Evaluation Primer

H. Stuart, PhD, Queens University
October 2007
Draft 3

Prologue

Concurrent Disorder and Schizophrenia: A National Awareness Strategy

There are limited resources and information on concurrent substance use disorders and schizophrenia, particularly from a Canadian perspective. Access to information about treatment programs and health professionals across the country is not centralized or coordinated. Given the complexities of concurrent schizophrenia and substance use disorder, from a treatment and diagnostic perspective, the need for centralized, up to-date information is critical.

In 2005, the Schizophrenia Society of Canada received a grant from Health Canada as part of a national awareness strategy to create resources and information on concurrent disorders specific to schizophrenia and substance use disorder for individuals (youth and adult) and families affected by the illness and professionals within the mental health sector. Information produced from this project is contained on a website: www.schizophreniaandsubstanceuse.ca

This primer contributes to the broader initiative by providing an evaluation framework that can be used by practitioners to assess the effectiveness of the services and programs they offer and, in so doing, build best practices for the field.

National Advisory Panel Members

The following individuals served as advisors to the national awareness strategy:

Barbara Beckett - Assistant Director, Institute of Neurosciences, Mental Health and Addictions, Canadian Institutes of Health Research, Ottawa, Canada BBeckett@cihr-irsc.gc.ca

David Boyce – Director of Concurrent Disorders Program, Georgainwood Mental Health Centre, 500 Church Street, Penetanguishene, ON. L9M 1G3 Canada
DBoyce@mhcp.on.ca

Sonia Chehil – Adolescent Psychiatry- Dalhousie University, 5909 Veteran's Memorial Lane, 8th Floor ,Abbie J. Lane Memorial Building, QEII Health Sciences Centre,Halifax, Nova Scotia, B3H 2E2
Canada Sonia.Chehil@Dal.Ca

Cathy Croucher – Family Counsellor – Early Psychosis Program – Waterford Hospital, Health Care Corporation Waterford Bridge Road, St. John's, NF. A1E 4J8 Canada
Cathy.Croucher@hccsj.nl.ca

Barry Fogg – Mental Health Services Development Specialist, Winnipeg Regional Health Authority, 4-189 Evanson Street, Winnipeg, MB R3G 0N9 Canada
BFogg@wrha.mb.ca

Stan Kutcher - Sun Life Chair in Adolescent Mental Health Department of Psychiatry Dalhousie University 5850 University Avenue P.O. Box 9700 Halifax, Nova Scotia Canada B3K 6R8
Stan.Kutcher@Dal.Ca

Juan Carlos Negrete – Addictions Unit, McGill University Health Centre, 1650, av. Cedar, Montréal, QC, H3G 1A4 Canada juan.negrete@mcgill.ca

Caroline O'Grady – Advanced Practice Nurse, Centre for Addiction and Mental Health, 33 Russell Street, Toronto, ON M5S 2S1 Canada Caroline_O'Grady@camh.net

Patrick Smith – Senior Advisor, Mental Health and Addictions, British Columbia Provincial Health Service Authority, Riverview Hospital, 2601 Lougheed Highway, Coquitlam, BC V3C 4J2 Canada psmith2@bcmhs.bc.ca

Shari McKee – Clinical Psychologist, Concurrent Disorders Program, Georgainwood Mental Health Centre, 500 Church Street, Penetanguishene, ON. L9M 1G3
SMCKEE@mhcp.on.ca

Laura Panteluk / Arlene MacLennan – Manitoba College of Psychiatric Nurses, 1854 Portage Ave, Winnipeg, MB R3J 0G9 Canada lpanteluk@crpnm.mb.ca;
amaclennan@wrha.mb.ca

Greg Purvis - President, Atlantic Association of Addictions Executives, P.O. Box 359, 199 Elliot

Street, Pictou, N.S. B0K 1H0 Canada
Greg.Purvis@pcha.nshealth.ca

Sharon Scott – National Coordinator, National Network
of First Episode Families, Box 395, J-1631 St. Mary's
Road, Winnipeg, MB. R2N 1Z4 slsm@rainyday.ca

Wayne Skinner – Director of Concurrent Disorders
Program, Centre for Addiction and Mental Health, 33
Russell Street, Toronto, ON M5S 2S1 Canada
Wayne_Skinner@camh.net

Chris Summerville – Executive Director, Schizophrenia
Society of Manitoba Chris@mss.mb.ca

Greg Zed – President, Schizophrenia Society of New
Brunswick greg.zed@gnb.ca

CONTENTS

Overview

Now, more than ever, it is important to evaluate mental health programs to ensure that clients receive services that are consistent with best practices in the field, that meet their needs, and result in positive therapeutic outcomes.

This primer is meant for people who have little or no prior experience with program evaluation but who would like to begin evaluating their own programs or participate as an advisory member in an evaluation project. Chapter I will take you through the five main steps in developing and conducting a simple quantitative evaluation:

1. Getting started – figuring out when to evaluate
2. Understanding how the program to be evaluated is supposed to work
3. Developing a strategy to assess whether the program is working the way it should
4. Determining what data collection tools to use to evaluate program functioning
5. Developing usable results

Chapter 2 will illustrate these concepts with real-life examples of evaluations of programs for people with concurrent disorders. Examples will be drawn from the existing scientific literature.

The Appendix contains auxiliary materials that may be helpful for individuals who are planning to evaluate a concurrent disorders program. It includes worksheets, examples of survey questions, scales, and other auxiliary materials that readers may need to assist them in conducting and using evaluation results. A reference list of further readers will be included here to direct readers to more advanced materials.

Chapter 1 - THE BASICS

1 Getting started – figuring out when to evaluate

1.1 Why Evaluate?

Evaluations, when done correctly, take time and resources. Even simple evaluations need to be well thought out and planned. Evaluations are done in the spirit of improvement. If there is no action to be taken as a result of the evaluation, then it is likely not worth pursuing. Thus, the first question to ask when considering an evaluation is: what improvements could be made once the results are in? All evaluations, even those that are mandated by funders to demonstrate specific outcomes have been achieved, should be taken with the goal of program improvement in mind.

First and foremost, evaluation is a management tool (rather than a scientific endeavor), undertaken to answer specific questions about the relevance, functioning, or performance of programs and to guide management decisions. Secondly, evaluation fulfills an important accountability function; demonstrating to funders and to those who use the program, that it is fulfilling its mandate in ethically and therapeutically appropriate ways. Most importantly, evaluation is a quality assessment tool to improve the quality of services provided and ensure that they adhere to best practice standards.

