

**Evaluating
Outcomes of
Community Food
Actions:
A Guide**



***Measurement and Analysis Companion
Document***



Funded by the Public Health Agency of Canada

June 2012

Table of Contents

The Purpose of this Document	1
Evaluation Data, Methods, and Tools.....	1
Qualitative versus Quantitative Data	2
Common Approaches and Tools for Data Collection.....	3
Using Surveys.....	4
Measuring Knowledge and Skills	6
Custom-Made or Existing Surveys	6
Creating Scale Measures for a Survey	7
Qualitative Questions in Surveys.....	8
Focus Groups	9
Advantages to Using Focus Groups	9
Limitations of Focus Groups	9
Tips for Designing and Conducting Focus Groups.....	10
Types of Focus Group Questions	10
Dealing with Challenging Group Members.....	10
Developing and Conducting Interviews.....	11
Constructing and Conducting Interviews.....	11
Sampling for Interviews and Focus Groups	12
Program Outputs and Participant Satisfaction	13
Program tracking and outputs.....	14
How to collect and record output data	14
Program satisfaction and feedback	15
Summary Concepts.....	15
Analysis of Your Evaluation Data	18
Quantitative Data Analysis	19
First Things First – Entering, Cleaning Up, and Coding Your Data	19
Statistics – What are they and what can they tell us?	20
Letting Evaluation Questions Guide Quantitative Analysis.....	21

Qualitative Data Analysis 23

- First things first – Managing Your Qualitative Data 24
- Coding Qualitative Data 24
- Analyzing Your Qualitative Codes 25
- Improving the Rigour of Qualitative Methods and Analysis 27

The Meaning of Data: What counts as success? What is a good finding? 27

Communicating and Reporting Your Findings 29

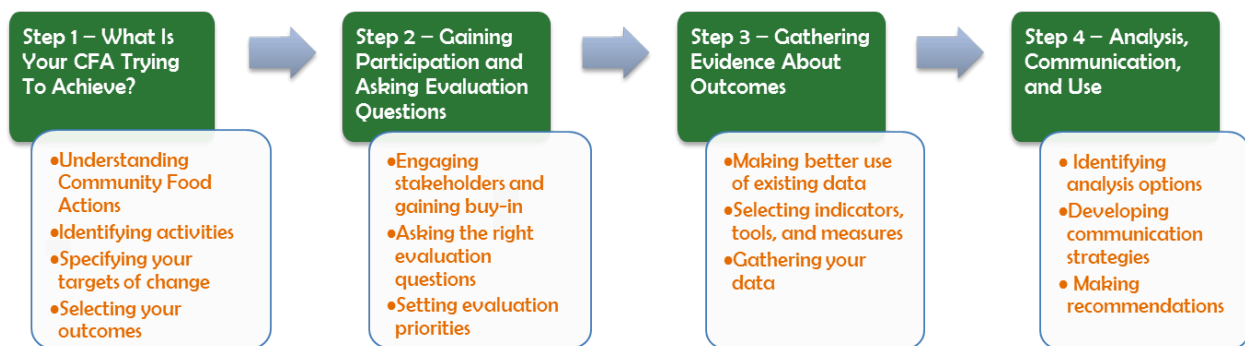
- Tips for Making Recommendations 31

The Purpose of this Document

Evaluating Outcomes of Community Food Actions: A Guide is a comprehensive evaluation resource developed by the Public Health Agency of Canada for wide distribution to the food security field in Canada. During the development of the Guide, it became apparent that some users may have a need for greater depth and detail regarding evaluation practice, most notably in the areas of evaluation measurement, design, and analysis. Rather than overloading the Guide with extensive content on evaluation methodology, a decision was made to move this content to this companion document.

If you work in the food security field and wish to investigate the impact of a Community Food Action (CFA), this companion document may be helpful in building your evaluation design, as a complement to the main Guide. The Guide is available at [\[url\]](#) and should be reviewed alongside the present document.

The content that follows supplements Steps 3 and 4 of the Guide, pictured below as two of the four major steps in the evaluation of CFAs.



Evaluation Data, Methods, and Tools

Effective evaluation requires the strategic collection, analysis, and interpretation of appropriate information. Your evaluation is only as good – useful, informative, accurate, insightful – as the information you collect. Volumes have been written about what is meant by good data and information and the best methods to get a hold of it. Steps 3 and 4 of the Guide examine the main points of this area of evaluation. What information are we talking about and what methods and tools are commonly used to acquire it? This companion document is primarily about *generic* evaluation tools and methods in the field of evaluation, such as what makes for a good survey or focus group, whereas the main Guide provides you some specific tools that can be used to evaluate CFAs.

At the end this companion document there is a useful table that summarizes a selection of core evaluation methods, options for administration, strengths and weaknesses, and when to use them.

Qualitative versus Quantitative Data

Some researchers suggest that quantitative research methods are generally more objective, credible, and rigorous and therefore superior to qualitative methods. Other researchers favour qualitative approaches, arguing that quantitative approaches fail to capture the importance of context, personal experience, and individual differences. The relative merit of both is a long-standing debate in the social sciences. Evaluation as a field has largely moved past this debate, however, often taking the position that quantitative approaches are better at answering some types of questions while qualitative approaches are better at answering others. Furthermore, both are viewed as necessary, complimentary tools that provide a fuller picture of a program or initiative and the associated impact. Evaluation is a discipline that has a rich history of using “mixed methods” to answer real world questions. The table below compares and contrasts the two data types in relation to the kinds of questions they are each useful in answering.

Comparing the Uses of Quantitative and Qualitative Data

Quantitative methods are good at answering questions about...	<u>Quantitative data:</u>
<i>Program outputs, program delivery:</i> How many program components were delivered (e.g., events, sessions, classes, meetings, etc.)?	<ul style="list-style-type: none"> • Represent information through numbers and quantities – ratings, counts, sums, averages, variances, etc. • Are useful to summarize groups on measured variables (e.g., average scores on a food security measure) and to generalize to similar groups. • Are useful to assess degrees (quantities) of change (e.g., greater knowledge after program participation as compared to before) • Are easier to use when things can be obviously counted (e.g., # of participants). Other things are more difficult to reduce to numbers (e.g., self-identity, personal experiences). • Allow you to conduct tests of statistical significance, which can tell you if group differences or change over time are due to chance or due to some real difference, such as a program effect.
<i>Program outputs and participants:</i> How many people participated in each component of our program?	
<i>Participant information:</i> What is the age, gender, family composition, income, etc. of participants.	
<i>Degree of change in outcomes:</i> What is the degree or quantity of change of outcomes over time?	
<i>Group differences:</i> How are the outcomes of groups of people different from one another in relation to the program? (e.g., grouping based on income, by program participation, etc.)	

