its services, whether the program has all the components needed to be successful, and to what extent it achieves the outcomes that are its goal.

---

[1] The term program is used for the set of activities that are carried out to accomplish a defined set of goals related to improving the well-being of people, groups, organizations or communities. A Primary healthcare team may be involved in delivering several programs (e.g. hypertension management program, Healthy You program)

How Does it Differ from Monitoring a Program using Administrative Data?

Program evaluation is time-limited while program monitoring is an on-going activity that aims to provide information to assist in managing the program. Most programs try to pick two to four key indicators that are routinely collected to provide on-going feedback about the program's operation. For example, a Hypertension Clinic program may monitor, on a monthly or weekly basis, the number of people seen, the percentage of on-going patients seen whose blood pressure readings remained within the normal range during the period monitored, the number and percentage of no-shows/cancellations among new referrals and on-going patients. A negative change on one of these indicators at one measurement period, in itself, is unlikely to reflect a problem. But, monitoring trends over time or rapid changes in an indicator may raise questions that deserve further exploration and may lead to program evaluation. In program evaluation, new information is gathered that is not readily collected as a by-product of the routine activities of the program. The questions posed require new information to be gathered. For example, "What are the outcomes of all the people who ever attended the program last year, including people who dropped out of the program?" cannot be answered by routine monitoring of key program indicators.

The manager of a program is usually tasked with watching the indicators about her/his program and may discuss them with the program staff to try to understand what is behind changes in an indicator. The stakeholders involved in an evaluation of a program are likely to be other members of the primary healthcare team staff, the patients attending or eligible to attend the program, the primary healthcare team board and the Ministry of Health and Long-Term Care, as well as those managing or working in the program. Key stakeholders may have questions about the program that current administrative data cannot address.

Examples of Program Monitoring
A 'Healthy You' program decided that one key question for the program to monitor was "Are we reaching the right people, that is, people who are likely to benefit from the program?" They decided to establish clear entry criteria and monitor program entry to ensure that criteria creep did not occur. They chose to only target people whose body mass index (BMI) at program entry was greater than 26. They now collect BMI information at program entry and monitor the distribution of BMIs of new enrolees.
The Respiratory Clinic monitors the proportion of patients who have COPD who have not required an intervention with additional medication or hospitalization in the previous month. They compare trends over time, looking at improvement. Since, in many of their patients, the exacerbation of breathing difficulties is seasonal, they compare two month blocks from year to year.

Example of program monitoring leading to an intervention whose success is shown by continued monitoring

The Mental Health Program of a Family Health Team decided to monitor the size of its waiting list and the number of days people waited to receive an initial call from the program for an appointment. They were disheartened to find that the number of people on the waiting list continued to grow, along with the time to first contact. Doctors and nurse practitioners (NPs) who had talked with the patients and recommended they would benefit from short term counselling added the patients' names to the wait list. In discussing the wait list problem, the counsellors noted that some people did not choose to enter counselling once contacted for an interview, dropped out quickly or really needed referral to other resources in the community. They brainstormed about ways to increase the likelihood that people on the wait list were appropriate for their service and were ready to engage in the counselling process. They developed a brief form which the doctors and NPs were asked to give people they had previously placed on the waiting list. It asked the potential client to describe their current problem and explain why they wanted counselling. In order to be seen, the person had to return the form which was given to them by the doctor or NP (A staff member was available to help those with literacy problems complete the form). Now the counsellors could screen cases and refer those whose problems were too severe or related to addictions or sexual abuse to other community resources. They found that if people completed the form and returned it, they were indeed interested in getting help. All of the indicators that they used to monitor the program improved. Wait times dropped as did refusals, no shows and cancelled appointments. More people received the help they needed more quickly.

