

# **Workbook 2**

# **Implementing Evaluations**

© World Health Organization, 2000



WHO  
World Health Organization



UNDCP  
United Nations International Drug Control Programme



EMCDDA  
European Monitoring Center on Drugs and Drug Addiction

This document is not a formal publication of the World Health Organization (WHO) and all rights are reserved by the Organization. The document may, however, be freely reviewed, abstracted, reproduced and translated, in part or in whole but not for sale nor for use in conjunction with commercial purposes. The views expressed in documents by named authors are solely the responsibility of those authors.

# Acknowledgements

The World Health Organization gratefully acknowledges the contributions of the numerous individuals involved in the preparation of this workbook series, including the experts who provided useful comments throughout its preparation for the Substance Abuse Department, directed by Dr. Mary Jansen. Financial assistance was provided by UNDCP/EMCDDA/Swiss Federal Office of Public Health. Cam Wild (Canada) wrote the original text for this workbook and Brian Rush (Canada) edited the workbook series in earlier stages. JoAnne Epping-Jordan (Switzerland) wrote further text modifications and ed-

ited the workbook series in later stages. Munira Lalji (WHO, Substance Abuse Department) and Jennifer Hillebrand (WHO, Substance Abuse Department) also edited the workbook series in later stages. Maristela Monteiro (WHO, Substance Abuse Department) provided editorial input throughout the development of this workbook.

Some of the material in this workbook was adapted from a NIDA publication entitled “How Good is Your Drug Abuse Treatment Program? A Guide to Evaluation.” Contributions drawing from this report are gratefully acknowledged.



# Table of contents

Overview of workbook series	6
Introduction	7
The six steps of implementing evaluations	8
Step 1. Prepare for data collection	10
Step 2. Colect data	22
Step 3. Analyse data	24
Step 4. Report results	40
Step 5. Make use of what was learned	44
Step 6. Start again	46

# Overview of workbook series

This workbook is part of a series intended to educate programme planners, managers, staff and other decision-makers about the evaluation of services and systems for the treatment of psychoactive substance use disorders. The objective of this series is to enhance their capacity for carrying out evaluation activities. The broader goal of the workbooks is to enhance treatment efficiency and cost-effectiveness

using the information that comes from these evaluation activities.

This workbook (Workbook 2) describes step-by-step methods for implementing evaluations. These steps span from starting the study, to collecting, analysing, and reporting the data, to putting the results into action in your treatment programme.



## **Introductory Workbook** Framework Workbook



## **Foundation Workbooks** Workbook 1: Planning Evaluations Workbook 2: Implementing Evaluations



## **Specialised Workbooks** Workbook 3: Needs Assessment Evaluations Workbook 4: Process Evaluations Workbook 5: Cost Evaluations Workbook 6: Client Satisfaction Evaluations Workbook 7: Outcome Evaluations Workbook 8: Economic Evaluations

# Introduction

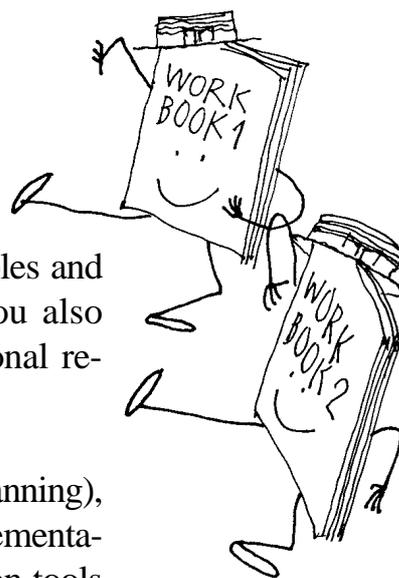
You should use the appropriate specialised workbook with this workbook as you move through the steps of evaluation implementation.

This workbook describes the general steps involved in conducting treatment evaluation, including selecting/preparing your data collection measure(s), deciding how to administer the measure(s), and developing a data collection and management plan. The workbook then provides advice and some technical details regarding analysing and reporting the results. Some sections should be read by all evaluation partners, whereas others are most appropriate for those trained as researchers and/or functioning as evaluation consultants. Each section is marked appropriately.

As the skills and resources that are available for evaluation vary widely, you may not be able to undertake all of the steps and procedures that are outlined here. However, you should try

to adhere to the basic principles and steps that are presented. You also should try to acquire additional resources you may need.

Workbook 1 (on evaluation planning), and this workbook (on implementation), will give you foundation tools and information to plan and implement programme evaluation. With an understanding of the material in these two workbooks, and some practice applying it, you will be prepared to undertake evaluations of different types. The remaining specialised workbooks in this series present the principles and practices of these different types of evaluation. You should use the appropriate specialised workbook with this workbook as you move through the steps of evaluation implementation.



# The six steps of implementing evaluations

As described in the framework manual, there are six steps that you need to accomplish to implement evaluations:



1. Prepare for data collection.
2. Collect data.
3. Analyse data.
4. Report results.
5. Make use of what was learned.
6. Start again.

Each of these steps is presented below. The discussion of these steps includes some exercises for you to follow that will help

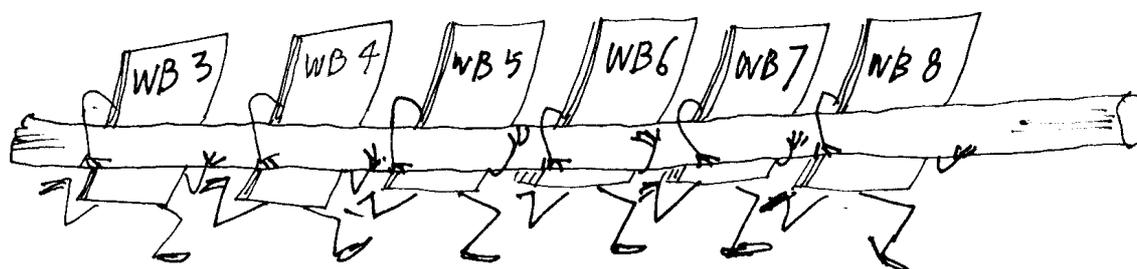
you learn and apply the material to your own situation.

## Use the foundation and specialised workbooks together, to help you make the most of the information that is presented.

Workbooks 1 and 2 provide a solid foundation of general information about conducting evaluations, whereas the specialised workbooks (Workbooks 3 through 8) present detailed information for different types of evaluation. If you already know what type of evaluation you are going to conduct, you should consult the

workbook that is applicable. Use the foundation and specialised workbooks **together**, to help you make the most of the information that is presented. If you do not know what type of evaluation you are going to conduct, wait until you have developed your evaluation questions (Step 5) to consult a specialised workbook.

If you are conducting a...	Then you should review...
Needs Assessment Evaluation	Workbook 3
Process Evaluation	Workbook 4
Cost Evaluation	Workbook 5
Client Satisfaction Evaluation	Workbook 6
Outcome Evaluation	Workbook 7
Economic Evaluation	Workbook 8



# Step 1

## Prepare for data collection



There are a few important tasks that need to be accomplished when starting your study. These include:

- 1 managing ethical issues
- 2 developing a data management plan

3 conducting a pilot test

4 writing an evaluation plan.

These tasks are explained in detail below.

### 1. Manage ethical issues

**No person should be forced or coerced to participate in the study.**

People being studied need to be protected from undue burdens on their time, invasion of privacy, and risks or other harms. You need to pay attention to these ethical issues!

For all studies involving people as participants (sometimes called “human subjects” studies), you must first ask approval from an ethics committee, and/or consult the WHO-CIOMS publication, **International Ethical Guidelines for Biomedical Research Involving Human Subjects** (1993). These guidelines are intended to protect the rights of participants. Some of the basic principles are outlined below.

**Informed Consent.** According to this principle, researchers must ask permission of individuals before enrolling them in the research study. When you do this, you

should explain to them the purpose, nature, and time of their participation. Give them, or have available upon request, detailed information about the evaluation in writing. No person should be forced or coerced to participate in the study.

The standard practice is to have a Consent Form that informs clients about these procedures and that asks for their written consent for data to be collected from them or about them. A consent form typically:

- describes the purposes and methods of the study explains study requirements, for example, completion of a questionnaire, the kinds of information requested, the time to complete the questionnaire, the amount of time involved

- describes the purposes and methods of the study explains study requirements, for example, completion of a questionnaire, the kinds of information requested, the time to complete the questionnaire, the amount of time involved
- explains that participation is voluntary and individuals can withdraw from the study at any time without fear of reprisal such as withholding of treatment that the person would otherwise be entitled to receive
- explains the extent to which confidentiality will be maintained
- explains all foreseeable benefits and risks that the study poses
- explains any monetary or other benefits (if applicable) that accompany participation
- uses simple words and concepts that can be easily understood by participants
- explains that they are entitled to free medical care for any harm that results from participation

The investigator must communicate all necessary information, give the prospective subject full opportunity to ask questions, and exclude the possibility of unjustified deception or intimidation in the consent process. The participant must sign the consent form before the information is collected, and must receive a copy of the signed form. If a participant is illiterate, you must read the consent form and obtain verbal consent. Note that consent forms also may be needed for evaluations that use archival records about individuals.