<i>Qualitative Data:</i>	Qualitative approaches are good at answering questions about...
<ul style="list-style-type: none"> • Represent information through words, phrases, text, and dialogue. • Are useful to summarize the experiences and perceptions of individuals or small groups of people (e.g., their experience of food security, their perceived impact of a program). • Are useful to understand qualitative change (<i>how</i> have things changed) but less useful to assess degrees or amounts of change (<i>how much</i> things have changed) • Are useful in examining individual differences, but not as useful when looking at group differences (especially larger groups). • Provide much more information than quantitative data, making data collection and systematic analysis more difficult and time consuming. • Allow you to provide rich, detailed narratives and descriptions. 	<i>Program process:</i> What goes on inside the program? What are participants' experiences of the program?
	<i>Outcome concepts:</i> What do "food security" and "food insecurity" mean to participants?
	<i>The experience of change:</i> How did the program help participants? How do participants think things have changed, and why?
	<i>Individual differences:</i> How are individuals different from one another in relation to their program experiences?
	<i>Systems change and community engagement:</i> How are partners working together? What are citizen perspectives on food security issues? What is the "story of systems change"?

Common Approaches and Tools for Data Collection

In the field of evaluation there are a few common, almost standard, methods and approaches to producing useful data. These are:

- Surveys
- Focus groups
- Interviews
- Program observations and outputs

Within each of these methods there is a tremendous amount of diversity, sub-methods, and strategies. There are also alternative approaches, mostly qualitative in nature, to gather useful information about program process and impact. These include:

- Goal attainment measurement (common in clinical and individualized programs)
- Participant (self-recorded) observations, such as logs and journals

- Role-playing, group simulations, and reflection (these are often part of a program’s activities, but can yield good evaluation information as well)
- Multi-media, such as video productions and photo journals
- Creative expression through theatre productions, poetry, and visual art
- Anecdotes and testimonials

We would like to point out that most methods do not automatically suggest the exclusive generation of quantitative or qualitative data. For example, focus groups usually are recorded and analyzed qualitatively, but one could also have focus group participants discuss an issue and then individually rate themselves on a scale. Surveys are effective at producing quantitative summaries through multi-item rating scales, but also tend to include written feedback from participants. An observational video of a program is highly qualitative in nature, yet researchers may quantify and count observed events. A transcribed interview is often organized into qualitative themes; yet these themes may also be coded and counted as a way to summarize the findings. In short, the production of both types of data really depends on the type of question you are trying to answer.

In this document, we provide brief summaries of each of the common approaches and their application to program evaluation contexts.



This document does not provide the depth of a methods textbook and we recommend further reading about the range of quantitative and qualitative methods and designs that are available. For example see:

Cresswell, J.W. (2009). **Research Design: Qualitative, Quantitative, and Mixed-Methods Approaches**. Thousand Oaks, CA: Sage.

Patton, M.Q. (2008). **Utilization-Focused Evaluation**(4th edition). Thousand Oaks, CA: Sage.

Patton, M.Q. (1987). **How to Use Qualitative Methods in Evaluation**. Thousand Oaks, CA: Sage.

Wholey, J., Hatry, H.P., & Newcomer, K.E. (Eds.) (2010). **Handbook of Practical Program Evaluation** (3rd edition). San Francisco, CA: John Wiley & Sons.

For resources specific to Food Security Issues and Community Food Action Programs, also see:

National Research Centre (2006). **Community Food Project Evaluation Handbook**. Boulder, CO:

Community Food Security Coalition:

Using Surveys

A survey is a generic term for a collection of tools, methods, and questions that are combined together for administration to program participants. They are especially useful for collecting a

lot of information at one time. A single survey can assess multiple outcomes, program participation, participant demographics, program satisfaction, and feedback. Most surveys combine qualitative and quantitative information. Often the main intent of a survey is to collect information from all program participants, or at least a good-sized sample. For this reason, surveys often collect different types of quantitative information, such as rating scales (e.g., self-ratings of coping skills regarding food insecurity), checklists (“do you shop for food at the following places?”) and other descriptive information (e.g., age, gender, level of program participation). Qualitative questions are often included to help provide some context to the quantitative pieces.

Surveys tend to compile a set of individual tools or measures that have been drawn from the literature or customized for the program to serve as indicators for selected outcomes. For example, a survey may ask participants to self-rate their knowledge of food preparation skills, their coping skills and strategies to deal with the stress of food insecurity, and their typical food buying and consumption patterns.

Survey design is a discipline in itself and there are many excellent resources if you want to learn more.¹ Some additional tips on survey design are provided here. You should note that there are competing pros and cons of different survey administrative methods, provided in the table below. The type of survey administration you select will depend on your specific program context, especially your ability to access participants.

Comparing Different Types of Survey Formats

Survey Type	Pros	Cons
Written survey, mailed to participants	No large advantages, but may be useful to reach people who cannot access an online version.	Must acquire full mailing address; time and resource intensive; often very low response rates.
Written survey while in program	A “captive audience” of program participants lends itself to a high response rate. Many participants can complete at once.	Must make time for administration within program; anonymity, if desired, is difficult to guarantee (although depositing sealed surveys in a “drop box” can be effective).
Telephone surveys	Provides for more personal engagement with participants; control over question delivery, opportunity to clarify, ask for elaboration.	People are less likely to provide phone number due to privacy; calling is time intensive; difficult to reach people at home; most people dismiss phone surveys as unwanted solicitation; anonymity is not possible.
Online survey	Time efficient; survey design and editing is easier; facilitates analysis; can reach many people quickly.	Misses people without computer & internet access; easier for people to decline, although ease of completion helps in this regard.

¹ For a great handbook, see Fowler, F.J. (1995). **Improving Survey Questions: Design and Evaluation**. Thousand Oaks, CA: Sage.

Measuring Knowledge and Skills

New knowledge and skills are often the first meaningful changes participants experience as a result of a program and are therefore important to measure. There are many ways to measure knowledge and it is worth knowing the options. Some are listed below from “most objective/reliable” to “least objective/reliable”:

- *Knowledge Tests* (e.g., quizzes, true/false, multiple choice items, etc.) objectively assess the knowledge of participants and are the most rigorous indicators of knowledge acquisition. However, such methods are sometimes not welcomed by participants, as few people like being “tested”.
- *Behavioural Indicators of Knowledge* ask participants to describe their behaviours in particular circumstances which strongly infer their knowledge of particular content, without the appearance of being tested. For example, the question “when you cook a healthy meal, what do you like to serve?” is less explicit than asking straight out “what are the components of a healthy meal?” Knowledge of healthy food preparation can be inferred by answers to the first question.
- *Self-Reported Knowledge* questions ask participants to rate their level of knowledge regarding the content of interest. This approach avoids the negative feelings associated with testing, but can create a “self-presentation bias” (people prefer to appear knowledgeable).
- *Self-Reported Knowledge Change*. Sometimes a pre-program survey is not possible to administer for comparison to a similar post-program survey (although it is recommended wherever possible). In such cases, participants may be asked to rate how much knowledge they feel they have gained in particular areas after the program is over. The self-presentation bias is still an issue here and suffers from requiring a retrospective (and less reliable) self-assessment.

In general, the best advice is to ask “testing” type questions in circumstances where assessing knowledge seems less threatening to people and to revert to behavioural indicators and self-knowledge ratings when testing questions seem inappropriate. Ultimately, what is more important is the wording of the questions (are they clear, concise, and strongly linked to the knowledge domains of interest?) and the design (can you compare answers before and after the program? Can you compare to some other group?).