1.2 The Nature of Evaluation

Evaluation involves making a judgment about the value or worth of a program (or program activity) based on information that has been systematically collected.

1.3 Who Should be Involved?

The starting point for any evaluation is determining the stakeholders that should be involved in the planning, execution, interpretation, and use of the results. In particular, it is important to create buy-in from the intended users of the evaluation. If they are not integrally involved in the process, then the chances of the results being used will be slim.

To identify appropriate stakeholders, it is important to include individuals who stand to gain from the evaluation. In mental health programming, clients and family members are the key stakeholders in the majority of evaluations, since they have the most to gain from program improvements and the most to lose if programs do not work effectively.

Evaluations can create considerable tension for program staff, who may fear that the results will demonstrate they are not working hard enough, or that the program is not effective. They may be asked to change the way in which they deliver services as a result of the evaluation. Also, they may be asked to do added work to help in the evaluation effort by providing data or helping to coordinate data collection from clients. Thus, staff may feel they have a lot to lose. Therefore, staff are a second key stakeholder group.

Program decision makers or managers are a third important stakeholder group. Not only are they in a position to facilitate the evaluation process, they will set the tone for the evaluation and determine how the results will be used to improve program functioning. They may also feel threatened by an evaluation, particularly if it is externally driven for accountability purposes.

Finally, mental health programs maintain many ties to other programs within the community.

Representatives from these programs may also have an interest in the results of the evaluation, particularly if they feel that program changes may alter the amount or quality of services they have been receiving.

1.4 What to do with so many stakeholders?

It is not difficult to identify a number of stakeholders who should be involved in mental health program evaluations. Thus, a key challenge is how to organize the various stakeholder groups to provide useful input into the evaluation without paralyzing momentum.

Figure 1 provides a useful model for organizing the various stakeholders that may wish to be involved in a program evaluation. At the centre is the evaluation team—the lead evaluator and the evaluation support staff. The lead evaluator may be a staff member who has been given responsibility for managing the evaluation or an externally contracted evaluation consultant.

Figure 1

In many cases, it is useful for the evaluator (or evaluation team) to have logistical support from a small number of individuals who are in a position to help implement the processes and procedures that are required to collect the evaluation data. Usually these are individuals within the program who have detailed knowledge of how the program functions. Without an Implementation Team to help in the design and the execution of data collection, the evaluation will risk being unwieldy, impractical, and an unnecessary burden to program staff and clients.

Some evaluations have a Steering Committee—usually composed of the individuals who have commissioned the evaluation and who will maintain an oversight function. If a Steering Committee exists, the Evaluation Team will be directly responsible to this group. The Steering Committee will usually meet regularly to set major milestones, approve the evaluation design,

and review progress. They will be the first group to receive the evaluation results.

Sometimes it is useful to develop a broad Advisory Committee composed of stakeholders who have a general interest in the results of the evaluation, and who give advice on the evaluation project, but who do not have an oversight role. Advisory Committees are usually composed of members of the wider service community, including members of local client and advocacy groups. They meet infrequently, at important project milestones, to receive progress reports and to provide general feedback.

Stakeholders will often express their preference for a particular level of involvement once the structure of the evaluation is known, and depending on the time commitment involved. While clients and family members often feel most comfortable participating on Advisory Committees, it is important to encourage them to become involved in all aspects of the evaluation.

Finally, the structure that is developed to support the evaluation should be match the size and scale of the evaluation itself. Small evaluations require small structures. Large evaluations are more costly and require more organization. Therefore, they also benefit from larger, more formal structures.

New Program

2 Understanding how the program to be evaluated is supposed to work

2.1 Matching program evaluation activities to program development

Young Program

There are many approaches to evaluation. It is important to match program evaluation activities (and evaluation questions) to the stage of the program's development. Evaluation activities involving newly established programs often focus on issues pertaining to program inputs, such as documenting new or unmet needs and determining

Mature Program

the correct mix of staffing or resources required to address these.

In a **new program**, program staff often have well developed goals and objectives, and a description of the services to be offered, but little else. Their first priorities are to get the program up and running, acquire and train staff, implement a program delivery model, and establish internal procedures. When programs are first established, they are not working at peak capacity or efficiency. The way in which services are delivered may change dramatically from one day to the next in response to problems that emerge. Unless a program has stabilized, it is difficult to measure or interpret outcomes.

An important focus for evaluation activity in a new program is on laying the groundwork to support future assessments of program processes and outcomes and identifying and eliminating stumbling blocks. This might include insuring that information systems to collect client data are appropriately organized and insuring that sufficient standardized baseline information is collected to support future evaluation activities. Secondly, it is helpful to use this data to describe the clients who are receiving services, and the services they receive, in order to evaluate whether the program is reaching the target population originally intended. In addition, detailed contextual information obtained from staff and from clients is important in determining how services are provided and for identifying trouble spots.

Evaluations of **young programs** typically focus on program processes, issues of fidelity to program models of change, and adherence to best practice standards. In a young program, most of the start-up problems have been worked out and routines have been established. Refinements may still be made to the processes of care, but these occur less frequently. The program is reaching peak capacity and efficiency.

Evaluation activities at this stage typically focus on the processes of care—determining how the program works. Statistical monitoring of client characteristics and utilization is useful to ensure that the program continues to serve the intended target population in the ways originally conceived by the program developers. Evaluating fidelity to best practices is a second focus of evaluation activity in a young program. Continued monitoring is required to ensure that the program continues to reach its intended client population with services that meet their needs and follow best practice guidelines. It is also important to understand whether clients are satisfied with the way in which the program delivers services. Unless these can be firmly established, it is premature to move to an evaluation of program outcomes.

Evaluation activities for **mature programs** may focus on processes issues, such as fidelity, but also examine the outcomes of care, thereby offering a comprehensive view of program functioning. Underlying this is the notion that programs must be correctly implemented before they can achieve their desired results. To be helpful, evaluation activities must pose questions, that assist programs through this evolutionary process. Mismatched evaluations rarely contribute to program development and can be destructive to program functioning.