**Take necessary precautions to protect the safety of your participants.**

## Risks of participation

Before beginning a study, researchers should thoroughly evaluate risks that might be present for participants. The typical risks in evaluating treatment programmes for PSU disorders are accidental disclosure of the information obtained about patients' PSU and illegal activities. In addition, evaluations involving biomedical interventions or confrontational tactics could harm clients physically or psychologically. Take necessary precautions to protect the safety of your participants.

Participants may be paid for their inconvenience and time spent, and may receive free medical services in the course of participation. However, payments should not be so large, or medical services so extensive, so that they would induce people to consent to participate against their better judgement.

## Confidentiality

All researchers have an obligation to protect the confidentiality of their participants.

Researchers must never reveal the name, address, or any identifying information about a participant unless they have written permission to do so. If researchers want to publish a case report of a participant or a group of participants, they must change enough details so that no one could discover the identity of the person or group.

You should develop procedures to ensure that all information collected about any person will remain anonymous. No names or other data should be kept that could identify an individual with the information gathered. If the data collection plan includes follow-up with clients over time to measure services received or individual outcomes, each client can receive a unique identification (ID) number. This number should be the only identifier used in data processing and analysis. Only authorised personnel should have access to the names and other information linked to the ID numbers for use during the evaluation. In addition, everyone who collects or prepares data should sign a pledge signifying they will adhere to all confidentiality procedures.

**You should develop procedures to ensure that all information collected about any person will remain anonymous or confidential.**



## Into action

Our evaluation team discussed all possible risk that might be present for participants and how they are going to ensure that all participant information will remain confidential. They also consulted the WHO-

CIOMS publication, International Ethical Guidelines for Biomedical Research Involving Human Subjects (1993) in order to write a consent form for their evaluation project.

They created the following consent form:

*Our inpatient opiate detoxification programme is conducting a study the effectiveness of our AIDS education intervention. Participants will complete a questionnaire at their initial assessment interview at arrival into our programme. The questionnaires will ask about attitudes, beliefs, HIV-risk behaviour and psychoactive substance use. Completion of the questionnaires will take about 15 minutes. Participants are free to skip any questions that they find troubling or overly personal.*

*The purpose of this research is to find out if our AIDS education intervention is useful for people with opioid dependence. This information will be helpful in providing better information on HIV for opioid dependent people.*

*All questionnaires and information collected from your initial assessment will be kept confidential; only a code number for each participant will appear on the questionnaires and initial assessment sheet. All information about clients in this service will be kept in a safe, locked location. The names of clients will be kept separate from the information itself.*

*There is no risk involved in this research, other than the unpleasant feelings that might arise when thinking about problems associated HIV and your opioid use. Your drug worker is available to talk to if you experience unpleasant feelings as a result of participating in this study. There are not likely to be any direct benefits to participants.*

*Participation in this research is voluntary. The decision to participate will not affect the nature of the treatment services participants receive. You are free to withdraw from the study at any time without giving any reason and without prejudice to your continued care.*

*I understand all the above information, and agree to participate in this study. I understand that I may withdraw from the study at any time.*

\_\_\_\_\_

*Participants Signature*

\_\_\_\_\_

*Date*



## It's your turn: (1A)

Your evaluation planning group should consider how to best manage ethical issues in your study. Use the examples provided below as a guide, then adapt them to your own evaluation project.

- 1 Each partner should individually write down possible safety issues (harm that could come to your participants) as a result of participating in your research. After each partner has finished, you should discuss these issues as a group. Write down a specific plan for how you will minimise or limit this harm.

**Example:** In a questionnaire survey study of individuals being treated for opioid dependence, possible risks include a) unpleasant feelings that might arise from thinking about problems associated with opioid use, and b) possible breaches of confidentiality. To minimise this potential harm: a) all participants will be instructed at the time of enrolling that they may have access to a trained mental health professional

to discuss any unpleasant feelings that might arise as a result of participation, and b) use of confidentiality policies and procedures (see below).

- 2 Working separately, write down a plan for how you will ensure that all participant information will remain confidential. Discuss your ideas as a group and write down a specific plan for how you will maintain confidentiality.

**Example:** The data will be kept in a safe, locked location; the names of participants will be kept separate from the data itself; all research staff will sign a pledge of confidentiality.

- 3 Working as a group, develop a consent form for participants to sign. This form should outline the study itself, as well as all risks of participation. Use the information provided in the preceding section and the example below as a guide. Also consult your institution's ethical guidelines on consent of participants, if applicable.

## 2. Develop a data management plan

Another important step of starting a study is to develop a data management plan. Ideally, a programme administrator should make specific staff responsible for the important task of data management - collecting and editing the data. Those who use and interpret the data should understand and appreciate the strengths and weaknesses of the information collected. Assignment of clear responsibilities for data quality is an important part of the process.

Establish well-defined objectives for your data collection system, and make a strong commitment to achieving the objectives. Your data management plan must also address:

- designing your record keeping system
- acquiring and processing the data
- training the people who will collect data

## Designing a record keeping system

**Development of a record keeping system or database design is an important part of a successful evaluation.**

Development of a record keeping system or database design is an important part of a successful evaluation. The following suggestions may be helpful in guiding you through this process:

- It is a good idea to delegate one person to monitor the data collection process. This person will ideally take charge of administering the instrument(s), to make sure that the data are collected properly, monitor the forms and questionnaires that are used, oversee the storage and retrieval of data, and supervise the data analyses.
- Keep all your records in one place. This includes any questionnaires, interview schedules, tapes, data collection forms, and archived materials. By storing all information associated with the evaluation in one place you can avoid problems of having to search multiple locations to find the information you require. Develop a system for how you will organise the records. (Organisation by ID number is usually a good idea.)
- Remember to use ID numbers instead of names, and to remove any identifying information from the central records.
- Assign a person (or group of people) for regular transferral of information on the records to a central “database.” This could be a computer file (see below), or could be a notebook containing ID numbers and summary data for each participant. This transfer of information should be completed routinely (once a week is a good guideline).
- Keep the database in a safe location. Make a “back up” copy if possible, in case the original data base is damaged in some way. You can back up computer data onto a separate diskette, or photocopy written records. Back up copies should be stored in a separate location from the central database, in case of fire, flood, theft, etc.
- Make a habit of routinely picking a few cases from the database to ensure that the information in the database corresponds with the information from the original records (e.g., questionnaires). This will help to spot errors and/or make corrections to your data transferring procedures.

---

## Designing and setting up an automated (computer) database

This will require extensive pilot testing to identify errors and shortcomings that almost always exist in the initial stages. It also requires careful design to make the best use of limited storage, and to keep data processing time to a minimum. For these reasons, it is important to use a skilled computer programmer to design and pilot test the database. It is equally important that the programmer have a good understanding of what you do, and what you need from the database.

Database software for microcomputers includes dBASE, Paradox, Rbase, Dataease, Access and Excel. Becoming proficient with any database software requires many hours of use. Consequently, databases should be developed only by computer programmers with extensive training or experience with the selected software. A spreadsheet programme might be a better choice for a treatment programme. Many of the latest spreadsheet pro-

grams, including LOTUS 123 and Excel, are easier to use. Database managers are capable of filling most of the data processing needs you are likely to have. They have data entry screens that make them easy to use, and they can be used for analysis.

If you are using a computer database, you may want to consider obtaining specialised software to conduct statistical analyses of quantitative data. A variety of options are available, ranging from general programs capable of performing a wide range of analyses (SPSS, SAS), to specialised programs that handle specific types of analyses or provide simple statistical calculations (Minitab, EPI-Info). EPI-Info is a specialised piece of software designed for organising and analysing data from evaluations, surveys and other studies of health and social services. It is shareware, which means it is available free of charge. You may order a copy from: The Division of Surveillance and Epidemiology, Centres for Disease Control (CDC) Atlanta, Georgia

30333., USA: Phone 404-728-0545; Fax 404-315-6440; Internet [www.cdc.gov](http://www.cdc.gov).

If you are using a computer system, each completed form should be edited for missing data and other problems before the data are entered in the computer. Depending on the number of forms involved and delays due to data problems, this phase can take up to a month to complete. After information from a batch of forms is edited, the data can be entered and processed. Routine processing should be done every two weeks, to allow enough data to build up for efficient operations. Exceptions can be made when the flow of forms warrants more or less frequent processing. Automated procedures for data entry should be fully explored before the task is delegated to someone to enter the information into the computer by hand.

If you are not using a computer, an independent person can be assigned to randomly spot check for data entry mistakes from the original data entry form to the central data notebook.

## Data collection staff

**Selecting knowledgeable, competent data collection staff is a key to success for high quality programme evaluation.**

Selecting knowledgeable, competent data collection staff is a key to success for high quality programme evaluation. Your data collection staff should be trained in routine or standardised procedures for interviewing and/or administering questionnaires. Training should include an overview of the purpose of the study, the purpose of each research measure and how it is used, and instructions for recording or filling in responses to each item. The training should also include some practice sessions to ensure that the staff members are comfortable and competent in using the measures. In addition, the staff should have a thorough understanding of participants' rights to confidentiality, and a plan for handling unusual or emergency situations (e.g., suicide attempts, "overdose,"

crises at home or work) that may arise in the course of data collection.