Custom-Made or Existing Surveys

In most content areas of evaluation there are pre-existing survey tools that can be used to measure your outcomes. It is up to you to ensure that existing tools are appropriate and

strongly linked to your outcomes. The most reliable and valid standardized tool is effectively useless if it does not measure what you want it to measure. The benefit of using existing tools is that you do not have to invest additional time in survey design and you usually have some information of their reliability and validity, and examples of how it has been used. Sometimes, however, there may be some expense to purchase the tool.

Customized tools that you create yourself are often necessary to measure shorter-term outcomes that are specific to your program content, such as assessing knowledge. Custom tools can be made to be more sensitive to program context, but may be time consuming to create and must be piloted to ensure the information you need is generated.

In sum, you may be more likely to find good existing tools to measure outcomes associated with general food security concepts, such as access, availability and consumption. These outcomes, it can be argued, are shared by most food security programs, even if they are reached in different ways. For this reason, existing food security evaluation tools are likely useful across many programs.

Creating Scale Measures for a Survey

Measurement tools can take a variety of forms (fill in the blanks, true/false, checklists, rankings), but the most common are measurement scales. These are collections of questions that each requires a rating by the participant on a scale that provides a response range from 1 to some upper number. Five-point scales are most common, although many scale ranges exist (rarely going higher than 10). Below are some tips on creating your own scale measures:

- Provide enough scale points for people to comfortably rate their answer. 3-point scales may be frustrating for some people because they prefer to make some distinctions between moderate, strong, and very strong ratings (in either the positive or negative direction). Five-point scales are usually sufficient for most people, although 7-point scales provide finer distinctions.
- Provide a middle point (i.e., by way of an odd-numbered scale.) This is preferable for most people as they may dislike being forced to one side when they feel equivocal or uncertain. Some researchers, however, create even numbered scales for this very purpose – to force an opinion one way or another. This distinction is less important than having enough scale rating options.
- Ensure the scale is designed to measure a single concept (e.g., “food access”) from a variety of different angles. Many scale items taken together (e.g., summed or averaged) will provide a better indicator than any single item taken alone. Make sure

you create multiple items that ask the question in different ways and cover various (but highly related) elements of the same thing.

- “Agreement scales” are fairly easy to create and helpful for respondents because the meaning of the ratings is consistent. With agreement scales you can have a wide variety of content in the form of individual statements, to which participants can “strongly disagree”, “agree”, “strongly agree” and so on.
- Avoid complex items that contain many ideas. Participants need to be clear on what, exactly, they are rating.
- Some researchers vary scale items so that participants are asked to rate their agreement with statements that are positively worded (“I like to go fishing”) and statements that are negatively worded (“I do not like to go swimming”). The reasoning is that this prevents participants from mechanically rating the items similarly without giving each a reasonable amount of thought. For this reason, there is some benefit to varying item wording, but do so sparingly. Note that negatively worded items will need to be recoded for analysis – in other words, a rating of “1” on a negatively worded item is equivalent to a rating of “5” on a positively worded item.
- Always avoid double negative items (“I do not think that swimming isn’t fun”), which only serves to confuse participants.

Qualitative Questions in Surveys

It is highly recommended that any survey you create includes some qualitative information. Remember that a survey should attempt to answer multiple evaluation questions at once. While quantitative information answers the *how much* questions (e.g., how much knowledge do participants gain regarding food budgeting strategies?), qualitative information can help you answer the *how* questions (e.g., how did the program help participants learn food budgeting strategies, which ones were most helpful, and how were they helpful?)

Qualitative questions (e.g., when you ask participants to provide written feedback), are most useful when open-ended yet directive. The survey question “please add any comments you might have” is notoriously common yet inconsistent in what it yields from participants. It is preferable to provide a bit more focus in qualitative questions, such as “In what ways did you benefit from the program” or “how do you think the program could be improved?” Questions should be simply worded and linked to the experience of participants in the program.

Acknowledging the goal of brevity for most surveys, you should consider at least asking participants to:

1. Provide feedback on what they liked about the program.
2. Provide feedback on how the program could be improved (or what they did not like about the program).
3. Elaborate on how program activities helped them to achieve particular outcomes.

Focus Groups²

Focus groups involve gathering together a group of participants to discuss issues regarding their personal experiences in relation to the program.

Advantages to Using Focus Groups

- They are quick and cheap, and relatively easy to assemble.
- They can generate rich information from participants' own words and help to develop deeper insights about the program activities and outcomes.
- People are able to build on one another's responses and come up with ideas they might not have thought of in a 1:1 interview.
- They are useful for obtaining information from children or people who can't read
- They provide an opportunity to involve people in data analysis and interpretation (e.g.: "as you look at this list of issues we have generated, which ones seem most important for the program?").
- Most often focus groups produce results that are easy to understand.
- There is an opportunity for new connections to form among people, which may be useful beyond the evaluation.
- Quality and interpretation of evaluation data may be improved as participants act as checks and balances for one another, such as identifying factual errors or extreme views.

Limitations of Focus Groups

- The sample of people is usually not random, and often comes down to who is available and willing to participate.
- If few groups are conducted and the total number of people is small, it is more difficult to generalize the findings.
- The responses of each participant are not independent and potentially influenced by other group members in unexpected ways.

² Adapted from "Some Tips for Running a Good Focus Groups" (Author: Centre for Community Based Research, no date).

- A few dominant focus group members can skew the results to their views.
- Focus groups require facilitation skills to be well conducted.
- The information may be rich, but may be difficult to analyze because focus groups (despite best intentions) can become unstructured and off-topic.

Tips for Designing and Conducting Focus Groups

- Try to get a mix of at least four people and no more than 12 for a single focus group.
- Be aware of the energy levels of your participants and manage your time accordingly, but do not go beyond two hours.
- Do not have too many questions. Prioritize the most important questions to answer.
- Ensure you have a comfortable space for participants that is easy for them to access.
- Use name tags.
- Be very clear about the confidential nature of the focus group, emphasizing that what people talk about should not be shared outside of the session.

Types of Focus Group Questions

- Very unstructured questions provoke discussion of an issue or a situation without identifying any particular aspect of it – e.g. “so what’s your perspective on the program? How is it going?” This may be useful to loosen up the group and get them talking, but you will need to become more focused.
- Keep your questions open-ended, but grounded in the evaluation questions you wish answer. Try to make qualitative links between what the program does and what it has achieved from the perspective of participants – e.g., “I’d like to hear about the ways in which you have used the community kitchen.”
- Question probes should be designed to draw out the details of responses. Check responses with the rest of the group – e.g., “Have other people used the community kitchen in this way?”
- Questions can also be more focused and concrete - e.g., “ What would you say were the two or three most useful things you learned from participating in the community garden?”

Dealing with Challenging Group Members

- THE (REAL OR SELF-APPOINTED) EXPERT who controls conversation or intimidates others: Set ground rules for speaking, and invite each person, in turn, to speak.

- **FRIENDS WHO SIT TOGETHER AND FORM A "CLIQUE":** Avoid interviewing friends in the same group if possible, or have them sit apart.
- **THE HOSTILE GROUP MEMBER:** Remain objective and avoid personal confrontation, and try to allow the group to police itself – e.g. "do others in the group feel this way too?"
- **THE QUIET PERSON:** Ask for their input directly. If this does not increase their participation, ask them afterwards if they would like to do an individual interview. Gently inquire as to why they did not speak during the focus group.
- **THE OUTSIDER (a person who has a very different perspective than other group participants):** Characterize their differences as useful and important contributions to the group. Use their ideas as a topic of discussion (as long as it is “on topic” – their ideas are reasonably connected to the questions) to solicit alternative views. Avoid taking sides.
- **OFF TOPIC GROUP MEMBERS (people who seem to miss the purpose and focus of the group):** Restate and clarify the purpose of the group and ask more direct questions.

