MINISTRY of HEALTH
REPUBLIC OF BOTSWANA

GOVERNMENT OF BOTSWANA

MINISTRY OF LOCAL GOVERNMENT

Self-Directed Learning Workbook

District Level M & E Training and Reference Material for Primary Health Care Programmes

Workbook 2

Doing the "M" of M & E

I-TECH International Training & Education Center for Health

UNIVERSITY of WASHINGTON

UCSF University of California San Francisco

BOT USA PROJECT

NACA NATIONAL AIDS COORDINATING AGENCY

Acknowledgements

These three workbooks were developed by I-TECH Botswana (International Training and Education Center for Health Botswana) in collaboration with:

- Ministry of Local Government - Monitoring, Evaluation, and Research Unit
- Ministry of Health – Department of HIV/AIDS Prevention and Care
- Ministry of Health – Department of Policy, Development, Monitoring, and Evaluation
- National AIDS Coordinating Agency (NACA) - Monitoring and Evaluation Unit
- BOTUSA (US Centers for Disease Control and Botswana Ministry of Health)

We would like to thank the district monitoring and evaluation (M&E) officers who were involved in the review and piloting of the materials. We would also like to thank the University of Botswana graduates and other volunteers from Tebelopele and NACA, who participated in the piloting of the materials.

Foreword

Monitoring and evaluation (M&E) is an important process for collecting and analyzing data on health programmes that can be used for evidence-based planning and decision making. The information garnered from M&E increases capacity to improve health service delivery and the health of the population.

This set of three self directed learning workbooks is designed to provide information and guidance to you in carrying out M&E of health programmes in your district. The workbooks can be used as training documents as well as reference material for district-level M&E officers. They can also be used by other programme officers in the district who are involved in M&E.

The first workbook, entitled ***An Orientation to District-Level Monitoring & Evaluation*** is focuses on tasks and information necessary for newly recruited M&E officers who are beginning work in the field. It provides an orientation including: an overview of HIV and AIDS, the national health programmes in Botswana, job description, core activities of district M&E officers, an introduction to M&E, and an introduction to e-reporting of district health data.

The second workbook, entitled ***Doing the "M" in M&E*** focuses on monitoring activities. This workbook provides information on basic M&E processes. It also provides a practical overview of data collection, data management, data quality, basic data analysis, as well as a guide on presentation skills.

The third workbook, entitled ***Doing the "E" in M&E*** focuses on evaluation activities. This workbook provides information on designing evaluation studies, collecting and analyzing evaluation data, and writing reports.

Together these workbooks create a comprehensive view of the role that M&E officers are expected to play in their districts and provide you with the necessary guidance and tools necessary for you to fulfil this role. By actively reading the workbooks and completing the associated exercises, M&E officers will be prepared to carry out M&E activities for HIV and AIDS and other health programmes in their districts.

Table of Contents

Table of Contents

Self-Directed Learning Workbook2:
Chapter One
Introduction to M & E Concepts

Understanding M&E is an important step to succeeding as a professional

Chapter 1:
Introduction to M&E Concepts

Estimated time needed for completion: 3 hours

Chapter Overview

In workbook 2 we will focus on understanding what monitoring and evaluation (M&E) is, why it is important, and the basics of what it entails. The role of a district level M&E officer includes developing a plan to closely monitor programmes, collecting data on variables, managing and maintaining the data collected, analysing data, and presenting data. In this chapter, we will start with specifying what we mean by M&E. You will learn how to develop M&E frameworks using the causal pathway and how to create indicators to measure objectives. Additionally, we will outline the steps in developing an M&E plan.

Learning Objectives

At the end of this chapter you will be able to:

- identify the basic purposes and scope of M&E,
- differentiate between monitoring functions and evaluation functions,
- identify and understand how to use various types of frameworks and the causal pathway,
- describe how the causal pathway is used for M&E planning,
- describe the functions of an M&E plan,
- identify criteria for the selection and creation of appropriate indicators, and
- describe how indicators are linked to frameworks.

1.1 Introduction to Monitoring and Evaluation Concepts

(Information in section 1.1 was partially adapted from *OpenCourseWare. Fundamentals of Programme Evaluation* (JHU n.d.), *Introduction to Monitoring and Evaluation Course (MEASURE EVALUATION n.d.), and RHRC Consortium Monitoring and Evaluation ToolKit: The Causal Pathway* (RHRC n.d.))

M&E is an essential component of any intervention, project, or programme. This chapter is intended to help you understand and create M&E frameworks using the causal pathway, create indicators to measure objectives, and outline the steps in developing an M&E plan. M&E should not be an afterthought, but an essential part to the programme planning process and be completely integrated into implementation. Investing in M&E strengthens programme design and improves implementation. M&E also helps to justify allocation of limited resources, improve use of allocated resources (thereby increasing cost-effectiveness), generate knowledge, identify factors (individual, community, programmematic) that influence health outcomes, and meet organizational requirements. A well planned and executed M&E plan will help you understand if programmes are reaching intended objectives.

1.1.1 What Is Monitoring and Evaluation?

When you read that the coverage of antiretroviral therapy (ART) in a country is 25%, have you ever wondered how this calculation was derived? Or when you hear that the percentage of women in need of prevention of mother-to-child transmission (PMTCT) services and received it rose from 21% to 27%? Do you wonder how this data is known?

These types of statistics and other similar information result from monitoring and evaluation, also known as M&E. M&E is the process by which data collection is designed, collected, analyzed, and presented in order to provide information to policy makers and others for use in programme planning and project management.

The scope of M&E is broad and can be applied to any and all of the following:

- programme - organized effort to respond to a broad social problem (this can be organized at national level),
- project - specific set of activities with specific objectives that contribute to the overall objective of a programme or institution (usually organized at sub-national level), or
- intervention – can be same as project, or only be focused on a specific sub-set of a programme or project activities.

The term programme can be used in many ways. For our purposes, programmes are large and overarching efforts with many components and many levels. A programme may have multiple projects within it.

Monitoring: What are we doing?

Tracking inputs and outputs to assess whether programmes are performing according to plans

(e.g. people trained or condoms distributed)

Evaluation: What have we achieved?

Assessment of impact of the programme on behaviour or health outcome

(e.g. condom use at last risky sex, HIV prevalence)

1.1.2 Why do you conduct M&E?

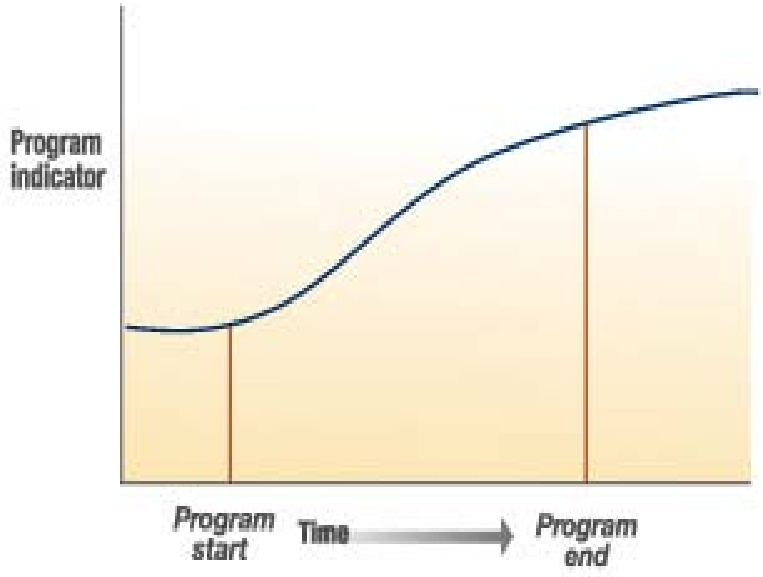
M&E are essential to programme success and tell you if programmes are reaching intended objectives. Some of the reasons for having a M&E system include to:

- see what has been achieved,
- measure progress against objectives,
- find out what works and does not work,
- identify strengths and weaknesses,
- identify ways to improve a programme,
- justify changes to work plans and budgets,
- identify areas for technical support or capacity building,
- share information and lessons learned, and
- manage programmes more effectively.

1.1.3 Monitoring Defined

Monitoring of a programme or intervention involves the collection of routine data that measure progress toward achieving programme objectives. Monitoring is sometimes referred to as *process evaluation* because it focuses on the implementation process. Monitoring provides regular feedback that measures change over time in any or all of the programme components. Programme monitoring can also be stated as the ongoing tracking of information about the activities and operation of a project or programme to determine what the project or programme is actually doing and whether activities are being accomplished. Its purpose is to permit stakeholders to make informed decisions regarding the effectiveness of programmes and the efficient use of resources.

Figure 1.1.3 A graphic illustration of programme monitoring over time.



In Figure 1.1.3, the vertical axis ("Y" axis) is measuring the programme indicator. The programme indicator could be any element of the programme that needs tracking. Some examples of programme indicators are the cost of supplies needed to run the programme, the number of times staff provide certain information to clients, or the percentage of clients who are pleased with the services they received. The horizontal axis ("X" axis) is measuring time. Therefore, the curved line indicates change in programme indicator over time.

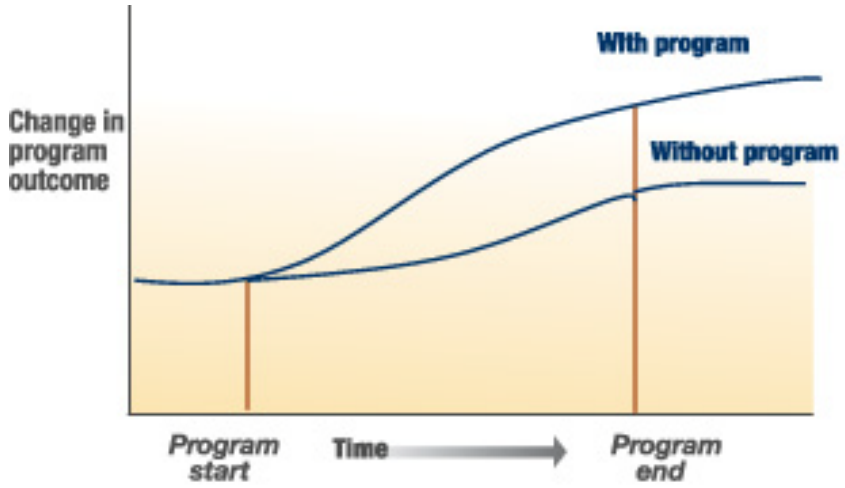
Examples: Monitoring answers questions such as:

- How well has the programme been implemented?
- How much does implementation vary from site to site?
- Did the programme benefit the intended people? At what cost?

1.1.4 Evaluation Defined

Evaluation is a systematic process limited in time of collecting, analyzing, and using information to assess the effectiveness, relevance, and impact of achieving your programme's goals. It provides feedback that helps determine the consequences, outcomes, and results of a programme. Evaluation can also be stated as the *periodic* systematic collection of information to make judgments, improve effectiveness, identify lessons learned, and determine whether a programme has had the desired impact.

Figure 1.1.4 A graphic illustration of programme evaluation.



Evaluations generally require:

- data collection at the start of a programme (to provide a baseline) and again at the end, rather than at repeated intervals during programme implementation;
- a control or comparison group in order to measure whether the changes in outcomes can be attributed to the programme; and
- a well-planned study design.

Examples: Evaluation answers questions such as:

- How much of the observed change was *caused by* the programme?
- What would have happened without the programme?
- Were programme goals achieved?

1.1.5 Monitoring versus Evaluation

In summary, monitoring activities measure inputs, processes, outputs, and outcomes usually with data from your programme and patients. Therefore, monitoring is an ongoing continuous process. It requires the collection of data at *multiple* points throughout the programme cycle, including at the beginning to provide the *baseline* status of an indicator at the beginning of the programme. Monitoring can be used to determine if activities need adjustment during the intervention to improve desired outcomes.

Examples: Monitoring usually pertains to *counting*, *tracking*, and *collecting*. Examples include:

- counting clients seen or health workers trained,
- tracking condoms distributed, and
- collecting data on clinic clients.

In comparison, evaluation activities tell you how well programme activities have met expected objectives and/or the extent to which changes in outcomes can be attributed to the programme or intervention. Evaluation is fundamentally an exercise to help decision-makers understand *how*, and *to what extent*, a programme is responsible for particular, measured results. This exercise usually aims to measure the effectiveness and efficiency of a programme or a project. Variables used to assess impact of programmes or projects overlap with indicators that are also used in programme monitoring.

M&E use many shared and overlapping data sources but the difference lies in the evaluation/research design and analytic techniques used. Monitoring and impact evaluation play different, but related and equally essential, roles in effective programme management.

To reach programmatic goals, monitoring and evaluation are both equally important components of effective programme management.

1.2 Frameworks

(Information in section 1.2 was partially adapted from *OpenCourseWare. Fundamentals of Programme Evaluation* (JHU n.d.), *Introduction to Monitoring and Evaluation Course* (MEASURE EVALUATION n.d.), *and RHRC Consortium Monitoring and Evaluation ToolKit: The Causal Pathway* (RHRC n.d.))

A framework can help you in the design of a programme and in the M&E process. Frameworks are visual representations of relationships.

In M&E, we use frameworks as communication and planning tools. Frameworks serve as a map for programme goals. They visually represent links between resources and activities. They also guide the systematic selection or creation of programme indicators. The purpose of their use is to clarify assumptions, goals, and interrelationships among factors relevant to the project or programme and to define desired results in terms of planned activities and realistic, objective impacts.

The **conceptual framework** is a graphical depiction of relationships between factors thought to influence the problem of interest. The **logical** or **results framework**, on the other hand, links the *goal* and *objectives* to the interventions. Figure 1.2.a depicts a framework of the factors thought to be determinants of health. Figure 1.2.b depicts a logical framework representing how the goal of improved health status can be achieved.

Figure 1.2.a A graphic illustration of a conceptual framework that shows the determinants to health **(Roberta 2005)**

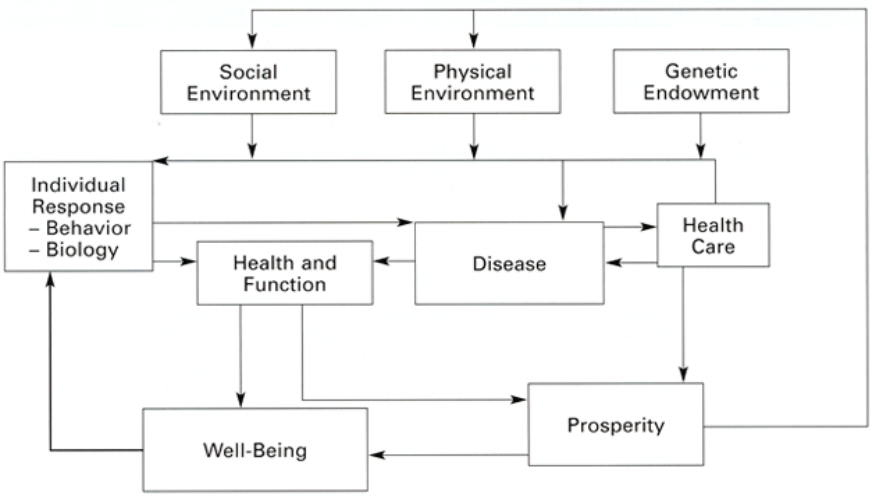


Figure 1.2.b A graphic illustration of a results framework that shows objectives involved to achieve improved health **(Marsha 2008)**

For the rest of this section, we will focus on a third framework known as the causal pathway framework. This framework is called causal because it is based on the premise that the activities you carry out should logically cause desirable results to occur and pathway because it is based on the idea that the causal links form a technically and programmatically sound logical progression. The causal pathway framework is also called the logical framework.

This causal pathway framework may not be very different from the way you already think about planning programmes, or you may already be using a tool or process that you find helpful. That is fine. In such cases, this section may serve as a reminder for you on what you already know; for others, it may be a new way of thinking about good programmes.

1.2.1 Using the Causal Pathway to Design Projects and M&E Plans

The causal pathway is divided into five main components: *1) impacts, 2) outcomes, 3) outputs, 4) activities, and 5) inputs.* Each of these components will be described.

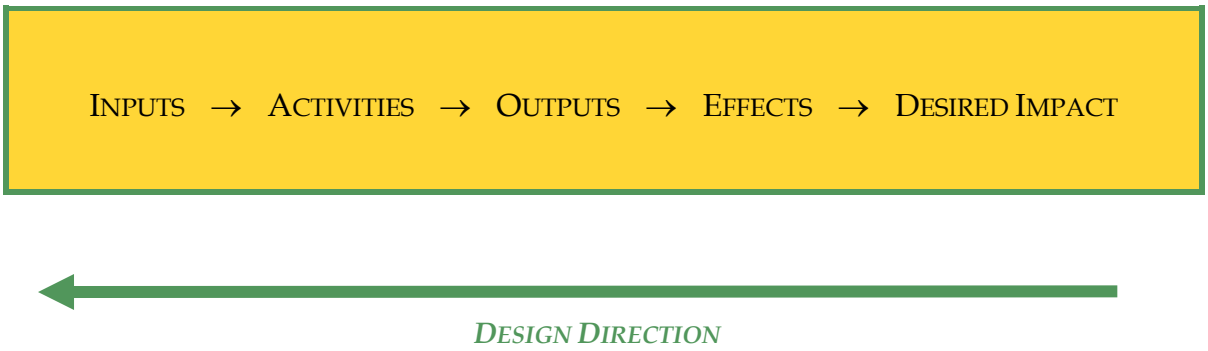
When thinking about a programme or project, most of us tend to think about the activities of the programme. Whether it is community education or clinical health services or stigma reduction, we often describe our work through the activities we do every day. However, we also need to keep the big picture in mind, to remember that our programmes have a broader purpose.

For example: we are handing out condoms not for the sake of handing out condoms; but to increase the availability and use of an effective method of sexually transmitted infection (STI) and HIV prevention in order to reduce transmission, which will contribute to reduced illness and death among the population in Botswana.

The ultimate big picture reason that we undertake many of our programmes is to improve the social, economic, or health status of the population with which we are working. In the language of the causal pathway, this ultimate purpose of the programme is the desired *impact*.

Figure 1.2.1.a depicts the components of the causal pathway. As we discuss the causal pathway, we will start at desired *impact* and work our way back towards *inputs*. When designing an intervention, it is useful to design the intervention in this manner starting with the *impact* and working back to the *inputs*.

Figure 1.2.1.a Causal Pathway



Impact

Impact is a change in the health, social, or economic status of the population of interest. In the condom distribution example used previously, the desired *impact* is reduced illness and deaths due to STI and HIV, which will contribute to better health status. In a visual display of the causal pathway, you will place the programme's *impact* at the very end of the pathway, since that is where we are ultimately headed.

Figure 1.2.1.b Causal Pathway-Desired Impact

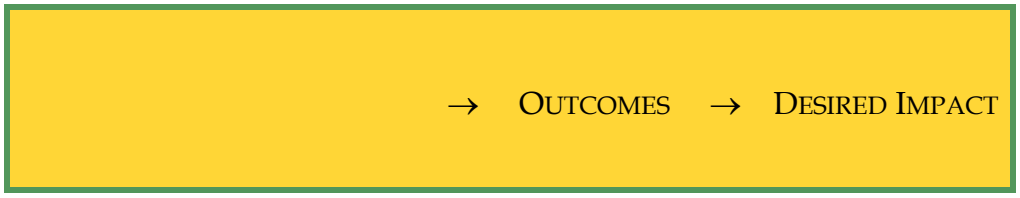
Now that you have an ultimate goal, you go through the process of carefully examining how to get there. You must ask yourself, "What has to happen before we can see the improvement that constitutes the desired *impact*?" In order for a programme to result in the desired *impact*, people must choose to change things about themselves, typically their knowledge, attitudes, skills, intentions, and/or behaviours. In the language of the causal pathway, we call these changes *outcomes*.

Outcomes

Outcomes are the changes in knowledge, attitudes, skills, intentions, and/or behaviours of the population of interest that contributes to the desired *impact.* Continuing with the condom distribution example, before you will observe a change in illness and death due to STIs/HIV (the desired *impact*), people must at minimum know that condoms prevent disease (knowledge), believe that condoms are effective (an attitude or perception), and actually use the condoms (a behaviour).

In the visual display, you place the *outcomes* before the *impact*, since these changes must occur BEFORE the *impact* can be achieved.

Figure 1.2.1.c Causal Pathway-Outcomes



In the next step on the causal pathway, you must once again ask, "What must be in place to enable people to make the changes described in the *outcomes*?" In service delivery projects, a set of products and services must be available if we expect people to use them. These products and services are known as *outputs.*

Outputs

Outputs are the products and services that must be in place for the o*utcomes* to be achieved. Continuing with the previous example, people cannot use condoms if none are available. One *output* needed to achieve condom usage (an o*utcome*) is a reliable, convenient source of condoms. However, making condoms available might not be enough. People might not use condoms even if they are available, if they do not believe there is an advantage to using condoms. In addition to providing the condoms themselves, other *outputs* may be necessary for services to achieve changes in knowledge and attitudes (*outcomes*). These services could be a community education programme, a mass media campaign, or school-based programmes. Thus, in our visual display *outputs* are placed before *outcomes*.