Mature programs have been in operation for a several years and have been functioning at peak capacity for some time. They have stable staffing and service delivery mechanisms, and they have demonstrated that they employ best practice guidelines to delivery services that meet the needs of their clients, and that these services are delivered in acceptable ways. Evaluation activities can now focus on whether, given these processes of care, the program is achieving its intended outcomes.

2.2 Program Logic

The logic of a program is the road map that links the program's resources to the processes of care and to the outcomes to be achieved. Understanding the logic of a program is important for establishing both process and outcome evaluation questions. The logic model spells out how the program produces results, so it helps focus the evaluation on the key elements and assumptions.

The process of developing a logic model typically involves a number of program stakeholders working together to understand the underlying rationale for the program; why and how it will meet client needs; what series of steps need to be implemented to meet short, intermediate, and long term objectives; how objectives interrelate to produce an overall effect; and the conditions under which success is most likely. The process of building a program logic model requires clarity of thinking that can highlight areas where important assumptions are tenuous, where service links are weak or missing; or elements in the broader service environment that may facilitate or hinder the program's success. The model becomes a focal point for discussion and consensus and provides program staff with an important tool for describing their program to others.

Outcome models link the program processes that produce the desired program outcomes. Connections are made between short term, intermediate, and long term objectives. However, a logic model may also be used to describe program activities, linking together the logical sequence of steps that must be put into place as part of the process of program implementation. Finally, a logic model may be theoretically based. This type of model links theoretical constructs from the scientific literature to explain the underlying principles and assumptions of the program, in an effort to explain why a particular outcome is achieved. A theory based model provides a template for making predictions about how a program should work. Often these

types will need to be combined to provide a comprehensive view of program activities.

2.3 Needs assessments

Needs assessments are a special type of evaluation focused on determining the extent of unmet needs in a population group in an effort to establish a new program designed to meet these needs, or expand an existing program. The target population of interest may already be receiving services from existing programs, but still have areas where important needs are poorly addressed, as when clients have multiple and complex problems. Conversely, the target population of interest may be entirely new, having never received appropriate intervention. An important outcome of a needs assessment is the identification of the size, location, and characteristics of the under serviced group, including any barriers to care that have prevented them from receiving services from existing programs. A second outcome is the description of a program or service delivery model that can address the unmet needs that have been uncovered or overcome the barriers to care that have been identified. This requires a detailed assessment of the scientific literature reporting on the efficacy and effectiveness of different approaches to care followed by careful reasoning explaining what a particular approach is likely to work in this case. It is also important to provide a budget that estimates operating costs and, in some cases, a detailed operational plan that can be provided to potential funders to outline how the new program will work and why this approach (as opposed to others that may be available) is likely to be successful in addressing the problem or filling the unmet need. Knowing the size and scope of the target population to be reached is a key part of developing the budget.

Needs assessments can also be used as an important tool for generating political support for a new program or initiative by creating buy-in and stakeholder participation in the

evaluation. An important by-product of a needs assessment may be increased community participation, increased motivation for networking among community agencies, or the creation of new networks or coalitions.

In summery, the main outcome of a needs assessment is a number, reflecting the size of the potential client population to be served by the new or expanded program. A second outcome is a description of the strategy that will be used to link potential clients to services that have been scientifically demonstrated to be effective. Finally, an estimate of cost of the expansion is also usually necessary.

3 Developing an evaluation strategy

3.1 Determining the evaluation question

The first and most important step in carrying out a successful evaluation is getting the question right—making sure that it is 'in sync' with the nature of the program being evaluated; that an answer to the question cannot be determined through some other route; that stakeholders can indicate how they will use the information to improve program activities; and that there are no important hidden objections that need to be addressed.

3.2 Making sensible design choices

Once the question has been successfully negotiated, the next step is to develop the evaluation strategy. In part, this should be a logical extension of the question (or questions) being posed. In addition to being well suited to the question(s), sensible choices will be suited to the resources available for the evaluation, the availability of program data, and the expectations of funders and other key stakeholders.

Many times the most useful (and usable) results will come from a simple design that can be easily implemented and understood. Large, overly complex, and overly "scientific" approaches often place too heavy a burden on program staff and take too long to provide useful results. In evaluation, an 80/20 rule often applies. This means that the evaluator should look for approaches that give 80% of the information for 20% of the effort, as these are often sufficient to guide decisions and reduce uncertainty about program functioning.

Two simple evaluation strategies that are widely used are the survey and the before/after (or pretest/posttest) design.

3.2.1 Surveys:

Surveys systematically collect data on a cross-section or a consecutive series of clients (or sometimes staff). They are used to provide a snap-shot of the program at a point in time. A benefit of surveys is that they can be used to obtain direct feedback from clients concerning their experiences in the program (process) as well as the program's perceived effects (outcomes). They may also be used to obtain feedback from other groups, including family members, program staff, and staff in other agencies. They may be conducted face-to-face, by telephone, or through the mail, making them one of the most flexible and cost-effective evaluation methods. Surveys can also be used to identify needs in groups not receiving treatment from the program to support program expansion or the development of new services.

3.2.2 Before/After Designs:

Before/after designs (also known as pretest/posttest designs) are used to measure change. The before measure is collected when the client first enters the program, before any services have been received, and the after measure is collected after important treatment

milestones have been reached or when treatment has been completed. Multiple after measures can be collected to create a more complex, repeated measures design. Measurement instruments used to assess the outcomes must be identical at each time point. Differences in the two measures then reveal the amount of change that has occurred while the client has been in the program. Any differences are typically attributed to the treatment(s) received. However, this assumption often requires justification, such as by ruling out factors other than the treatment (spontaneous remission; maturation, auxiliary treatments outside of the program) that may have come into play to create improvement.

3.3 Data Analysis:
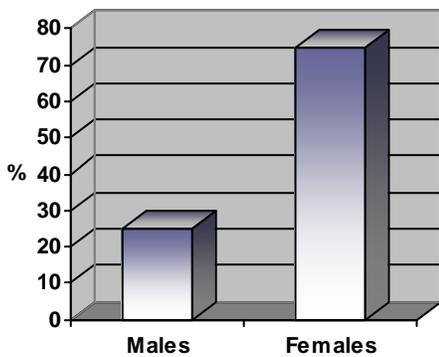
Statistical analysis is not for the faint of heart and is often the greatest stumbling block to evaluation work. It typically requires specialized skills in data management and analysis that many smaller programs do not have. Therefore, before embarking on a statistical analysis, it is important to reflect on what information would be provided that couldn't be obtained using other means. If statistical analysis seems inevitable, then a few simple skills can go a long way.