Developing and Conducting Interviews

Interviews are another common approach to gathering evaluation information. The essential difference between an interview and other methods is that an interview provides a one-on-one context for data collection. Technically speaking, the act of directly asking questions of a particular individual – be they open ended qualitative questions or quantitative rating type questions – is an “interview”. However, interviews are commonly understood to be conversational and qualitative in nature and for our purposes this is how we are using the term.

Interviews are useful when you are seeking to gain rich and detailed information about participants’ experiences in relation to a program. The disadvantage of interviews is that they are time-consuming to arrange, conduct, and analyze. Many evaluation designs attempt to achieve a degree of breadth with surveys of many program participants complemented by the depth achieved with individual interviews.

Constructing and Conducting Interviews

In some areas of social research there is a belief that interviews, if they are to be objective and systematic, must be highly structured and consistent. This means that questions are to be asked using identical wording for each interviewee and that elaboration, clarification and probing are discouraged. In evaluation practice, very few interviews actually proceed in this

manner, as some of the most important learnings are gained through following up and expanding upon the ideas of participants. Below are a number of tips to construct and deliver a useful set of interview questions.

- Structure your interview questions in relation to your evaluation questions and your program model. Make linkages between key concepts:
 - between activities and outcomes
 - between program experience, satisfaction, barriers & outcomes
 - between short and long-term outcomes
 - the circumstances under which people benefit and do not benefit
- Ensure your questions are not leading. For example, “What was your experience in participating in the community garden?” is less leading than “How has the community garden benefited you?” Once people start talking about benefits (or lack of benefits), you can then probe to get them to elaborate.
- The rule to avoid close-ended questions is probably overstated by many textbooks. Closed-ended questions (e.g., “did the program help you?”) are natural in conversations and normally lead to elaboration on their own. However, you must probe to ensure that they do (e.g., “Can you tell me how the program helped you?”)
- Do not ask questions about “quantity” (e.g., “how much did you benefit from the program”) as this can be easily gathered from people using a survey. Stick to qualitative questions that ask about participant experiences of the program – the *how* questions – because these help to shed light on your theories on how the program is supposed to work.
- Use probes to clarify meaning, confirm responses, contrast present versus ideal states of affairs, and follow up on new ideas that you may not have expected.
- Try using “event-centred” questions to illustrate key examples (e.g., “tell me about a time when you felt the members of the community garden were really working well together”).
- Do not make the interview overly long; get to the most important content quickly.
- Avoid hearsay, such as when interviewees comment on and evaluate the experiences of others. Focus on personal and concrete information.

Sampling for Interviews and Focus Groups

Sampling for interviews is improved when it is purposeful and strategic, rather than random, because the sample size is usually small (because interviews and focus groups are resource intensive). Small random samples may lead to groups of participants that do not capture the

full range of experiences with the program. It is even more problematic when a small random sample leads to the overrepresentation of extreme cases – i.e., people who absolutely adore the program or people who really dislike it. Sampling, then, is best accomplished by selecting individuals with an eye toward diversity. For example, you will want to ensure that the group sample is diverse in the following areas:

- Gender
- Family composition
- Income
- Level of program participation
- Neighbourhood
- Level of need
- Satisfaction with the program
- People who showed change in outcomes versus those who did not.

This presupposes of course, that you have access to this information. Having such information demonstrates the power of mixed methods. Creating a sample from survey results is a very useful strategy to link qualitative and quantitative information and to investigate possible group differences in relation to your program.

In some cases, the purpose of the interviews/focus groups is not to access a varied cross-section of program participants but to gain deeper insights from “community experts”. These are sometimes called “Key Informant Interviews”. Key informants are selected because they bring to the evaluation a level of knowledge, expertise, and/or experience that is unusual and potentially very useful. It is more common to use key informant interviews and focus groups in systems and policy level initiatives, since the content area and the nature of change is complex and multi-leveled. For example, key informant interviews or focus groups with leading service providers or community advocates may be used to gather information about how food security policies are being interpreted and implemented on the ground.

Program Outputs and Participant Satisfaction

We have already mentioned that effective evaluations need to collect information about program process (how the program is delivered and to whom) and program outcomes. The emphasis in this Guide is on the outcome side of the equation, but it is worth making some basic points about process data as well.

Program tracking and outputs

All programs need to gain an understanding of how their activities are delivered to participants. Outputs – which are countable measures of service delivery or activity – are usually already collected by programs to some extent. Outputs are particularly useful to get a handle on who is participating and to what degree, but they also provide a basis for investigating outcomes. For example, individuals who have low levels of participation cannot be expected to demonstrate the same degree of change as compared to those who participate at higher levels. The “high participation” group is the group that comes closest to representing your intended program delivery. Outputs can include things like:

- # of people participating in the program
- # of group activities offered
- # of people participating in each offering of a group activity
- # of people participating in multiple program components over time (i.e., level and consistency of participation)
- # of people requesting more information about the program
- # of promotional materials sent out to community members or partners (flyers, pamphlets, etc.)
- # of website visits, downloads, activity on discussion boards, etc.
- # of meetings and attendees (e.g., when tracking community partnerships)
- # of new deliverables, resources, papers, etc., that are created as a result of the program

In addition to tracking numbers of people, it is also useful to know *who* is participating. Wherever possible, try to collect information about participants, such as basic demographic information; this can include things like name, age, gender, family composition, income, neighbourhood, and so on.

How to collect and record output data

Since collecting output data is an ongoing data collection activity (as compared to outcome data, which is usually collected only a few times at specified periods), programs need to put a consistent and clear system in place. Staff play a crucial role in collecting output data. Consider the following methods:

- If possible, all program participants should complete some sort of simple intake or sign-up form that collects descriptive information about themselves. This should happen as early as possible during program contact. Assign a unique number to each person who comes into the program.
- If the program is anonymous in nature (such as a farmers market, where people are not “in” a program, but rather avail themselves of a service when desired), track the number of people who use it.
- Staff should record participation in distinct program activities, such as attendance at a community kitchen. Sign-in sheets are useful in this regard. Wherever possible, track individuals (who participated) rather than merely the total number who participated.
- Staff can complement program outputs by completing logs or ratings of the quantity and quality of participation, from their own observations of program activities.
- Create some master file or database of this information in a spreadsheet. The database should combine individual names (if possible) with program participation details (which activities they participated in, how many times, etc.).

Program satisfaction and feedback

Evaluation should always be focused on options and opportunities for improvement and there is no more valuable information than the perspectives and experiences of program participants. We should note that participant satisfaction and feedback can be collected using all the previous methods described – surveys, interviews, and focus groups. We highlight it here because it is a unique and important type of data. While this type of information can be collected quantitatively and qualitatively, we recommend the latter at some point nearing the end of program participation (or at some sensible point after a certain amount of program activity). A mix of the two can be very helpful. For instance, a survey could ask participants to rate their level of satisfaction on a series of scale items, followed by an opportunity for written elaboration. Every program will at least want to try to answer the following types of questions:

- What did you like or find most useful about the program?
- What did you dislike or find least useful about the program?
- Were there any barriers to fully participating in the program?
- What changes or improvements would you like to see made to the program?