Figure 1.2.1.d Causal Pathways - Outputs

→ OUTPUTS → OUTCOMES → DESIRED IMPACT

Examples of *outputs* include:

- volume of drugs distributed to regional storage facilities,
- number of facilities using TB directly-observed therapy,
- number of clinic staff trained in rapid testing, and
- number of TB cases tested for HIV.

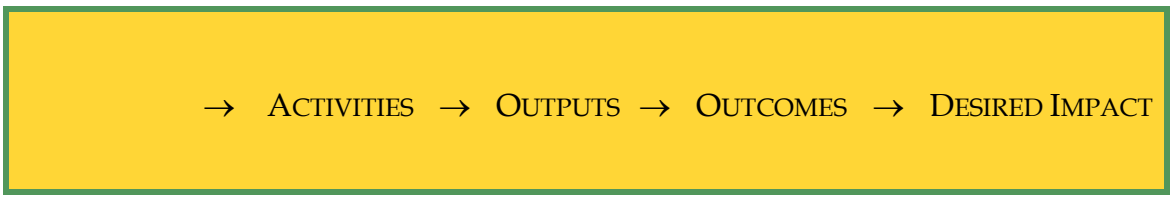
It can be easy to confuse *outcomes* and *outputs*. *Outputs* are measured at the programme level such as activities conducted (number of brochures printed) or service utilization (number of client visits, number of condoms sold, number of HIV tests performed). *Outcomes* are measured at the population level (percent of TB patients successfully treated) or (percent of youth using condom at last sexual encounter).

The next question you must ask is, "What does the project have to do in order to produce a reliable condom supply and a community education network (the *outputs*)?" It must do all of the *activities* that we often think of as the core of the programme such as mobilize and involve the community; develop curricula and community education materials; select and train workers to set up and maintain the condom supply system and to educate the community; supervise and support workers; and many other tasks. In the causal pathway, these are obviously called *activities*.

Activities

Activities are the technical and support tasks required to produce the *outputs*. *Activities* that would lead to the output of increased condom use could include an education campaign targeting youth on the benefits of condom use; meetings with elders and community leaders to answer questions and dispel myths they may have about the dangers of condom use; and distribution of condoms to each clinic in the area with a request that they be set in public areas and on reception desks.

Figure 1.2.1.e Causal Pathway - Activities



Inputs

Inputs are the resources required to support your *activities*. Before you can begin your *activities* the necessary resources must be available. These typically include funds, staff, materials, sites, and community goodwill. With the condom example, inputs include funds for education campaigns, buying, condoms and distributing condoms; staff for training and distribution; and education and distribution sites.

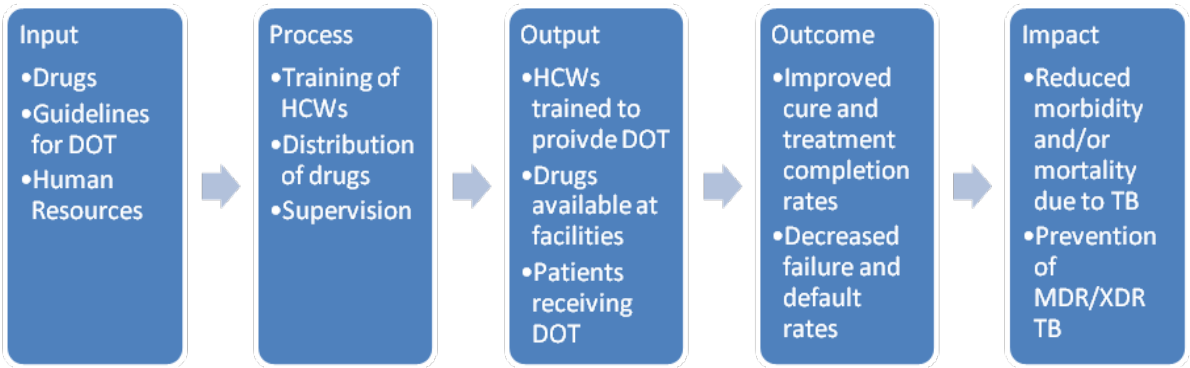
Figure 1.2.1.f Causal Pathway- Inputs

INPUTS → ACTIVITIES → OUTPUTS → OUTCOMES → DESIRED IMPACT

Unless you can trace precisely how your *activities* are linked to your desired *impact*, you risk spending time, effort, and money on activities that are not in the causal pathway. If you find that there is no clear link from an *activity* to an *output* that is needed to obtain the intended *outcomes* and the ultimate *impact*, then you should question whether that particular *activity* is needed at all.

In figure 1.2.1.f you will find an example of a very simple causal pathway for TB DOTS, a TB treatment programme in which TB patients are directly observed taking their TB medication by a clinician rather than given medication to take at home. This programme was designed to reduce treatment failure and default rates.

Figure 1.2.1.g Sample Causal Pathway



1.2.2 Using the Causal Pathway to Monitor and Evaluate Projects

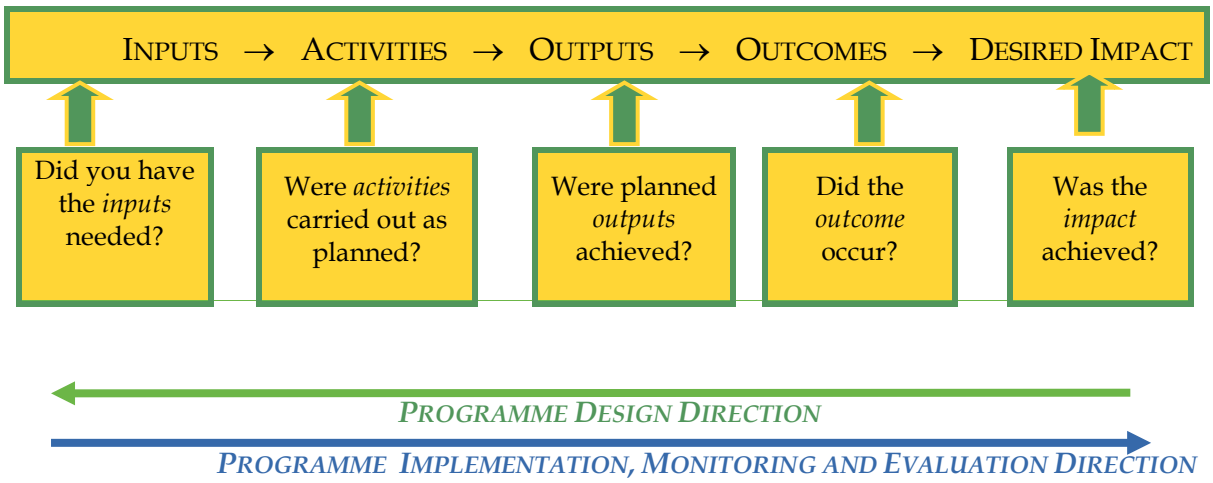
When you summarize your programme in the form of a causal framework, it pushes you to ask questions about how you will know whether your expectations are met. How will you know if condom supplies are adequate and consistent? How will you know if condoms are used? How can you tell if HIV and STI morbidity and mortality have declined?

After developing a causal pathway, it is useful to develop a causal hypothesis that summarizes the proposed factors and relationships. A causal hypothesis indicates that *this* set of *inputs* and *activities* will result in *these* products and services (*outputs*), which will facilitate *these* changes in the population (*outcomes*), which will contribute to the desired *impact*.

Causal hypothesis example: Intensive, individualized HIV education and condom distribution by peers will increase awareness of HIV, willingness to use condoms and consistent use of condoms by young adults in Chobe District, thereby improving their overall health through a reduction in HIV and STI incidence.

Figure 1.2.2 depicts the questions guiding the causal hypothesis. The purpose of the project's M&E system is to help you answer these questions. You must decide, at the design stage, what information you need so that the information collection can be built in from the very start of the project. While the programme design direction is backwards from *impact*, the monitoring & evaluation direction is forward from the *inputs*.

Figure 1.2.2 Questions Guiding the Causal Hypothesis



1.2.3 How Do We Know if the Causal Pathway Steps Occurred?

To know whether the causal pathway is happening as you expected and stated in your causal hypothesis, you need to measure the steps along the pathway. This will tell you if a specific link in the pathway is occurring as planned, or if the link is breaking down and causing a dead end on the pathway. Either way, it is valuable information. If it shows you are on track, it reinforces what you were already doing. If a link is not occurring as planned, then you have a problem and have to do something about it.

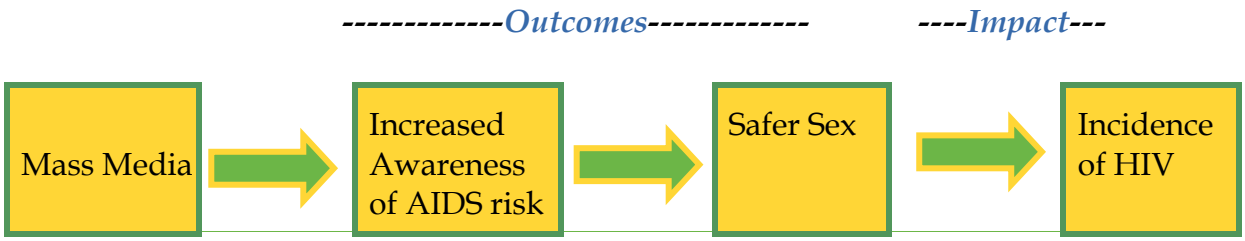
Example: if you are training workers (*activity*), but the workers are performing no better in their jobs (*output*), then that link (i.e., that training leads to better performance) is not working. If your causal hypothesis depends on good workers to encourage the public to use your services (an *outcome*) to improve their lives (the *impact*), then the broken link can cause the rest of the pathway to fail.

The example able is why it is important to look beyond *activities* and very basic measures of *outputs*. In this example, if you were satisfied with simply counting the number of workers trained, you would not see any problem since, by that measure, the project is on track. Yet, by looking a bit further – at how they are doing their jobs after training – you see that the training is not causing the kind of change you need in the workers.

Outcomes should always be thought through, planned for, and specified by your *causal pathway*. However, outcome evaluations and impact evaluations are more often implemented at the national level on a periodic-basis due to the financial, logistical, and technical expertise needed. For example, the *Botswana AIDS Impact Study (BAIS)* is conducted every four years or so, which provided important data on sexual behaviours and HIV rates.

In determining whether your outcomes were successful it is important to be aware of the difference between *theory failure* and *implementation failure*. To explain the difference, let us consider an example of a mass media programme proposed to reduce HIV transmission (Figure 1.2.3).

Figure 1.2.3 A conceptual model depicting of how a mass media programme may reduce HIV transmission



Implementation failure is when a programme is not implemented as planned. When implemented as planned, a mass media campaign should be able to increase people's knowledge as well as change attitudes. Implementation failure, however, will occur in instances in which the population finds the mass media campaign to be interesting, but it is not delivered in such a manner to actually increase knowledge and change attitudes.

Theory failure is when the programme is implemented as planned but the intervention does not produce intermediate results and/or they do not produce the desired outcome(s). For example, the mass media campaign may be effective at increasing people's knowledge of HIV (i.e. people understand about the dangers of HIV and how they can protect themselves). Increased knowledge alone may not, however, lead to safer sex. Based on the mass media campaign, a man may understand that using condoms is important to reduce his risk of acquiring HIV, but if condoms are not readily available, he will not use them. Similarly, a woman may learn the importance of using condoms, but not have the skills necessary to negotiate safer sex with her partner. These are examples of theory failure.

We have to be careful in our assumptions. We may not be able to assume that just because we make information available, the population is more knowledgeable and that this knowledge will translate to behaviour change. What we cannot assume, we should measure.

1.2.4 How Do You Measure the Framework Steps?

You measure the steps in the causal pathway using indicators. You need to focus on developing indicators for *outputs, outcomes, and impact* since they are what you are trying to create from the *inputs* committed to a project and the *activities* carried out. All projects should measure their o*utputs*. Many will measure o*utcomes*, but very few will measure *impact* due to the resources needed.

Indicators

Indicators are clues, signs, or markers that measure one aspect of a programme and show how close a programme is to its desired path and outcomes. They are used to provide benchmarks for demonstrating the achievements of a programme. One of the most critical steps in designing an M&E system is selecting appropriate indicators. The M&E plan should include descriptions of the indicators that will be used to monitor programme implementation and achievement of the goals and objectives. Quantitative or qualitative measures of programme performance are used to demonstrate change and detail the extent to which programme results are being or have been achieved. Indicators can be measured at each level: input, process, output, outcome, and impact.

Indicators should be formulated precisely so that the measure is consistent from one time to the next. Indicators are typically formulated as numbers or proportions (percent, per 1,000, per 100,000, etc). In the development of your M&E plan, make the effort to replace general concepts. For example, "Prenatal care" is not an indicator with precise definitions a better example would be, "number of prenatal visits to clinic X in period Y" or "percent of women with children under 5 who had at least 3 prenatal visits during last pregnancy."

When developing indicators it is important to keep in mind the quality of the data sources, the frequency of the data collected, and the cost of collecting additional data. These could all impact the type of indicator that you choose. In addition it is important to map out who is responsible for data collection, data entry, and data analysis in order to trouble shoot potential data problems.

Output Indicators

Output indicators measure products and services provided by the programme as well as the quality of these products and services. Usually, project records are the main source of information for output indicators. Any programme should track very basic measures of *outputs*, like "number of condom posters printed" and "number of workers trained." It can also be useful to express the numbers as proportions, by comparing what has been achieved to what was planned. To calculate the indicator, "percent of planned posters printed," you would divide the number printed so far by the total number the project plans to print. Since the quality of the products and services produced is often key to the success of programmes, it can also be very useful for projects to measure quality.

Choosing o*utput indicators* is an essential part of every project design. It is important to note that as important as they are, output indicators only tell us about the work *the programme* is doing. They do not tell us anything about the o*utcomes* these products and services have on the population.

Remember!

An important note about indicators: Two different projects might use the same *output* indicator "number of workers trained" but have very different *targets*. One project might plan to train 1000 workers and the other one 50 workers. So if you are converting it into a proportion of number of services planned, you must be aware of this difference.

Learning Activity 1.2.4

Indicators for a Causal Pathway

Directions: Consider the following three indicators for a project in which good staff performance is an important part of the causal pathway.

Which indicator is the best measure of the implementation of the activity "to train X health outreach staff to give high quality reproductive health education to mothers in the PMTCT programme?"

a. Number of health outreach staff trained.

b. Percent of trained health outreach staff who received rating of "good" or "excellent" on reproductive health training exercise.

c. Percent of trained health outreach staff who perform reproductive health responsibilities "well" or "very well".

Discussion 1.2.4

Indicators for a Causal Pathway

1. Which indicator is the best measure of the implementation of the activity "to train X health outreach staff to give high quality reproductive health education to mothers in the PMTCT programme?"

The best answer is **C**.

a. Number of health outreach staff trained.

This indicator is very common in all kinds of projects, and it is important to track but it does not tell you anything at all about staff performance, the aspect that is of interest.

b. Percent of trained health outreach staff who received rating of "good" or "excellent" on reproductive health training exercise.

This indictor introduces an element of quality. In addition to knowing the number of workers trained, you could measure the proportion (percent) who achieved a desired competence level by the end of training. You would have to define specifically what you mean by "good" and "excellent" and then systematically measure it.

C. Percent of trained health outreach staff who perform reproductive health responsibilities "well" or "very well".

This indicator is the strongest because it does not measure just whether staff were trained, or whether the just 'know' the material on the last day of training. It measures how well they actually do their jobs on the ground. This indicator has an element of quality that is beyond the scope of a single activity, but incorporates multiple activities such as training and follow-up supervision. Again, the project would have to clearly define the standards for performing "well" or "very well"; these standards could then be incorporated into a checklist that the supervisor uses during visits to the field. The workers would be observed as they carry out their jobs and rated. If these field observations were done regularly over time the project could track whether quality is at an acceptable level, is declining, or is improving.

It clearly takes more effort to measure quality. It is up to you at the project design or redesign stage to decide whether having this information warrants the extra effort.

Outcome Indicators

Outcome indicators (also called effect indicators) measure the knowledge, attitudes, skills, intentions, and behaviours of the population you are trying to help. It is usually a good idea to measure *outcomes* in your project if you can. It can be difficult, however, because it can require measurement at the population level (such as with a community-based survey). This is technically harder, more expensive, and takes more time than using project-based records.

Like all indicators, *outcome* indicators must be formulated carefully. In many cases, they can follow the form: The [number or percent] of [population group of interest] who [know/believe/do] [specific knowledge/attitude/skill/behaviour]

Examples of *outcome* indicators:

- percent of adolescents aged 15-24 who know at least 3 correct sources of condoms in their community (an indicator of knowledge),
- percent of adolescents aged 15-24 who correctly demonstrate how to put on a condom using a wooden penis model (an indicator of skill), and
- percent of adolescents aged 15-24 who report having used a condom the last time they had sex (an indicator of behaviour).

The examples above are good *outcome* indicators because they are precise and meaningful. Notice the element of quality in the second indicator: we are measuring whether the adolescents know how to put on a condom *correctly*. Be precise and specific as you formulate your *outcome* indicators.

Avoid formulations like the following:

- "Adolescents who know about condoms." It does not specify number or percent and it is not clear what they are supposed to know about condoms (i.e., that they exist, how to use them, where to get them, etc.).
- "Number of adolescents who use condoms." It is best to be more specific. For example, do you want to know if the adolescent has ever used a condom? Used it at last sex? Used it consistently during sex for the last 3 months?

To measure the indicator, "percent of adolescents aged 15-24 who know how to correctly put on a condom," your project would have to conduct a community-based survey with a random sample of this group (i.e., adolescents aged 15-24). You may ask why not just interview young people at schools, which would be far easier than going door to door? The answer is that you need a *representative* sample of the

population of interest – in this case, adolescents aged 15-24. Young people in school are often very different from those young people not in school, and those differences will affect your results. If you interview young people in schools, then you cannot conclude anything about the broader population of adolescents. Your findings will be limited to school-going youth.

If you are able to measure o*utcome* indicators at baseline and follow-up, you will know if there was any change in knowledge, attitudes, skills, or behaviour in the population; but not necessarily whether your project caused the change. However, collecting such population-based measures may require setting funds aside to hire a survey consultant, or hiring programme managers with survey experience, which can be costly and time consuming and necessitates early planning. Nevertheless, *outcome indicators* are often more meaningful than output indicators since they tell us information directly about the population we are trying to help. This is closer to the end of the Causal Pathway, and that is something to aim for.

Remember!

We reserve the term o*utcome* for the population we are trying to help. It does not include our staff or other workers, even if they are volunteer community workers and therefore part of the community. So, while it may be important in our causal pathway to change the knowledge and behaviours of nurses, lay counsellors, and volunteers, that is an *output* because they are the ones delivering the services to the broader population. When the general population of men and women gain knowledge and alter *their* behaviour, *then* we have achieved an *outcome*.

Impact Indicators

Impact indicators measure the health, social, or economic status of the population of interest. They are often rates or ratios and are virtually always population-based measures.

Examples of *impact* indicators:

- crude mortality rate (also called crude death rate) is the number of deaths per 1000 population in a given time period,
- maternal mortality ratio is the number of maternal deaths per 100,000 live births, and
- total fertility rate is the number of children a woman will bear in her lifetime at current age-specific fertility rates.

CHECKLIST FOR SELECTING INDICATORS (RHRC Consortium, no date)

Before you finally determine the indicators for your project, consider the following criteria. Indicators (and the methods used to collect them) should be:

- *Ethical:* Information should be ethically obtained and managed. When you question people about personal information, it is important to respect their confidentiality and security. You should tell them how you will be using the information as well as who will have access to it. Use informed consent.

- *Useful:* Do not collect "nice to know" information. Collect only the information that is directly related to your causal pathway so it will inform you about your progress and guide you in decision-making.

- *Scientifically robust*
  - Indicators must be *valid*. Indicators are substitutes for the truth, and we want to choose indicators that get us as close to the truth as possible. If we want to measure "students' knowledge" in an education programme, we could choose, "% of students who pass the final exam." But is this a valid measure of knowledge? Some students are better at taking tests than others. Some students will not feel well the day of the exam. Their scores might not reflect their true knowledge.

  - *Reliability* refers to the consistency of measures over time or over measurers. If two outreach workers measure "household cleanliness" on a 4-point scale, they should both come up with the same rating. And, if they come back in 6 months and the house is in the same condition, the rating each gives should be the same as it was on the previous visit.