Creating tabular or graphic descriptive summaries depicting the characteristics of clients who have received services is a useful starting point. Ideally, this information should be easily available from an agency's computerized information system through a standard built-in reporting system. Examining differences in the percentage distributions of clients according to demographic characteristics (such as age group or sex), social characteristics (such as marital status or living arrangements), and clinical characteristics (such as diagnosis or length of illness) can provide the detail needed to

demonstrate that the program is reaching its intended target population.

When socio-clinical profiles of clients are augmented with information depicting the services that were provided, a much better picture of the match between services and needs emerges. Looking at characteristics of sub-groups can identify potentially high-needs populations. Regular monitoring of this information can identify program drift. Program drift occurs when the program drifts away from its original mandate and target population—often toward a more easily serviced group. Descriptive information will help keep the program on track and assure funders that the intended client population is being served.

The same principles apply to data collected through questionnaires or chart reviews. Simple descriptive summaries using percentages, combined with sub-group analyses go a long way. If the major relationships aren't visible using these simple techniques, then one must be suspicious of more sophisticated approaches that reveal otherwise "hidden" patterns in the data.

Even when the planned analysis is simple, it is crucial to plan out the statistical analysis in advance of the data collection. This can be done using dummy tables and graphs that are populated with 'expected' results. Not only does this encourage those involved in the evaluation to articulate their hypotheses and expectations about the way in which the program functions, it flags areas of controversy (where more detailed data may be required), highlights flaws in logic, and provides a practical template for the data analyst. It is also a way to engage stakeholders in the process of data interpretation and learn more about their expectations. It is better to make mistakes with test data, rather than with the real data!

3.4 Data Management:

Many evaluations—particularly outcome evaluations—require data that is not available from routinely collected program information systems. This raises the issue of data management. Raw data (collected via surveys and questionnaires) must be entered into a computer for analysis and cleaned to remove entry errors and logical inconsistencies. This can be a laborious process and will often consume the lion's share of the time set aside for analysis.

The availability of computerized statistical and data management programs means than someone with minimal data entry and management skills can develop a database and conduct a descriptive analysis. However, entering data is a finicky job that cannot be done amidst other work pressures. A common mistake is to collect data with no real plan for how it will be entered, managed, or analyzed. A second common mistake is to expect busy clerical staff to serve double-duty as data entry or statistical personnel.

By jeopardizing the quality of the data, the entire evaluation is placed at risk. If the data are worth collecting in the first place, then they are worth managing properly. Consequently, it is important to plan for protected time for someone with data management and analysis skills to undertake this important function, or protect resources to contract out.

Again, the 80/20 rules seems to apply. Approximately 80% of the time allocated for analysis will be used in entering and preparing the database and 20% will be used to make the actual calculations.

3.5 Ethics Clearance:

It is important to keep in mind that many organizations require all studies involving human

subjects (including program evaluations involving client data) to be vetted by a local Research Ethics Committees. It is the responsibility of the evaluation team to identify local policies and procedures and to ensure that their evaluation plan has been appropriately cleared.

3.6 Budgeting for an evaluation:

Evaluations take time and resources. Thus, it is important to plan an evaluation budget before beginning to collect data—even if the resources that will be contributed to the evaluation come from internal funding mechanisms or from existing staff time. A formalized budget is a must when external funds are being requested.

Budget items can be considered to fall into a number of generic categories:
- Staff salaries and benefits
- Consultants
- Travel for data collection and stakeholder meetings
- Communications (postage, telephone)
- Printing and duplication (survey questionnaires, final reports)
- Supplies and equipment (including data entry and analysis software)
- Other expenses (library fees for articles)
- Contingency expenses (5% or 10% of total operating budget)

4 Determining what data collection tools to use to evaluate program functioning

The point of any evaluation is to use systematically collected data to tell a story about the program and how it is functioning. Therefore, the acquisition of high quality data is key to an evaluation's success. While it is always preferable to use existing evaluation instruments, often these are not available or not appropriate for the questions being asked. Thus, many evaluations must include time to develop and test new measures. Before undertaking this time-

consuming task, it is important to be absolutely sure that existing measures will not work.

4.1   Judging the performance of existing measures:

Measurement instruments can be thought of as scales and indices that provide an aggregate quantitative (or numeric) score which combine values across a range of individual items. The aggregate score depicts where each individual lies on a range of values pertaining to the construct being measured and its various sub-dimensions. The scientific literature is rife with instruments that can be used to measure changes in:

- Symptoms,
- Behaviours,
- Diagnoses,
- Functioning,
- Well-being,
- Social integration,
- Quality of life,
- Perceptions of recovery, and
- Experiences of stigma,
- And many more…

The performance of existing measures is typically reported in two ways: reliability and validity. Reliability refers to the repeatability of a measure. When used the same way, or by different people, it will give the same results. Validity, the more important of the two, means that the instrument measures what it is supposed to measure. In addition, when conducting evaluation studies, sensitivity to change is important. Some measures reflect traits, or ongoing conditions that are stable, while others are capable of detecting changes that may occur as a result of program interventions. A common mistake is to use a measure that is insensitive to change to try to assess program outcomes.

The development of a new scale or index requires considerable skill and time; since the performance properties (or psychometric properties) must be established over a range of

studies. As such, the creation of a new scale or index should never be considered as part of the evaluation plan. This is because it is impossible to test the psychometric properties of a new measure and use it to evaluate the effects of an intervention at the same time. If the program does not perform as expected, it will never be clear whether it was the measure or the intervention that was deficient. Poorly conceived measures can place a program in considerable jeopardy.

A systematic search of the literature, which can be performed using web-based search engines such as Google Scholar, will provide information on what others have done and how instruments have performed. Many articles will be available in a downloadable format, but others will need to be ordered from local libraries (usually at a cost). Local reference librarians can also help develop systematic search strategies to ensure that all relevant material is identified.

4.2 Developing a questionnaire:

While the development of new scales and indices are not recommended for evaluations, it is often necessary to create a questionnaire to survey client's or other stakeholder's perspectives about the services provided or their unmet needs. Questions may ask respondents to indicate the presence or absence of something, or to rate something according to a set of ordered categories, such as from *strongly disagree* to *strongly agree*.