Summary Concepts

Now is a good time to sort out some distinctions and overlaps between common measurement terms. We have used a range of terms that seem similar, such as data, indicators, measures,

tools, and methods. It is a fair to say that a lack of clarity in the field is frustrating. We try and make sense of some of these terms in the two tables below. The first provides some definitions and the second provides a summary and overview of the major approaches to data collection, including common administration methods, strengths and weaknesses, and usefulness for interpreting particular evaluation questions.

Common Measurement Terms and Definitions

Term	Definition and Detail
Data	Data is a general term that refers to the concrete information that is collected for a range of different purposes. Data are numbers, averages, percentages, words, categories, stories, and so on. Thus, data can be quantitative and qualitative. Information may be called data because of a desire to denote that it has been gathered systematically to answer a particular question.
Indicator	Information that helps to determine the degree to which your outcomes have been achieved. Indicators are derived from data that have been collected with the specific purpose of assessing outcomes. Note, however, that the evaluation field also talks of indicators of other things, such as program process, satisfaction, cost-effectiveness, etc. – essentially anything that you can gather data about. Indicators are <i>observable</i> . An outcome, such as “social support” cannot be observed until you decide what will <i>count as social support</i> . “Self-reports of feeling socially supported” may be an indicator. So might “number of friends” or “number of social engagements”. Indicators may be very specific, such as “average scores on a coping scale.” All things can be measured in different ways, from various angles, so there are always multiple indicators.
Measures (the noun) and Tools	The surveys, instruments, focus groups, checklists, etc., used to gather your indicators (e.g., a pre- and post-survey, the Coping Strategies Index, The Whole Measures Evaluation Rubric).
Survey	We would like to note that some <i>specific tools</i> are called “surveys” (e.g., the <i>Household Food Security Survey</i>). In this usage of the term, survey refers to a specific tool designed to measure something in particular. In this document, survey refers to the generic method of gathering multiple types of information through a set of questions. In other words, a survey often contains multiple tools and sections, each getting at something in particular.
Measure (the verb)	To systematically assess something, usually quantitatively. Measure strongly implies assessing <i>quantity</i> , which is why people don’t say that qualitative data “measure” something.
Methods	This is a broader term that refers to the general ways in which indicators and other data are collected. Methods describe the overall process and format of data collection. Surveys, focus groups, interviews, video observations, etc. are all methods to gather data.

Table 7 – Review of Major Approaches and Tools for Data Collection

Method	Type of data	Methods of administration	Strengths & Weaknesses	Useful to interpret...
Surveys	Mixed data, but emphasis on quantitative data (scale ratings, checklists, demographics) supplemented by short written feedback	<ul style="list-style-type: none"> Completed by staff in a face-to-face survey interview with participants Paper copies completed during program participation Email and online surveys (hosted on a website) Interviews completed by staff by telephone Paper copies by mail to be completed and returned to the program 	<ul style="list-style-type: none"> Access to many program participants Ability to summarize groups Can be used “pre-post” to detect change Not as useful to gain depth and details of individual experiences Response rates can be low when participants are “off-site” 	<ul style="list-style-type: none"> Participation in the program Changes in outcome measures Satisfaction with program
Focus Groups	Usually qualitative	<ul style="list-style-type: none"> Face to face small group (4 to 12 people, typically) Group discussion of key issues, experiences 	<ul style="list-style-type: none"> More efficient than interviews without sacrificing too much depth Some people may dominate the conversation, others may be shy People may be leery of sharing life circumstances with others or disagreeing with group consensus 	<ul style="list-style-type: none"> Satisfaction with program The relationship between program activities and outcomes
Interviews	Qualitative, although surveys may sometimes be administered in an interview format	<ul style="list-style-type: none"> Face to face By telephone Provide written questions for participants to complete (hybrid of survey and interview) 	<ul style="list-style-type: none"> Yields detail and depth of experience Small sample size; hard to generalize to full range of program participants Time intensive to conduct and analyze 	
Program Outputs	Quantitative “counts” of service delivery	<ul style="list-style-type: none"> Intake forms Sign-up sheets Staff logs and records 	<ul style="list-style-type: none"> Very useful to summarize participation levels and consistency Key ingredient in interpreting outcomes (but cannot assess outcomes on their own) Significant time required from staff Challenging to collect and record data in a consistent manner over time 	<ul style="list-style-type: none"> Who is using the program Participation levels Group differences on outcome measures (when combined with outcome data)

Analysis of Your Evaluation Data

You have conducted an evaluation, systematically collecting information from program participants and other stakeholders to answer important questions about your program's operation and impact. At this point program staff may become overwhelmed with varied and detailed data and information; how to analyze and interpret it is no easy task.

All analysis, be it qualitative or quantitative, involves some form of *data reduction* – no set of data can be understood in any meaningful way without somehow reducing it down to manageable bits, groupings, and summaries. Think about a baseball game. In every game, there are at least 10 players on each team and 9 innings of play. The game is composed of many pitches, hits, base running, fielding plays, and outs. How are games described? Qualitatively they may be described as “exciting”, “dull”, “back and forth”, “skilled”, and so on – these are qualitative summaries of many observations. Quantitatively speaking, it would be difficult to make sense of what happened in a game by listing in succession every pitch, pitch count, hit, play, base runner, and so on. Instead baseball games are summarized with a brief set of informative numbers and statistics: the total score, hits, runs, and errors. More specific information can be used to answer more specific questions, such as “how well did the pitcher perform?” Indicators, such as the proportion of strikes to number of pitches thrown, hits, walks, and strikeouts, help to answer this question from different angles. How well did a pitcher pitch *this whole season*? Sums and averages of various indicators can summarize many observations (i.e., games) into one snapshot of performance, and observations taken at different time periods can demonstrate upward and downward trends.

Analysis of evaluation data is hardly any different in principle and strategies can range from simple description to complex analytic techniques. Statistics can be straightforward and familiar to most people, but they can also become a bit more complicated, especially when using the sort of inferential statistics that help determine the “significance” of change. On the qualitative side, many long and detailed interviews can be burdensome and complex to summarize together. Quantitative and qualitative analysis are both disciplines in their own right and entire courses and textbooks have been devoted to them. We encourage you to read more. However, many basic approaches to analysis are straightforward, informative, and extremely useful to answering evaluation questions.

Quantitative Data Analysis

All quantitative data *represents* information, such as categories, quantities, frequencies, and averages. It is primarily used to summarize many observations together. With a few simple and sensible operations, you can capture a lot of information (e.g., food security scale ratings of 50 program participants) to arrive at a general interpretation of status or change.

First Things First – Entering, Cleaning Up, and Coding Your Data

Quantitative data is best organized and analyzed using an electronic database. Microsoft Excel is the most commonly used spreadsheet program and can handle all the descriptive statistics that you need. If you hope to use inferential statistics (e.g., a “t-test” to see if changes over time are statistically significant) then a statistical software program such as SPSS is more appropriate.