  - An indicator that is both *sensitive* and *specific* gives you accurate measures and not false results. For example, if a woman reports that she "has enough food," it should mean just that, and not that she is too proud or too ashamed to accept food. If she reports that she "does not have enough food," it should mean just that, and not that she wants extra rations.

  - The indicators you choose must be *accessible.* That is, you must be able to collect the information with a reasonable level of effort. To gather population-based data, it is necessary to conduct a population-based random sample survey. If your project does not have the time, money, or expertise for this undertaking, then you cannot use this indicator in your M&E plan because it is not accessible.

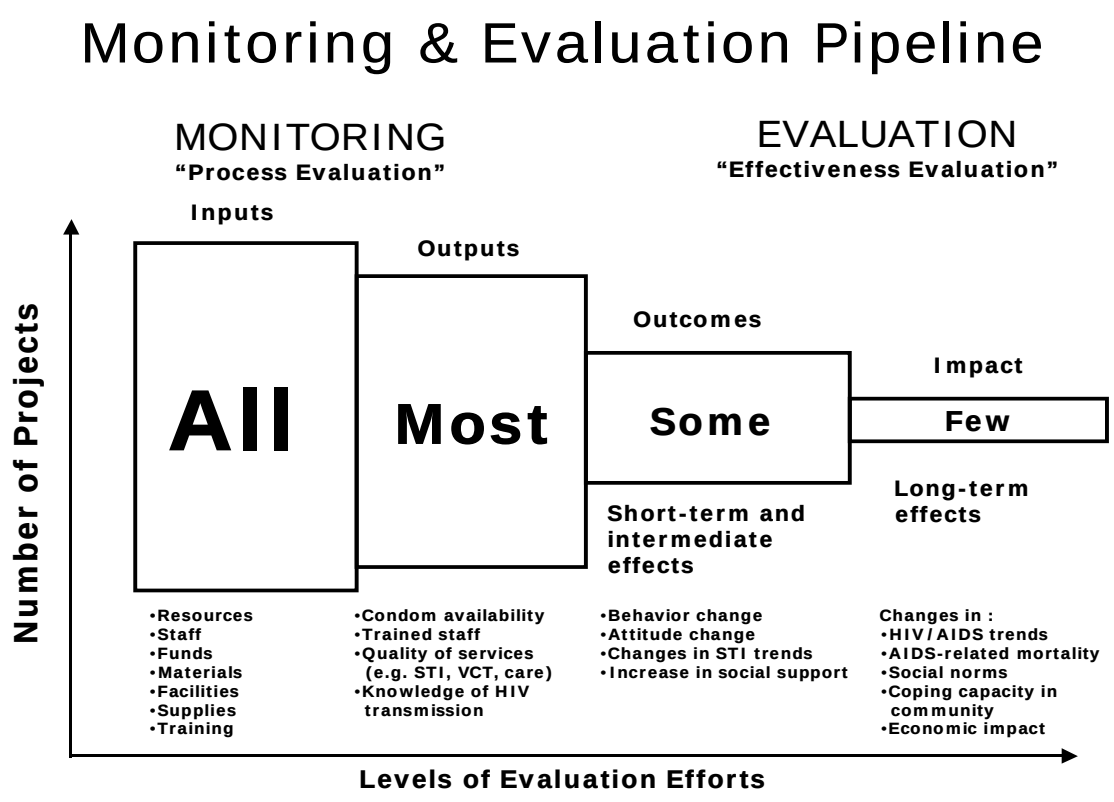
1.3 PROJECT OBJECTIVES

(Information in section 1.3 was partially adapted from *OpenCourseWare. Fundamentals of Programme Evaluation* (JHU no date), *Introduction to Monitoring and Evaluation Course* (MEASURE EVALUATION no date), *and RHRC Consortium Monitoring and Evaluation ToolKit: The Causal Pathway* (RHRC no date)).

When you have completed your programme design and are developing the M&E plan, you will have to decide if you are going to measure *outputs* only or if you are also going to measure *outcomes* and/or *impacts*. This is an important decision because it determines what you will choose as your project objectives.

In the causal pathway framework, the project objectives are what the project promises to accomplish and measure. You can think of it as the farthest point along the causal pathway for which you will have evidence or data. In most projects, the objectives will be one or more of the *outputs* or *outcomes* you have specified in your causal pathway; in some cases, it will be the *impact*.

Figure 1.3 M&E Pipeline



SMART Objectives

The objectives listed in the programme description should be **SMART-** an acronym that stands for:

S **Specific:** *The project's intended accomplishments must be clearly specified.*

M **Measurable/quantified:** *The intended accomplishments must be quantifiable and measureable.*

A **Attainable:** *This is a reality check: consider the context and resources you have, and whether the size of the planned change is feasible given the available resources.*

R **Relevant:** *Your objective (and your programme) must address a problem identified as important in the needs assessment and the objective is appropriately related to the programme's goal?*

T **Time-bound:** *Specify a time limit for your objective and programme.*

Learning Activity 1.3.a

Deconstructing Objectives

Directions: Review the two objectives below. Determine which objective is the best and list your reason(s) why.

a. To train 18 secondary school teachers in HIV education in this province within 6 months.

b. To increase the number of well-qualified teachers providing good quality HIV education in secondary schools in this province from 32 to 50 within 6 months.

1. Which is objective is written the best? _____

2. Why?

Discussion 1.3.a

Deconstructing Objectives

a. To train 18 secondary school teachers in HIV education in this province within 6 months.

b. To increase the number of well-qualified teachers providing good quality HIV education in secondary schools in this province from 32 to 50 within 6 months.

1. Which is objective is written the best? ___***B***_____

2. Why?

Objective A is actually an activity, something you have to do to produce an output, which is reflected in the second objective. To have simply trained school teachers does not tell you anything about whether education is being provided, or about the quality of that education.

Try to avoid stating your objectives as activities. Focus on the output – the product or service you expect to occur because of that activity – or, to move even further down the pathway, to the outcomes that you expect will occur.

Objective B forces you further along the causal pathway; it focuses on the education being provided which is why you are training teachers in the first place. It also focuses on the quality of the education being provided.

Output Objectives

Examples of *output* objectives:

- To provide high quality HIV education and condoms in the project area through a network of 200 community health workers, to begin within 12 months of project start-up.
- To increase the number of schools in the project region providing standardized secondary HIV education to adolescent men and women from 0 to 10 in 12 months.

Notice that the above objectives focus on delivering a particular service, rather than on what you expect to happen as a result of the service. In the first example, you are promising to provide "good" quality HIV prevention services and you specify the means by which the project will do that. Thus, you must measure quality of the service you will provide. You have not promised anything about how many people will use condoms (*outcome*), or that mortality from HIV will decline by a certain amount (*impact*). The emphasis is on the service delivery, because it is an *output* objective.

The same points can be made for the second example. You are promising that there will be 10 functioning schools (a service), but you are not specifying the change you expect in the students' knowledge (*outcome*), or the *impact* you expect the HIV education to have on their lives.

Outcome Objectives

Examples of *Outcome o*bjectives:

- To increase use of STI testing and treatment services by men aged 15-24 in the project area to 20% in 1 year.
- To increase the proportion of enrolled secondary school students in all schools who pass the HIV knowledge test from 50% to 75% in academic year 2012.
- To increase the proportion of commercial sex workers who report using a condom the last time they had sex from 50% to 90% by the end of the project.

These objectives tell us something about the people's knowledge, attitudes, skills, or behaviour. They go further along the causal pathway than the *output*. So, for example, we not only know that we have functioning STI services, but we know that more men are using them. We still do not know, however, if and how those men's lives have been changed.

Impact Objectives

Impact objectives are less common in projects, but they may be appropriate in certain cases where measurement is feasible.

An **example** of an *impact* objective is to reduce refugee camp mortality to 1 death per 10,000 population per day by the end of the 2012.

Because *impact* is defined as a change in the health, social, or economic status of the "population of interest," this generally refers to the broad population with which we work, or large subsets such as women or students. When you can show a status change among these large groups, your programme has achieved *impact*.

Since *impact* is what we want to achieve in our programmes, you may wonder why we even consider using *output or outcome* objectives instead of always measuring *impact* objectives. There are four good reasons why you might choose to use *output* objectives in your project:

1. You can only measure the changes that will occur in the time the project will be present, funded, or functioning. Depending on the starting point, time frame and likely pace of change, the project might not be in a site long enough to observe a change in *impact*.

2. Some desired *impacts*, such as reduction in HIV transmission or psychosocial health status, are not easily measurable.

3. Measurement requires resources – funds, staff, time, and expertise. *Output* data are usually relatively easy to collect. *Outcome* data collection generally requires more resources, and collecting. *Impact* data may require even more intensive efforts.

4. The causal pathway between outcomes and impact are strong. A strong causal pathway has good scientific evidence or documented programme experience that it works. For example, the causal link between correct contraception use and prevention of unwanted pregnancies is very strong. We know this from decades of scientific research and plenty of programme experience. If you have a good *outcome* measure of contraception use, you can have confidence that unwanted pregnancies will be prevented, even if you do not measure it.

Learning Exercise 1.3.a

SMART Objectives

Directions: Do the Outcome/Effect Objectives below meet the SMART criteria? Why or why not? Place an X in each of the criteria it meets.

OBJECTIVE	S	M	A	R	T
1. To increase use of STI testing and treatment services by men aged 15-25 in the catchment area by 40% in 2 years.					

How would you revise the objective to make it better?

2. To increase use of family planning services.	S	M	A	R	T

How would you revise the objective to make it better

3. To reduce the rate of low birth weight infants born in the local hospital in the year 2007.	S	M	A	R	T

How would you revise the objective to make it better?

4. To increase the use of voluntary counselling and testing services by first time testers of ages 18-35 in the MOH hospital by 15% in the next year, through an education campaign in the catchment area.	S	M	A	R	T

How would you revise the objective to make it better?

Discussion 1.3.a
SMART Objectives

Review the answers below.

OBJECTIVE	S	M	A	R	T
1. To increase use of STI testing and treatment services by men aged 15-25 in the catchment area by 40% in 2 years.	X	X	?	?	X

How would you revise the objective to make it better?

You could include how it is achievable. What would be done to achieve this objective? You could also include to whom and how this is relevant.

	S	M	A	R	T
2. To increase use of family planning services.	X	?	?	?	?

How would you revise the objective to make it better?

In this objective you know what you want to achieve but it is not clear where, how or when. In addition, "increase" is not specified. We do not know what the baseline for the increase would be. These would all have to be added to the objective as well as the means of measurement.

	S	M	A	R	T
2. To reduce the rate of low birth weight infants born in the local hospital in the year 2007.	X	?	?	X	X

How would you revise the objective to make it better?

The objective of a reduction in the rate of low birth weight infants is clear and the timeframe is clear. We can also assume that reduction of low birth weight infants is relevant in any context. However, in order for this objective to be SMART you need to specify what reduction you would like to see (measurable) and how this will be achieved (programme or intervention).

	S	M	A	R	T
3. To increase the use of voluntary counselling and testing services by first time testers of ages 18-35 in the MOH hospital by 15% in the next year, through an education campaign in the catchment area.	X	X	?	?	?

How would you revise the objective to make it better?

This objective is SMART. We know why and what we are measuring, we know the expected change, we know what the intervention will be to achieve the objective, we can assume that everyone should be tested and therefore it is relevant, and we know the timeframe for achieving the objective.

1.4 Data Quality and Data Use

Data used for M&E can be routine clinic data, special programme data, or special study data. We usually use routinely collected health service data from the healthcare facilities and other stakeholders such as faith-based organization and schools. Data for evaluation or more in-depth monitoring can be produced using any of the standard research techniques, which will be discuss in other chapters. These standard research techniques include:

- surveys,
- focus groups,
- in-depth interviews,
- client-provider observation,
- key informant interviews, and
- surveillance.

With any data, it is important to consider data quality. If quality is low, health service statistics may provide unreliable data on health interventions or provide inflated or erroneous results.

Chapter Summary

M&E is an essential component of any intervention, project, or programme. In this chapter, you reviewed definitions of M&E, why it is important, and the basics of what it entails. Understanding the definition of causal pathway elements (inputs, activities, outputs, outcomes, and impacts) and how they interrelate is a necessary step in developing a strong M&E plan. Aligning a programme's causal pathway to an M&E plan consisting of SMART objectives and measurable indicators requires using critical thinking to apply the processes outlined in this chapter and knowledge of the public health programmes being implemented. Allotting appropriate space, time, and resources to M&E should be done in the planning phase of a programme or intervention, as indicators should be aligned with each step of a programmes causal pathway. A well-planned and implemented M&E plan will help you to understand if programmes are reaching the intended objectives.

Understanding M&E is an important step to succeeding as a professional

Self Assessment Quiz

1. Indicators should be vague so that people can interpret them to meet their own needs, making them more useful.
   a) True
   b) False

2. Which of the following is an example of a routine data source?
   a) Clinic service statistics
   b) Household surveys
   c) National censuses
   d) All of the above

3. Population-based surveys, such as the *Botswana AIDS Impact Study (BAIS)* or a national population census, are examples of non-routine data sources. These sources are useful because:
   a) They provide information on a timely basis.
   b) They can be used for calculating coverage indicators.
   c) They include health outcomes for only those using health facilities.
   d) All of the above

4. A data collection plan should include the following:
   a) The timing and frequency of collection.
   b) The person/agency responsible for the collection.
   c) The types of information needed for the indicators.
   d) All of the above

5. When should the M&E plan be created?
   a) During the design phase of a programme.
   b) At the midpoint of the programme.
   c) At the end of the programme.
   d) After all of the data have been collected but before they are analyzed.

6. Evaluations measure:
   a) The timeliness of a programme's activities.
   b) The outcomes and impact of a programme's activities.
   c) How closely a programme kept to its budget.
   d) How well the programme was implemented.

7. At what stage of a programme should monitoring take place?
   a) At the beginning of the programme.
   b) At the midpoint of the programme.
   c) At the end of the programme.
   d) Throughout the life of the programme.

8. Once an M&E plan is made, it is set in stone and cannot be changed.
   a) True
   b) False

9. Which of the following is NOT considered "monitoring"?
   a) Counting the number of people trained.
   b) Attributing changes in health outcomes to an intervention.
   c) Tracking the number of brochures disseminated.
   d) Collecting monthly data on clients served in a clinic.

10. When it is said that objectives should be SMART, the "S" stands for:
    a) Simple
    b) Strategic
    c) Silly
    d) Specific

11. Which of the following is a characteristic of a good indicator?
    a) Is clearly defined in unambiguous terms.
    b) Produces the same results when used repeatedly to measure the same condition or event.
    c) Measures only the condition or event it is intended to measure.
    d) All of the above are characteristics of good indicators.

12. Data quality can be negatively affected by:
    a) Sampling bias (the sample taken is not a representative sample)
    b) Non-response rates
    c) Subjective measurement (data influenced by measurer)
    d) All of the above
    e) a and b only

References & Bibliography

Johns Hopkins University (JHU) no date. *OpenCourseWare. Fundamentals of Programme Evaluation.* Available at: <http://ocw.jhsph.edu/courses/fundamentalsprogrammeevaluation/lectureNotes.cfm> [Accessed 16 November 2010]

Marsh DR et al. 2008. A Results Framework Serves Both Programme Design and Delivery Science, *American Society for Nutrition, 138:630-633*

MEASURE EVALUATION, no date. *Introduction to Monitoring and Evaluation Course* Available at:

<https://www.cpc.unc.edu/measure/training/mentor/materials/hiv-english/intro-to-m-e-and-course-overview> [Accessed 29 November 2010]

Reproductive Health Response in Conflict (RHRC), no date. *RHRC Consortium Monitoring and Evaluation ToolKit: The Causal Pathway.* Available at: <http://www.rhrc.org/resources/general_fieldtools/toolkit/index.htm> [Accessed 29 November 2010]

Roberta, P 2005. Economic Considerations of Health Literacy: Model and Health Production Function. Jannetti Publications, Inc; *Nurs Econ. 23(4):173-180*

Available at: <http://www.medscape.com/viewarticle/511636_4>

Self Assessment Quiz Answer Key

1. Indicators should be vague so that people can interpret them to meet their own needs, making them more useful.

   a) True

   b) False

   Answer explanation: *Indicators should be very specific so that anyone looking at the indicator will know exactly what, when and how information will be collected. For example, instead of "women enrolled in PMTCT" your indicator could be "% pregnant women, diagnosed HIV+ during 2011 and 2012 at the Lolongwe VCT centre, recorded as enrolled in that centres PMTCT programme".*

2. Which of the following is an example of a routine data source?

   a) Clinic service statistics

   b) Household surveys

   c) National censuses

   d) All of the above

   Answer explanation: *Clinic service statistics are collected on a routine (usually monthly) basis and therefore are a routine data source. Household surveys and national censuses or not routine and may take place only once (as in the case of a specific household survey) or only every three or four years (as in the case of a national census).*

3. Population-based surveys, such as a Demographic and Health Survey (DHS) or a national population census, are examples of non-routine data sources. These sources are useful because:

   a) They provide information on a timely basis.

   b) They can be used for calculating coverage indicators.

   c) They include health outcomes for only those using health facilities.

   d) All of the above

   Answer explanation: *While these surveys and censuses usually cannot provide timely information and often cannot detail specific health outcomes for specific health facilities, they do provide information about the population as a whole that can be used to determine the coverage needs of a programme, or the baseline population data to compare against your specific facility. For example, if you know that the prevalence of HIV+ patients in your clinic is 8% but it is 12% in the population in the clinics catchment area, than you can assume that your clinic is not reaching a significant percentage of the population.*

4. A data collection plan should include the following:
   a) The timing and frequency of collection
   b) The person/agency responsible for the collection
   c) The types of information needed for the indicators
   d) ***All of the above***

   Answer explanation: *The data collection plan should include all of this information. It will be a guide to data collection and therefore everyone should know their specific roles, what type of information needs to be gathered, and when and where that data gathering will take place. This plan should be specific enough so even someone unfamiliar with the project could ensure that data collection occurred.*

5. When should the M&E plan be created?
   a) ***During the design phase of a programme***
   b) At the midpoint of the programme
   c) At the end of the programme
   d) After all of the data have been collected but before they are analyzed

   Answer explanation: *An M&E plan is an integral part of programme design. Often, what data you need and when you need it will influence how programme implementation occurs. The M&E plan should be developed at the same time and in conjunction with the design of the implementation of a programme.*

6. Evaluations measure:
   a) The timeliness of a programme's activities
   b) The outcomes and impact of a programme's activities
   c) How closely a programme kept to its budget
   d) ***How well the programme was implemented***

   Answer explanation: *While outcome evaluations measure the outcomes/effect and impact of a programmes activities, evaluations in routine M&E measure how the programme was implemented and whether the objectives of implementation were met.*

7. At what stage of a programme should monitoring take place?
   a) At the beginning of the programme
   b) At the mid-point of the programme
   c) At the end of the programme
   d) ***Throughout the life of the programme***

7 - Answer explanation: *Monitoring is the on-going measurement of the programme. Therefore, monitoring should take place throughout the programme and information from the monitoring process should be used to ensure that the programme is being implemented as intended.*

8. Once an M&E plan is made, it is set in stone and cannot be changed.

   a) True

   b) ***False***

   Answer explanation: *An M&E plan is a guide to the monitoring and evaluation of programme implementation. However, as programme implementation priorities or programme design changes, the M&E plan should also change. Routine monitoring could inform programme changes and therefore modify the M&E plan.*

9. Which of the following is NOT considered "monitoring"?

   a) Counting the number of people trained

   b) ***Attributing changes in health outcomes to an intervention***

   c) Tracking the number of brochures disseminated

   d) Collecting monthly data on clients served in a clinic

   Answer explanation: *"Monitoring" is an on-going process looking at short-term objectives in a programme. All of the other answers to this question are examples of monitoring. However, attributing longer-term change in health outcomes to an intervention goes beyond the routine monitoring of a programme. This involves long-term evaluations and often additional time and resources. This would be considered an outcome evaluation.*

10. When it is said that objectives should be SMART, the "S" in the acronym stands for:

    a) Simple

    b) Strategic

    c) Silly

    d) ***Specific***

    Answer explanation: *S in SMART stands for specific and is a reminder that you need to be very clear about the change you would like to see when developing an objective.*

11. Which of the following is a characteristic of a good indicator?

a) Is clearly defined in unambiguous terms

b) Produces the same results when used repeatedly to measure the same condition or event

c) Measures only the condition or event it is intended to measure

d) All of the above are characteristics of good indicators.