Questionnaires are much easier to develop than scales and indices, since the individual questions describe some important aspect of the program. Questions do not need to be combined (numerically) to provide a scale score. Evaluation questionnaire are often developed with considerable input from stakeholder committees. Each question should be mapped onto a program evaluation issue. This will insure that every question has a place in the evaluation—that there

are no superfluous questions—and that everyone
will be clear on how they will be used to make
evaluative judgments.

Before using the questionnaire in the evaluation,
it is important to test it out with a small
sample of the intended targets. Usually this
will be program clients, but it may also be
representatives from other agencies (for an
agency survey). A practical way to test the
questionnaire is to have several intended users
meet as a group to discuss each question. These
individuals should be specifically chosen to
maximize the quality of the feedback received
(i.e. those who are motivated, interested,
articulate, thoughtful, and critical). The
developer leads the group through each question
and asks them to provide their understanding of
what the question means. When questions are
poorly worded or unclear, group members will
typically provide different interpretations.
They will also identify words or phrases that may
be reactive (or offensive) and help provide
alternate wordings for questions that are
ambiguous. Members should be encouraged to
provide feedback on the flow of questions, the
appearance of the questionnaire, and any other
aspect that may affect the quality of the data
obtained. Consumer and family advisory groups
are an invaluable source of support for survey
development.

4.3   The importance of including qualitative data:

Most often when we think of evaluation, we think
of collecting quantitative data (such as surveys
or scales) that can be subjected to statistical
analysis. However, qualitative information,
collected through interviews, anecdotal accounts,
or observation, can often tell a stronger story.
Qualitative information is important for making
sense out of statistical patterns and providing a
broader context for the findings. Qualitative
data adds richness and detail to the evaluation
that would not otherwise be available.
Qualitative data may also shed light on important

aspects of the program and program functioning that cannot be measured using conventional statistical approaches, such as how a program works, why, and why it is successful—arguably some of the most central questions in an evaluation. In short, thoughtful interpretation requires one to look beyond the statistical findings to what the result mean, what led to the findings, and whether the findings are important. Qualitative data is vital for this process.

4.3.1 In-depth Interviews:

In-depth interviews with key stakeholders, such as program staff, clients, family members, community members, or staff from other agencies are among the most common of all qualitative approaches used in evaluation. The interviewer is typically guided by a series of broad questions that are designed to elicit stakeholder opinions, perspectives, and experiences. Information may be taped and transcribed, but more often the evaluator keeps detailed notes. The aim of the analysis is to highlight the key themes and issues that emerge, identifying areas of convergence and divergence across informants. Informants are typically chosen because they are information-rich. They may represent a range of different experiences, and be purposefully chosen for this reason, to gain insight into a particular program process or problem. Conversely, they may be chosen because they have something in common that is of interest to the evaluation team.

In-depth interviews are particularly useful when external evaluators are asked to assess a program's performance, since they may know little about the program or its context. A series of in-depth interviews with agency staff and clients may help them focus the evaluation in ways that would not have been possible otherwise.

4.3.2 Focus Groups:

Focus groups typically bring 6-8 individuals together to discuss a particular issue. They are facilitated by one or two individuals who seed the discussion with general questions and probe for more detail when interesting points emerge. The purpose of a focus group is not to achieve consensus, but to elicit a full range of experiences or perceptions surrounding a given topic. As with all qualitative techniques, focus groups work best when the goal is to obtain rich descriptions that provide insight into another's reality. As with in-depth interviews, focus groups may be taped and transcribed, however it is usually more expedient if one facilitator takes detailed notes while the other manages the group. The success of the focus group will lie with the skill of the facilitator so it is important to select someone with prior group experience who is a good communicator and a good listener.

Focus group data are generally analyzed to highlight the various dimensions that underlie a problem so as to bring multiple perspectives to the table. As previously discussed, focus group data can be particularly helpful in guiding questionnaire development to insure that all relevant dimensions are covered in ways that will be acceptable to the target respondents. Focus groups can also play a vital role in pilot testing questionnaires to identify poorly worded, confusing, or offense questions.

5 Developing usable results

The raison d'être of any evaluation is using the results to improve program delivery and/or program outcomes. Often this means writing a report outlining the evaluation strategy, findings, and recommendations. Evaluation reports come in many forms; thus it is important to ensure that the report that is produced meets the needs of those who are intended to use it.

Often several reports of various sizes and detail will be required to meet the information needs of all of the different target audiences who have an interest in the evaluation results.

5.1 Report Writing:

A good evaluation report will tell a story about a program and will not be overly technical or difficult to read. Technical details, when required, may be contained in an appendix or a technical companion report so that they don't interfere with the main message.

Reports need to be tailored to their key target audiences. High level government decision makers and potential funders will need completely different information than program staff. Reports must be sensitive to these different needs and address them in different documents. The type, number, and size of reports should be planned at the outset. This is particularly important when contracting an external evaluator as each report will take time and resources to produce. These need to be negotiated in advance.

When targeting a report, it is important to consider how much technical detail to include. If the audience is non-technical, they will want to know what the evaluation found and the recommendations. They will be much less interested in the methodological details, statistics, or procedures. If the audience is technical, they will want to know how the evaluation was conducted before they will place any confidence in the findings and they will be on the look-out for flaws. Depending on the mix of readers, the report writer may choose to provide a non-technical, content-oriented synopsis of the findings as well as a comprehensive report containing both technical and non-technical information. Within the comprehensive report, the technical details may be contained in an appendix, where they won't interrupt the flow for content-oriented readers.

5.2 Characteristics of a Good Report:

Good reports are clearly written and concise. Large and unwieldy reports which are data dense and contain many scores of recommendations are unlikely to be too helpful. More is not better! Generally, reports need be no longer than 25-35 pages. A 1-3 page executive summary should be included to provide an easy-to-read synopsis of the evaluation approach, major findings, and recommendations. Executive summaries are always advisable. In some situations a 1-page bulleted summary of the main findings may be helpful; particularly for senior decision makers who have limited time and need to grasp the essence of the report in order to appropriately delegate it to someone for more detailed review and commentary.