Proper analysis and interpretation rests entirely on the accuracy of your data. This requires clarity of surveys and tools for people completing them, followed by accurate data coding entry. Data coding is the simple transformation of responses (words, checkmarks, circling of ratings, and so on) into numbers, because numbers are much easier to work with when using a database. Here are some tips:

- Assign every participant a code number that is consistently used across all data collection. Ensure this number is on every page of every survey, interview transcript, intake form, etc. This number is used in the first column of your database and when all data associated with that person is entered in subsequent columns.
- Decide on your coding scheme for categorical questions. Coding categorical data is arbitrary because the numbers merely tell you group membership, rather than quantity. For example, if gender is recorded, you may want to specify F=1 and M=2. The coding scheme could be F=7 and M=3000 and it would make no difference. All that matters is that the coding is consistently applied. Pick codes you will remember.
- Decide on your coding scheme for scale questions. For scale questions, quantities are being measured and the values of the codes matter. Often the scale numbers are presented in the measure (e.g., 1 to 5 for each item) but sometimes only the “value labels” are provided (strongly disagree, disagree, neutral, agree, strongly agree). In this case, you have to code them for data entry (e.g., 1=strongly disagree to 5=strongly agree).

- As a general rule, scales tend to be scored as “higher is better” (more desirable and beneficial), so higher ends of scales are the “better” ends of the scales for later interpretation.
- Make sure you reverse code negatively worded items. For example, if most of your 5-point scale items in a measure are positively worded (e.g., “I feel knowledgeable about how to prepare healthy meals”), any negatively worded items (e.g., “I do not feel confident knowing what to buy at the grocery store”) need to be recoded. A rating of “1” on a negative item is equivalent to a “5” on a positive item. This is a requirement if you want to produce an average scale value, and generally makes all analysis a little easier.
- It will help to take blank copy of your survey and write in your codes on it next to each question. This will provide a useful guide for data entry.
- When entering your data it helps to have two people. One person reads off the data points while the other keys in the data. This is much faster and more accurate.
- When all your data is entered, generate some descriptive statistics of the data to make sure all the data is in range. For example, any averages over five in relation to a 5-point scale indicate data entry errors.

Statistics – What are they and what can they tell us?

Statistics are numerical summaries of information and most of us are familiar with some of the common terminology. We constantly hear about the “average” (called a “mean” in statistics) of one thing or another, such as average income, grades, age, and so on, in relation to innumerable social issues. “Frequencies” are also common, which are numbers of cases or observations that fall into a particular category. For example, the number of smokers versus non-smokers in a population or the number of males versus female voters are frequencies. Frequencies can also be expressed as percentages (“20% of the city population buys food on credit at least once a month”). Other statistics include the “median” (the middle score of a group) the “mode” (the most common score in a group), and the standard deviation (the average “spread” of scores in group). In answering simple evaluation questions, the use of frequencies and group averages is most common, although looking at other descriptive statistics can be helpful.

Another useful statistical technique is correlation. This method is simple to use in statistical software programs and provides an indicator of the strength of the relationship between variables. For example, it is often useful to correlate participation rates (e.g., number of hours attending the community garden) with outcome scales (e.g., average score on food affordability

measure). This will tell you if your outcomes vary with the *amount* of participation. A correlation (abbreviated as “r”) ranges from 0 to 1 (or -1), where 0 represents no relationship, and 1 means a perfect positive correlation. A -1 means a perfect negative correlation, in which high scores on one variable are related to low scores on another variable. The higher a correlation in either positive or negative direction, the stronger it is (over .70 or .80 is “high”). Remember that a correlation does not equate with cause, although there could be a causal relationship. For example, there is a good theoretical (and common sense) reason to presume that greater participation in a program “causes” improved outcomes.

Inferential statistics are more complex and require more advanced statistical training and ability to use statistical software packages. A full description is beyond the scope of this document. However, inferential statistics can be very useful in answering certain questions about difference and change. The theory is that groups of people – including the same group at one point in time compared to a later point in time – are either different due to chance or different because they are *truly* different. Inferential statistics provide the probability that differences are real and not random. For example, a researcher can say that “a difference between two groups is a *true difference* 95% of the time, which leads us to believe this particular difference is probably real”. This is what is meant by “statistical significance”.

The fact that two groups are truly different does not speak to the nature or cause of that difference. Sometimes it’s plainly obvious from your evaluation design – we measured this group on content knowledge, taught them the content knowledge, and then measured them again and there was an increase. It is fair, in this case, to attribute the change to the intervention. However, you might find group differences based on gender – the difference may be significantly different, but without other information, the nature of that difference is unknown and merely conjectural.

Letting Evaluation Questions Guide Quantitative Analysis

Once you have all your data entered, there may be a desire to analyze everything at once and to become overwhelmed with the possibilities. When this happens, go back to your evaluation questions and set about answering them first. This will provide greater focus and may in turn lead to other questions in a more systematic way. Begin by describing your sample, by generating frequencies for your demographic data, such as ages, gender, income, etc. This might come in handy later on as you examine your outcomes.

In the instances where you have collected quantitative data, look at your evaluation questions. There are a limited number of question *types* that can be asked in relation to quantitative data.

Below is a table that provides some of the more common types of questions, some analysis options, and examples.

Analysis Options for Quantitative Question Types and Data

Example questions to be analyzed quantitatively	This is a question about...	Analysis options	Example
We have lots of demographic data and want to understand our participants better.	Sample description	Create frequencies & percentages for data categories.	Males: 29 (46%) Females: 34 (54%)
We have a 10-item food insecurity coping measure and want to know where participants stand in their level of coping.	Results description	Create response frequency table for each item, with item and scale average	See example table below.
We got a sense of a lot of variation in the level of participation in the program. We want to sort out participation levels and relate participation to outcomes.	Relationship between groups (in this case, groups based on participation rates) and outcomes.	Divide your group into low and high participation and compare outcome measures. Run a correlation of participation to outcomes. Conduct a t-test to compare group averages.	High participation group average = 4.21 versus low participation group average = 3.13. Correlation: $r=.72$, suggesting a moderately strong relationship.
We collected data using a food security measure before and after our program and we want to know if there is any improvement.	Change in outcomes	Compare pre- and post-test average scores. Conduct a “paired t-test” to examine statistical significance of difference.	Difference of averages is statistically significant at $p < .05$ (the difference is due to chance only 5 times in 100).
We want to know if people think food budgeting skills is related to perceptions of food affordability.	Relationship between outcomes	Run a correlation of outcome to outcomes.	$r=.43$ shows a positive but somewhat weak relationship.

A frequency table is a common analysis tool, as well as a handy way to combine and present information especially when looking at multi-item scale measures. An example is provided below. One can immediately see patterns of scores, and the relationships of different items to one another. If you create a table like this at pre-test and post-test you can also see patterns of change over time, in response frequencies and item averages.