Answer explanation: *Indicators should be specific. In addition, data from this indicator should be able to be collected multiple times with the same results. This means that the indicator is only measuring what it is intended to measure and that it is unambiguous; even someone unfamiliar with the programme could collect this data and it would produce the same results.*

12. Data quality can be negatively affected by:

a) Sampling bias (the sample taken is not a representative sample)

b) Non-response rates

c) Subjective measurement (data influenced by measurer)

d) All of the above

e) a and b only

Answer explanation: *In all of these cases the data can be negatively affected. Sampling bias can lead to bad results because only one portion of the population is represented and collecting data on another portion of the population could lead to completely different results. High non-response rates could mean that only some types of people responded (thus leading to more bias in the sample) and it could also mean that not enough responses were collected to be able to see a change. Subjective measurements could lead to poor data quality because, in the analysis of the data, it would be impossible to determine if answers were based on the respondents or were influence by the measurer. For instance, if you were measuring clinical performance of nurses, some measurers could be hard graders and others could be easy graders. Therefore, any one nurses' score on the measurement would be more a reflection of who the measurer was than what the nurse could do.*

Self-Directed Learning Workbook2:
Chapter Two
Data Collection and Data Management

Collecting data through surveys can supplement data received from data facilities

Chapter 2: Data Collection and Data Management

Estimated time needed for completion: 2 hours

Chapter Overview

In Chapter 1 of this workbook you learned more about what monitoring and evaluation (M&E) means. You were presented with various tools that can be used in M&E. You also learned that M&E is, for the most part, conducted using data. Let us now take a closer look at what is involved in collecting data, particularly in relation to data that you are managing in your district. We will begin by clearly defining the terminology used in data collection. We will then demonstrate how data is collected, using a sample data collection tool from the HIV Testing and Counselling (HTC) national programme.

Learning Objectives

At the end of this chapter, you will be able to:

- explain the role of an M&E officer in data management,
- describe methods used for collecting data for a national programme,
- discuss challenges that can occur in data collection for programmes,
- define data management, and
- list tips for good file management.

2.1 Data Collection for National Programmes

Your key responsibility as an M&E officer will be to help manage the health data that is generated at the different service delivery sites in your district. The primary source of the data you collect will be from clients seen at the health facilities (hospitals, clinics, health posts) in your district. Data will also be collected on the health related activities that are carried out in the communities in your district. Different types of data are collected, depending on the different national programmes.

2.1.1 Data vs. Information
(This section is adapted from Quynth 2005)

The two words *information* and *data* are frequently used interchangeably in everyday conversation. However, in the context of M&E *information* and *data* have distinct separate meanings.

Data is defined as individual facts, statistics, or measurements. Data by itself has little utility. Data that has been analysed and synthesised provides us with information. Information is defined as knowledge concerning a particular fact or circumstance that is communicated or received. Once data has been analysed and has become information, it has meaning and utility.

Data is often obtained as a result of recordings or observations. Health data are facts that when analysed, provide information. This information is then used by health care planners and providers in order to maintain effective and efficient public health services (Last n.d.).

For example: the CD4 cell count of a person living with HIV is periodically measured to assess health status. The number of CD4 cells counted is *data*. The doctor or nurse records the CD4 count in the patient card as soon as they receive it. The trends in the CD4 count for this patient are analysed and the doctor or nurse uses this *information* to make conclusions about the need for Antiretroviral Therapy (ART) or the effect of ART. The CD4 count pattern across many different clients may be analysed against changes in health and used by health care planners to determine when most clients should begin ART.

2.1.2 What is data collection?

Data collection means gathering information for a particular purpose. There can be many reasons for collecting data. For example, data may be collected to monitor a particular programme, to conduct an evaluation, or to answer a research question. The type of data you collect and the method you employ to collect it will depend on your reason for collecting the data. It is important to think about the reason you want to collect data and the questions you want answered when planning data collection. You also need to consider what the potential data sources are: whether the data you want is available, from where, and from whom you can get the data (PIH 2006). One other thing you should consider is data use - how the data collected will be organised, analysed, interpreted, and disseminated (PIH 2006). In the chapter that follows (chapter 3), you will learn about data quality and data auditing. Chapter 4 will discuss how you can analyse the data you collect to produce information that can be used for health care policy, planning, and programmes.

As a district level M&E officer you will not be directly involved with collecting data from clients at the health facility level. However, understanding this process and the tools used to collect data by different programmes will help you when you perform data analysis as well as when you conduct data quality audits.

Data Collection at the Health Facility Level

In the health facilities in your district, you will find health records of individuals who have received services there. A *health record* is a collection of personal health information over time, usually based on the individual and is organised primarily to support efficient and high quality health care (Quynth 2005).

Examples of data collected as part of a health record include (Last n.d.):

- personal identifying data: name, date of birth;
- socio-demographic data: sex, age, occupation, place of residence, and education level;.
- clinical data: medical history, clinical investigations, diagnoses, and treatment regimens;
- administrative data: referrals, sites of care;
- behavioural data: adherence to the recommended regimen, condom use, and number of partners.

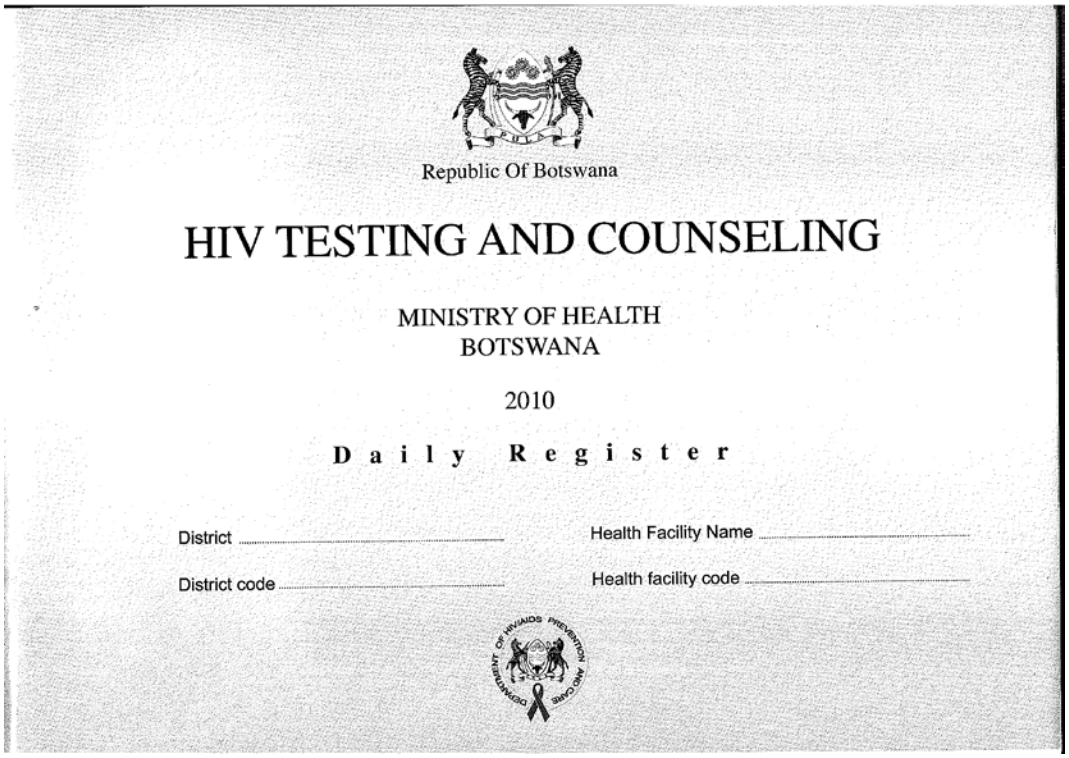
Health records are mainly in the form of paper files or registers, although we are slowly moving towards having electronic data systems in place. You will not find all data for one individual in the same register. There may be several different registers or data collection tools used to collect data for a single individual, as they receive multiple health services. There may also be multiple registers for a single national programme. For example, the Prevention of Mother-to-Child Treatment (PMTCT) programme uses a counselling register, an Antinatal Care (ANC) Register, and a

maternity register. Usually health care workers at health facilities collect information directly from the clients and document it in registers. Sometimes you will not find all data for one individual in the same register. For example, PMTCT uses a counselling register, an ANC register, and a maternity register.

You will be shown an example of the patient data collection process at the health facility level, using the HTC programme data collection tool.

Figure 2.1.2 Patient Data Collection Process Example

1. The first step in collecting HTC data is to ensure you have selected the appropriate register to document information you collect from the client. The HTC register, as with other registers, has identifying information on the outside cover page, as shown below:


Republic Of Botswana

HIV TESTING AND COUNSELING

MINISTRY OF HEALTH
BOTSWANA

2010

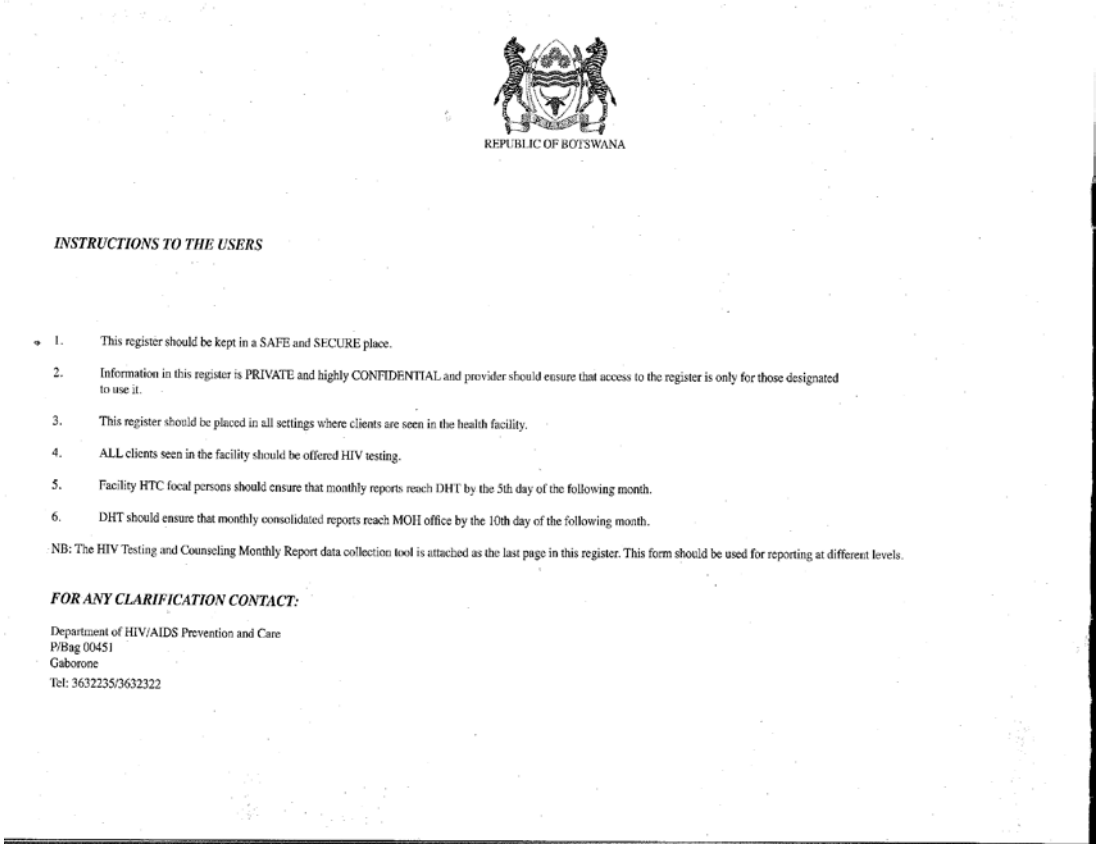
Daily Register

District ..... Health Facility Name .....

District code ..... Health facility code .....

- Check the name of the register to ensure it says: HIV Testing and Counselling.
- Confirm you are using a standardised national data collection tool by checking that the register has the Government of Botswana Ministry of Health emblem.
- Ensure you are using the most recent version of the register: 2010 revision.
- Ensure you are using the register for the correct facility by checking the district name and code, as well as the health facility name and code.

2. On the inside cover of the register, you will find instructions guiding you on how to utilize the register (see figure below). Included in this information are instructions for those compiling the monthly reports using this register, to submit the reports to District Health Management Teams (DHMT) by the 5th of the month following the reporting month. Your role as an M&E officer is to receive monthly reports from your health facilities; this should be no later than the 5th of the next month (see sample of monthly report form later in this section). Also included in the instruction page is that the DHMT, (here referring to you, as a district M&E officer) should ensure that the consolidated reports reach the MOH office by the 10th of the month following the reporting month. At the bottom of the inside cover page you will find details of who to contact for clarifications on the HTC register should you require assistance.



REPUBLIC OF BOTSWANA

INSTRUCTIONS TO THE USERS

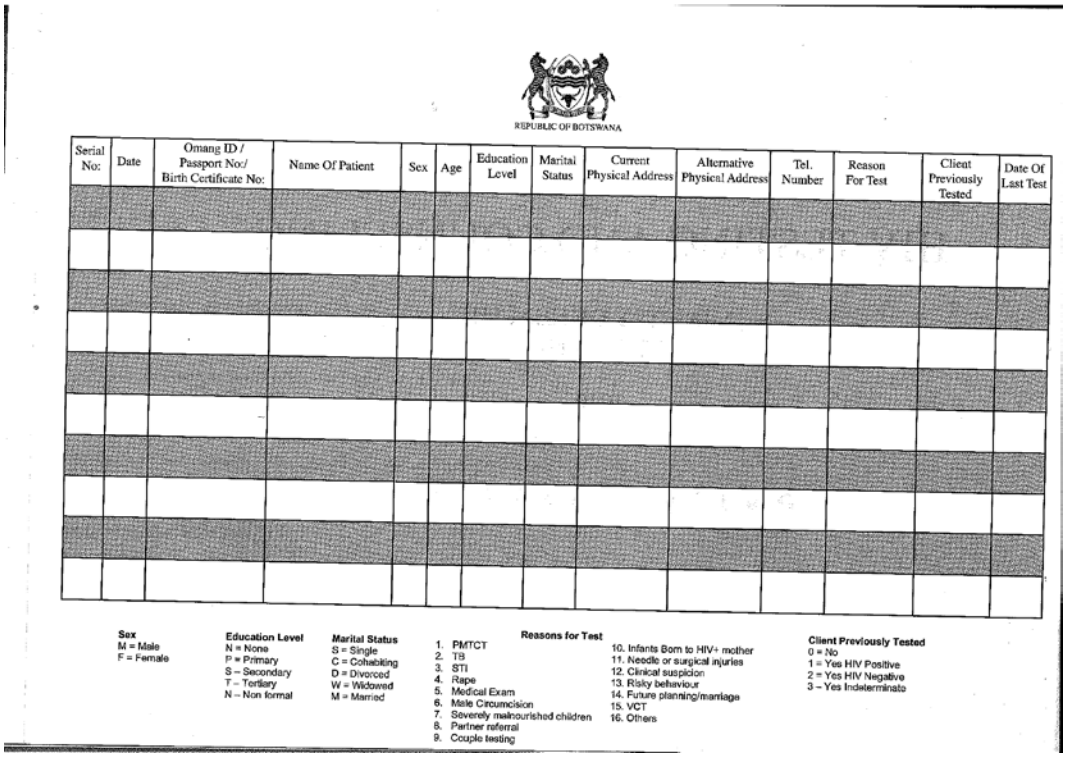
1. This register should be kept in a SAFE and SECURE place.
2. Information in this register is PRIVATE and highly CONFIDENTIAL and provider should ensure that access to the register is only for those designated to use it.
3. This register should be placed in all settings where clients are seen in the health facility.
4. ALL clients seen in the facility should be offered HIV testing.
5. Facility HTC focal persons should ensure that monthly reports reach DHT by the 5th day of the following month.
6. DHT should ensure that monthly consolidated reports reach MOH office by the 10th day of the following month.

NB: The HIV Testing and Counseling Monthly Report data collection tool is attached as the last page in this register. This form should be used for reporting at different levels.

FOR ANY CLARIFICATION CONTACT:

Department of HIV/AIDS Prevention and Care
P/Bag 00451
Gaborone
Tel: 3632235/3632322

3. Write the information about the health client in the register page shown below. When collecting the information from the client, ask them to provide documentation as evidence. Fill out each of the fields, moving across to the adjacent page to complete the row for each client. It is important to follow the key when filling out this information to ensure that it is filled out well and consistently.



REPUBLIC OF BOTSWANA

Serial No:	Date	Omang ID / Passport No:/ Birth Certificate No:	Name Of Patient	Sex	Age	Education Level	Marital Status	Current Physical Address	Alternative Physical Address	Tel. Number	Reason For Test	Client Previously Tested	Date Of Last Test

Sex
M = Male
F = Female

Education Level
N = None
P = Primary
S – Secondary
T – Tertiary
N – Non formal

Marital Status
S = Single
C = Cohabiting
D = Divorced
W = Widowed
M = Married

Reasons for Test
1. PMTCT
2. TB
3. STI
4. Rape
5. Medical Exam
6. Male Circumcision
7. Severely malnourished children
8. Partner referral
9. Couple testing
10. Infants Born to HIV+ mother
11. Needle or surgical injuries
12. Clinical suspicion
13. Risky behaviour
14. Future planning/marriage
15. VCT
16. Others

Client Previously Tested
0 = No
1 = Yes HIV Positive
2 = Yes HIV Negative
3 – Yes Indeterminate

- *Serial No.*: write the number depending on what number client you are seeing for the month. E.g. if it is the first client of the month, give the serial number 1, and the next client will be entered as number 2, and so on.

- *Date:* write the date you are seeing the client, indicating the day, month and year using this format: *dd/mm/yy*.

- *Omang ID / Passport No. / Birth Certificate No.:* Write the national identification (Omang) number for citizens and passport number for non-citizens.

- *Name of Patient:* Write the full name and surname of the client.

- *Sex:* Follow the key at the bottom of the register page and write 'M' for male or 'F' for female.

- *Age:* Verify the date of birth using the ID and write the age of the client on that day you are seeing them. i.e., If the client is turning 25 on the day following the one you saw them, write '24' under the 'age' field, and not '25'.

- *Education Level:* Follow the key at the bottom of the register page and write 'N' for none, 'P' for primary education, and 'S', 'T', and 'N' for secondary, tertiary, or non formal education respectively.

- *Marital Status:* Follow the key at the bottom of the register page and write 'S', 'C', 'D', 'W', and 'M' for single, cohabiting, divorced, widowed, or married respectively.

- *Current Physical Address:* Write the house number, street name, and ward, where the client is currently staying and can be contacted if needed.

- *Alternative Physical Address:* Write the house number, street name, ward, and village/town where the client can be contacted in case they are not available at the address documented in the 'current physical address' field.

- *Tel. Number:* Write the landline and/or mobile telephone number the client can be reached if the need arises.

- *Reason for Test:* Follow the key at the bottom of the register page and write the number(s) corresponding to the reason given for testing given by the client.

- *Client Previously Tested:* Follow the key at the bottom of the register page and write the number '0' if the client has not had a previous HIV test done. If they have had a previous test, write '1' if they had tested HIV positive, '2' if HIV negative, and '3' if had an intermediate HIV test result. N.B. Ask client to show you the written HIV test result, or documentation indicating ARV clinic attendance as proof.

- *Date of Last Test:* If you entered numbers 1, 2, or 3 in the previous field (client previously tested), then write the date the test was done, in the format: dd/mm/yy).

- *Pre test education counselling:* Follow the key at the bottom of the register page and write down 'GE', 'IE', 'IC', 'CC', or 'O' to indicate whether the counselling type the client received was group education, individual education, individual counselling, couple counselling, or other, respectively.

- *Test 1: Batch No. and Test 2: Batch No.:* Follow the key at the bottom of the register page and write down 'N', 'P', or 'I', to indicate whether HIV tests 1 and 2 results were negative, positive, or invalid, respectively.

- *Final results:* Follow the key at the bottom of the register page and write down 'N', 'P', or 'U', to indicate whether the final HIV test results were negative, positive, or indeterminate, respectively.

- *Post test counselling:* Follow the key at the bottom of the register page and write down 'I/PC', 'C/PC', 'or 'F/PC' to indicate whether the client received individual, couple, or family post test counselling respectively.