Good reports are pleasing to look at. This means that there is sufficient white space between paragraphs, tables, and figures so that the text does not appear cramped. Fonts should be readable (usually 11 or 12). Figures and tables should be embedded in the text for ease of reading. Tables and figures should be clearly titled and contain sufficient information to make them understandable (interpretable) without referring to the text for an extensive explanations. A common mistake is to exclude the number of subjects from graphs or tables.

5.3 Report Contents:

An evaluation report should contain the following information:

- **Executive Summary:** A brief (1-3 page) overview of the reason for the evaluation, the approach taken, the main findings, and the list of recommendations.

- **Table of Contents and Acknowledgements:** Include a table of contents to help readers navigate through the main sections and sub-sections of the report. Also an acknowledgement page should be included listing the names of all individuals who contributed to the evaluation, such as the

evaluation team; Steering Committee members; other committee members; and all sponsors.

* **Purpose:** A statement of the issue that the evaluation is intended to address. This might include the decision that the evaluation results are expected to inform or the question that the evaluation is expected to resolve.

* **Background:** This section includes any background information that bears on the evaluation issue. This could be a description of the historical development of the program that is the setting for the evaluation; the results of previous evaluations; the way in which the evaluation issue has been approached in other settings; and a review of the pertinent scientific literature describing the scope of the problem. If the evaluation is to assess the effectiveness of a new program, then a detailed description of the program (including the logic models) should be included.

* **Evaluation Design:** This section should contain enough information pertaining to the way in which the evaluation was conducted to allow an outside person to replicate the study. If the evaluation is complex, a brief overview of the design and data collection may be included here with more detailed technical information in an appendix. Issues that should be covered include the organizational structure of the evaluation (including a description of all committees); the evaluation design (survey, pretest/posttest, etc.) the measures used along with their reliability and validity; how surveys were designed and tested; how data were collected (face-to-face, telephone, questionnaire); the nature of the sampling plan (how subjects were selected) and any procedures used to recruit subjects and define study groups; the number of subjects on which the evaluation results are based;

response rates (if using a survey); how data were managed and analyzed; and any ethical considerations (including ethics clearance). Clear statements of limitations and strengths of the various design decisions made help critical readers place more confidence in the results.

* **Results:** This section should contain the main results of the evaluation using tables and graphics to illustrate the most important findings. Auxiliary information can be included in an appendix and briefly summarized in this section.

* **Conclusions and Recommendations:** Summarize the main findings of the evaluation and assess their implications for program functioning. Outline recommendations that flow from these findings being careful to include only those recommendations that are supported by the data. A common mistake is for committees to expand the recommendations into a "wish list" of items that may not flow from the evaluation results.

* **Technical Appendices:** These should contain all auxiliary information that the reader may need to fully understand the evaluation and the evaluation results. Consider including this information as a technical companion to the main report if it turns out to be extensive.

5.4 Writing Recommendations:

Since the main point of an evaluation is to guide program development, recommendations are the focal point of most readers' interest. They must be clear, easy to understand, feasible, and few in number. Recommendations must also flow from the evaluation results. A common mistake is to include suggestions for program improvements that lie outside of the scope and findings of the evaluation. It is a good idea to keep the list of recommendations small—small enough for a person to remember if they don't have the report handy. This will ensure that the recommendation are usable.

It is best for recommendations to outline the "what", rather than the "how". Often there a multiple ways of implementing a recommendation and these need to be carefully considered by program staff.

6.0 Closing:

With this brief overview of the evaluation process, readers should now feel more confident if they are asked to participate in an evaluation project, either as an evaluation team member or as an advisor. The next chapter provides examples of how these principles may be applied to the field of concurrent disorder programming.

CHAPTER 2 – EVALUATION & CONCURRENT DISORDERS

This chapter draws on principles and approaches described in the previous chapter to give examples of how these evaluation principles may be applied to different programming issues in the field of concurrent disorders.

2.1 Assessing Needs

From a review of the epidemiologic literature, you have learned that people who have a serious mental disorder, such as schizophrenia, have a rate of substance use that is three or more times higher than the general population. People who have concurrent mental and substance abuse disorders have poorer outcomes and use more services. In addition, the literature suggests that many people with concurrent disorders fall through the cracks. If they receive services in the mental health system, their substance abuse problem may not be identified or addressed. If they receive services in the addictions system, their mental illness may not be well addressed. Increasingly, it is thought that hybrid programs, which address both mental health and substance abuse issues, provide the best practice approach.

You have been commissioned by the Director of a community-based program for people with serious and persistent mental disorders. The Director is concerned that the staff may not be picking up on or adequately addressing clients' substance abuse needs. You are asked to determine how many of the agency's clients have a concurrent substance abuse problem and how many of these have been correctly identified by staff. You decide to conduct an assessment of clients' needs.

Because you are concerned that clients may not wish to disclose illegal activities, such as illicit drug use, you will focus on alcohol misuse. You recognize that alcohol is the most common substance of misuse among people with schizophrenia and alcohol misuse is often correlated with misuse of other substances so consider alcohol misuse as a proxy (or indirect) indicator of substance related need.

Based on your discussions with the Director, you will use two questions to guide the evaluation:

Question 1: How many incoming clients have co-occurring alcohol misuse problems? You decide to narrow your focus to incoming clients as each client undergoes an intake interview that is designed to identify their presenting problems. You consider that it would be practical to ask clients to complete a brief screening scale at the close of their intake assessment. Further, you consider that the results of this component of the evaluation will tell you whether the intake assessment is thorough enough, or whether it needs to be improved with the addition of standardized screening scales for substance problems. So as not to introduce an interviewer bias (by sensitizing staff to the nature of the evaluation and changing their regular behaviour), you will give clients the clinical screen at the close of their intake assessment, before they leave the agency.

CAGE Questionnaire*

1. Have you ever felt you should **C**ut down on your drinking?
2. Have people **A**nnoyed you by criticizing your drinking?
3. Have you ever felt bad or **G**uilty about your drinking?
4. Have you ever had a drink first thing in the morning to steady your nerves or to get rid of a hangover (an **E**ye opener)

* Ewing JA: Dececting alcoholism: The CAGE Questionnaire. JAMA 252(14): 1905-1907, 1984.