Data collectors should be trained in the importance of **standardisation of procedures**. This means that they should collect data the same way with each participant. Why? Standardisation helps to ensure that differences in participants' responses are not due to differences in how the data are collected. For example, if a data collector interviews some participants in a group situation, but others alone, then it is possible that differences in responses are due solely to this different interview format.

The same principle is true between data collector: if one data collector is very

**Standardisation helps to ensure that differences in participants' responses are not due to differences in how the data are collected.**

friendly to participants, while another is distant, then participants may give different types of responses. Clear training in standardised procedures helps to prevent these possibilities.

When considering standardisation, think about the following issues:

- Will data be collected individually, or in the presence of others?
- Will participants complete questionnaires on their own, or receive guidance from the data collector?
- How “friendly” should data collectors be with participants?

- How should data collectors respond to questions from participants?

There is no right or wrong answer to these questions: the key to success is ensuring that everyone does it the same way every time. This is the essence of standardization.

Data collection also should involve extensive communication between those collecting the data and those managing it. Issues inevitably arise in the data reporting practices, and these must be understood by those using the data. Ideally, select a data coordinator who has enough time to obtain, edit, and co-ordinate collection procedures, and who can communicate effectively about the complexity of your data collection.



## Into action

Chris C. was selected as the *data coordinator*. He would supervise the data collection and will train all other drug workers at the service in standardised procedures for administering the questionnaire. All data would be stored in locked file cabinets. Chris C. would have the keys for these cabinets. He would also collect questionnaire and file them in a “data to be entered” file within the cabinet. Once a week, Chris C. would enter these records into the main record database using the database software SPSS. Once every two weeks, Sue R. would double check Adam’s work to ensure that errors are not being made. A separate list, linking data code numbers to names, would be stored in the Sue R. office in a locked file cabinet. She would be the only one to have the keys to this file cabinet.

The team decides to conduct a 3 hour-workshop with all data collector (drug workers at the service) in order to train them in standardised procedures of ad-

ministering the questionnaire. This training would include information about purpose of the study, about the chosen measurements and instructions for filling in responses to each item as well as some practice session. Further, the data collection staff would be informed about the standardised procedures:

- 1** The questionnaire will be administered individually
- 2** Participants will complete questionnaires by themselves
- 3** In case of difficulties in completing the questionnaire the client will receive guidance from the data collector
- 4** The data collector should respond to questions as discussed in the training workshop
- 5** The data collector provides time for questions about the study, after the clients filled in the questionnaire.



## It's your turn (1B)

Working as a group, complete the following tasks:

- 1 Write down where you will store your data, how it will be organised to ensure clarity and protect participants' confidentiality.

**Example:** Data will be stored in locked file cabinets in the main storage area. Only the data co-ordinator will have the keys. A separate list, linking data code numbers to names, will be stored in the investigator's office in a locked file cabinet. Only the investigator will have keys to this file.

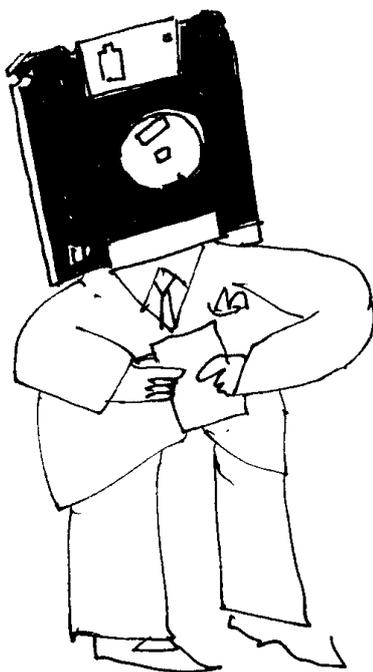
- 2 Write down how you will maintain central records of the data, and who will be responsible for maintaining the records.

**Example:** A data co-ordinator (Jon) will collect questionnaires and file them in a "data to be entered" file within the cabinet. Once a week, Jon will enter these records into the main record database. Once every two weeks, Serene will double check Jon's work to ensure that errors are not being made.

- 3 Write down standard procedures for your data collectors, and a plan for how you will train them in these procedures.

**Example:** All data collectors will receive training in how to obtain consent, administer the questionnaires in a standardised fashion, and deal with emergency situations. Michel (senior clinician) will conduct this training. Before being allowed to collect data, data collectors will be tested and approved by the study investigator.

- 4 If you plan to use clinicians as data collectors, consult with them to ensure that your data collection plan is feasible from their perspective. (Because clinicians sometimes feel too busy to use standardised forms or participate in research, gaining their co-operation at the outset is essential.) Work with them to achieve a mutually agreeable plan.



## 3. Conduct a pilot test

**The reason for a pilot test is to identify any flaws in the data collection plan so they can be corrected before a full-blown implementation of the evaluation gets underway.**

(Note: This section is most applicable for researchers. If you have a researcher among your evaluation partners, have that person read this section and then explain its contents briefly to the others.)

Once you have reached this point, you are ready to try out your data collection instrument(s) in a small pilot test. The basic idea behind conducting a pilot test is to try out your data collection instrument(s) in a small, preliminary sample in order to answer the following questions:

- Does the data collection instrument provide useful information? For example, are variables being measured accurately?
- Can the instrument be administered properly, given constraints on time and staff? For example, is the instrument too long or too complicated to be completed?
- Can the information be easily managed? For example, can the form(s) be integrated with the data management system easily?
- Does other information need to be collected? How did people react to completing the instrument? What is missing? How should the questions be changed, and why?

If a formal pilot test is not possible, have staff, community members, or patients complete the draft data collection forms to see which parts may be unclear, confusing, or otherwise misleading. If a more formal pilot test is feasible, consider the following suggestions in testing different types of forms:

- 1 Self-administered questionnaires:** Select 10-15 people who are similar to your clients; ask them to complete the form; and analyse the data. Review the completed forms carefully, looking for potential problems: Is there a particular question several respondents did not answer? Are respondents answering questions they should have skipped? Are some checking or circling more than one response category when instructions say to give only *one* answer? If such problems appear (and they usually do), talk to some respondents who failed to answer properly. Ask them why they answered as they did. The wording may be unclear, or the format may need improvement. Using special type face features might help, such as italics, underlines, or bold type for instructions.
- 2 Interviews:** Conduct several pilot interviews to test the interview instrument. Consider having a group discussion with pilot test respondents after they have completed their individual interviews. A moderator would lead the people being interviewed through the form as a group, encouraging them to talk about any problems they had with the questions or the interviewer's style, and to suggest changes. Group discussions can be an effective way to identify problems and gather suggestions. They also may point out differences in item interpretation that would remain hidden in individual interviews.
- 3 Record Reviews:** Pilot testing your record review forms can be done efficiently as part of staff training for this activity. Problems in wording or format should become clear after a few trials or exercises.

The reason for a pilot test is to identify any flaws in the data collection plan so they can be corrected before a full-blown implementation of the evaluation gets underway. Thus, the last step, in the pilot test phase is to summarise findings and to revise the data collection plan as needed. If the changes are significant, it may be advisable to conduct another pilot test of the revised procedures before proceeding with the evaluation.

You may want to have a staff meeting to introduce the evaluation and the data collection plans to everyone else who will be involved. This meeting would explain the reasons for the evaluation, the questions to be addressed, and the importance of each person's role in answering the questions.

After you have pilot-tested your data collection instrument and have incorporated any changes that are needed, you are ready to formally start collecting data.



## Into action

The next task was to conduct a pilot test. Each of the drug workers administered the baseline questionnaire to two clients. Afterwards the collectors and the evaluation team discussed identified flaws. The data collectors reported that given the emotional and physical state of some clients, it seemed not wise to administer the questionnaire at the first face-to-face interview with the client. They suggested that the questionnaire should be

given to the client during the second day of his stay at the unit during his second individual contact with his drug worker. The evaluation team agreed to administer the questionnaire during the suggested second interview with the drug workers. Apart from some wording problems, no other difficulties arose from administering the questionnaire. To administer the questionnaire took on average 15 minutes.



## It's your turn (1C)

- 1** The researcher in the group should explain the key points of this section to his/her evaluation partners. For example, explain the important functions that pilot tests serve, and how to conduct them properly.
- 2** As a group, write down when you will conduct a pilot test, and how many people will participate in each portion of it. Use the information in the preceding section as a guideline for choosing the number and type of people for your pilot test.

## 4. Write an evaluation plan

The final step in preparing for data collection involves the preparation of a written evaluation plan. Writing a structured evaluation plan will help you to organize your thoughts clearly. It also serves as a reminder to everyone involved about the purpose of the evaluation and what questions and decisions the results are intended to address.

The basic elements of an evaluation plan are:

### 1 Background and General Purpose

This is a summary of your evaluation plan. It should include a brief description of the purpose of the evaluation, how you plan to conduct the evaluation, who will receive the results, and how results will be used.

### 2 Programme Logic Model (refer to Step 4 of Workbook 1 for more information). The model should describe:

- what it is that you plan to evaluate (e.g. services, network of services);
- your process or implementation objectives — the short-term and long-term goals of various parts of your programme/service/system.

### 3 Evaluation Team

This section lists the people who will be involved with your evaluation project, along with a brief description of their experience and anticipated roles in the evaluation.