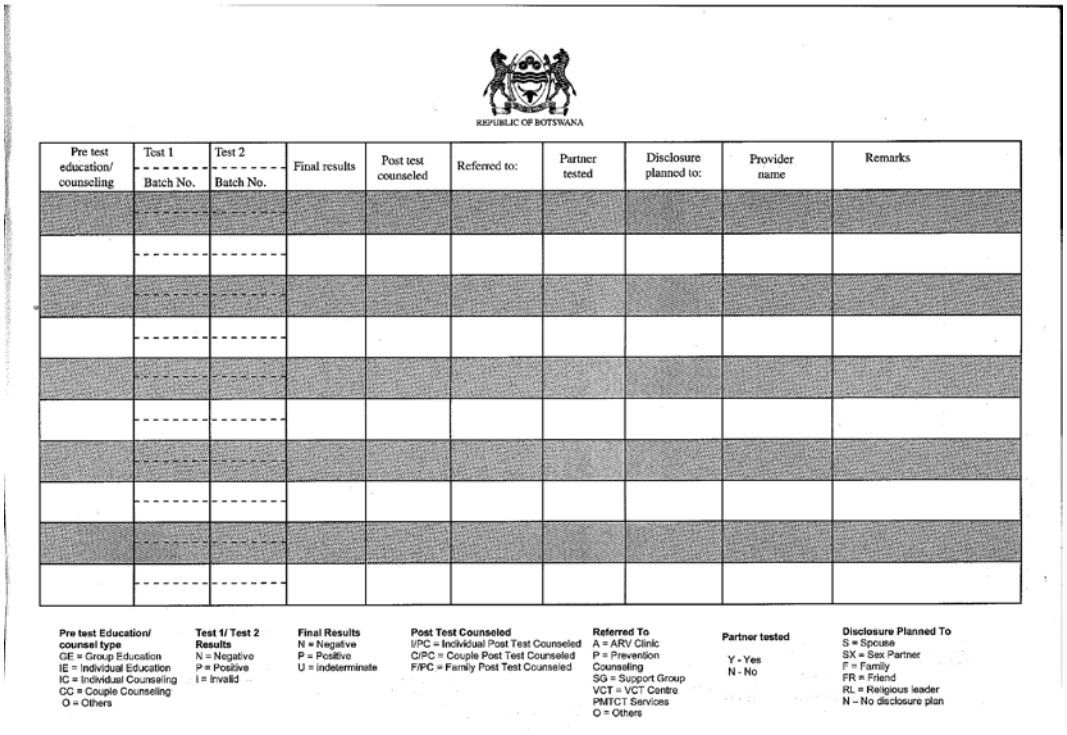
- *Referred to:* Follow the key at the bottom of the register page and write down 'A', 'P', 'SG', 'VCT', 'PMTCT services', or 'O' to indicate whether the client was referred to the ARV clinic, for prevention counselling, to a support group, to a VCT centre, for PMTCT services, or other place/intervention respectively, after being tested for HIV.

- *Partner Tested:* Follow the key at the bottom of the register page and write down 'Y' for 'Yes' if the client's partner has been tested for HIV or 'N' for 'No' to indicated they have not been tested.

- *Disclosure planned to:* Follow the key at the bottom of the register page and write down who the client is planning to disclose their HIV status to: 'S', 'SX', 'F', 'FR', 'RL', or 'N', to indicate spouse, sex partner, family, friend, religious leader, or if there is no disclosure plan respectively.

- *Provider name:* Write your name, as the person who provided the counselling and testing service to the client.

- *Remarks:* Write comments to explain any discrepancies in the entries made, e.g., 'client asked to bring previous HIV test results' if they had reported previous HIV testing without providing the evidence.



REPUBLIC OF BOTSWANA

Pre test education/ counseling	Test 1 Batch No.	Test 2 Batch No.	Final results	Post test counseled	Referred to:	Partner tested	Disclosure planned to:	Provider name	Remarks

Pre test Education/ counsel type
GE = Group Education
IE = Individual Education
IC = Individual Counseling
CC = Couple Counseling
O = Others

Test 1/ Test 2 Results
N = Negative
P = Positive
I = Invalid

Final Results
N = Negative
P = Positive
U = Indeterminate

Post Test Counseled
I/PC = Individual Post Test Counseled
C/PC = Couple Post Test Counseled
F/PC = Family Post Test Counseled

Referred To
A = ARV Clinic
P = Prevention Counseling
SG = Support Group
VCT = VCT Centre
PMTCT Services
O = Others

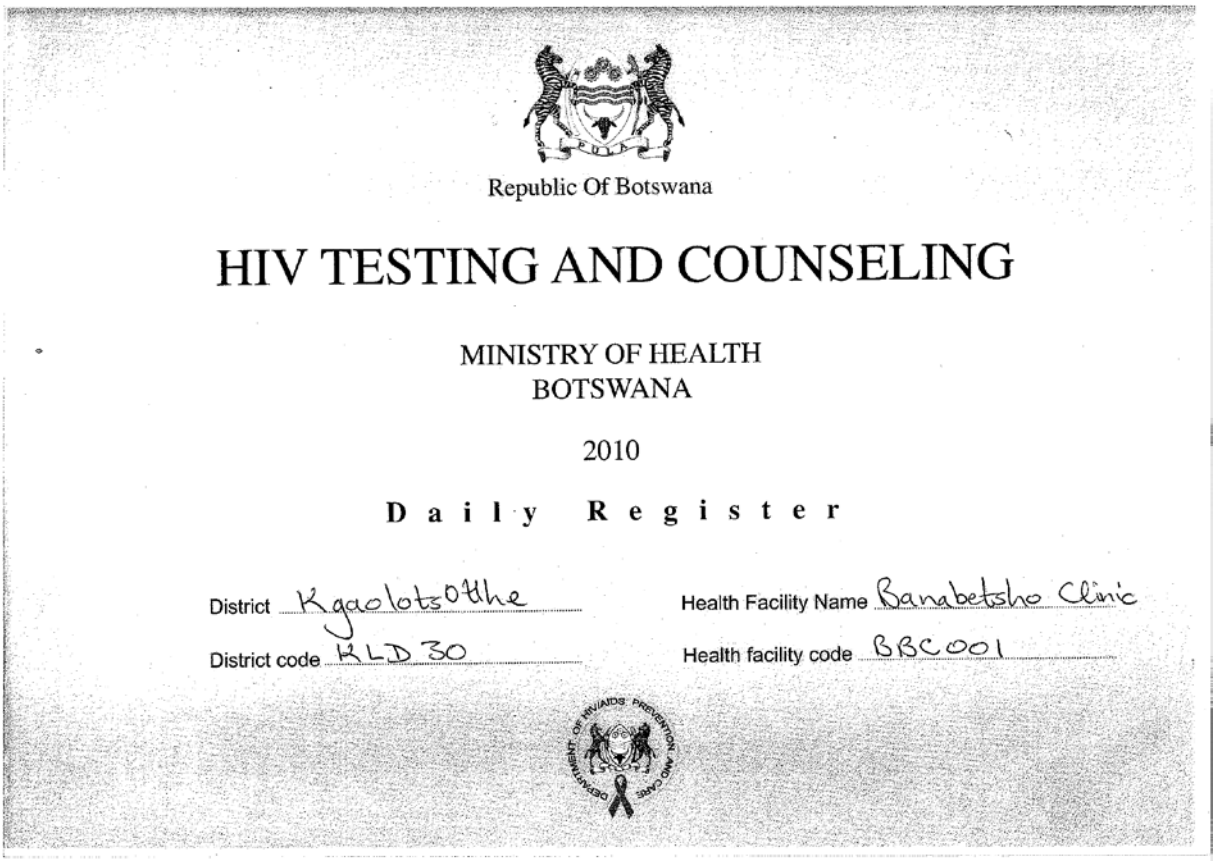
Partner tested
Y - Yes
N - No

Disclosure Planned To
S = Spouse
SX = Sex Partner
F = Family
FR = Friend
RL = Religious leader
N – No disclosure plan

Learning Activity 2.1.2

Completing the HTC Monthly Report Form

Directions: Use data from the filled out HTC register shown below for the month of October 2010, to complete the HIV Testing & Counselling Monthly Report form (found at the end of this activity box) for the fictitious health facility called Banabetsho Clinic (Facility code: BBC001) in the fictitious District called Kgaolotsotlhe (district code: KLD30). You will find a copy of this form on the inside back cover of the HTC register.



Republic Of Botswana

HIV TESTING AND COUNSELING

MINISTRY OF HEALTH
BOTSWANA

2010

Daily Register

District Kgaolotsotlhe

Health Facility Name Banabetsho Clinic

District code KLD 30

Health facility code BBC001

REPUBLIC OF BOTSWANA

Serial No:	Date	Omang ID / Passport No:/ Birth Certificate No:	Name Of Patient	Sex	Age	Education Level	Marital Status	Current Physical Address	Alternative Physical Address	Tel. Number	Reason For Test	Client Previously Tested	Date of Last Test
01/10	5/10/10	505939707	Mpho Etsile	F	18	P	S	Hse No 7 Selwadi Rd Ext 4 Kgaolo	Hse No 32 Nata Rd Serowe	72000091	3. STI	0	N/A
02/10	5/10/10	793525000	Batho Bateng	M	25	T	C	Plot 20 Naledi Str Kgaolo	Hse No 66 Phala Rd Good Hope	323909	6. SMC	2	01/09/10
03/10	18/10/10	400669001	Bakae Kukoo	M	45	T	C	Hse No 3236 Marang Rd Ext 2 Kgaolo	Hse No 3236 Marang Rd Ext 2 Kgaolo	75323999	2. TB	2	20/10/10
04/10	20/10/10	Birth certificate 00132994	Modisa Otsile	M	7	N	N/A	Hse No 133 Sebele Kgaolo	Plot 8 Samaneh Rd Gaborone	3932601	10 & 12	0	N/A
05/10	29/10/10	392353707	Atla Modise	F	33	T	M	Hse No 5 Kota block Kgaolo	Hse No 13 Marang Rd Gaborone	75223911	1. PMTCT	2	01/04/10

Sex
M = Male
F = Female

Education Level
N = None
P = Primary
S – Secondary
T – Tertiary
N – Non formal

Marital Status
S = Single
C = Cohabiting
D = Divorced
W = Widowed
M = Married

Reasons for Test
1. PMTCT
2. TB
3. STI
4. Rape
5. Medical Exam
6. Male Circumcision
7. Severely malnourished children
8. Partner referral
10. Infants Born to HIV+ mother
11. Needle or surgical injuries
12. Clinical suspicion
13. Risky behaviour
14. Future planning/marriage
15. VCT
16. Others

Client Previously Tested
0 = No
1 = Yes HIV Positive
2 = Yes HIV Negative
3 – Yes Indeterminate

REPUBLIC OF BOTSWANA

Pre test education/ counseling	Test 1 Batch No.	Test 2 Batch No.	Final results	Post test counseled	Referred to:	Partner tested	Disclosure planned to:	Provider name	Remarks
IE	P T3500	P K2201	P	I/PC	A	N	SX	Nmaniso	Encouraged to talk to partner about testing
IE	N T3500	N K2201	N	I/PC	P	Y	SX	Nmaniso	
GE	P T3500	P K2201	P	I/PC	A	N	SX	Nmaniso	Encouraged to talk to partner to test
CC	N T3500	N K2201	N	C/PC	N/A	N/A	N/A	Setengm	
CC	N T3500	N K2201	N	C/PC	P	Y	S	Setengm	Encouraged to test from 36 weeks gestation

Pre test Education/ counsel type
GE = Group Education
IE = Individual Education
IC = Individual Counseling
CC = Couple Counseling
O = Others

Test 1/ Test 2 Results
N = Negative
P = Positive
I = Invalid

Final Results
N = Negative
P = Positive
U = Indeterminate

Post Test Counseled
I/PC = Individual Post Test Counseled
C/PC = Couple Post Test Counseled
F/PC = Family Post Test Counseled

Referred To
A = ARV Clinic
P = Prevention Counseling
SG = Support Group
VCT = VCT Centre
PMTCT Services
O = Others

Partner tested
Y - Yes
N - No

Disclosure Planned To
S = Spouse
SX = Sex Partner
F = Family
FR = Friend
RL = Religious leader
N – No disclosure plan

N.B. the two register pages with patient entries lie side by side in the actual register. i.e. each patient entry cuts across 2 pages

HIV Testing & Counseling Monthly Report

District: District Code: Month: Year:

Facility: Facility Code: Date of Reporting

Reporting Officer:

Reason for testing

PMTCT	TB	STI	Rape	Med Exam	Male circumcision	Severely malnourished children// suboptimal growth	Partner referral	**Couple Testing**	Infant born to HIV+ mother	Needle / Surgical Injury	Clinical Suspicion	Risky behaviour	Future planning / marriage	VCT	Other

	Sex	< 1	1 - 4	5 - 9	10 - 14	15 - 19	20 - 24	25 - 29	30 - 34	35 - 39	40 - 44	45 -49	50 - 59	> 60	Total
Offered	M														
	F														
	Tot														
Done	M														
	F														
	Tot														
HIV +	M														
	F														
	Tot														

Elisa	Rapid Test	PCR	First time test	Post test counseled	Patients seen in Facility	No. of staff trained in rapid testing	No. of Facilities providing HTC	No. of Facilities reporting HTC

Discussion 2.1.2

Completed HTC Monthly Report Form

HIV Testing & Counseling Monthly Report

District: KGAOLOTSOTHE District Code: KLD30 Month: OCTOBER Year: 2010

Facility: BANABETSHO CLINIC Facility Code: BBC001 Date of Reporting 01 NOVEMBER 2010

Reporting Officer: KABELO TSAME

Reason for testing

PMTCT	TB	STI	Rape	Med Exam	Male circumcision	Severely malnourished children// suboptimal growth	Partner referral	Couple Testing	Infant born to HIV+ mother	Needle / Surgical Injury	Clinical Suspicion	Risky behaviour	Future planning / marriage	VCT	Other
1	1	1			1				1		1				

	Sex	< 1	1 - 4	5 - 9	10 - 14	15 - 19	20 - 24	25 - 29	30 - 34	35 - 39	40 - 44	45 -49	50 - 59	> 60	Total
Offered	M			1				1				1			3
	F					1			1						2
	Tot			1		1		1	1			1			5
Done	M			1				1				1			3
	F					1			1						2
	Tot			1		1		1	1			1			5
HIV +	M											1			1
	F					1									1
	Tot					1						1			2

Elisa	Rapid Test	PCR	First time test	Post test counseled	Patients seen in Facility	No. of staff trained in rapid testing	No. of Facilities providing HTC	No. of Facilities reporting HTC
			2	5	5			

2.1.3 Challenges in Data Collection

Below is a list of some of the challenges with data collection that you may come across in your district.

- Limited time available to collect all the data required due to staff shortage and work overload. This may mean that all the data required is not collected in a timely manner.

- Lack of understanding of what is being asked in the data collection tools leading to conflicting or incorrect information. For example, if the data collection tool is asking for the birth date to be written as dd/mm/yy and the data collector writes it as mm/dd/yy, this would result in incorrect information.

- Insufficient skills to obtain data (particularly sensitive data) from clients. For example, the data collector may not be able to explain what they mean by anal sex and, so may incorrectly mark the health client has engaging in anal sex.

- Insufficient support for data collection, e.g., lack of adequate data collection tools or registers in which to enter the data.

- Lack of standardised and harmonised systems for data collection and reporting (Data Use Network 2010) so that organisations are reporting on different items within the programme. For example, one programme collects information about counselling and testing. However counting only the number of clients counselled in one clinic and separately collecting information on those who were tested, could leave a gap in the data collected. Others may be collecting at different times so that some reports cover month to month, while others cover mid-month to mid-month, giving an inaccurate picture for any time period.

- Insufficient understanding of how data can improve programmes or inability to see clear links between the items in the data collection tools and their value for decision making (Data Use Network 2010). For example, if the data collector does not realise the importance of knowing what current medications the health client is on, then they will not note it. This may result in the doctor giving a prescription for a medicine that should not be taken with current medication.

- Poor attitudes or low motivation of health care workers, often due to feeling overworked.

- Lack of feedback on the quality of data collected resulting in continued errors or lack of incentive to continue with good practice.

- Many paper registers to complete because clients receive interventions at different points at health facilities, therefore, a different register is used at each point. E.g. when collecting PMTCT data, counselling and testing data will be collected by the lay counsellor using the PMTCT Counsellor Register while the midwife will use a different register when providing ANC interventions. This may mean duplication of data, such as age and number of children, which are necessary for identification. It also means that the midwife does not have access to the counselling and testing data for the woman being served by her.

- Other challenges include having incomplete data and communication problems.

2.2. Data Management

Now that you have learned about data collection at the health facility level, you need to know about the management of this data. In the following section, data management will be discussed, with particular emphasis on file management.

2.2.1 What is Data Management?

Data management is the process of managing data as a resource that is valuable to a programme. This is done with development, execution, and supervision of plans, policies, programmes, and practices that control, protect, deliver, and enhance the value of data and information assets (DATA INTERNATIONAL 2007). For you, data management means keeping the records of programmes and facility level data, safe, secure, and accessible. Data management is important for easy retrieval of data. It increases efficiency because it:

- prevents time being wasted in looking for records,
- prevents time or effort wasted in using a non-current version of a record, and
- prevents data from being lost or deleted.

To have an efficiently working data system, you do not need to destroy your existing system and start a new data management system from scratch. You can review your existing system and build on your strengths and re-strategize to streamline the weak points.

You can start with making an inventory of the forms, log books, and other tools used to record and summarize data at different levels. You can then assess the quality of the data being collected using the existing forms at different levels. Among the aspects to be included in the assessment are:

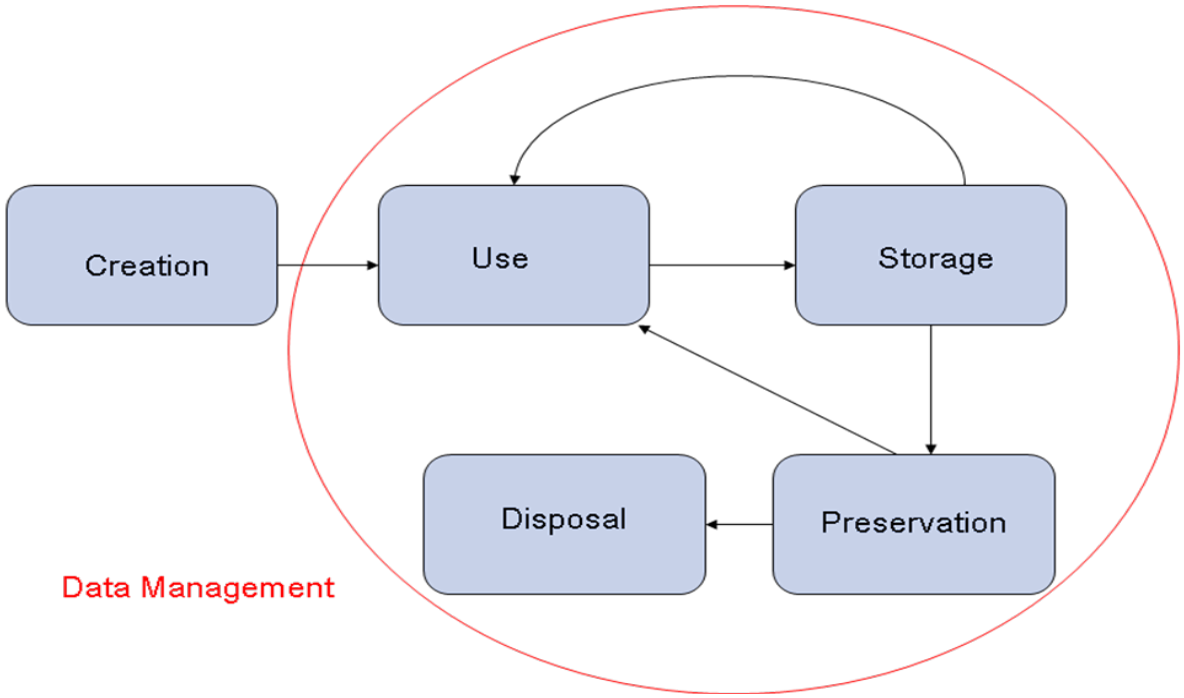
- accuracy,
- completeness,
- adequacy, and
- timeliness.

When reviewing your data systems, look at how the current system is working in respect to data processing, data analysis, data dissemination, supply and logistics, coordination, cooperation, and communication within and between different units in the facility, region, or with MOH. You can then ask the questions listed below:

- Are there aspects of this system that are beneficial to retain?
- What needs to be modified?
- What can be abolished?

2.2.2 The Data Life Cycle

Figure 2.2.2 The Data Life Cycle



Data management is concentrated in the four stages of the data life cycle: data use, storage, preservation, and disposal. Once data is created it needs to be organised for use. Even after the data has been used to its full capacity, it needs to be archived in an accessible location for possible review. Over time, which can be years to decades, most data loses its value and eventually it needs to be preserved for rare uses or disposed of in order to make way for incoming data. Depending on the facility or the organization, there can be rules for certain data to be archived for a few years or permanently. For example, morbidity data is usually kept for ten years. You should check with the facility or the ministry you are working with to find the rules of data retention.

How you *use* the data will inform you how you manage (store, preserve, and dispose) the data. There are a few questions you can ask yourself to help guide your data management practices. Ask yourself these questions:

• Location	Where will documents be stored? Where will people need to go to access documents? How safe will your data be?
• Filing	How will documents be filed? What methods will be used to organise or index the documents to assist in later retrieval?
• Retrieval	How will documents be found?
• Distribution	How can documents be available to the people that need them?