You have decided to use the CAGE Questionnaire which is a brief screen for clinically significant alcohol problems that is in the public domain (i.e. not copywritten). It is the most widely used screen and has been demonstrated to have good validity. It contains 4 items that are answered either "yes" or "no". A cut-point of 2 (reflecting two affirmative

responses) has been demonstrated to identify 78% to 81% of clinically significant alcohol-related disorders (with less than 24% false positive diagnoses). A score of two is the one that has been conventionally used throughout the literature.

Question 2: How many of these clients alcohol problems were identified by staff as a presenting problem? Your agency maintains a client database that includes several fields for presenting problems. Drop down menus allow staff to identify "Alcohol" as a presenting problem. The evaluation assistant will collect data from the database on presenting problems for all of the clients that you screen for alcohol use. Ideally, you would expect that 100% of clients who exceed the screen threshold representing a clinically relevant problem will have "Alcohol" flagged as a presenting problem. To allow for the possibility that up to a quarter of those identified by the screen may be false diagnoses, you decide to accept a benchmark of 75% as your best practice standard.

You successfully screened 200 clients, representing a response of 80%. Results showed that 35% (70) of your incoming clients met or exceeded the screen threshold reflecting a potentially clinically significant alcohol problem. After matching the screen data with the presenting problem fields from the database, you found that less than half 23% (46) had 'Alcohol' identified as a presenting problem requiring further clinical assessment. This fell 52% below your benchmark of 75%. You conclude that the staff are missing a significant portion of alcohol-related problems among incoming clients and infer that the clients are likely not getting the full range of addictions services they need.

In order to narrow this gap, you recommend that the agency undertake a series of educational sessions designed to sensitize staff to the high

prevalence of alcohol related problems among
their clients. You further recommend that the
agency consider including a standard alcohol
screen (such as the CAGE) as part of the intake
assessment. Finally, you recommend that the
agency consider mounting a series of continuing
professional education seminars and workshops
designed to improve clinical skills pertaining to
the identification and treatment of alcohol and
other substance related conditions.

2.2 Creating a Program Logic Model

You have just returned from a seminar on developing program logic models and your agency has asked that you lead a discussion on this topic with the program directors. You decide that an interesting way to approach the issue would be to have them complete program logic models for their respective program areas.

Program Logic

Expected Outcomes	Program Activities	Resources
We expect the following changes to occur in participant's lives as a result of our program	We will have to do the following activities in order to achieve these results:	We will need the following resources in order to undertake these activities:
•Immediate Changes (when the program is finished)	•What intervention •To whom •For how long •In what intensity •How often	•Staff •Volunteers •Money •Materials •Space •Equipment
•Short term changes (within the first year following the completion of the program)		
•Lasting changes (long term changes)		

You begin by asking them to consider their programs as a series of logical connections between inputs (the resources that they have), the processes (or activities that they undertake) and the outcomes (or changes that they bring about). Using a program logic template that you received from the seminar (**contained in the Appendix B**) you first ask them to specify their goals in terms of three time frames: what changes should be evident in clients immediately upon completing the program; what changes should be evident over the short time (such as in the first year) and what long-term changes the program is supposed to achieve. For each outcome that is identified, you ask the directors to consider how they would know whether the change has been achieved. You want them to think about specific types of measures that they could use to document that a change has occurred.

The next job is to identify the strategies and activities used by the program to achieve these ends. You ask the Program Directors to consider what they must do to achieve these results. In this way they will link program activities to specific outcomes. Finally, you ask them to specify their resources in terms of staff, equipment, space, money and materials and have them link these to specific activities so that each activity (or cluster of activities) has a well-defined budget.

36

Once all the linkages have been made, you ask the Program Directors to consider whether the program models they have developed make sense in light of what is known about theories of change and best practices in the literature.

2.3 Assessing the Implementation of a New Training Program

The research literature has clearly indicated that people with a serious mental illness and a concurrent substance abuse problem are best treated in an integrated way. Unfortunately, mental health services may refuse treatment to a person with an active substance abuse problem and addictions specialists may believe that a person cannot recover from a substance abuse problem until their mental illness has be treated. People with serious mental illnesses and substance abuse problems often fall through the cracks, sometimes being refused treatment by both systems.

In an effort to solve this problem, mental health, addictions, agencies in your area have agreed to work together to provide more integrated care to people with concurrent substance abuse and mental disorders. As a first step, they have developed a training program to build capacity among mental health and addictions professionals to identify people with concurrent disorders.

The training program would consist of a series of sessions where staff from both addictions and mental health agencies would jointly attend seminars. Because the program is new, evaluation of the extent to which it could improve client outcomes was considered to be premature. Instead, the evaluation would focus on shorter term objectives, defined as whether the joint training sessions had improved staffs' knowledge concerning the identification and treatment of people with concurrent disorders.

A pre-test/post-test design will be used to assess changes in participant's knowledge, attitudes, and feelings of competency before and after the program. Six months following the completion of the seminars a second post-test survey will be given to assess the extent to which any gains in knowledge are maintained over time. The evaluator thought that this information would help planners decide on the need and timing of "booster" sessions.

A questionnaire has been developed to address the specific educational goals of the training program. The trainers plan to conduct test runs of all program sessions with a small group of eligible staff prior to making it more generally available. This will provide an ideal opportunity to pilot test the evaluation questionnaire as well.

When the evaluation design was presented to the Project Steering Committee, two items of concern were raised.

- If the training sessions were offered on a voluntary basis, only those staff members who were already amenable to learning about concurrent disorders (and willing to change) would attend. Results of the evaluation would then tell little about whether the program would be effective teaching tool for all staff. After much deliberation, the Steering Committee decided that the training sessions would be mandatory for all agency staff. They also asked the evaluation team to ensure that the sample would be representative of all staff participating in the training.

- Secondly, Steering Committee members recognized that the issue of concurrent disorders was receiving increasing attention in the mental health and addictions systems in their district and a number of different initiatives had begun to improve services to this client group. They worried that these activities, not the educational program,

could account for changes in staff knowledge and attitudes. A researcher on the committee suggested that the only way to control this problem would be to include a comparison group composed of people from similar agencies in a different part of the district who did not receive the training sessions, but who completed both pre and post-test surveys. To ensure that the groups would be equal at the outset, the researcher recommended that participants be randomized to receive either the training session or a placebo control (no training). Comparison of change across the 'treatment' and 'comparison' groups would then tell the evaluators whether the changes they observed in the 'treatment' group could be attributed to the educational program or whether broader contextual factors might also be involved.