Example of program monitoring leading to an evaluation which informs the development of an intervention whose success is shown by continued monitoring

A Family Health Team decided to use their Hypertension Clinic nurse to do repeat blood pressure measurements for patients whose latest blood pressure reading was high. This was seen as a measure to save doctors' time and to get the patients acquainted with the Clinic so that they would return for monitoring of how well the drug regimen was controlling their hypertension. Patients were asked to come to the clinic two more times to ensure that they were truly hypertensive. If the patient continued to have readings in the hypertensive range, they were referred back to the doctor to initiate treatment. The success of this initiative was evaluated by keeping track of the number of patients seen and the number of patients who were identified as hypertensive. They monitored the length of time it took before the newly diagnosed hypertensive patient re-entered the Hypertension Clinic for continued monitoring. They noticed the median length of time before these patients returned to using the clinic was 6 weeks after the diagnosis had been confirmed.

They decided to explore why it took so long for patients to begin treatment by doing some limited program evaluation. Since there could be many reasons why treatment was delayed, they chose to do semi-structured telephone interviews. Patients were asked about their satisfaction with their experience in the Hypertension Clinic, the convenience of its hours of operation, the extent to which their questions about their condition had been answered, the ease of getting a subsequent appointment with their doctor and any other things the patients wished to share. The answers of those who took more than 4 weeks to come back to the Hypertension Clinic on medication were compared with people who returned rapidly. Patients reported that they liked the clinic and found the nurse informative about hypertension. The problem seemed to be in scheduling a follow–up appointment with their doctor. As the appointment was seen as not urgent, there was often a delay of several weeks before they were seen. Some patients forgot to make an appointment with their doctor and only returned to the doctor when an urgent problem arose. Some also took considerable time to get their prescription filled. Sometimes they returned to the doctor for another appointment because they were unsure of their drug regimen.

Now, using a medical directive, the nurse in the Hypertension Clinic counsels the patient and begins treatment once the diagnosis is clear, reducing the time from diagnosis to active treatment and on-going management.

How Do I Decide Whether a Program should be Comprehensively Evaluated?

There are several reasons that you might decide that one of your programs, or a program that you are planning, requires a comprehensive program evaluation rather than a brief examination of one or two issues.

1) *The program is new and novel.* There is no information about whether it will work as you expect. For example, you may be concerned about the number of families whose financial difficulties are spilling over into relationship problems. You would like to forestall these consequences with a new couples' group program where 4-6 couples talk about managing feelings of loss and anger with the help of a therapist. At least one of each couple will have lost their job recently to qualify for the group. You can't find any descriptions of programs like it in the literature. But, the idea "makes sense" to you. Some of your staff say that people won't come to it if offered in a group. Others are sceptical that such a group would be helpful. Two staff members are sure it will work.

2) *You have chosen to substantially alter an existing program* by choosing a very different target group for the program, and/ or making major changes in the program or the types of workers who deliver the program. For example, you decide to target teenagers who are just beginning to smoke for a smoking cessation program that has been reported in the literature as producing a high and sustained quit rate among adult smokers. You also altered some of the messages and materials in hopes of making the program more successful with teenagers. There is uncertainty that the altered program will be as effective as the original one because of these changes.

3) *The program is very expensive* and you wonder if it is worth having or perhaps should be replaced by other less expensive programs. The primary healthcare team has an ultrasound machine that is used for a variety of purposes. The equipment is now a bit dated and the person trained on the machine is moving to another community. You wonder whether this expensive program should be part of your services. Other patients in your community currently travel out of town for ultrasound examinations. The staff is divided as to whether this program should be kept or whether several other less expensive programs that would benefit more (but different) patients should be mounted.

4) *Major decisions need to be taken about the program*. There are major questions about the program that cannot be answered by existing monitoring tools or limited questions. There is disagreement among the stakeholders. For example, your primary healthcare team has had a smoking cessation program for several years and recently, you note that fewer people are using it. You are not sure why. But, you wonder if it is time to close down this program and offer a different kind of program with the resources it currently uses. There are strong proponents who believe the program is highly effective and should be maintained, while others think it should be abandoned and say that they think the relapse rate among former smokers after leaving the program is high. Here, comprehensive evaluation is needed to address the issues raised.