Example Frequency Table for a Scale Based Measure

Survey Item – In the last 7 days...	Strongly Disagree		Neither Agree or Disagree				Strongly Agree		Average
	1	2	3	4	5	6	7		
I/we had very little food in the house at any one time.	0%	0%	0%	2 (10%)	4 (21%)	5 (26%)	8 (42%)	6.00	
I/we had to go other places in the community to get food because we didn't have enough food in the house	0%	0%	0%	4 (21%)	4 (21%)	4 (21%)	7 (37%)	5.74	
I/we had to borrow money or buy food on credit in order for us to have enough food in the house.	1 (5%)	0%	0%	5 (26%)	8 (42%)	4 (21%)	1 (5%)	4.84	
I/we had to borrow food from family or friends because I/we did not have enough food in the house eat.	4 (21%)	3 (16%)	2 (11%)	3 (16%)	5 (26%)	1 (5%)	1 (5%)	3.47	
Overall scale average:								5.07	

Items are excerpted from the *Promising Practices Food Security Tool* (Glacken, 2010). The 7-point response labels and range have been altered for illustrative purposes only.

Qualitative Data Analysis

Qualitative analysis can be a very detailed, in-depth process, and we sometimes hear of research projects in which there are long-running “case studies” that follow individuals with personal interviews over long periods of time. But for most evaluation projects, qualitative analysis has many of the same concerns as quantitative analysis – to reduce and summarize a lot of information and to do so systematically and consistently. The biggest difference is that qualitative analysis makes far fewer assumptions about what is important to participants and

does not presuppose that everything to be known about a program can be guessed in advance and measured. Qualitative analysis, then, is more inductive and open and is concerned about processes, experiences, and relationships between things. And, yes, it can also tell us a lot about outcomes.

First things first – Managing Your Qualitative Data

How you record and store your qualitative data will depend on how it is collected. Written feedback questions in surveys, if not too lengthy, can just as easily be entered in your database alongside your quantitative data. In fact, this can be especially useful when a qualitative question is linked to quantitative data, because the former can provide useful information to interpret the latter (e.g., a low rating on a satisfaction scale followed by a comment that explains this low rating). Copying and pasting excerpts from your spreadsheet into tables is also straight forward.

However, if your data is gathered from focus groups or interviews, the data require some form of storage and organization. Wherever possible, we recommend audio recording participant responses. Even the fastest note-takers miss information and this is even more of a problem if the person taking notes is also conducting the session (not advisable!). Having an audio recording provides a back-up to on your notes and, if you have adequate time and resources, transcribing the audio recording for analysis is even better. Transcription can be time-consuming but worth the effort if interviews or focus groups are long and complex.

As with quantitative data collection, make sure you properly identify your participants with their unique code numbers. Audio record the code number and date at the beginning of the interview or focus group.

Coding Qualitative Data

The coding of qualitative data is meant to summarize ideas and information within and across participants. There are many ways in which to code qualitative data and this is a reason why the method is criticized by some as overly subjective. However, there are many strategies to make coding more systematic and consistent. You first have to decide the format you for coding the data. You can print hard copies of transcripts or notes and write in the margins or use the comment function in your Word processing program (e.g. MS-Word). Highlighting text with different colours that correspond to different code categories is also helpful. Another

option is to use qualitative software programs (e.g., Nvivo or AtlasTi) that are effective for organizing and coding your data, especially when there is a lot of it.

Here are some helpful tips for qualitative coding:

Begin with a start list of general codes. Before you even start your interviews, generate a list of code areas you are interested in. These should be general codes that follow the intent and content area of your evaluation questions. For example, if you are interested in barriers to participation in a program, “Barriers” could be a code title. When making a first pass of your data, you can quickly apply this high level code to examples of barriers. More detailed codes (e.g. “BARRIER – TRANSPORTATION”) can then be applied to each general code. And of course, you should add any other general codes that emerge that you had not identified in your start list.

Generate more specific subcodes. There are many ways to code verbal information and much of your learning will come by actually doing it. Some suggestions include coding for context and conditions, interactions between people, actions and consequences, reports of behaviour, personal perspectives, and so on. For example, you may get a variety of responses to the interview question “How does participation in the community garden benefit you?” The start code would probably be something simple like “BENEFITS”. Subcodes may include things like “BENEFITS-HEALTHY FOOD”, “-ACCESS FOOD”, or “-MEET PEOPLE”.

Create detailed definitions of your codes. Code labels are only meaningful to the coder and it is therefore important to create code definitions so the meaning is clear and accessible to others (and to yourself, as you code). The code “BENEFITS-ACCESS FOOD” may be defined as “instances where participants report a benefit of having greater access to food for their household”.

Write “memos” to yourself about the coding process. This is a very important feature of coding and creates a link between coding and further analysis. When you see interesting things in the data – things you don’t expect, that you find confusing, or that are particularly informative – write a little memo to remind yourself to follow up on the idea later on.

Analyzing Your Qualitative Codes

The distinction between coding and analyzing data is a blurry one. Certainly a degree of analysis is happening as you apply codes. The distinction, if it helps, might be to say that the coding phase is a process of *describing* a lot of information in a systematic way, whereas the

analysis phase is to make sense of the information by summarizing, examining consistent patterns, and paying attention to exceptions. This is quite similar to the quantitative approach – people responding to 5-point scale items are essentially coding themselves. Analysis and interpretation happens after this coding, through summary and interpretation. Here are some tips for analyzing your data once coding is complete:

- Do counts of general and specific coding instances. Counting is a useful first step in understanding patterns. It will be useful for your program to discover that one type of barrier is consistently mentioned by participants or that one program component is seen as most helpful.
- Investigate exceptions in your data. For example, if you find that one or two people are saying something completely different or unique about their experience in the program, finding out why could be useful and provide new insights.
- Look at *conditional* information. In other words, the opinions of participants may only apply under certain circumstances, which may be very important. For example, users of a mini-market may only find it useful to them in summer months when vegetables are plentiful, as compared to winter months.
- *Pose “if-then” questions.* Qualitative analysis in part involves the testing of ideas. If a program component is designed to lead to a particular set of outcomes, you need to test this relationship (provided this was a focus of the interview/focus group in the first place). Ask if-then questions and investigate whether they hold up in the data.
- Compare and contrast different people, groups, and/or codes by creating matrices. A matrix is simply a table that crosses two or more types of information. For example, you may take the codes of participants from one neighbourhood and put them in one row to compare to the codes of another neighbourhood, in a second row. Two columns could be “Most helpful part of program” and “Least helpful part of the program”.

Table 13 – Compare and contrast of qualitative codes

	Most Helpful Part of Program	Least Helpful Part of Program
Neighbourhood 1 (n=6)	Learning gardening techniques (4) “Our food system” panel (3) Food Basket (3)	Food budgeting workshop (6)
Neighbourhood 2 (n=7)	Food Basket (5) Food budgeting workshop (4)	“Our food system” panel (4)

Improving the Rigour of Qualitative Methods and Analysis

Qualitative analysis is vastly improved when the method to gather the information is well thought out and executed. We spoke about strategies to improve data collection in previous sections. Let's assume that you conducted comprehensive interviews with a suitable cross-section of program participants. How do you ensure the analysis itself is credible and "trustworthy"? There are several strategies to improve your analysis in this way:

- Make sure you have a way to track where you drew information from to reach your conclusions. For example, if you are asked about what data contributed to a particular code, you should be able to find that data quickly. Qualitative software programs can assist with this, but it is otherwise helpful to list participants that have been coded on each code.
- Clearly define each code.
- Have more than one person code the same interviews using the same code definitions. Do they arrive at the same coding? Debrief and challenge one another on your interpretations and modify code definitions if necessary.
- Link your qualitative analysis to quantitative findings and other information. If you have quantitative evidence of group differences or change over time, can it be supported and explained by your qualitative data?
- Do not overstate your findings. If a finding is drawn from only one or two participants, do not generalize to others.
- After interpreting the data, consider feeding back this interpretation to participants. Does it make sense to them? Did you capture their perspectives fairly?
- Memoing serves another function beyond reminding you and prompting you about your ideas. Memoing demonstrates that you were comprehensive in working with your data. Describe this process of memoing and where it led you in your analysis.
- Do not ignore contradictory or disconfirming data. Make sure you report it and do your best to explain it. Why were one or two people very different? What do you know about their experience that could explain this difference?