The evaluation team considered that a design involving a comparison group would be too difficult to implement given the time and the budget allowed for the evaluation. Instead, the evaluation team decided to stick with a one-group, pre-test/post-test design but incorporate a question on the post-test survey that would directly ask program participants if they had received information about the identification or management of clients with concurrent disorders from any other sources while in the program. A sub-group analysis comparing those with and without this additional information would provide some insight into the role of broad contextual factors. They also decided to include a question on the post-test survey asking whether the staff member believed that their knowledge and competencies had improved *because of* the educational program. The Steering Committee agreed that the proposed solution was a reasonable compromise between scientific rigor and practicality.

2.4 Assessing the Outcome of a Mature

Program

Five years ago, the hospital in which you work received additional funding to develop a group treatment program for people with concurrent disorders. An outcome evaluation has now been commissioned to assess the progress of clients who received the new program. Specifically, funders wish to know whether clients have demonstrated improved functioning and quality of life compared to people with concurrent disorders who did not take the program.

You decide that a pretest/post-test design with a comparison group of individuals with concurrent disorders who did not receive the group treatment program would be the best way to proceed. You consider that all individuals enrolled in the group program during the next fiscal year will comprise your 'treatment' group. You obtained permission to recruit a consecutive series of new clients with concurrent disorders receiving usual community based care from another community mental health agency in your district. These will be your 'comparison' group. Client ratings of their substance use, social and occupational functioning, and quality of life will form the main outcome measures.

Because the groups come from different agencies they may be non-equivalent in terms of factors that may predict positive treatment outcomes. You are particularly concerned that the two groups be comparable in terms of their overall prognosis so you will also collect extensive social and clinical data so that the groups can be compared on these potentially prognostic factors such as:
- Employment
- Housing
- Income
- Age
- Gender
- Diagnosis
- Type of Substance Abuse
- Previous hospitalizations
- Previous arrests

- Types of services being received
- Physical health problems
- Social support

If the groups are unequal on one or more of these factors, then it will be difficult to tell whether it is the new program, or the imbalance in prognostic factors that is responsible for the outcomes. Because special statistical analyses will be required to mitigate any imbalances, you decide to consult with a statistician to ensure that you have appropriate statistical back-up on the evaluation team.

Treatment drop-out is a second issue of concern. You wonder if the frequency of participation in the group is predictive of a positive outcome so you decide to analyze the outcome data by the number of group sessions attended. If the group is effective, you should see a 'dose-response' pattern where the level of improvement increases by number of sessions. If a dose-response pattern occurs, you believe this will lend stronger evidence to the conclusion that the program has caused the effect.

2.5 Sharing Evaluation Results

You have been asked to disseminate the results of an evaluation to potential users. Before doing this, you want to be sure that the evaluation has provided credible, timely results and that the recommendations are well-founded. You don't want to be in the position of disseminating the results of a poor evaluation. Thankfully, this evaluation has been conducted under the auspices of a broadly based Steering Committee composed of the major stakeholders in the area. They have provided a cover letter indicating their support for the findings. A technical advisory committee has also endorsed the results.

You decide to use different dissemination channels because you think that allowing for wider access to the findings will increase their credibility and their use. In addition to the

detailed evaluation report, which you have mounted on your organization's web page, you are considering the following additional approaches to dissemination:

- Press conference to alert local media of the evaluation results.
- A 3 page briefing note summarizing the main findings that can be distributed widely, directing people to the full report on the web page.
- A summary of the main findings to be included in the agency's annual report.
- A public meeting with agency clients and staff to present the evaluation findings and discuss how the agency will implement the evaluation recommendations.

Appendix A	**Further Readings and Useful Websites**

Further Readings About Program Evaluation:

Patton, Michael Quinn. **"Utilization- Focused Evaluation – The New Century Text"**, 3rd Edition, Sage Publications, 1997.

Rossi, Michael H., Freeman, Howard E., and Lipsey, Mark W. **Evaluation – A Systematic Approach"**, 6th edition, Sage Publications, California, 1999.

Scriven, Michael. **"Evaluation Thesaurus – Fourth Edition"**, Sage Publications, 1991.

Useful Websites:

American Evaluation Association. **"The Programme Evaluation Standards"**. Available at:
http://www.eval.org/EvaluationDocuments/progeval.html

Aubel, Judy. **"Participatory Program Evaluation Manual – Involving Program Stakeholders in the Evaluation Process"**, Catholic Relief Services, Child Survival and Technical Support Project, Second Edition, December 1999. Available in English, Spanish and French at:
http://www.childsurvival.com/features/bookmarks/pemanual.cfm

Centres for Disease Control and Prevention (CDC). **"Steps in Program Evaluation"**, CDC Evaluation Working Group, September 1999. Available in English on the web at: http://www.cdc.gov/eval/steps.htm

ILO. **"Guidelines for the Preparation of Independent Evaluations of ILO Programmes and Projects"**, Evaluation Unit, 1997. Available at http://www.ilo.org/public/english/bureau/program/guides/evalmenu.htm

UNICEF. **"What goes into a Terms of Reference (ToR)"**, UNICEF Technical Notes, Issue Number 2, April 2002. Available at http://www.unicef.org/evaluation/TechNote2_TOR.pdf

UNICEF. **"Writing a good Executive Summary"**, UNICEF Technical Notes, Issue Number 3, August 2002. Available at http://www.unicef.org/evaluation/TechNote3_Exec_Sum.pdf

U.S. Centres for Disease Control and Prevention (CDC). **"Framework for Program Evaluation in Public Health"**, 1999. Available at http://www.cdc.gov/eval/over.htm

USAID. **"Selecting Performance Indicators"**, Performance Monitoring and Evaluation TIPS Number 6, 1996. Web site: http://www.dec.org/pdf_docs/pnaby214.pdf

Program Logic

Expected Outcomes

We expect the following changes to occur in participant's lives as a result of our program

•Immediate Changes (when the program is finished)

•Short term changes (within the first year following the completion of the program)

•Lasting changes (long term changes)

Program Activities

We will have to do the following activities in order to achieve these results:

•What intervention
•To whom
•For how long
•In what intensity
•How often

Resources

We will need the following resources in order to undertake these activities:

•Staff
•Volunteers
•Money
•Materials
•Space
•Equipment