Planning for Comprehensive Program Evaluation

Careful planning for program evaluation is crucial to its success and eventual usefulness. Overlooking aspects of the planning process and neglecting to think through the entire evaluation ahead of time can lead to serious problems later. It can jeopardize the usefulness of the evaluation by producing findings that are inaccurate and misleading. Even valid, trustworthy findings may be disregarded by key stakeholders, or they may try to sabotage the evaluation in subtle ways if careful attention is not given to their concerns during the planning stage. Thus, whoever is responsible for overseeing an evaluation needs to spend a significant amount of time in negotiation, discussion, briefings, prioritizing and presenting information about the evaluation during the planning phase.

Questions to Ask While Planning a Comprehensive Program Evaluation

1)   *What are the goals and objectives of the program and how does it meet them?*

Program evaluation requires clear delineation of the program, its objectives, target population, components, outputs and the intended outcomes and impacts. The first step is to develop a logic model to describe key features of the program and how it is supposed to work.

Logic models[2] were developed as tools to clarify the nature of a program and its intended effects. A logic model of the program is a useful planning tool that provides a diagrammatic description of a program depicting its goals and objectives, the target group for the program, the component activities needed to accomplish the program's goals, their outputs (countable by-products of each component), short and long term outcomes (direct results or accomplishments) and impacts (effects for which the program can claim only partial responsibility). It can help identify all the tasks that need to be accomplished to implement the program. It allows examination of assumptions about how a new program (or changes in a program) will produce the effects intended. The logic model may reveal that no activities have been planned that would allow the program to have some of the effects intended. Conversely, it may point out that some planned activities are not well linked to program objectives.

A logic model provides a program overview that helps everyone understand how the program works and the assumptions that underlie the delivery of the program. It is useful in program planning and review as it helps ensure the goals are clear and that the objectives and activities undertaken allow the program's mission to be achieved.

2)   *What are the goals and objectives of this evaluation?*

---

[2] There are now numerous resources on the Internet that can help people learn how to complete a logic model. For example, please see http://www.qiip.ca/user_files/logicmodelrg.pdf.

The range of questions that can be asked about a program and be addressed in program evaluation is broad. It is important that the evaluation objectives are clearly stated, so that the scope and intent of the evaluation are clear. Again, making sure a logic model is available for the program to be evaluated is important as it facilitates examining its features. You need to agree on the most pressing questions to address. You may want to know: "Is the program being delivered as planned? Is it having the effects expected? Are there any unintended, negative effects? How cost-effective is it?"

3)   Will the evaluation be done using existing staff or will an external evaluator be retained?

 A decision must be taken about who will conduct the evaluation. Program evaluation can be carried out by outside evaluators or staff. The resources you have available, the time frame for the evaluation and your comfort level with performing evaluations will influence your decision.

There are advantages and disadvantages to each option. Having an external evaluator can ensure that the person leading the evaluation has the expertise to conduct program evaluation. The primary disadvantage to using outside consultants is cost. It may also take the evaluator time to understand the program as well as internal people do. However, a fresh pair of eyes looking at the program can provide objectivity by offering an outsider's view. The main advantage is that this person is knowledgeable about evaluation issues and techniques. Choosing someone who has done previous evaluations also allows you to do some reference checking. Ask how they learned their evaluation skills and whether they are a member of the Canadian Evaluation Society.

Depending on the size and resources of an organization, staff may perform program evaluations. If program evaluation becomes an in-house activity, the other work responsibilities of staff members conducting the evaluation must be reduced. Large teams may have well-trained staff that devote all or a major portion of their time to evaluation activities.

 If staff are inexperienced in research methods and measurement issues and/or lack the skills to develop consensus about the key issues the evaluation should address, the evaluation will suffer. Staff may not have the time to perform an evaluation in addition to their regular responsibilities. You need to weigh the pros and cons and decide what will be the most effective approach for you.