The Meaning of Data: What counts as success? What is a good finding?

"When the coyote bounces after falling off the cliff, does the second time he hits the ground count as a second incidence of violence?" (author unknown)

Evaluation, by definition, is about *interpretation and judgement of value*. Analysis will tell you what you found, but the final challenge is coming to conclusions on what it all means: is your information meaningful, important, and actionable? It is very important to establish clear definitions of what certain types of data mean. For example, if you are measuring food security, what data scores represent an acceptable level of food security? What participant scores or comments reflect “poor coping skills and strategies”?

A central part of interpretation involves deciding on what counts as a “good finding” and this is best done in advance of analysis and in collaboration with the project stakeholders. Failing to have this discussion can lead to explanations of the data after the fact, to discount, mitigate, or dismiss less positive findings. For example, let’s say a local food system resource centre tracked visits of local food vendors to their office and found that 85% of vendors only visited once, and 15% visited two or more times. Staff may make “after the fact” interpretations like:

Staff 1: *That’s about what I anticipated*

Staff 2: *Plus, remember, the data do not include vendors in our workshops and special classes.*

Staff 3: *I think the observation time was really too short*

Staff 1: *I agree. January and February are bad months, you know, everyone depressed by winter....³*

However, if desired standards are agreed upon in advance of analysis, then you have established consensus in the interpretation and greater credibility in the findings. This discussion may also alert you to additional data that need to be collected for improved interpretation (e.g., reasons for visits). Example standards for interpretation, selected in advance of analysis, appear below.

Pre-decided Judgements Based on Potential Findings (adapted from Patton, 2008)

Judgement	% and number of food vendors who have contact with the office 3 or more times. Total number target is 25.
We’re doing an outstanding job of engaging food vendors.	At least 20 vendors, 80% of total vendors visiting
We’re doing an adequate job of engaging food vendors.	At least 15 vendors, 60% of total vendors visiting
We’re doing a poor job of engaging food vendors.	10 or few vendors, 40% or less of total vendors.

³ Adapted from Patton, M. Q. (1997). Utilization-Focused Evaluation. 3rd Edition Sage: Newbury Park, CA.

Establishing such standards is easier for things that can be counted. For example, you may set a standard for the number of people who report that a particular program component was valuable to them, and so on. When examining change or group differences, statistical significance may be used as the criteria for difference (and therefore “success”).

Communicating and Reporting Your Findings

Reporting your findings should be much more than writing a technical report for people to read if they are so inclined. On the contrary, the reporting and dissemination phase of evaluation is a continuation of your findings and interpretations, as others begin to digest, work with, and sort through the implications of the results.

There are numerous ways to present and disseminate your findings. Consider the following options:

- Technical reports
- Academic Publications
- Executive Summaries
- Fact sheets
- Newsletters
- Web pages
- Formal and informal presentations
- Community forums
- Debates and panels
- Narratives, stories, pictures, video.

The options you select will depend upon the needs and expectations of your audiences, thus knowing your audience is very important. A program funder may require a full technical report while community members may only need a fact sheet or summary. In all cases, clarity is paramount. Most people only have a limited amount of time (and memory!) for your findings. Ask yourself the following questions:

- What are the three most important things about your evaluation?
- If you had 30 seconds in an elevator to tell someone what you learned through your evaluation work, what would you say?

While analysis is a structured and technical activity, reporting introduces an element of creativity. This does not mean “creatively telling a story people want to hear” – you must remain true and fair to the findings – but creatively engaging your audience so they care to listen.

Dense technical details should be avoided (although available in an appendix or separate document if requested) in favour of graphics, tables, pictures, and quotes. When summarizing data, ask yourself what the audience needs to know. Do they need to know the frequencies and percentages of every scale item of measure, such as we presented in Table 11? They might not. If so, what would be a meaningful summary of this data?

Here is an example of building a summary data display with clarity in mind. Below are the levels of usefulness of different components of a food security program, listed by percentage of responses. The items are listed as they appear in the survey. This is lot of information to take in and it is difficult to quickly assess the priorities as expressed by participants.

Program Component	Extremely useful	Quite Useful	A little useful	Not Useful
Food sharing program	35	36	13	16
Community freezer	33	38	19	10
Coupon program	42	28	9	21
Food budgeting workshop	26	45	25	4
Food basket	58	13	6	23
Recipe sharing	47	22	15	16
Community gardening skills	33	38	10	19
Community kitchen nights	40	31	12	17
“Your Local Food System” Panel	29	39	16	16
Food Matters newsletter	11	58	17	14

This is raw data and it is a necessary first step. But it may not be necessary, in this form, to communicate to many audiences. If your goal is to communicate what is most useful in the program, the display can be simplified into a rank order of the “extremely useful column”. This clearly displays participants’ priorities.

Program Component	Extremely useful
Food basket	58
Recipe sharing	47
Coupon program	42
Community kitchen nights	40

Food sharing program	35
Community freezer	33
Community gardening skills	33
“Your Local Food System” Panel	29
Food budgeting workshop	26
Food Matters newsletter	11

Tips for Making Recommendations

Stakeholders of an evaluation often want some solid commentary of the bottom line – what are the main recommendations or implications of your evaluation? Recommendations represent the transition point between evaluation and action and are therefore an important stage in reporting.

There is a real art to crafting good recommendations. They need to be challenging enough to prompt meaningful change, but they also need to be closely tied to the evaluation data, constructive, and achievable.

Below are some tips in developing and reporting your recommendations.

1. Ask your audiences what kinds of recommendations they would find most useful.
2. Try to include a few small recommendations that are easy to achieve, as well as a few bigger, more ambitious ideas.
3. Avoid the temptation to put the responsibility for change on people, organizations, or systems that have not been part of your process.
4. Recommendations should be linked to and follow clearly from evaluation findings (this is not as obvious as it might seem).
5. Distinguish different types of recommendations
 - a) major versus minor
 - b) short-term versus long-term
 - c) changes to program implementation (“let’s do things a bit differently, or more consistently”) versus program model or approach (“let’s reconsider how the program is structured and delivered”)
 - d) based on strong evidence versus less strong evidence
6. Consider multiple options in implementing recommendations, with pros and cons associated with each.
7. When dealing with major changes, include a discussion of costs and benefits.
8. Limit yourself to actions that are within the control of program stakeholders.

9. Be politically sensitive; find out political implications of recommendations.
10. “Pilot” recommendations with others before finalizing them in a report.
11. Don’t forget that inconclusive evaluation findings can often lead to very good recommendations. They can be used to identify questions for the next wave of evaluation, for example.