### 4 Evaluation Questions to be addressed

Your questions should follow logically from your programme logic model and should be linked to one of your programme's implementation objectives or goals in a specific way.

### 5 Data Collection Strategy

This section is very important, as it explains how you will get from your evaluation questions to results. Refer to Steps 6 and 7 of Workbook 1 for information and suggestions about creating your data collection strategy.

For each evaluation question you listed above, you should indicate:

- whether you will use a standardized data collection or develop your own, and what the specific variables you plan to use are how you will collect the data;
- who will administer the data collection instrument(s);
- how many participants you will assess-your sampling strategy (how you will choose participants)-your time line for data collection.

### 6 Data Management Plan (refer to Step 1-B of Workbook 2 for more information)

In this section, explain how you will store your data and maintain central records.

### 7 Staff Training (refer to Step 1-B of Workbook 2 for more information)

Describe how you will train all evaluation staff to conduct their jobs in a standardized way, and whether additional staff or outside consultants will be needed.

### 8 Pilot Test (refer to Step 1-C of Workbook 2 for more information)

What type of pilot test will you conduct, and how many people will participate in each portion of it? If you are using a newly developed data

collection instrument, you should include it in your pilot test and then make changes as necessary after the pilot test.

## 9 Strategy for Using Results

Explain how your evaluation results can be applied to the long-term goal of improving substance use treatment. Explain who will receive results, and how this information could be used to make changes to improve treatment.

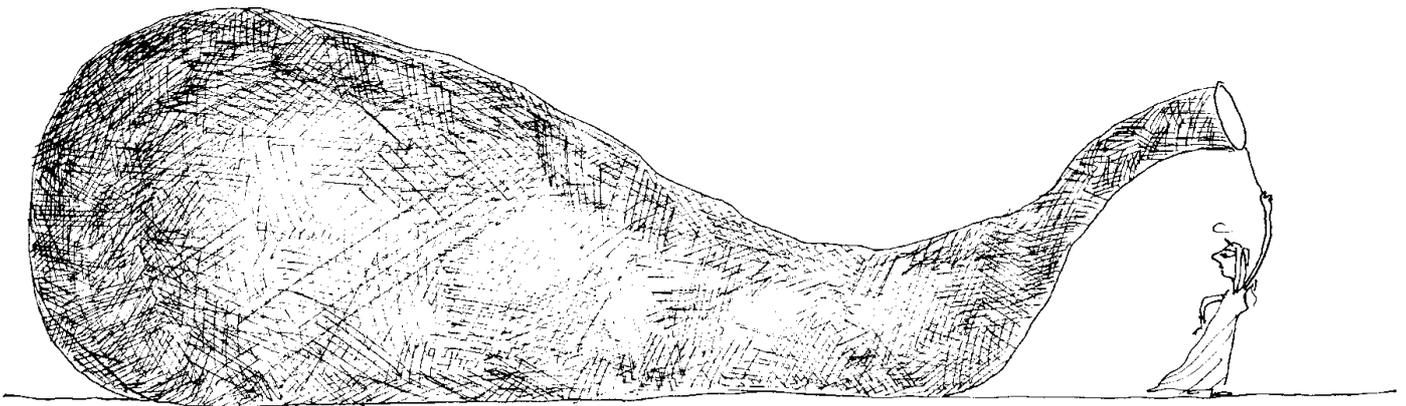


## It's your turn (1D)

- 1 Compose a written evaluation plan. Use the exercises that you have completed so far to help you. Components should include:
  - Background and general purpose of the evaluation (Step 3, Workbook 1)
  - Programme logic model (Step 4, Workbook 1)
  - Evaluation questions to be addressed (Step 5, Workbook 1)
  - Data collection strategy (Steps 6 and 7, Workbook 1)
  - Data management plan (Step 1B, Workbook 2)
  - Staff training (Step 1B, Workbook 2)
  - Pilot plan (Step 1C, Workbook 2)
  - Strategy for using results
- 2 Review your written evaluation plan with the expected user(s) of the results. Modify as needed.

# Step 2

## Collect data



**In resolving data collection problems, keep the instruments and procedures consistent unless you must make changes to solve fatal flaws in the project.**

(**Note:** This step is most applicable for researchers. If you have a researcher among your evaluation partners, have that person read this section and then explain its contents briefly to the others.)

If you have planned carefully during the preceding stages, data collection should proceed smoothly and with little need for readjustment. Nonetheless, a few guidelines may assist you:

Periodic meetings with staff involved in the data collection can be very helpful. The topics can include questions about data collection. Write down how these questions are resolved. You also may need to discuss results of reliability and validity

checks on the indicators, and preliminary findings.

In resolving data collection problems, keep the instruments and procedures consistent unless you must make changes to solve fatal flaws in the project. Even minor changes in forms and procedures can cause more problems than they solve. For this reason, once you have started your data collection, attempt to make it as systematic and as free from bias as possible.

Over time, data collectors' skills tend to erode. Periodic retraining of data collectors, even every few months, can help ensure quality control during your study.



## Into action

Our evaluation team decided to meet twice a month to discuss problems with data collection and entry. Sue will be respon-

sible for organising them. She will also take responsibility for recording the decisions about the evaluation project.



## It's your turn

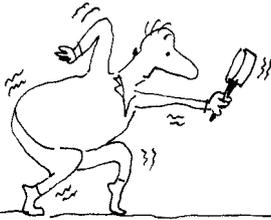
The researcher in the group should make the following decisions:

- 1** Write down how often you will have research team meetings, and who will be responsible for organising them.
- 2** Nominate a person to take responsibility for recording the decisions about research procedures. Have this person maintain a notebook of all these decisions, so they can be referred to later if questions arise.

Discuss your plan with your evaluation partners. Make changes based on their input as appropriate.

# Step 3

## Analyse data



(**Note:** This step describes technical information about data analysis procedures. It is most applicable for researchers and or data consultants. If you have a researcher among your evaluation partners, have that person read this step and then explain its contents briefly to the others.)

Data analysis is the reward for all of your planning and data collection efforts. If you carefully plan and carry out data collection, the analyses will help you provide answers to questions about what efforts your programme is making and what effects those efforts are having.

Data analysis is not simple, and requires calculation, interpretation, and judgement. Do

not hesitate to obtain technical assistance, if available, during this phase of the evaluation.

A general guideline is that the methods of analysis you choose should be tailored to the specific evaluation questions with which you started. These choices also depend on the size of the sample you collect, the quality of the data, the resources available to you (e.g., transcribers, interpreters, computers with statistical programs installed), and the effort you are willing to put into the analysis. This section will introduce you to a number of general concepts and procedures in data analysis and presentation.

**Descriptive analysis summarises the measurements you have made on the variables included in the data collection instrument. Explanatory analysis attempts to explain relationships among the variables in which you are interested.**

### Decide whether you are addressing descriptive or explanatory questions

Before you begin your analysis, it is very important to be clear about what general *type* of analysis you wish to conduct. **Descriptive analysis** summarises the measurements you have made on the variables included in the data collection in-

strument. **Explanatory analysis** attempts to explain relationships among the variables in which you are interested. The figure below summarises the differences between these two different approaches to data analysis.

Descriptive analysis	Explanatory analysis
<p>Goal: to summarize the measurements for each variable</p> <p>Examples: How many clients reported? How much time did staff spend? How much variability was there in ...?</p>	<p>Goal: to explain relationships between variables and groups</p> <p>Examples: Do male and female clients differ? Compared to clients assigned to a control group, were clients exposed to ... more likely to...?</p>

## Different types of data analysis

As the figure shows, descriptive analyses summarise the data collected for each individual variable. If your indicators are quantitative in nature, descriptive analyses can be used to provide information for questions such as “how many clients in the programme are over the age of 50?”, or “how much time did staff spend completing administrative forms in the assessment process?”, or “how much variability was there in the education levels of our clients?”. If your indicators are qualitative in nature, descriptive analyses can be used to provide information for questions such as: “what types of responses were reported when clients were asked about the most beneficial aspect of the programme?”

The figure also shows that explanatory analyses describe relationships between

or among variables included in your data collection instrument(s). Explanatory analyses can also use either quantitative or qualitative data. For example, if your indicators are quantitative in nature, explanatory analyses can be used to provide information for questions such as “do male and female clients differ in their relapse rates in our programme”, or “compared to clients who didn’t receive a motivational interview, were clients in the interview programme more likely to show up for group therapy sessions?”. If your indicators are qualitative in nature, explanatory analyses can be used to address questions such as “did clients who reported liking the harm reduction orientation of our programme report more positive benefits in our interview?”.