Often, your budget is the major factor restricting choice. Using a combination of internal and external evaluators may allow you to build internal evaluator skills among interested staff. For example, you might retain an outside evaluator to help you set up the evaluation and monitor it, but then use staff members to implement the study. The staff member charged with the task of overall coordination of the evaluation should be able to consult with the external evaluator regularly and as questions arise. However, this option assumes that staff exist who can be freed up to take on the evaluation and have the expertise to guide and monitor the evaluation on a day-to-day basis, oversee data collection and analysis to ensure that it is done properly and can write interim and final evaluation reports. Dividing up these responsibilities in creative ways can help achieve a cost-effective evaluation.

4)   *Who are the stakeholders and audiences of the program?*

Many questions can be posed about programs. Often, the type of questions we are interested in depends on who we are, or in the language of program evaluation, what type of stakeholder we are. Stakeholders are people with an interest in the program and can include funding bodies, program administrators, staff, patients and their families, etc. Audiences are groups that are interested in the answers to questions posed but are not directly affected by the questions or their answers (e.g., the community, Health Canada, health professional associations). It is useful to have representatives from major stakeholder groups represented when discussing the plans for an evaluation.

5)   *What roles will stakeholders play in the evaluation?*

The roles that various stakeholders expect to play need to be clarified as part of evaluation planning. When, how and to what extent they expect to be involved in the project should be discussed and agreed upon. Different stakeholders may have different perspectives on a program and different concerns about it. They may not agree on what the most important outcomes of a program are or see the need for specific program components. Some stakeholders may be as concerned about possible negative unintended effects of a program as they are in intended effects. Unanimous agreement about which questions are most important is rarely seen; but, addressing only the concerns of one group will make it difficult to convince everyone that the evaluation is fair and balanced.

If stakeholders disagree about what they consider important to evaluate, time should be spent to build consensus or, at least, to acknowledge that the limited scope of the evaluation proposed cannot address some issues important to one or more stakeholder groups. Engaging stakeholders in priority setting and or consensus-building exercises can be helpful.

Remember that evaluation is a political activity. Not acknowledging stakeholder interests and ensuring that key stakeholders agree the evaluation should proceed may spell disaster for the evaluation. The concerns about the evaluation raised by stakeholders should be treated seriously and addressed in a transparent, straightforward fashion.

Stakeholder concerns, when examined carefully, usually highlight unresolved problems and issues that may hamper the evaluation if not addressed. Listening carefully can lead to improvements in the design of the evaluation and allow the evaluation to be implemented successfully.

Who will be affected by answering the question(s) an evaluation is to address and how they might be affected needs to be thought through carefully before the evaluation proceeds.

6)   *Why is this evaluation to be done now?*

The reasons for the evaluation should be transparent. If there are hidden agendas (e.g., the decision has already been taken to make certain changes and the evaluation will be used as a pretext for the changes) or important questions are left unanswered, the evaluation will be discredited.

Quality Improvement & Innovation Partnership

7)   How will we answer our questions in a transparent and unbiased fashion?

A. *What methods will be used?* There is no perfect evaluation design. Evaluations must balance timing, methodological strength, values of stakeholders and cost. The evaluation methods chosen should fit the values of key stakeholders. If key stakeholders think that only a quantitative design with control groups can provide high quality evidence to address the questions, the results of a carefully planned qualitative study that involves participant observation may be ignored. If only high-quality evidence is acceptable, information gathered through quick but less reliable means (which some may see as sufficient to answer the question) will be ignored while evidence which has been gathered over a longer period with more sophisticated measurement instruments will be accepted, even if they arrive at the same conclusions. Similarly, if a key stakeholder (e.g., the funder) thinks that the evaluation plan is too complex and time-consuming; you may be asked to scale down evaluation activities, even if this sacrifices some methodological rigour. The design trade-offs made must be clear to and accepted by all stakeholders.

Quantitative methods and measures are useful to assess outcomes, particularly for comparisons across time or by sub-programs. Quantitative methods such as randomized controlled trials and quasi-experiments are useful in examining the magnitude of the effects seen and comparing them with no service or more limited services. Surveys can also be used to collect data that allows quantification of outcomes (e.g., the proportion of clients who are satisfied with services) although a comparison group may be lacking.