Source: Adapted from *Document Management System*
http://en.wikipedia.org/wiki/Document_management_system (DMS 2007)

Answering the above questions will help you to develop your document management system. You will now be taken through principles of file management, he main practice related to data management, in more detail.

2.2.3 File Management

(This section is adapted from *The Art of Filing and Managing Your Time (Mindtools website n.d.)* and *10 File Management Tips – File Management Basics for Electronic Files (Ward n.d.)*)

Are you able to locate files in your office and on your computer easily and quickly? If your colleague needed to access your computer, would they be able to locate a file easily?

Look at the pictures below. Ask yourself, which one more looks more like your current office space?

Effective file management will save you time and make it easier for you and others to locate information. File management is about the organisation of information stored in paper files or on your computer. On computers, data are usually organised using drives, folders, and files. The drive is the physical storage portion of computer; the 'folder' is tool for organising on your drive and the files are the basic unit of organisation – the document itself. Similarly, paper-based documents are also stored in files and folders, usually in filing cabinets or well-organised stacks, if cabinets are not available.

TIP:

No matter how you choose to file your documents, information must be in folders. The folders can be categorised in a manner that makes sense to you. Stick with your system for categorisation.

Organising Files

There are many different tips and tools on how to manage documents and files, both on paper and on your computer. This chapter lists a few tips, but is not all inclusive. Here are some tips to help manage your files and data:

- ***Integrate your paper and electronic filing systems.*** Using similar systems for both paper and electronic files makes it easier to remember where files are stored and easier to retrieve them in both systems. You can also store an outline of your paper-based filing system on your computer, so you know where you can expect to find paper files.

- ***Avoid saving unnecessary documents.*** Do not make a habit of saving everything. Save only what is relevant to your work. Having too many pieces of paper or data on your computer adds to clutter and makes it harder to find things in the future. With paper files, you can quickly run out of physical space to store documents securely. For computer files, it may, over time, slow down your computer's performance. Be selective about what you keep.

- ***Store related documents together, whatever their type.*** For example, on a computer, store Word documents, presentations, spreadsheets, and graphics related to a particular project in a single folder rather than having one folder for presentations for all projects, another folder for spreadsheets for all projects, and so forth. This way, it is much quicker to find, open, and attach documents for a particular project. You can use a similar system with paper-based files, storing all documents related to a particular programme or facility in one location.

- ***Nest folders within folders.*** Create other folders within your main folders as need arises. For example, you can divide a main folder into subfolders for quarter, programme, or clinic. You can give a different appearance or look to different categories of folders, such as using different coloured files for your paper-based documents. This can make it easy to tell them apart at first glance. On your computer, try to limit you folder depth to four levels. When there are too many folders within folders, files can become difficult to find. The goal is to have every file in a folder rather than having a bunch of orphan, or loan, files listed.

- ***Separate ongoing work from completed work.*** Some people prefer to save current or ongoing work in a convenient location, such as on their desk or on their computer's desktop until a job is completed. Then, once it is done, they move it to the appropriate location, where files of the same category are stored, such as in a filing cabinet. At periodic intervals (for example, weekly or every two weeks), move files you are no longer working on from your desk or computer desktop to the folders where your completed work is stored.

- ***Cull your files regularly.*** This is just a fancy way of saying throw away, delete, or thin, what is no longer needed. This helps to keep things uncluttered and keep only relevant documents on hand. However, be careful to make sure you do not delete important files that you still need or may need to reference later. This is especially useful with your paper files as well as you electronic documents.

- ***File as you go.*** Make sure that you make the time to file as you go. It is much easier and takes less time to file as you go instead of waiting until you have a stack of hundreds of documents and data to file.

- ***Make sure your filing system and files are backed up regularly.*** Make sure that your desktop and laptop are backed up regularly and, secondly, that the backup includes the directories where you file information. This can be done with hard drives, flash drives or even CDs. This should be done on a regular basis (see the next section for further details on backing up data.)

Naming Files

Take the tips of file management just presented a step further by using a standard naming protocol for customising your file management. This can help you prioritise your work, which in turn can lead to better efficiency.

- ***Create rules for naming files and folders.*** The goal when naming files is to be able to tell what the file is about without having to open it and look. So if the document is data from a health facility in August, call it something like "KwenengHP0808" rather than something like "data". How will you know what data it is without opening it? Also, use plain language to name your folders. For example you do not want to be looking at this list of folders in the future and wondering what "TFK" or any other unclear abbreviation means. Two other file naming tips include: 1) do not use spaces in file names and 2) keep file names under 27 characters. This will keep documents stable and facilitate their recovery.

- ***Follow a consistent method for naming your files and folders.*** Once you decide on a system – stick to it! If you are consistent about classifying and naming your files, then you will consistently be able to find them when you need them most, without spending a lot of time searching. If you end up breaking any of the rules you have made, consider changing your rules so that you can consistently follow them.

- ***Organise files by dates***. Incorporate a date into the file name. This will help you determine which is the most recent document in the folder, without having to open the file and read through the content. For example, a file named "Guidelines-12Oct07" would indicate a version of the guidelines file dated October 12, 2007.

- ***Use version numbers*** to distinguish between documents that have been reworked or changed. Examples would be "PMTCT_Q4_v1" and "PMTCT_Q4_v2." This also makes it easier to pick out the most current file. If your document is going to be looked at, used, or amended by several people, you need to be particularly careful about version control: people can get very annoyed if versions are mixed up and their work on the document is lost.

Make Data Management a Part of Your Life

Clear an hour in your schedule somewhere in the next 7 days, and set your filing system up!

Keep up the filing…it is an ongoing process.

Backing-Up Electronic Data

Making sure that data is safe and accessible is one of the most important aspects of your work as an M&E officer. We mostly save our data on our laptop or desktop hard drives. Using your laptop or desktop alone can be unreliable. There is a saying that "In the world of computers, there are only two types of hard drives: Those that have failed and those that are going to fail." Keep this in mind as you consider the various methods of backing up data. Backing up data, as you may recall, was our last tip eight under file management. This section will go into more detail about methods for backing up and storing data.

Backing up data is crucial in order to ensure you have the most up-to-date data in the event that your hard drive fails. Electronic databases can serve as a backup to paper files. External hard drives, CDs, and memory sticks (also known as a USB, thumb drive, or flash drive) are means to backup your electronic databases and files. When using a memory stick to back up your data regularly, it is necessary to ensure that they are virus free and uncorrupted. Therefore, you should regularly scan for viruses when opening the memory stick. You can also use CDs to back up data. CDs are generally inexpensive and have the benefit of being incapable transmitting or harbouring viruses. An external hard drive can also be used for backing up. These are more expensive versions of memory sticks drives that have a larger memory. You can also back up your data on more than one computer. For example, the DHMT has a desktop for your use and you also have a laptop. You can use one computer as your primary storage and the other for back up.

The process of backing up critical files is one of the most overlooked essentials of owning a computer even though everyone realises how important it is. As M&E officers you should ask how the back-up system works for your site. This should include some form of off-site data sets in case of fire, theft or any number of interruption possibilities. Always make sure you have multiple copies of your data, just in case.

Backing up your data is only one consideration in safeguarding your data. Storage of your backed up data is something you need to consider as well. Backing up helps protect your data in case technical issues arise, such as your computer crashing. However, it is also important to carefully consider *where* you will store your backed up data. It may seem like a good idea to store your memory stick or CDs containing your backed up data with your laptop in its case. However, if your laptop is in its case and it is stolen, you have also lost your back up devices and data. There are countless stories of people who were backing up regularly and even ensuring their files were password protected, only to lose it all because the backed up data was in the same place as the original data and it was all lost or stolen at one time. In order to safeguard against losing data in the event of losing your laptop, make sure you have a copy of the information saved on your laptop and another location that is separate from your laptop.

Below are some basic steps on how to back up data from your computer to other sources, such as memory sticks and CDs.

How to Back Up Data to a CD/DVD using Windows XP

You can back up files or folders to a CD using a feature of the Windows XP operating system. You do not need to format the CD before using this procedure.

1. Insert a blank, **unformatted** CD/DVD into the drive on your computer that allows you to write to a CD. (The drive is usually labelled as **Recordable** or **Rewritable**).

2. If a dialog box appears that asks how to open the CD, click **Cancel**.

3. Open **My Computer** as you normally do.

4. Click the files or folders that you want to copy to the CD. (To select more than one file, hold down the **CTRL** key while you click the files you want.)

5. In the top-left dialog box labelled **File and Folder Tasks**, click **Copy the selected items**.

6. In the **Copy Items** dialog box, click the drive into which you inserted the CD.

7. Click **Copy**. You will see a copy progression window.

8. In the lower-right corner of the window, a pop-up balloon will appear. The balloon will read "**You have files waiting to be written to the CD. To see the files now, click this balloon**."

9. Click the balloon.

10. A dialog box will appear that indicates the files and folders you selected to copy. Check that the files and folders that you want to copy to CD are displayed under **Files Ready to Be Written to the CD**.

11. In the top-left dialog box labelled **CD Writing Tasks**, click **Write these files to CD**.

12. The **CD Writing Wizard** will open.

13. If the **CD name** field is blank, type a name for the CD. (Note: By default, Windows names the CD with the current date.)

14. Click **Next**. You will see the copy progression window. The time required to copy the files and folders you selected depends upon the amount of data they contain.

15. After the files have been copied, the CD automatically eject from the drive.

NOTE: After you have copied files or folders to a CD/DVD, double-click the CD/DVD icon to confirm that the files were copied.

Editing Rights & Restricted Access to Files

As an M&E officer you are dealing with confidential data and may be working in a shared office or even using a shared computer or laptop. Even if you work alone in your office, you need to take measures to ensure the confidentiality of your data. You can protect individual documents by using a password to help prevent unauthorised access to view and/or unauthorised access to make changes. These instructions apply to all Microsoft Office programmes including Word, Excel, and PowerPoint. Any data that has links to individual names should be password protected to ensure patient privacy. In addition, you may want to password protect documents with data sets that someone could unintentionally modify.

Password Protecting Files

This section is adapted from *Create Strong Passwords* (MS Office online 2010)

Go to **Save As**; select which format you want to save your file as.

1. Click on **Tools** in the bottom of the save box.
2. Select **General Options**
3. Enter password under **Password to Open**A password here will require anyone wanting to open the document to view it to enter a password.
4. And/or Enter Password under **Password to Modify.** A password here will allow people to view the document, but require a password to make changes or modify the data and document.
5. Select **read only** - this is recommended by Microsoft and helps maintain the security of the data
6. Confirm password
7. Select **Save**

Your document is now protected! If you lose the password there will be no way to access the file. Choose passwords that you will remember or keep a list of passwords somewhere safe.

PASSWORD TIPS:

- Passwords are case-sensitive. Make sure that your caps lock key is turned off when you enter your password for the first time.
- Make sure your password is kept in a safe place, away from the computer.
- If you forget the password Microsoft cannot recover your password.

Chapter Summary

Your key responsibility as an M&E officer will be to help manage the health data that is generated at the different service delivery sites in your district. In this chapter, strategies and techniques for collecting, storing, filing, and backing up of data have been described. Utilising these strategies and techniques will enable you and others to effectively access and utilise district health data when needed.

Collecting data through surveys can supplement data received from data facilities

Self Assessment Quiz

1. Saving a file with a name ending in the words "v1" or "v2" is an example of which of the following data management techniques?

   a. Version Control

   b. Paper and electronic file integration

   c. Folder classification

   d. Management of the data life cycle

2. Name one reason why data management is important.

3. Describe the relationship between 'data' and 'information'. Provide a real-life example of this relationship from your work experience.

4. How can you confirm that you are using a standardised national health data collection tool?

   a. Check whether the tool is available in all facilities

   b. Look for the Government of Botswana Ministry of Health emblem

   c. See whether the tool collects data on individual clients

   d. Check whether the title of the tool has the word "national" in it

5. The HTC monthly report should reach the MOH office by which date?

   a. 5th of the following month

   b. 10th of the following month

   c. 15th of the following month

   d. 30th or 31st of the following month

References & Bibliography

Data International 2007. *Data Management Body of Knowledge Guide: Introduction & Project Status*. Data Management Association. Available at: http://www.dama.org/files/public/DI_DAMA_DMBOK_Guide_Presentation_2007.pdf

Data Use Network 2010. *Overcoming Constraints to Using Data in Decision Making*. September

Dictionary.com no date. Available at: http://dictionary.reference.com/browse/data [Accessed 24 November 2010]

Last, JM no date. *Data Sources and Collection Methods*. Encyclopedia of Public Health. Available at: http://www.answers.com/topic/data-sources-and-collection-methods

Microsoft (MS) Office Online 2010. *Safety & Security Center: Create Strong Passwords*, Available at: <http://www.microsoft.com/security/online-privacy/passwords-create.aspx>

Mindtools website. *'The Art of Filing and Managing Your Time'. Mind Tools, Essential Skills for an Excellent Career*. Available at: www.mindtools.com/pages/article/newHTE_85.htm

No author, no date. *Section 4: Methods of Data Collection*. Cited in: a)Brinkerhoff, DM *et al.* 1983. *Programme Evaluation: A Practitioner's Guide for Trainers and Educators.* Kluwer-Nijnoff Publishers, Boston, 1983 and b)Atcher, TM 1988. *EDGE Guide to Evaluation: Analyzing Qualitative Data*, prepared by Atcher, T.M., Shelby Country Extension, The Ohio State University and c) Leeds, C 1992. *Edge Guide to Evaluation: Constructing a Questionnaire*. Union Country Extension, The Ohio State University. Available at: http://ohioline.osu.edu/b868/pdf.

Partners in Health (PIH) 2006. *Monitoring and Evaluation: Data Collection, Record Management, and Electronic Telecommunications. Monitoring and Evaluation Chapter 4*. Partners in Health Model Online. Available at: http://model.pih.org/files/Manbuals/2006_AIDS/Chapter_4.pdf

Quynth Lê 2005. *Issues on Health Data Collection*. University Department of Rural Health, University of Tasmania, Australia. Available at: http://www.aare.edu.au/05pap/le05644.pdf

Ward, S no date. *'10 File Management Tips – File Management Basics for Electronic Files'*. Available at: About.com. http://wbinfocanada.about.com/od/datamanagement/a/computerfilemgt.htm

Wikipedia.org no date. *Document Management System*. Available at: http://en.wikipedia.org/wiki/Document_management_system

World Health Organization (WHO) regional Office for Western Pacific 2004. *Developing Health Management Information Systems: A Practical Guide for Developing Countries (2004)*Available at http://www.wpro.who.int/NR/rdonlyres/3A34C50D-C035-425A-8155-65E8AD3CB906/0/Health_manage.pdf [Accessed on 24 January 2011]

Self Assessment Quiz Answer Key

1. Saving a file with a name ending in the words "v1" or "v2" is an example of which of the following data management techniques?

   a. Version Control

   b. Paper and electronic file integration

   c. Folder classification

   d. Management of the data life cycle

2. Name one reason why data management is important.

   Possible answers:

   - *It prevents time being wasted in looking for records.*
   - *It prevents time or effort wasted in using a non-current version of a record.*
   - *It prevents data from being lost or deleted.*

3. Describe the relationship between 'data' and 'information'. Provide a real-life example of this relationship from your work experience.

   Possible answer:

   Describe relationship *- Data is made up of characters, symbols, facts, and images. By itself it has no meaning. When data is analysed and interpreted it becomes information, which can be used by health care planners and providers to make decisions about public health services.*

   Example – *When you receive the monthly reports from the sites that can be considered 'data', numbers that have little meaning in and of themselves. When you analyse the data by comparing it to previous months, use the data to calculate indicators, and present the results to health care planners and providers, then the data becomes information. For example, in the Weekly Notifiable Diseases form, you may have six facilities with 7, 10, 21, 15, 3 and 8 cases of acute diarrhoea listed for children under 5. Only by totalling the data (64 total cases) and comparing it to the previous week (which showed only 10 cases, can you obtain the information that there has been an increase in diarrhoea within your district.*

4. How can you confirm that you are using a standardised national health data collection tool?

   a. Check whether the tool is available in all facilities

   b. ***Look for the Government of Botswana Ministry of Health emblem***

   c. See whether the tool collects data on individual clients

   d. Check whether the title of the tool has the word "national" in it

5. The HTC monthly report should reach the MOH office by which date?

   a. 5th of the following month

   b. ***10th of the following month***

   c. 15th of the following month

   d. 30th or 31st of the following month

Self-Directed Learning Workbook2:
Chapter Three
Data Quality

Data cleaning is critical to data quality

Chapter 3:
Data Quality


Estimated time needed for completion: 2 hours

Chapter Overview

In chapter 2 of this workbook you were introduced to some data collection tools and data management practices. In this chapter we will present and describe the 7 components of data quality. We will also discuss the processes and tools involved in conducting basic routine data audits and more advanced in-depth data quality audits. This chapter will also provide you with information on how to correct data as part of the data auditing process.

Learning Objectives

At the end of this chapter, you will be able to:

- define data quality and explain its importance,
- list and describe the 7 components of data quality, and
- explain how to conduct a routine and in-depth data quality audit in your district.

3.1 Data Quality

Data quality is defined as the extent to which data is valid, reliable, precise, complete, and timely, as well as the extent to which the data has integrity and is kept confidential. Ensuring high quality data is critical to conducting monitoring and evaluation (M&E) activities. Taking steps to ensure data quality is one of the most important things you will do as an M&E officer. High quality data is essential in order to inform decisions at the national level. Data quality is a key component in ensuring that appropriate conclusions are drawn from the information captured at the service delivery sites and integrated into national level reports. High quality data is important for programme decision making, sharing programme information and reporting, and accountability.

There are seven components of data quality: 1) validity, 2) reliability, 3) precision, 4) completeness, 5) timeliness, 6) integrity, and 7) confidentiality. In this section we will discuss each of these components in detail. For each section we end with a set of questions to ask yourself when reviewing the data you receive.

3.1.1 Components of Data Quality

Validity

Validity is defined as the extent to which the data actually measures what it was intended to measure. For example, if we wanted to know whether the orphans and vulnerable children (OVC) in the orphan care programme were regularly attending primary school, we might calculate a percentage as follows:

$$\frac{\text{Total number of registered OVCs regularly attending primary school this quarter}}{\text{Total number of registered OVCs this quarter}}$$

Would this be an accurate way to get the percentage of OVCs regularly attending primary school? Given that OVCs could be up to 17 years old, maybe not. The calculation above would include all registered OVCs, even those who are not at the age to attend primary school. This may lead to calculating a lower percentage than actually exists, and thus the programme may look less effective than it actually is. A better, more *valid* indicator might be:

$$\frac{\text{Total number of registered OVCs in primary school this quarter}}{\text{Total number of registered OVCs between 5-11 years old this quarter}}$$

This revised indicator more accurately captures how many of the registered orphans who are eligible for primary school are actually attending.

With the data M&E officers work with, there are three main types of threats to validity:

1) The indicators in the data collection tools do not accurately capture the information that we intend to measure.

As described in the example above, the initial indicator calculated above did not accurately capture what programme officials intended to measure. Only by revising the indicator do we capture the information we are interested in.

2) The data is incorrectly entered.

A common threat to validity is when data is not entered into a form or database correctly. Errors in data entry might occur when data are transferred from one form to another or from a paper-based system to an electronic system. For example, in hand-written forms a number seven could be misread as a one or a zero could be accidently omitted from a number. This would result in data being inaccurate, and thus invalid.

3) The data is incorrectly summarised.

Mathematical errors can also contribute to data being invalid. For example, when data is summarised, it can be incorrectly added or subtracted. Think of a Health Care Worker (HCW) manually adding up data from a register. Going through line by line, adding data by hand; it is easy to see how errors could be made during this process.

Questions to ask yourself:

- Is there a relationship between the activity or programme and what you are measuring? Can you think of other, more accurate ways to measure the activity or programme which would also be possible given the resources you have available?

- What is the process for transferring data from registers to monthly report forms? Is there potential for error?

- Are steps being taken to limit errors when transferring data? For example, entering the data twice and then checking the two sets of data against each other, built-in checks in electronic systems which provide error messages when the data falls out of specified ranges, or conducting random checks for errors?

- If the data need to be manipulated, are the correct formulas being applied?

- Are final numbers reported accurately (e.g., do the totals add up)?

Reliability

Reliability is defined as the extent to which a particular technique or measure, applied repeatedly, would produce the same result each time. In more basic terminology, reliability is the "consistency" or "repeatability" of your measures. Data is reliable when it is being reported the same way by all individuals across sites and districts. This means that the same tools and processes' for collecting the data must be used by everyone who collects data across all sites/districts. There are several threats to reliability of data. These include inconsistency in the data collection tools themselves, inconsistency in designated procedures for collecting data, and inconsistency in the processes used to collect data.

For example, think of a HCW collecting data in a clinic. Based on your experience, does the same HCW always complete the registers and compile the monthly reporting forms? Is everyone at that clinic likely to fill out the forms in the same way (i.e. do they have the same understanding of each of the indicators)? These same questions can apply on the district level as well. Does everyone in the district complete the forms in the same way?