## Into action

The research questions regarding the process evaluation would be analysed with the following types of analyses

Questions	Description
<b>1</b> Has the number of referrals increased from the previous year?	Descriptive analysis with quantitative indicators
<b>2</b> Did the characteristics of the clients change in comparison to last years clients?	Descriptive analysis with quantitative indicators
<b>3</b> Has the number of self-discharges decreased from the previous year?	Descriptive analysis with quantitative indicators
<b>4</b> Has the number of clients who were referred to long-term treatment increased from the previous year?	Descriptive analysis with quantitative indicators

For the outcome evaluation they decide on the following analyses:

Questions	Description
<b>5</b> Did attitudes in favour of HIV low-risk practices improve among clients?	Descriptive analysis with quantitative indicators
<b>6</b> Did client's HIV-risk behaviour change in favour of low risk practices after treatment?	Descriptive analysis with quantitative indicators
<b>7</b> Did clients knowledge about AIDS increased after the AIDS education intervention?	Descriptive analysis with quantitative indicators
<b>8</b> Did clients self-efficacy regarding their ability to use skills maintaining AIDS harm reduction behaviours increase?	Descriptive analysis with quantitative indicators



## It's your turn

- 1 Write down 4 research questions related to your evaluation project. Descriptive analysis with qualitative indicators (Section B)
- 2 Decide what type of analysis (or analyses) you will conduct for each question: Explanatory analysis with quantitative or qualitative indicators (Section C)

Descriptive analysis with quantitative indicators (Section A)
- 3 Review the appropriate sections below.

## A. Descriptive analysis with quantitative indicators

Recall from Workbook 1 that quantitative indicators can be arranged into different types, based on the **measurement scale** that they use (nominal, ordinal, interval, ratio). Each type of scale can be analysed in different ways, and lend itself to different presentation strategies, as indicated in

the following subsections. The subsections below (frequencies, measures of central tendency, measures of variability) provide information on how to calculate **statistics** that summarise the variables in your evaluation.

### Frequencies<sup>1</sup>

**If you count how many observations fall into each category, you obtain a statistic known as a frequency.**

Simple frequency tables are used to see patterns in groups of nominal and ordinal data. **A frequency distribution provides a record of the number of individuals located in each category of the quantitative indicator.** If you count how many observations fall into each category, you obtain a statistic known as a *frequency*. If you know the total number of observations that were measured for that variable, you can calculate the percentage of observations in each category by simply dividing the

number of cases in each category by the total number.

A frequency table displays these results. As the next figure shows, a frequency table lists the number of observations in each category. The table also presents the total number of cases and the percentages.

In the example presented on the next page, measurements have been taken to assess severity of PSU problems for 125 hypothetical clients.

<sup>1</sup> Note that all of the major statistical packages for computer-aided data analysis have routines that will calculate frequencies

<b>Title: PSU problem severity for a Sample of 125 hypothetical clients</b>			
<b>Labels</b>	<b>Category</b>	<b>Frequency</b>	<b>Percentage</b>
	None reported	42	33.6
	Mild	13	10.4
	Moderate	50	40.0
	Severe	15	12.0
	Very severe	5	4.0
<b>Totals</b>		125	100.0

In some cases, especially with small sample sizes, you might want to show the frequency of every measurement value. In most cases, however, (and especially for interval and ratio data) you will want to

group these measurements into a smaller number of categories called a **grouped frequency table**. The next figure shows an example, this time using client age as the variable.

A grouped frequency table describing age of clients

<b>Age (years)</b>	<b>Frequency</b>	<b>Percentage</b>
60-69	5	3.1
50-59	5	3.1
40-49	20	12.5
30-39	30	18.8
20-29	60	37.5
10-19	40	25.0
<b>Totals</b>	160	100.0

Each grouping is called an **interval**, and the number of scale values in each interval is called the **interval width**. There are no strict rules for deciding on the number of intervals or their width. But in general, there should not be so many groupings that you end up

with only one or two observations in each group. On the other hand, there should not be so few groupings that the highest and lowest observations are not distinct from each other. For convenience, the interval width should be a whole number.

## Measures of central tendency

Another useful way to summarise the measurements you have made for a variable is to report **the average or typical measurement** in your data for that variable. We

call this the data set's **central tendency**. There are three ways to describe central tendency: mode, median, mean. These are shown in the next figure.

### Measures, definitions, data types, and example of central tendency

**The mode is the score that occurs most frequently.**

Measure	Definition	Data-types	Examples
Mode	Score which occurs most	Nominal Ordinal	Gender Client Satisfaction
Median	Moderate	Ordinal	Client Satisfaction
Mean	Average	Interval	Numbers of Counseling

**The median score is the one that separates the upper half of the observations from the lower half of the observations.**

The mode is the score that occurs most frequently. This measure of central tendency can be used for nominal, ordinal, or interval scale data. Not that a data set can have more than one mode if the maximum frequency occurs equally for two or more observations for a variable.

The median score is the one that separates the upper half of the observations from the lower half of the observations. In other words, 50% of the scores fall above the median, and 50% of the scores fall below

the median. To calculate a median, you must be able to rank order the scores from highest to lowest. This measure of central tendency can be used with both ordinal and interval scale data. One important property of the median is that, unlike the mean, extreme scores do not affect it.

The mean is the arithmetic average for the observations on that variable. It is the measure of central tendency that is used most often to describe a set of interval-scale observations. To calculate the mean

**The mean is the arithmetic average for the observations on that variable.**

for a variable, simply add up all of the scores and divide that sum by the total number of scores. The next figure summarises these measures of central tendency, using a hypothetical example of 11 observations made on a client satisfaction variable, which uses a 5-point interval scale. A mean score can be influenced significantly by one or two extreme scores in the set of observations.

sure of central tendency that is used most often to describe a set of interval-scale observations. To calculate the mean for a variable, simply add up all of the scores and divide that sum by the total number of scores. The next figure summarises these measures of central tendency, using a hypothetical example of 11 observations made on a client satisfaction variable, which uses a 5-point interval scale. A mean score can be influenced significantly by one or two extreme scores in the set of observations.

The mean is the arithmetic average for the observations on that variable. It is the mea-



Client satisfaction data (5 point interval scale)	Rank-ordered data
5	5
1	4
1	4
2	3
3	2
4	2
1	1
5	5
1	4
1	4
2	3

Mean

$$= \frac{\text{Sum of scores}}{\text{\# of scores}} = \frac{24}{11} = 2.18$$

Median

$$= 2 \quad (\text{divides observations into upper \& lower halves})$$

Mode

$$= 1 \quad (\text{most frequently occurring observation})$$

Note: 1=hated program; 2=disliked program; 3=neutral; 4=liked program; 5=loved program

## Measures of variability<sup>2</sup>

**The range determines the most extreme observations in your data set and the difference between them.**

A measure of central tendency is not enough to summarise or describe a quantitative variable. You also need to know how widely a set of observations varies around its average. Another way to think about variability is that **measures of variability provide a useful summary of how different the observations are from each other**. There are several measures that can be calculated.

The range determines the extreme observations in your data set and the difference between them. For example, a series of observations made on a 5-point client satisfaction interval scale might look like this: 2, 1, 4, 3, 3. The range of scores here is 1-4 (lowest observation to highest observation). The range is useful because it tells us about minimum and maximum observations for a variable. Unfortunately, it only tells us about extreme scores and doesn't take into account all of the observations at once.

**The standard deviation describes how far the observations deviate from the mean in a data set.**

The standard deviation describes how far the observations deviate from the mean in a data set. It is the most important and most widely used measure of variability. It is always reported in conjunction with the mean for a variable (e.g., "mean client satisfaction in the sample was 3.2, with a standard deviation of 1.7"). How do we calculate a standard deviation? There are two important properties of this measure:

- if all of the observations are equal to the mean, then there is no variability in the observations;
- the more the observations differ from each other, the greater the standard deviation for that variable.

To calculate standard deviations, we need to introduce the concept of **deviation scores**. Here is the way we represent deviation scores:

$$X - \mu,$$

where any single observation (X) is subtracted from the overall mean of the observations for that variable ( $\mu$ ). To calculate a standard deviation, we first have to square the summed deviation scores for all observations:

$$\text{sum } (X - \mu)^2 = SS = \text{sum of squares}$$

This value is called the "sum of squares" because it simply reflects the sum of the squared deviation scores for our data.

The variance of our variable gives us an "average" of the variability of our observations, and is calculated as follows:

$$S^2 = \text{variance} = SS (\text{sum of squares})/N$$

where N = the number of observations in our data set for that variable.

Finally, the standard deviation of our variable is simply the square root of the variance:

$$S = \text{standard deviation} = \text{square root of variance}$$

You can see on the next page an example, again using client satisfaction data, but this time for only 7 observations:

<sup>2</sup> All of the major statistical packages for computer-aided data analysis have routines that will calculate measures of variability for you

<b>X</b> (Observations)	<b>X - <math>\mu</math></b> (Deviation Scores)	<b>(X - <math>\mu</math>)<sup>2</sup></b> (Squared deviation scores)
3	-1	1
3	-1	1
4	0	0
4	0	0
4	0	0
5	1	1
5	1	1

sum = 28

sum = 0

sum = 4 = sum of squares

mean =  $28/7 = 4 = \mu$ Variance =  $SS/N = 4/7 = .57$ 

Standard deviation = square root (.57) = .76

Remember the following steps when calculating standard deviations:

- Calculate the mean for the variable;
- Subtract the mean from each score; put these values in a separate column;
- Square each deviation score; put these values in a separate column;
- Calculate the variance and standard deviation, after summing the columns.

## B. Descriptive analysis with qualitative indicators

**Decide whether your analysis will try to find examples of what interests you already. This is referred to as a top-down, deductive method. Your alternative will be to construct types based on what you find. This is a bottom-up, inductive method.**

Whenever you have qualitative (detailed observation-based or language-based) indicators in your data set, instead of statistics, descriptive analyses provide a series of responses that summarise the indicators and describe their characteristics.