Qualitative research methods, such as the use of in-depth interviews, focus groups and participant observation can be useful tools in evaluations. They allow new insights to emerge from the data. They are particularly useful when not enough is known about program effects to choose quantitative measures well. For example, if it is not clear what it is about the services they receive clients value the most or in what ways their ability to cope with changes and the impact of these changes, focus groups of former clients asking open ended questions about such issues may provide information that can later be used to develop quantitative measures of these outcomes. Qualitative measures are also helpful when unanticipated effects are of interest. Such methods may be especially useful during program implementation because they often allow quick feedback to program managers who are eager to make corrections early in the implementation process. They can help us understand what is happening as well as provide ideas regarding improvement.

More evaluation designs are using mixed methods to deal with the complex purposes and questions addressed by program evaluation. This approach combines qualitative and quantitative research methods and uses them in a complementary fashion.

B. *What measures or indicators will we use?* If a quantitative design has been chosen, the next question is whether there are measurement techniques available to capture key variables of interest. The terms indicators and measures are frequently used interchangeably, although measures often refer to scales (a collection of questions whose answers are summed to obtain the result) while indicators often refer to a discrete countable metric (e.g., the number of people seen) or something that can be answered in a yes/no fashion (e.g., had to visit the emergency room). Process indicators or measures refer to the quantity and quality of the program activities you are delivering, while the quantity and quality of the outcomes that your program is achieving are commonly referred to as outcome indicators.

It is preferable to use existing measures that are known to be reliable and valid. If such indicators do not exist, they must either be developed, the question must be discarded or re-focussed, or a different approach to the evaluation (i.e., qualitative methods) must be adopted in order for it to proceed.

The outcome indicators most frequently used in program evaluation have been knowledge gained, attitudes changed, and satisfaction with services. These measures provide a somewhat restricted view of how well a program accomplished its objectives because they do not address changes in behaviour, health status or functional capacity. Operational definitions are needed for the many different outcomes and impacts sought by the program aside from knowledge and attitude change or immediate assistance provided. This task is not an easy one. Yet, we should not be satisfied with substituting easy-to-measure outcomes for important outcomes.

8)   What steps need to be taken to ensure privacy of respondents and confidentiality of responses?

Ensuring that the evaluation methods and measures respects the privacy of individuals and are carried out in an ethical fashion that ensures confidentiality of participants' responses is also part of managing an evaluation. Legislation requires us to tell patients about how information supplied will be used.[3] Having a local research ethics board review the planned evaluation activities is one way of ensuring that the necessary safeguards are in place.

9)   How quickly are the answers required?

If the results of the evaluation cannot be available in a timely fashion to support good decision-making, the evaluation may not be worth doing. The scale of the evaluation may have to be changed and some compromises made in the quality of the evidence that can be brought to bear on the issues of concern if decisions must be taken in the next few months. Such trade-offs must be clear and everyone must agree that the resultant information will still be important to have. Alternatively, such discussions may lead to a greater appreciation for the need to delay a decision until important information can be made available to inform decision-making.

10)   What resources do we have for the evaluation?

An evaluation requires resources (both people and monetary). If these resources are lacking, the evaluation may be compromised. Ascertaining beforehand that sufficient resources will be available to support an evaluation is important to its successful execution. Most evaluations report that resources were a constraint. Evaluation may require staff time in completing questionnaires or audits as well as time from a paid external evaluator. From 1-10% of the cost of the program is often needed to evaluate it. Lower costs (proportionately) are possible when a large program is being evaluated and when monitoring systems already exist.

11)   How will we interpret and use the results of this evaluation?