Data obtained from facility-based registers and records can be unreliable if HCWs make errors in categorising services and treatments given to clients. It may be the case that not all of the HCWs have been trained on the data collection tools, or despite having been trained, they may still have different interpretations on how to fill out the tools.

For example, when the MOH was going through the process of revising the prevention of mother-to-child transmission (PMTCT) tools they found that HCWs had different interpretations of several of the indicators. In the antenatal care (ANC) section of the PMTCT tool, some HCWs were confused about the indicator "Women tested for HIV" – should it count only those who were tested for HIV at that ANC visit at that ANC centre or those who were tested before that visit? When documenting the results for "Women tested HIV-negative" HCWs were unclear as to whether they should count all tests done within the past year or only those done within the past three months.

Within the Routine HIV Testing (RHT) programme, officials found that after initiating RHT, the vast majority of patients were being recorded as being voluntary and based on the patient's wish and few were referred through the new provider initiated RHT process. This worried officials that perhaps the RHT programme was not effective in increasing HIV testing. However, after looking at historical information the number of people who had received an HIV test overall had greatly increased. Upon further investigation officials found that within the health facilities, the nurses would indeed prompt patients to get an HIV test, but would then refer them to the Lay Counsellor for testing. When the Lay Counsellor then received the

patient they did not know that the patient had been referred by a nurse, thus they would record them as 'VCT' (Voluntary Counselling & Testing) rather than 'clinical suspicion' on the RHT form.

Changes in data collection tools can also affect reliability, as the data would not be collected "consistently" over time and may not be comparable between sites if the sites are not using the same form. After the MOH revised the PMTCT tools, some of the indicators in the newly revised tools may not be comparable to those in the previous version.

Questions to ask yourself:

- Is the same instrument used from year to year and site to site?
- Is the same data collection process used from year to year and site to site?
- Are there procedures in place to ensure that data are free of significant errors (such as instructions, indicator-information sheets, training, etc.)?

Completeness

Completeness is defined as the extent to which all the information is available, with no element or part left out. It is a measure of the amount of available data compared to the amount that was expected to be obtained. Think about data forms you have received from various sites. Are they always completely filled out with a number in every space? Of the data that is reported, how well do you think it represents the total amount of services provided (i.e. are all clients, events, activities recorded)?

For example, sometimes HCWs get busy with a high patient load and may not fully record all patient data in the registries. When this happens, information on all of the patients cannot be compiled and reported in the monthly form for that facility. This incomplete data is then included in the district report and gets reported nationally.

Another example to consider is: imagine a situation where patients' TB treatment outcome (i.e. cured, completed, treatment failure) is not being recorded at a facility. If the facility has only 10 patients, you might imagine that this might not effect national reporting much. However, imagine the same lack of recording happening on a large scale, where hundreds of facilities across the country are not recording all the TB treatment outcomes at their facilities. Without recording all cases, the actual TB treatment outcomes could be far different than those that are recorded. For example, you could have a situation where the recorded cure rate is only 40%, but in reality it may be closer to 80%. As the Botswana National Tuberculosis Programme's (BNTP) goal is to reach an 80% cure rate, such a difference would have serious implications for programme planning.

Questions to ask yourself:

- Is there any missing data?
- Are the data all encompassing?
  - Are data from all clients being captured?
  - Are data from all sites being captured?
  - Are all relevant pieces of data being collected?

Precision

Precision is defined as the extent to which the data collected have the necessary detail in order to be meaningful. For example in our orphan care case above, if we did not collect data on the age (or better yet birth date) of each orphan as they registered then we would not have the data available in order to record the number of registered orphans who are of primary school age. Without this information we would not be able to calculate our indicator, percentage of registered orphans in primary school.

Another example with the PMTCT tool was that in the previous version of the register there was no place to record whether patients were on cotrimoxazole, although this was an indicator that needed to be recorded in the monthly report. We would say that the register did not have the necessary *precision* for completing the monthly report.

As in the cases above, the precision of data is usually a reflection of the data collection tools or data entry systems. Namely the tools and data entry systems need to be refined enough to capture the data with enough detail in order to be able to address the key questions that programme officers and policy makers want answered.

Questions to ask yourself:

- Do the data collection tools collect the data with enough detail to answer the key questions?
- Does the data entry or compilation system maintain the necessary level of detail so that key questions can be answered?

Timeliness

Timeliness is defined as the extent to which data are collected, collated, and reported within the time period necessary to inform the decision-making process. Without timely data, it would be difficult for programme officers and administrators at the district and national levels to make decisions on how to best utilise programme resources in order to have the greatest impact. For example, if the national ARV programme did not receive data on ARV patients from some districts in a given month, they would not know how many patients are on ARV, and thus would not know how much of each drug stock regimen to order for the Central Medical Stores.

Additionally, programme officers and administrators would not have timely information on key quality of service indicators, such as how many (and what percent of) patients received CD4 testing after testing positive for HIV and how many are lost to follow-up for districts not reporting data by a specific deadline. They also would be unable to calculate cumulative variables, those variables that track the total number of patients since the start of the programme, or future time periods, which would make it difficult to see the progress of the programme as a whole. Given the large number of districts who did not submit data within the given time period, until that data was received, programme officers would not be able to compare the data to previous national reports to determine whether the programme is maintaining its quality or improving nationally. Thus, it is critical that data is delivered within an appropriately short time period so that it can be useful to influence decision-makers.

Questions to ask yourself:

- Is a regular schedule of data collection in place to meet programme management needs?
- Are data collected, compiled, and reported while they are still relevant and able to inform programme management decisions?
- Are data from within the policy period of interest (i.e., are the data from a point in time after the intervention has begun)?

Integrity

Integrity is the extent to which data are free from manipulation or bias introduced by human or technical means, either wilfully or unconsciously. It can also refer to the extent to which information can be accessed and modified only by those authorised to do so.

For example, imagine a situation where the codes for the TB register are changed. In the previous version of the tool the code "TC" meant "treatment cancelled", but in the new register "TC" now means "treatment completed." If computer systems, such as District Health Information System (DHIS) or the TB e-register, is not changed to reflect the new coding system, the results would be biased towards a lower number of treatment completions and a greater number of treatment failures.

In a few cases, individuals intentionally misreport data. For example, a HCW may be under pressure to report data and in the rush they may fill in fake numbers in order to submit the report.

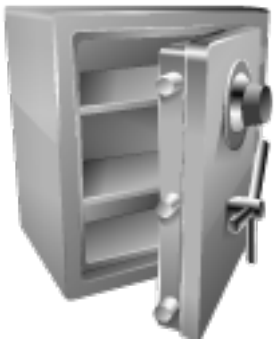
Questions to ask yourself:

- Are there risks that data might be manipulated for personal or political reasons?
- What systems are in place to minimise such risks?
- Has there been an independent review of the data sources?

Confidentiality

Confidentiality is the extent to which patient information is safe and secure as specified by national and/or international standards for patient privacy. There are at least four types of confidentiality to consider; the confidentiality of patients in the facility setting; storing direct patient information in registers; entering patient information into data files; and sharing the data files and performing analysis on the data.

Confidentiality in the facility setting is important to ensure that patients feel comfortable talking about their condition and others, who are not directly responsible for their care, should not able to hear this information. Patients should only be asked questions about their health in a private setting, preferably in a private examination room with the door closed and no interruptions.



Storing patient information on registers is another critical piece in ensuring confidentiality. Sensitive information, such as HIV status, is often coded on registers and should remain coded so that non-authorised people should not be able to determine someone's status by glancing at their register.

Once the information is on the register and is sent for data entry it is very important that the data entry staff understand the importance of confidentiality and review tips for storing confidential data, including storage in a secure location; such as a locked office, filing cabinet, or a password protected computer, where only the necessary staff have access to it.

When sending that data for analysis it is important to determine ahead of time if the people who will receive the data need to know the identity of the patients or just need basic information. For example, if I want to analyse data from the PMTCT programme, I would not need to know individual names but I might need to know ages, or reading level. In this case, it is important to take all of the names out of the data before sending it for analysis.

Questions to ask yourself:

- Is data collected from patients in a way that ensures patient confidentiality?
- Is patient information kept in a secure location, where only the necessary staff has access to it?
-

3.1.2 Comparing the Components of Data Quality

Table 3.1.2 below summarises the seven data quality components, with definitions, examples of how data quality could be compromised for each component, and the results when the data quality is compromised.

Table 3.1.2 Data Quality Components

Components	Definition*	Examples of when Data Quality Errors may Occur	Consequences of poor data quality in Data Analysis & Presentation
Validity	*The right information is collected and is accurate.* The data collected properly measure the quantities of interest. Errors such as recording or interviewer bias, transcription error, and sampling error (to be explained in Workbook 3, Chapter 4) are negligible.	• **Collection:** Tallies are not added correctly or indicators are misinterpreted. • **Entry:** Transcription errors are made during data entry.	Analysis/Presentation results in wrong information and can mislead programme and policy decisions.
Reliability	*The information is consistently measured and collected.* Standard protocols and procedures are in place and observed, ensuring that data are measured and collected consistently.	• **Collection:** Each person who collects data records it differently; a different number of facilities report each month. • **Storage:** Files are not consistently saved in the same location, making it difficult to access them.	Analysis/Presentation does not result in useful information because data is not consistent.
Precision	*Data are available with sufficient detail.* Aggregate information is recorded by subgroups of interest (e.g. sex, age, location).	• **Collection:** Data collection tools do not allow recording of data by age (i.e. only recording patients' age as over/under 14, rather than by specific age). • **Entry:** Rather than entering data by gender, age, etc. as it is recorded, only more general data is entered.	Analysis/Presentation is not as useful as it could be because data is not available at the necessary level of detail to understand how different groups are affected differently by a condition or disease.

Components	Definition*	Examples of when Data Quality Errors may Occur	Consequences of poor data quality in Data Analysis & Presentation
Completeness	*All of the necessary information is included.* The information system from which results are derived is appropriately inclusive, representing the complete list of eligible persons or units in its area.	• **Collection:** Data on some indicators or at some facilities is not collected. • **Entry:** Some indicators are omitted during entry. • **Storage:** Some files are not saved.	Analysis/Presentation is not useful because key information is missing.
Timeliness	*Information is up-to-date (current), and available on time.* Data are up-to-date (current) and information is available on time. Timeliness is affected by: 1) the rate at which the programme's information system is updated, 2) the rate of change of actual programme activities, and 3) the schedule of information use.	• **Collection:** Data collected at facilities is not sent in a timely manner for reporting. • **Entry:** Data is not entered in a timely manner. • **Storage:** Lack of storage. Process for storing does not allow it to be retrieved in a timely manner.	Data analysis/ presentation is not completed in time to inform evidence-based decision-making.
Integrity	*The information systems are secure.* The information system is protected from deliberate bias or manipulation for political or personal reasons.	• **Collection:** Facility does not consistently fill the register, so forms are filled in with fake numbers in order to complete the summary report. • **Entry:** The district wants to show a decrease in particular disease areas, so the data is changed to show a decrease. • **Storage:** Proper protections for the data are not set, so files are borrowed and not returned or mistakenly deleted.	Analysis/Presentation results in the wrong information and can mislead programme and policy decisions.

Components	Definition*	Examples of when Data Quality Errors may Occur	Consequences of poor data quality in Data Analysis & Presentation
Confidentiality	*Data are maintained according to patient privacy standards.* Data are maintained according to national and/or international standards for patient privacy. Personal data are not disclosed inappropriately and are stored securely (e.g. hard copies kept in locked cabinets and electronic information in password protected files).	• **Collection:** Sensitive data is collected from patients in an open, public space. Patients may not feel comfortable providing information. • **Storage:** Patient data is stored in a location where anyone can access it.	Analysis/Presentation has inaccurate data, as patients were not comfortable and did not provide full information.

* *Adapted from: Measure Evaluation (2010)*

3.2 Improving Data Quality – Data Quality Audits

In the previous section we presented the seven components of data quality and provided examples of the various threats to data quality. As a district level M&E officer, one of your main responsibilities is to improve the quality of data reported in your district. In this section we will provide instructions, tools, and ideas to assist you in improving data quality.

There are many techniques that you can use to improve data quality within your district. The first step in addressing threats to data quality within your district is identifying what some of the potential threats might be. This can be done by conducting a data quality audit (DQA). A data quality audit is a structured assessment of the quality of your data across the seven components of data quality. There are two types of data quality audits which are recommended for M&E officers like yourself to use in your work: a routine data quality audit and an in-depth data quality audit.

3.2.1 Routine Data Quality Audit

A routine data quality audit (R-DQA) should be conducted *every* time you receive data. It can be thought of as a "desk review" of the data, meaning you only need to review the data that you receive and follow-up on any questions or concerns you have about the data. This is in contrast to actually physically checking the data sources.

Based on the data you collect and analyse in your district you should develop standard operating procedures (SOPs) for conducting routine data quality audits. These will help you make sure that you use a systematic method for checking the data quality on a regular basis. The SOPs can also help programme officers and other M&E officers in conducting their checks.

Below are some common steps which could be used to develop your SOPs for conducting an R-DQA.

Routine Data Quality Audit
(R-DQA)

1. Record when the data was submitted.
2. Review reports for completeness.
3. Compare data with reports from previous reporting periods.
4. Check to see if the numbers "add up".
5. Follow-up with and provide feedback to the reporting sites and facilities.
6. Correct and report data.

Conducting the Routine Data Quality Audit

Step 1. Record When the Data was Submitted

This will help you to track the *timeliness* of facilities in different programme areas over time. As soon as a report is received you should write the date you received the report in a standard location – such as the upper right hand corner. Then you can use a paper-based form or Excel to record timeliness across all facilities. Below is an example of a timeliness tracking tool.

Table 3.2.1.a Example of a Timeliness Tracking Tool.

#	**Quarter:** Oct-Dec 2010	**PMTCT Monthly Report**			**TB Monthly Report**			**…**
			Deadline Met?			**Deadline Met?**		
	Facility	**Submission Date**	**YES**	**NO**	**Submission Date**	**YES**	**NO**	
1								
2								
3								
…								

Step 2. Review Reports for Completeness

The simplest method of checking for *completeness* is to scan over the report forms that you receive to look for blank or empty spaces. Record the areas that have been left blank. If any spaces are completed as "Not Applicable" in the form, review the data to ensure that they truly not applicable. For example, on the BHRIMS-003B Section 2.9, it asks for the number of STI cases at health posts. If the district does not have health posts, you might expect to see not applicable, whereas in district that does have health posts this should be completed.

You should then record which fields are completed and which ones are not. A tool, such as Table 3.2.1.b "Example of a Correct & Complete Tracking Tool", could be used to track this information. You will notice that the tracking tool includes all the same fields as the MH1049 form, with a breakdown of each disease by gender and age. Within each of the spaces where data would be filled in on the regular form, a

letter from the key is filled in, which records whether that space has been correctly filled on the form.

Table 3.2.1.b Example of a Correct & Complete Tracking Tool

Quarter: Oct-Dec 2010		**Form: MH1049**							
Key: F = Correctly Filled Inc = Incorrectly Filled PF = Potentially Incorrectly Filled UF = Unfilled N/A = Not Applicable									
Report Submitted? Y/N									
		Facility 1				Facility 2			
Indicators		Total	<1	1-4	5+	Total	<1	1-4	5+
Unconfirmed Malaria	M	F	F	F	F				
	F	F	F	F	F				
Malaria w/ severe anaemia	M	PF	U	U	U				
	F	PF	U	U	U				
Uncomplicated Malaria lab-confirmed	M	…							
	F								
Typhoid Fever	M								
	F								
….	M								
	F								

Alternatively, if most spaces are usually filled in completely, you may decide instead to just keep a list of errors by facility. This is usually less time consuming, but has the disadvantage of focusing on a few errors rather than overall success. However, you can calculate the total percent of correctly filled data fields by subtracting the number of incorrectly filled from the total number of spaces and dividing by the total number of spaces.

Step 3. Compare Data with Reports from Previous Reporting Periods

When comparing new data with data from previous reporting periods you should look for the following:

- Whether the data seems to be within an appropriate range. You would expect the numbers for the current reporting period not to differ too much from previous reporting periods, unless there was a special reason. Look to see whether the numbers seem to be what you would expect. For example, if the number of people who receive HIV tests at a given facility is usually around 50 per month but during one reporting period you notice it suddenly decreases to 5 people a month, you would want to check with the HCWs at that facility to see whether there was a reason for the sudden change or whether it was a data entry error. Did they run out of HIV test kits? Did they accidentally forget to add a 0 to the report?

- Whether the numbers change from month to month. Are the number of HIV tests performed exactly the same, each month? In some instances, people may report data for the previous month simply because the current month's data has not been collected of captured.

Once found any errors or 'suspicious' data should be recorded and followed-up with the individuals who submitted the reports.

Step 4. Check to See if the Numbers "Add Up"

Some indicators when added up should equal each other. For example in the PMTCT tool, the number of 'Women HIV tested' and the number of 'Women declined testing' should add up to the number 'Women offered RHT'. Similarly, the number of 'Women HIV tested' should be the sum of 'Women tested HIV-positive' and 'Women tested HIV-negative'. Making a list of all of these variables that should correlate for each tool you report on and then checking the reports you received against this tool can help to improve the quality of the data. Record any potential errors on a tracking tool.

Step 5. Follow-up With and Provide Feedback to the Sites

Once you have reviewed the tools for any potential errors, the next step is to contact the sites to check whether these are truly errors or whether the data are in fact, correct. The most common method of following up on data submissions is through phone calls. During these phone calls, be sure use a ***feedback sandwich***, where you begin with praise for a good work done, followed by feedback and a summarized praise. You can think of it as the beginning and ending praised as the breads and the middle feedback as the meat of your feedback. To give a feedback sandwich, you can start with:

1. providing praise for something specific that the site has done well,
2. addressing questions or clarifications on the potential errors you have found, and then
3. ending with another point of praise on something else they have done well in reporting.

Receiving feedback that is focused only on problems or errors can be difficult. Using the feedback sandwich helps individuals to receive what could be perceived as 'negative' feedback more positively, as it also focuses on what they have done well.

Learning Activity 3.2.1.a

Feedback Sandwich

- *Practice using the feedback sandwich with a friend or co-worker for anything you would like to provide feedback on. In the space below, write down what you said and how the person responded.*

- *Do you think using the feedback sandwich helped? How did you feel about using this approach?*

In addition to verbal feedback, it is also good to provide a written record to sites on a regular basis. While there is no specific guidance on how often the feedback should be provided, a quarterly or semi-annual basis is recommended. Whatever frequency you decide, it is important that the feedback be provided regularly, so staff know when to expect it and can begin to monitor their own changes in data quality over time.

Table 3.2.1.c, "Data Quality Summary Tool", provides an example of a written record. Similar to the verbal feedback sandwich described above, this should include positive comments and feedback in addition to recommendations for improved data quality.

Table 3.2.1.c Data Quality Summary Tool

Data Quality Summary						
Reporting Period: Oct-Dec 2010				Facility Name:		
Tools	Submitted? Y/N	On-Time? Y/N	% of Fields filled correctly	# of Unfilled Fields & Errors	List of Unfilled Fields & Errors	Comments/ Recommendations
MH1049	Y	Y	93%	4	1) Rabies exposure, F, 5+ 2) STI Contacts, M, 0-14 3) AIDS, M, 0-14 4) AIDS, F, 0-14	When there is no data, fill in a 0 instead of leaving it blank
Total			N/A			

Additional Comments:

[Here could include progress since previous quarters, how this site compares with other reporting sites, etc.]

Having 93% of fields filled in correctly is an improvement over last quarter, where 88% of fields were filled in correctly. It also puts Mushongwe Facility in 3rd place for data quality across the district as a whole. The number of women who received an HIV test at ANC now matches the number that test positive and negative. Great job!

Step 6. Correct and Report Data

After following up with those who have submitted reports and clarifying any errors, it is time to correct the data. With the notes taken during your call or visit with the data capturers, update the tracking tools to reflect the latest data. Places where you had "potentially incorrect" should now be marked as either "Filled" or "Incorrectly Filled".

To correct the forms that have been turned in, attach a cover sheet to the form. On the cover sheet write a note describing what fields are being changed and why they are being changed. You should then sign and date your statement. For example you may write something like this:

> *"Rabies exposure, F, 5+ from [blank] => 0*
>
> *Rabies exposure, F, 5+ was unfilled when the form was received. Called to confirm with the Matron at Mushongwe facility and there were no rabies exposures for females in that age group.*
>
> *Signed: David Thusani, District level M&E officer, 1 November, 2010"*

The statement should clearly state the change being made, followed by a brief explanation of why the change is being made. This cover page allows a reader to scan through the changes while also being able to look at additional background information if needed.