Several preliminary steps are required to conduct descriptive analyses of qualitative data. The steps include the requirements:

- Make sure that each of the indicators has received an accurate **transcription**. In other words, you must transfer taped interviews to computer files containing the qualitative information. Further, checks should be made in order to ensure that whoever was delegated to transcribe interviews accurately recorded the information. If you are using open-ended written descriptions, transfer them to computer files

so that they can be easily sorted and manipulated.

- Make sure that you have allocated **sufficient time and staff resources** to conduct the descriptive analyses of the qualitative data. Descriptive analyses of quantitative data, can be accomplished easily using widely-available computer programs. However, qualitative analyses require people to interpret the indicator. This takes extensive time and effort. Do not undertake these analyses unless you are prepared to devote the resources to them.
- Decide whether your analysis will try to find examples of what interests you already. This is referred to as a top-down, deductive method. Your alternative will be to construct types based on what you find. This is a bottom-up, inductive method.

### Top-down (deductive) analysis

One prominent approach to descriptive qualitative analysis is to **use the transcripts to find examples of types that already interest you**. Consider, again, the variable used earlier in this workbook to illustrate qualitative data: “of the programme most liked.” You might approach this qualitative analysis in a top-down manner if you have **preconceived ideas about the types of responses that might be given** (e.g., if you think, in advance, that clients will report that three aspects of the programme are most liked: (1) harm reduction philosophy, (2) staff (such as case therapists), and (3) follow-up support groups. If you adopted this approach, the following steps could be followed:

- assign a reader who reviews and interprets each of the observations

- instruct the reader to read through the transcripts three times. The first reading will identify all responses that mention the harm reduction philosophy. The second reading will identify all responses that mention staff, and the third reading will identify all responses that mention the follow-up support groups
- After each reading, the interpreter collects the responses for each type and collates them into one document. The result will be three documents: one describing all of the interview responses for each type of interest to the programme evaluator (harm reduction, staff, follow-up groups)
- A second reader will duplicate the first reader’s interpretations, in order to determine whether instances of the three types were missing

## Bottom-up (inductive) analysis

Another approach to describing qualitative data is used when you don't have a preconceived idea of the types of responses that are given when the indicator is used. This approach attempts to describe the

variety of responses that are given. Your goal is to develop the types from the information you have, rather than confirm or disconfirm their existence. The figure on the next page provides an example:



### (Hypothetical Data)

Indicator	What is the aspect of the programme you liked best?
Observations	<p>“For me, it has to be the amount of good information I got and the support from other members of the group” (Client #43)</p> <p>I liked the fact that I could set my own goals regarding drinking and work in group therapy to achieve them.” (Client #89)</p>
Types	<p>“I didn't have to fill out too many forms” (Client #112)</p> <p>group therapy [43, 89]    ☞ group therapy (information) [43]</p> <p>☞ group therapy (own goals) [89]</p> <p>☞ group therapy (other clients) [43]</p> <p>forms [112]</p>

This figure presents a rudimentary analysis of three clients' responses to the open-ended qualitative question. In the bottom-up approach, the reader or interpreter attempts to develop a record of the types expressed in the observations. In this example, two main types were reported: group therapy (by clients 43 & 89) and forms (by client 112). Note that **within** the group therapy type, the interpreter identified three different **types**. These subtypes reflect different aspects of group therapy that were viewed positively; the fact that it gave good information (client 43), the flexibility of goals (client 89), and support from other clients (client 43). As you can see, the inductive approach yields new types of responses, as opposed to finding examples of pre-existing types.

In general, it is better to use both inductive and deductive approaches to control for one another. The following guidelines should be kept in mind:

- The interpreter must systematically read **all** of the transcripts, not just interesting ones or ones that show the data in a particular light.
- The interpreter must systematically **document the analysis** by recording **where the examples can be found**, and what aspects of the transcript support the interpretation.
- The interpreter's work should be **checked by another reader**, who doesn't know the results of the first analysis. This will allow you to determine whether there are any biases in the interpretative process.

## C. Explanatory analysis with quantitative or qualitative indicators

Descriptive analyses cannot fully answer all types of questions, such as: Do clients abusing alcohol and cocaine differ in the number of times they have been arrested for PS-related problems; or compared to clients assigned to a comparison group, were clients who received a motivational interview more likely to attend group therapy sessions? To make these determinations, you need to go beyond simply describing the frequencies, central ten-

dency, and variability of the variables you collected in your programme evaluation. **Inferential statistics go further to examine relationships among variables** in your evaluation. The following subsections outline two of the basic inferential techniques that are available for you. However, there are many more techniques that can be used, and you should seek technical help, if available, to conduct more advanced types of explanatory analysis.

### Chi-square ( $X^2$ ) test

The  $X^2$  test is a technique that can be used to test the **strength of the relationship between two variables**. This test is used with nominal data (e.g., gender, types of programmes), but also can be used with ordinal or grouped interval data. For example, you might want to compare the proportion of clients who relapsed following motivational interviewing with the proportion of clients who relapsed in your comparison group who received no motivational interviewing.

To begin, cross-classify each observation across the two variables: e.g., mo-

tivational interviewing versus no motivational interviewing, and relapse versus no relapse. Then, count the number of observations that fall into each of the categories created. For example, maybe you have a total of 200 observations (clients): 100 received motivational interviewing and 100 did not. Within the “motivation group,” 68 did not relapse, while 32 did relapse. Within the no-intervention group, perhaps 45 did not relapse while 55 did relapse. You can arrange this information in a contingency table, as shown in the figure below.

### $X^2$ Test of association between two variables

	Observed frequencies		
	Didn't relapse	Relapsed	(Row totals)
Motivational interview	45	55	100
No interview	68	32	100
Column Totals	68	32	100

The  $X^2$  test is based on the idea of finding the difference between the **observed frequencies**, displayed in the figure, and

the frequencies that you'd **expect**, if the two variables are independent of each other.

Here is how to calculate the expected values of the frequencies for each cell:

$$f_e \text{ (expected value for the cell)} = (\text{row total}) \times (\text{column total})/N$$

Therefore,

$$f_e \text{ (interview + no relapse cell)} = (100)(113)/200 = 56.5$$

$$f_e \text{ (interview + relapse cell)} = (100)(87)/200 = 43.5$$

$$f_e \text{ (no interview + no relapse cell)} = (100)(113)/200 = 56.5$$

$$f_e \text{ (no interview + relapse cell)} = (100)(87)/200 = 43.5$$

Now that we have calculated the expected values, we will test whether or not these expected frequencies differ from the actual, observed frequencies in our data.

The chi-square ( $X^2$ ) statistic provides a way for us to compare the actual and expected frequencies, and is calculated as:

$$X^2 = \sum (of - f_e)^2 / f_e, \text{ where}$$

of = observed frequencies,

and  $f_e$  = expected frequencies

If there is very little difference between our observed and expected frequencies,  $X^2$  will be a small value. If there is a great difference between our observed and expected frequencies chi-square will be a large value.

Category	of	fe	(of - fe)	(of - fe) <sup>2</sup>	(of - fe) <sup>2</sup> /fe
Interview/no relapse	45	56.5	-11.5	132.25	2.34
Interview/relapse	55	43.5	11.5	132.25	3.04
No interview/no	68	56.5	-11.5	132.25	2.34
No interview/relapse	32	43.5	11.5	132.25	3.04
<b>sum (of - fe)<sup>2</sup>/fe</b>					<b>= 10.76</b>

Hence, our observed value of  $X^2 = 10.76$ . Our question is now: is this observed value of chi-square sufficiently large to reject the hypothesis that motivational interviewing and relapse are independent? If there is no association between the two variables, what is the probability of obtaining an  $X^2$  value this large just by chance? Remember, motivational interviewing may not be associated with decreased relapse rates. If this is so, you may have happened by

chance to test a sample of clients for whom this is true.

To answer these questions, you must consult a **table of critical values** for the  $X^2$  statistic. This table reports the likelihood of obtaining various  $X^2$  values by chance alone. Any statistical textbook (see references) will provide a table of critical values. To use the table, you need to know the appropriate “degrees of freedom” and

**“level of significance”.** Most of the time, a level of significance of .05 is set, which declares that the difference between the groups is probably real if the likelihood of obtaining an  $X^2$  that large by chance is less than 5 percent. With two variables, a one degree of freedom test is indicated. At the .05 level of significance with one degree of freedom, the critical value of

$X^2$  is 3.84. That is, any observed  $X^2$  value greater than 3.84 would be considered statistically significant, while any observed value below 3.84 would be considered chance. In the present example, since the observed value of  $X^2$  is 10.76, we conclude that there is a statistically significant relationship between motivational interviewing and relapse rates.

## Independent samples t-test

The t-test is used to test whether two independent\* groups have truly different mean scores on some variable of interest. For example, you might want to know whether methadone clients who receive take home privileges use heroin fewer days per month than clients who do not have these privileges.

The t-test examines the difference between the two group means on a variable. It takes into account the variability of the scores and the number of observations in each group. A test statistic is computed ( $t'$ ) that shows whether the difference in the means across the groups reflects a true difference or random fluctuations. As in the chi-square statistic, a large t value indicates that the groups are quite different from each other (e.g., if you repeated the evaluation again, you'd find the same difference between the two groups). A small t value indicates that the difference between the means is likely due to chance.