---

[3] Readers should visit the Ontario Ministry of Health and Long Term Care website for information about the Quality of Care Protection Act that is Schedule B of the Health Information Protection Act. http://www.e-laws.gov.on.ca/DBLaws/Statutes/English/04q03_e.htm

Quality Improvement & Innovation Partnership

A very important part of evaluation planning is rehearsing possible answers to evaluation questions with all stakeholders and examining their implications for action. This procedure can identify whether or not the evaluation should be undertaken (do not undertake evaluations that have potential results you are unprepared to hear).   Rehearsing possible answers and their implications also provides a safeguard against producing ambiguous results than can be explained in numerous ways. If the evaluation is poorly framed, leaves out issues seen as crucial to key stakeholders or can be interpreted in a variety of ways, it is unlikely to be useful. Thinking about all plausible answers (not just the ones hoped for) may lead to rejection or refinement of the questions or to gathering some additional information. A multi-method approach that provides evidence from a variety of sources may be adopted. Such an approach is more likely to collect information that helps rule out alternative explanations for the findings.

Rehearsing possible outcomes and their implications can prepare the stakeholders to take action on findings. Most importantly, it ensures that key stakeholders are not surprised by the findings and key implications. From the early stages of the project onward, they are aware that the evaluation could produce such information.

How the information will be used should be clear. Even maintaining the program as it is currently structured and carried out is an action that has consequences. If the evaluation is being done to support decision-making, the decisions to be taken must be clear.

Ensuring Useful Evaluation

Evaluation of programs must be done in an ethical fashion and be sensitive to political issues. Usefulness of an evaluation is enhanced when it produces timely information that can inform the decision-making process. The activities chosen should be relevant and credible; they should lead to findings that are objective and presented in a clear and balanced way, ensuring that the conclusion and recommendations link directly to the findings. Evaluations must be sensitive to the differing constituencies that will be affected by the findings and their implications. They must be affordable.

Focussed evaluations often are useful to answer questions raised about a program through monitoring for program management purposes. Usually they address questions of concern to the program staff and manager.

Comprehensive program evaluation involves a wider range of people and tries to build consensus about what questions the evaluation should address. Reports about comprehensive evaluation should include information on both processes and outcomes to be maximally useful. Without information about what was done, how the findings produced were achieved is not clear. Useful new approaches may be discarded as not helpful when in fact they were never implemented as planned. As well, important program components needed to achieve the results obtained may not be understood if contextual information is lacking.

Conclusions

Program evaluation has matured as a discipline in the last 50 years. Generating information about a program using a wide range of valid research methods is not good enough. The procedures used should ensure that the information is also used to guide and improve programs. New approaches to program evaluation have been developed that are more inclusive and involve key stakeholders from its very inception.

Program evaluation's image has changed from something dreaded and imposed to something important to continued program development and accountability. It is rare that a program is scrapped based on its evaluation (unless it was a pilot project). Rather, the information gained is often used to make changes that improve programs. Such incremental changes are much more common than program closure.

Quality Improvement & Innovation Partnership

Web-based Further Reading about Program Evaluation

Canadian Evaluation Society website has useful materials as well as when they will have workshops on different program evaluation topics. http://www.evaluationcanada.ca/site.cgi?s=1
They have also published **Program Evaluation Standards**.
http://www.evaluationcanada.ca/site.cgi?s=6&ss=10&_lang=en

The Treasury Board of Canada is responsible for federal program evaluation and has published **Program Evaluation Methods: Measurement and Attribution of Program Results**, **3rd edition**, on line at http://www.tbs-sct.gc.ca/eval/pubs/meth/pem-mep_e.pdf. At the end of each chapter, references to books and articles about program evaluation are provided. There is also a useful glossary of terms in appendix 2 and reference lists in appendix 3 and 4.

The free encyclopedia Wikipedia has useful information and also links to other sites that have program evaluation materials. **Program Evaluation**. http://en.wikipedia.org/wiki/Program_evaluation

McNamara, Carter. 2008   Basic Guide to Program Evaluation
http://managementhelp.org/evaluatn/fnl_eval.htm

(current as of July, 2009)

Quality Improvement & Innovation Partnership