A conceptual formula for t is as follows:

$t = (\text{mean for group 1} - \text{mean for group 2}) / (\text{standard error})$

or ...

$$\frac{(x_1 - x_2)}{\sqrt{\frac{(N_1 - 1)S_1^2 + (N_2 - 1)S_2^2}{N_1 + N_2 - 2} + \left(\frac{1}{N_1} + \frac{1}{N_2}\right)}}$$

As you can see, if the numerator of this equation is zero, the overall value of t will be zero, indicating that there are no differences between the groups. But if the difference is large, relative to the variability within the groups (the standard error), then t will also be large, suggesting that there is a true difference between the group means.

The following example presents data on the number of abstinent days reported during a 20 day methadone treatment programme, for two different client groups (there were 9 clients in each group):

\* The values of the observations in one group are independent of the values of the observations in the second group. For example, we may be comparing a group of clients treated with behavioural therapy vs. drug therapy. These two groups are clearly different from each other. There are differ-

ent types of t-tests for situations where the groups are not independent (e.g., when making comparisons for the same group of clients at two different points in time). Consult a statistical text and/or a research professional to decide what types of test are appropriate for your evaluation question(s).

	<b>Group 1</b> (clients without privileges)	<b>Group 2</b> (clients with privileges)
	13	11
	13	9
	10	12
	14	10
	14	14
	10	7
	12	9
	12	10
	10	8
	—	—
Means:	12	10
<b>Standard deviations:</b>	<b>1.66</b>	<b>2.12</b>

The question becomes: Are the mean number of abstinent days reported in the two groups significantly different from each other? (Or, are clients with privileges actually more abstinent than clients without privileges?).

This formula can be expanded as follows:

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{(N_1 - 1)S_1^2 + (N_2 - 1)S_2^2}{N_1 + N_2 - 2} + \left(\frac{1}{N_1} + \frac{1}{N_2}\right)}}$$

$$= (12 - 10) / [(8 \times (2.76) + 8 \times (4.49)) / 16] \times [1/9 + 1/9] = 2.23$$

Our observed value of  $t = 2.23$ . Our question is now: is this observed  $t$  value sufficiently large to reject the hypothesis that there is no difference in the number of abstinent days between the two groups? If there is no difference between the two groups, what is the probability of obtaining a  $t$  value this large just by chance?

To answer these questions, you must again consult a **table of critical values** for the  $t$  statistic, which reports the likelihood of obtaining various  $t$  values by chance alone. Any statistical textbook (see references) will provide a table of critical values. As in the case of chi-square, you need to know the appropriate **degrees of freedom** and **level of significance**. Most of the time, a level of significance of .05 is used, which declares that the difference between the groups is probably real if the likelihood of obtaining a  $t$  that large by chance is less than 5 percent. In the present example, the degrees of freedom are 16 (see a statistical text to determine degrees of freedom for a  $t$ -test). At the .05 level of significance with 16 degrees of freedom, the critical value of  $t$  is 2.12. That is, any observed  $t$  value greater than 2.12 would be considered statistically significant, while any observed value below 2.12 would be considered chance. In the present example, since the observed value of  $t$  is 2.23, we conclude that there is a statistically significant difference in the number of abstinent days reported by the two client groups.



## Into action

After describing the type of analysis for each research question, Adam S. gave a brief overview about the most appropriate data analysis for each evaluation (process and outcome evaluation).

### Process evaluation

For research question investigating the service coverage at the activity level number of clients would be calculated and compared to last years figure.

### Outcome evaluation

Data analyses procedures for the outcome evaluation were proposed as follows:

In order to compare pre- and post-intervention knowledge about AIDS, degree of engaging in high risk behaviours and attitudes towards AIDS a t-test for dependent samples would be carried out, because variables taken into account were all measured at an interval scale. The level of significance of .05 would be used, which declared an observed difference between pre- and post-intervention could have arisen by chance in 5% of cases.



## It's your turn (3C)

1. The partner who read this section should briefly describe the most appropriate data analysis procedures to his/her evaluation

partners. Remember, this is highly technical information, so a brief overview is all that is needed for your partners.

# Step 4



## Report results

(Note: This section should be reviewed by all evaluation partners.)

Your evaluation is not complete until the results are communicated to the user(s) of the research. You can present the results in writing, with pictures or graphs, or in oral presentations.

If your results are favourable to your current programme or treatment network, then you should plan carefully for how you

will present them, in order to maximise their impact. On the other hand, if your results indicate that your programme is not performing as expected or desired, then you have an opportunity and a challenge. You can decide what changes are needed in your programme, then present the results with the proposed changes at the same time.

## Written report

A detailed, written report of the evaluation is a good place to start. This report can be distributed within your evaluation team, and/or to the users of the results. Components should include:

### Title page

The title page should include the name of the report, the name and address of the programme, the time period of the evaluation, and the date the report was completed.

### Summary

The summary is very important because many people do not have time to read the full report and may read this only. Summaries are usually 1-2 pages long, and

contain the purpose of the evaluation, a very brief description of the programme, how the evaluation was conducted, important results and conclusions, and recommendations.

### Table of contents

The table of contents will help readers to find information quickly.

### Description of programme

This section should briefly describe the programme that was evaluated.

### Evaluation questions

Start by presenting your evaluation questions to the reader.

## Description of evaluation

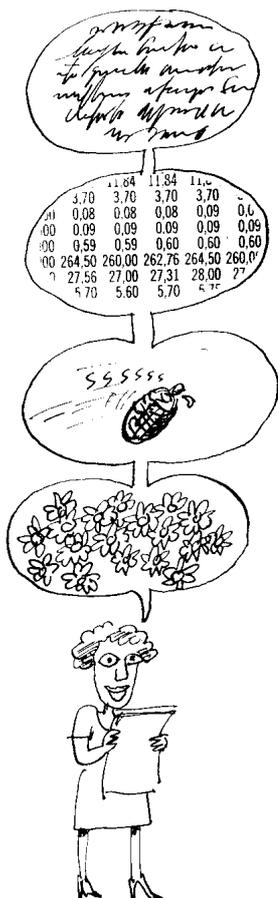
This section explains the methods used for the evaluation. It should describe who conducted the evaluation, and where, and what measures were used to collect data.

## Results

This section presents the results of the data analysis.

## Conclusions and recommendations

This section is very important, because it is where you can make recommendations about what to do in response to the evaluation findings.



## Oral report

If you are presenting information orally, it is important to remember that the average listener will retain only 3 or 4 main messages from your presentation. Be selective. Use visual aids to help orient listeners and keep their attention. Keep visual aids simple and to the point.

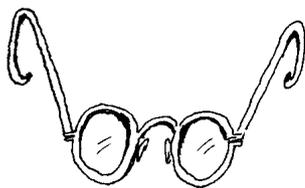
Think about highlighting certain results for different audiences. For example, administrators may want to know about efficiency or cost issues, while clinicians may be interested in hearing about the views of patients. Some audiences may be familiar with your programme, while others will not be familiar, and will need

background information about your organisation. In each case, it is important to tailor your presentation to your audience.

Consider giving oral presentations (and making written presentations available) in the following settings:

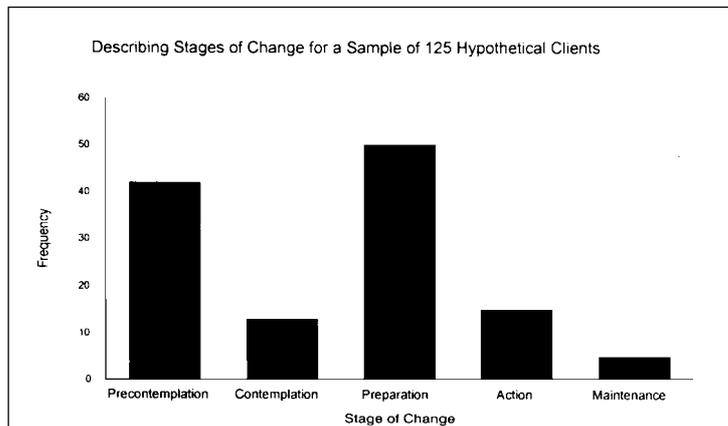
- A programme staff meeting
- A board meeting
- A meeting for funder(s)
- A meeting for patients/clients

## Visual aids

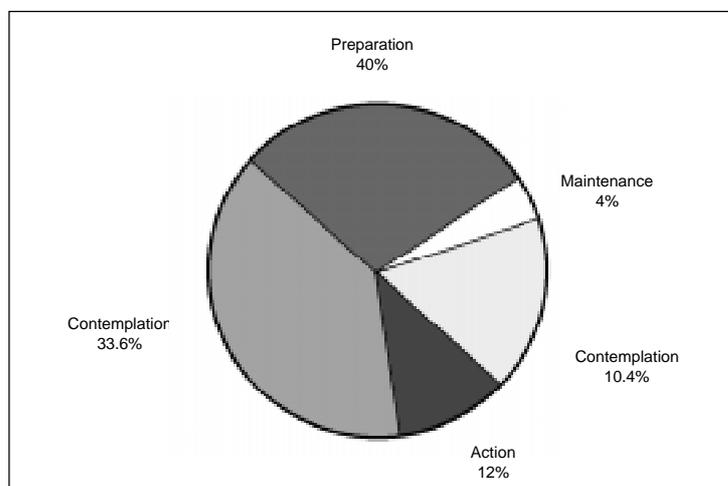


Bar graphs and pie charts are useful for summarising data in written and oral reports. A good bar graph has the following features:

- Both the categories and the frequency labels are clearly indicated.
- The length of each bar in the graph shows the frequency for that particular category
- The corresponding percentage of observations are presented at the top of each bar in the graph.
- All of the bars are of the same width.
- Adjacent bars are separated by a space.

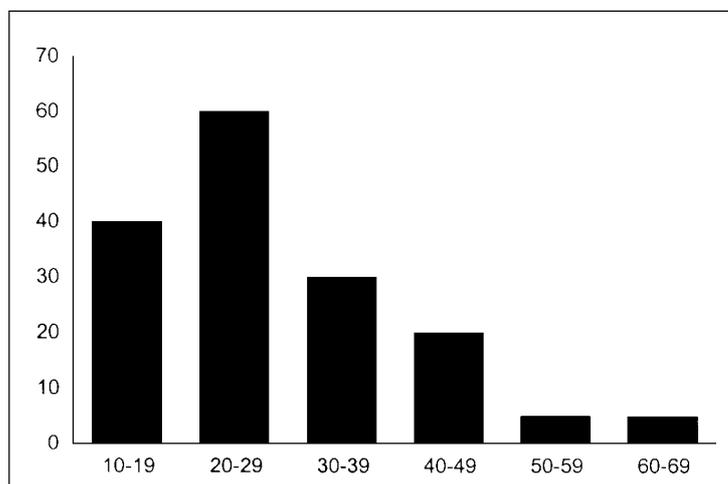


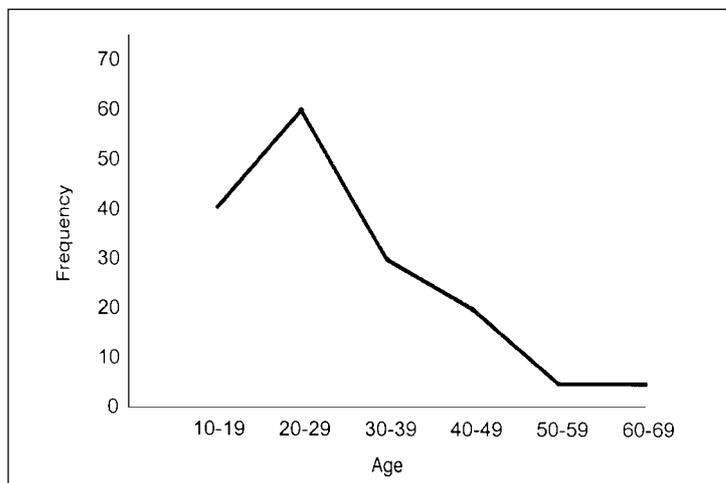
These same data can be displayed in a pie chart:



Grouped frequency data can be presented in two ways: histograms and line graphs. Histograms are like ordinary bar graphs, except that they (a) use interval-scale data, and (b) the data are grouped in intervals.

An alternative to the histogram is a line graph. Instead of drawing a bar for each interval, put a point above the centre of each interval marker and connect the adjacent points by straight lines.





## It's your turn (4)

- 1** Among your partners, nominate a person to take responsibility for preparing the written report. Use the format described above to create it.
- 2** Nominate another person to take responsibility for preparing oral report(s). Remember to tailor your reports to the intended audience. If you have multiple audiences, you may want to assign different people to prepare different reports:
  - Funder(s)
  - Expected users of the research
  - Clinicians
  - Patients

# Step 5

## Make use of what was learned

(Note: This section should be reviewed by all evaluation partners.)



After months or even years of time invested in conducting your evaluation project, you want to ensure that your results are put to good use. One way to accomplish this is to report your results in written or oral form (Step 4). It is equally important, however, to explore what the results mean for your programme and its service to patients/clients. Discuss your findings with expected user(s), community members, funders, and other important groups. Get their point of view on what should lie ahead for your programme. Do changes need to happen? If so, what is the best way to accomplish these new goals?

Before making changes in your programme, it is important to consider the limitations of your evaluation. If you have followed the steps outlined in this workbook series, you can feel reasonably confident in your results. Nonetheless, it may be prudent to attempt to replicate the study's results before making large-scale changes.

Following discussions, return to the expected user(s) of the research with specific recommendations based on your results. You should have done some of this work when preparing the written report, but now it is even more important for you to be clear and precise. List your recom-

mendations, link them logically to your results and mutual objectives, and suggest a period for implementation of changes. The examples below illustrate this technique.

Based on the finding that clients have to wait an average of 3,5 months before receiving an appointment, we recommend that the programme re-examine its assessment processes, with the goal of shortening assessments and/or moving to a group assessment format. The review could happen in January, and changes to the assessment process could happen by mid-February.

The results indicate that treatment B is superior to treatment A in terms of reducing PSU among patients. Therefore, we recommend that treatment B is offered to all clients, and that treatment A is reserved for those who do not show improvement with treatment B. A follow-up replication study, starting in April with a new sample of patients, is also recommended to confirm results.

In addition, consider taking your research results to the public. This can be accomplished through the local media, or through community groups. If approaching the media, prepare a one page summary statement of your study and its main findings.

If presenting at a community meeting, highlight the main findings and their implications using an oral presentation. Either way, be sure to explain your programme, the research, and the results using simple, non-technical language. Don't forget to make recommendations, if appropriate. For example:

Our findings indicate that most people with PSU problems do better in treatment when supported by a community member or family member. A trial programme that included family members in certain parts of treatment showed favourable results when compared to standard, individual PSU treatment.



## Into action

The following key recommendations were formulated by the evaluation team.

- 1** Based on our findings that the number of female drug users who enter treatment increased, the service needs to re-examine if the service meets the need of female clients.
- 2** The results indicate that less clients were referred to long term treatment in comparison to the previous year. It is recommended that the service re-examines their referral process and analyse possible barriers for clients to seek further treatment and how the service could provide help to overcome these barriers.
- 3** The findings show that the number of self-discharges decreased in comparison to the previous year. This represents a confirmation that the service provides good withdrawal management. Accountability requirements, safe withdrawal from opiates are met.
- 4** The results support the sufficiency of the AIDS education intervention. It is therefore recommended to continue the intervention and thus prevent further introduction and spread of HIV among intravenous opiate users and to prevent HIV transmission to their sexual partners. A randomised control trial, starting next September is recommended to confirm results.

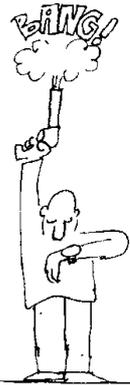
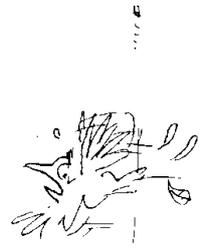
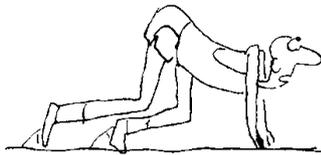


## It's your turn (5)

- 1** Working individually, think about your recommendations, based on your treatment research results. Write down the recommendations you want to make for each target audience (e.g., funders, clinicians, patients).
- 2** Discuss your ideas with your evaluation partners. Arrive at 1-3 key recommendations for each target audience.

# Step 6

## Start again



(Note: This section should be reviewed by all evaluation partners.)

**Remember, evaluation must be an attitude of continually questioning and gaining information.**

In the introduction to the Framework Workbook, you learned about the importance of creating a “healthy culture for evaluation” within your programme. You learned that a healthy culture for evaluation is one in which feedback loops are woven into the fabric of the treatment service or system. Some feedback loops serve the purpose of providing basic accountability data to programme or system funders, clients and the general public. Other feedback loops are better viewed as the means by which a programme or treatment system seeks to continuously improve their services and outcomes. The steps of planning and implementing evaluations provide you with the tools to make this happen in your own setting.

This final step, Step 6, serves as a reminder that feedback also should be directed toward new research and ongoing evaluation of your programme. Remember, evaluation must be an **attitude** of continually questioning and gaining information. Accordingly, you may want to consider replicating or extending this evaluation, or planning a new evaluation to answer a different question. Results from your current study will help guide you to your next evaluation activities. The steps outlined in Workbooks 1 and 2 can help guide future research efforts. Most important, now that you have established a culture for evaluation within your setting, you should take the necessary action to ensure that this culture remains healthy and strong for the future.



## Into action

Based on their research experience and their results the evaluation team had the following ideas for future evaluation projects:

- 1 to conduct a randomised control trial in order to confirm the effectiveness of the AIDS education intervention.
- 2 to plan and implement a process evaluation regarding coverage at the activity level in order to strengthen the culture for evaluation within the service.
- 3 to plan an outcome evaluation regarding illicit substance use, retention, transfer to further treatment.



## It's your turn (6)

- 1 Based on your overall research experience and results, what can you do next? Working individually, write down one new idea for your next evaluation research project. Discuss these ideas with your evaluation partners and make a plan for how to proceed.