It is important not to just make a change directly on the form. If there is a question later, it will be important to be able to track where the data came from and why the changes were made. If changes are made directly to the form, it may be difficult to determine how the data came in from the facility and what changes you made as the M&E officer.

Once the notations are made on the cover sheet, you can then correct the data in data entry system, either DHIS, eBHRIMS, or your paper-based compilation form. If there are major data issues that may affect interpretation of the data at a higher level, such as missing data from a facility or a major unsolved discrepancy, you may also want to include a cover sheet on your compiled report to notify M&E officers and programme staff at the national level.

Learning Activity 3.2.1.b

Create Your Own Tracking Tool

Choose one of the tracking tools above and adapt it to fit the data that you collect and monitor in your district. Fill in data from the previous month and practice by calculating some of the following for two facilities:

- *Percent of reports submitted*
- *Percent of reports submitted on time*
- *Number of unfilled fields and errors*
- *Percent of fields correctly entered*

3.2.2 In-depth Data Quality Audit

Whereas an R-DQA can be thought of as a "desk review" of the data, an in-depth data quality audit (ID-DQA) involves going to the sites where the data is collected in order to check data sources and talk to data collectors. An ID-DQA is conducted in order to better understand the data collection process. This type of data audit is more resource intense, but has the advantage of being able to assess more of the data quality components. This is why ID-DQAs are only recommended to be done on an annual basis, or semi-annually if possible.

The R-DQA is primarily focused on timeliness, completeness, and some measures of validity. An ID-DQA allows an M&E officer to also assess reliability, integrity, and confidentiality. You can think of an ID-DQA as including all the steps of an R-DQA while also collecting additional information in order to conduct a more complete data audit.

Similar to the R-DQA, you should develop SOPs for conducting ID-DQAs in your district so that a systematic method is recorded and used for ID-DQA.

Planning the Audit

Before going on site visits to conduct an ID-DQA, there are three main steps you should take: getting approval, focusing your audit, and getting organised, in terms of setting up the logistics and communicating the visits to sites.

Step 1. Get Approval

In order to conduct an ID-DQA you first must work with the district leadership and programme officers in order to build support for conducting an audit. You may need to sensitise district leadership to the purpose and importance of an ID-DQA. One method of doing this is to share with them the results of your R-DQA from previous reporting periods to provide them with information on what you already know about the current data quality. You can then describe the additional information and improvements in data quality that may result from an ID-DQA. You should also have an agreement with the district leadership and programme officers on a plan for documenting and disseminating the information collected during the audit.

In conjunction with the district leadership and programme officers you should decide who will be a part of the data audit team. This may be an iterative process with deciding how to focus your audit. Your audit team may depend on what sites or programme areas you will be visiting (see Step 2 below). An audit team could be small, with just one or two M&E officers. The auditing team could also be a subset of a larger team that is going on an overall programme review visit.

For example, one challenge that ID-DQAs may face is difficulty in securing transport. To address the transport issue, as well as to reduce the burden at the clinic, you could try to schedule your visits to coincidence with other programmatic visits to clinics. This may also help the programme officers visiting the clinics to collect data that is useful to them, as well as helping them to better understand the reporting systems and the importance of data quality.

In-depth Data Quality Audit

(ID-DQA)

Planning the Audit:

1. Get Approval
2. Focus the audit (site or programme)
3. Get organised

Conducting the site visit:

1. Brief the site leadership
2. Conduct audit using ID-DQA tools
3. Develop an action plan
4. Discuss the findings with leadership

After the site visit:

1. Write and disseminate data audit report(s)
2. Follow-up on the implementation of data quality improvement action plans

Another challenge in conducting an ID-DQA is getting adequate time away from your regular work in order to conduct such as audit. To reduce your own workload, you could schedule ID-DQA visits at "off-times", such as between reporting quarters for example, because quarterly reports are due January 31st, you might schedule an ID-DQA in November rather than January. Or you might schedule an ID-DQA visit between the 15th-30th of the month, after the previous months reports have been submitted and before the next round starts.

Step 2. Focus Your Audit

In addition to getting agreement that an audit should be conducted, another key element is deciding what your audit should focus on. You may want to check the quality of all the data at all the sites within your district. However, this will likely require staffing and resources well beyond what is at the disposal of most M&E officers. Therefore, it may be necessary to narrow the focus of your audit.

There are two main ways of focusing your audit – by site or by programme:

Site-focused audit

In a site-focused audit you would only conduct an audit of a few sites, while looking at a greater number of programmes. In choosing which facilities to visit there are several strategies you could use:

- Choosing those sites that most frequently send incomplete or late reports. This is the most obvious choice as these sites are likely in need of your support the most.
- Choosing a mix of poor performing and well-performing sites. Choosing a mix could help you to learn what strategies work best for ensuring good data quality. These lessons could then be shared with poor performing sites in order to help them identify way that they might improve.
- Choosing sites that you have not had a lot of contact with or have not visited in a long time. Maintaining relationships is a key part of ensuring consistent quality reporting. Visiting a site that you have not been to in a while can let the HCWs know that you are paying attention to the data they are submitting and encouraging them to continue with their good work. While it may be natural to want to focus on those sites that you notice are performing poorly, the "quiet" sites, which routinely report, may also have problems with data quality that are not apparent in the data you receive. Visiting these sites gives you a chance to look more in-depth at the data collection procedures.

Programme-focused Audit

In a programme focused audit, you would visit a wider number of facilities, but would only focus on one or two programmes. This is a good choice when a programme officer is a part of the data quality audit team, if a new data collection tool has been rolled out for a programme, or if you have noticed a lot of errors with a certain form.

The best information to use in deciding which programmes to focus your audit will come from the R-DQAs you have conducted and from the preferences of the district leadership and programme officers.

Step 3. Get Organised

An effective audit must be well-organised. To conduct an ID-DQA you will need to schedule the days that you will be away from your regular office, ensure that there is transport available, and communicate with the sites you are visiting to let them know when you will be arriving.

Each district has its own systems for how best to communicate clinic visits. When planning an ID-DQA it is best to follow those processes. Check with the other M&E officer or the programme officers to learn more about the procedures. Generally, a letter or savingram will be sent to the site to notify them of the visit. This letter should include the dates of the visit, who will be visiting, and the purpose of your visit, (i.e. an audit of the PMTCT registers). Details regarding who should send the letter and who it should be addressed to at the facility will differ by district. Following the letter, it is a good idea to follow-up with a phone call to ensure that they received the letter/fax and also call to confirm your visit a few days before you arrive.

Before you go on site visits, you should also put together a package of support materials. This package could include:

- Feedback to the site on recent reports, both data quality and a summary of recent data they have submitted. You will use this to help them identify trends in their data overtime (this will be discussed in Chapter 4 on Data Analysis).
- Copies of the most recent data collection forms and registers.
- Instructions for completing forms.
- ID-DQA tools.

Finally, in order to conduct an ID-DQA you will need to develop a tool to guide the audit. You can adapt parts of the R-DQA tools you developed. Another resource is the guidelines and tools developed by MEASURE Evaluation. MEASURE Evaluation is an organisation which provides technical assistance for M&E in many countries all over the world. The guidelines and tools relevant to conducting an ID-DQA are listed below. Note that definitions used by MEASURE may be different from those we use here in Botswana. The main point here is that you can use or adapt the resources listed below to conduct an ID-DQA:

- Routine Data Quality Assessment Tool: Guidelines
- Routine Data Quality Assessment Checklist
- Multi-Indicator Routine Data Quality Assessment

See Appendix B – D for a copy of these tools. To access these tools online visit: http://www.cpc.unc.edu/measure/tools/monitoring-evaluation-systems/data-quality-assurance-tools

Conducting the Site Visit

There are four steps for conducting a site visit: meeting with site leadership, using the ID-DQA tools to conduct the assessment, developing an action plan with data collection staff, and debriefing with site leadership. A site visit can take between two hours to one day to complete, depending on the size of the facility, the extent of the audit and data quality challenges encountered.

Step 1. Brief the Site Leadership

Upon arrival at the site, the ID-DQA team should visit the site leadership to inform them that you have arrived and remind them of the purpose of your visit which is to review the data collection and compilation processes. If the site is a health facility, you may also want to ask them to relieve the main data collection staff of some of their clinic duties that day so that they can spend time to conduct the review with you. In return you may offer to assist with some of the data recording at the site for the day.

Step 2. Conduct Audit Using the ID-DQA Tools

Using the ID-DQA tools you developed or adapted, conduct the data quality audit. The audit tools should cover the following areas:

- Observe how the standard forms are used and ask about routine data collection procedures at the site.
- Examine the appropriateness of data storage procedures; e.g. security, confidentiality, and accessibility.
- Identify parties responsible for the data generation; e.g. who captures/enters the data, how data is reported.
- Review registers for
  - Incompleteness
  - Recording errors
- Review summary reports for:
  - Incomplete entries
  - Recording errors
  - Whether reports match register entries
  - Whether formulas have been used correctly to aggregate data

During the review process you will also want to fill in any missing data and correct any errors in both registers and summary report forms as you would during an R-DQA and following the same guidance (see R-DQA Step 6. Correcting Data and Reporting).

Because health facilities are typically busy in the mornings, you may want to observe clinic operations for confidentiality, accuracy, and completeness first. Later in the day, when staff is less busy with patients, you can conduct the data audit and review policies and procedures with them.

Whenever possible, have data collection staff from the facility complete the ID-DQA with you. This is an opportunity to sit side-by-side with the data collection staff and learn more about their processes. Provide the data collection staff with a copy of the tool you are using. Sitting side-by-side you and the data collection staff member should independently fill out the tool. After you and the data collection staff member have each completed the tool, go over their results. Ask them to clarify anything that does not make sense or is not clear to you. Share your results with the staff member. Discuss any differences in the way you and the data collection staff member completed the form, explaining the rationale you used to complete the form. Having them go through this process with you can help to build their understanding of the important elements of data quality and can help build their buy-in for improving data collection and data entry.

Step 3. Develop an Action Plan

Once the policies, procedures and data have been reviewed, it is time to sit with the data collection staff to review the preliminary findings and collaboratively develop a data quality improvement action plan. At some sites the site leadership may also want to be involved in this process. For ideas on what types of activities can be included in the action plan see Table 3.2.2 Data Quality Components & Possible Solutions to Data Quality Threats below.

A space for recording action plans should be included within your ID-DQA tools (see Part 3. Recommendations for Service Site in the Measure Evaluation tools). This should be done collaboratively with the data collection staff. Recording these plans is useful tool which can help build the commitment to carry out the action plans. It also helps to be able to track progress on these action plans over time.

Step 4. Discuss findings with Site Leadership

Once you have reached an agreement with the data collection staff on an action plan, you should provide a final verbal summary to the site leadership. You should also provide them with a short written summary of key findings and the action plan for them to keep. Having the data collection staff complete a duplicate copy of the ID-DQA forms can help ensure that a copy is readily available on site.

Table 3.2.2 Data Quality Components & Possible Solutions to Data Quality Threats

Components	Possible Activities for a Data Quality Action Plan
Validity	• Thoroughly acquaint all stakeholders in the data management process with proper indicator definitions. • In case of electronic data entries, use built-in validation checks. • Check 10% of data fields against their source to determine whether the data has been entered/transferred correctly. • Run tabulations and look for data that is outside of the expected ranges.
Reliability .	• Conduct periodic review of data collection to determine whether practices are similar across data entry staff and facilities. • Develop policies for consistent filing. • Distribute instructions for completing the forms.
Precision	• For standard MOH data collection tools, advocate for changes in data collection tools. • For district level tools, discuss with DHMT members to modify tools and make changes.
Completeness	• Check to ensure all forms are turned in and all data fields are filled. • If possible, check other sources to determine missing values. • Follow-up with those who entered the data to verify values in incomplete data forms.
Timeliness	• Have a calendar of when data is due from facilities and to MOH/MLG. • Provide regular feedback and clarify dates of data collection periods when presenting the results or giving feedback. • Make realistic time frames (due dates) for data collection/ reporting allowing ample time at each level of the data management process.
Integrity	• Perceive data as official GOB property and keep it secure and under a password or in a locked cabinet. • Develop a routine system for randomly visiting site/facilities to check reported data against registers. • Maintain objectivity and independence in your work during data collection, management and assessment procedures.
Confidentiality	• Perceive data as official GOB property and keep it secure and under a password or in a locked cabinet. • Separate identifying information from health related information.

After the Site Visit

Conducting the site visits is only one part of the ID-DQA process. Once you return to your office you will need to compile the data collected into an ID-DQA Summary Report, disseminate the information, and follow-up with sites on implementing their data quality improvement action plans.

Step 1. Write and Disseminate the Data Audit Report(s)

Following a site visit, you should complete two reports: an individual site report and a summary report. The individual site report should be 1-2 pages in length and should summarise the data collected using the ID-DQA tools. This report will be similar to the R-DQA Data Quality Summary Tool described earlier in this chapter. However, the individual site reports should also include information about the level of data quality in the registers and the policies and procedures the site uses for data collection. If you have conducted several site visits, you should prepare a separate report for each site.

Copies of the site report(s) should be shared with the site leadership and programme officers.

Once you have prepared the site reports, you should compile them into a single summary report. Your summary report will be shared with your direct supervisor, programme officers, and the district leadership. For example, a summary report could include an introduction (why you conducted a data quality audit), a description of the sites visited and tools used, the findings of the assessment, and recommendations for improving data quality. A template for a summary report can be found in Appendix A.

In addition to creating and sharing a written report, you should also be prepared to present the information included in the report with the district leadership and programme officers. Such presentations can be a powerful tool for building buy-in to continue data quality improvement efforts and securing additional resources if needed.

Step 2. Follow-up on the Implementation of Data Quality Improvement Action Plans

The ID-DQA process is meant to be a catalyst to improve data quality within your district. In order to ultimately improve data quality, the most important step is to follow-up with the sites you visited and record their progress on their action plans. Following-up will help keep the facility staff motivated and the action plan moving forward. Following-up on the action plan can be done during regular follow-up phone calls to check in on monthly reports as a part of the R-DQA process. You can document progress on the action plans as a part of the R-DQA Data Quality Summary.

Chapter Summary

High quality data is essential in ensuring that appropriate conclusions are drawn from the information captured at the service delivery sites and integrated into national level reports. High quality data is the foundation for making programmatic decision on the district and national levels. Ensuring that data is of high quality is one of the most important things you will do as an M&E officer.

As an M&E officer, you should be well acquainted with the seven components of data quality: validity, reliability, precision, completeness, timeliness, integrity, and confidentiality. You should also be aware of the potential threats to data quality and how these threats can be addressed. Conducting data audits, both R-DQAs and ID-DQAs, is a powerful strategy for better understanding the threats to data quality within your district and can help in developing action plans to improve the quality of data in your district.

Remember:

The data you collect are only useful in generating meaningful information if they are of the highest possible quality!

Data cleaning is critical to data quality

Appendix A: DQA Report Template

Introduction

- Background about your district (location, population, etc.)
- Number/type of facilities
- Number/type of programmes

Methods/Tools

- Describe how you decided to focus the assessment – why did you choose specific facilities or programmes?
- Describe the tools used to conduct the assessment (attach copies of the tools as appendices).

Findings

- Summary of your assessment findings.
- Be sure to call out both strengths and weaknesses in the current data quality system (what is working well? what needs improvement?).
- Analyse information from tools by Facility and by Programme
  - By Facility - Ideas:
    - Table of facilities with percent of reports that are turned in on time and are complete.
    - Qualitative information on problems/issues that facilities have in correctly completing forms (could also include number of errors per facility).
    - Provide information on any solutions that have been implemented.
  - By Programme - Ideas:
    - Table by programme of the percent of reports that are turned in on time and are complete.
    - Qualitative information on problems/issues that facilities have in correctly completing forms by each programme area (could also include number of errors per facility).
    - Provide information on any solutions that have been implemented.
- Include a one to two paragraph conclusion that summarises the findings.

Next Steps/Recommendations/Way Forward

- Explain how the information gathered in this report will be used to improve data quality.
- Describe what steps you will take or recommend that others take in order to improve data quality.
- Describe any assistance that you will need from the mentoring team to be able to improve data quality.
- Describe how/when follow-up on these issues will be done.

Reflection

- Document any best practices or challenges you experienced in doing this data quality assessment.
  - What went well?
  - How would you improve this next time?

Self Assessment Quiz

Match the data quality component to its definition

A. Validity
B. Reliability
C. Completeness
D. Precision
E. Timeliness
F. Integrity
G. Confidentiality

1. ____ the extent to which data are collected, collated and reported within the time period necessary to inform the decision-making process.
2. ____the extent to which the data actually measures what it was intended to measure.
3. ____ the extent to which patient information is maintained according to national and/or international standards for patient privacy.
4. ____ the extent to which the data collected have the necessary detail in order to be meaningful.
5. ____ the extent to which all the information is available, with no element or part left out.
6. ____ the extent to which data are free from manipulation or bias introduced by human or technical means, either wilfully or unconsciously.
7. _____ the extent to which a particular technique or measure, applied repeatedly, would produce the same result each time.
8. Which of the seven data quality components does an R-DQA address? Which ones is it not able to address and why?

For each case below, write which of the data quality components is being affected. Include an explanation of why you chose that data quality component and then write a possible solution to improve data quality.

9. During a site visit you notice Goitse, the nurse charged with data collection at Boletswe clinic, has been using a different method than nurses at other sites to tally the number of Sexually Transmitted Infections (STI) contacts on the MH1049 form.

10. During a site visit you notice that the PMTCT register is being kept open in the patient waiting area, next to the patient benches. The clinic staff insists that this is the best location because they can all access it as needed.

11. During a phone call with the data entry staff at the hospital, he complains “Eish, I’m having a hard time completing this special report that the district leadership requested. We don’t collect information on how many patients have had a CD4 count in the past six months.”

References/Bibliography

International Development Research Centre. Designing And Conducting Health Systems Research Projects: Volume 1.
Available at: <http://www.idrc.ca/en/ev-56604-201-1-DO_TOPIC.html>
[Accessed: November 2010]

Measure Evaluation 2010. *Data Quality Assurance Tools*. Carolina Population Center at the University of North Carolina at Chapel Hill. Available at: <http://www.cpc.unc.edu/measure/tools/monitoring-evaluation-systems/data-quality-assurance-tools> [Accessed November 2010]

Self Assessment Quiz Answer Key

Match the data quality component to its definition

1. __*E*__ the extent to which data are collected, collated and reported within the time period necessary to inform the decision-making process.
2. __*A*__the extent to which the data actually measures what it was intended to measure.
3. __*G*__ the extent to which patient information is maintained according to national and/or international standards for patient privacy.
4. __*D*__ the extent to which the data collected have the necessary detail in order to be meaningful.

A. Validity
B. Reliability
C. Completeness
D. Precision
E. Timeliness
F. Integrity
G. Confidentiality

5. __*C*__ the extent to which all the information is available, with no element or part left out.
6. __*F*__ the extent to which data are free from manipulation or bias introduced by human or technical means, either wilfully or unconsciously.
7. __*B*___ the extent to which a particular technique or measure, applied repeatedly, would produce the same result each time.
8. Which of the seven data quality components does an R-DQA address? Which ones is it not able to address and why?

   Answer: *The R-DQA is really focused on timeliness, completeness, and some measures of validity. It is not able to assess some aspects of validity, reliability, integrity, and confidentiality as to assess these you would need to be able to view the data collection processes.*

For each case below, write in which of the data quality components is being affected. Include an explanation of why you chose that data quality component and then write a possible solution to improve data quality.

9. During a site visit you notice Goitse, the nurse charged with data collection at Boletswe clinic, has been using a different method than nurses at other sites to tally the number of STI contacts on the MH1049 form.

   Answer: *Reliability because the data is not being collected in the same way across all sites. Solution: provide Goitse with instructions on how to complete this field in the MH1049 form and explain them to her.*

10. During a site visit you notice that the PMTCT register is being kept open in the patient waiting area, next to the patient benches. The clinic staff insists that this is the best location because they can all access it as needed.
    - ***Answer:*** *Confidentiality because sensitive patient data is kept in the open where any patient could access it. Solution: remind staff about the importance of confidentiality, work with them to find another convenient location to keep the PMTCT register where patients cannot access it as easily.*
    - ***Another possible answer:*** *Integrity because the register is available for patients to make changes. Solution: same as above.*

11. During a phone call with the data entry staff at the hospital, he complains "Eish, I'm having a hard time completing this special report that the district leadership requested. We don't collect information on how many patients have had a CD4 count in the past six months."

    Answer: *Precision because the data is not collected with the necessary detail for this report. Possible solution: Explain situation to district leadership, advocate for changes in the ARV Site manager tool if this additional data is truly needed, or work with data entry staff to see if this information, or similar information, could somehow be pulled from the IPMS/PIMS II system.*

Self-Directed Learning Workbook2:
Chapter Four
Basic Data Analysis

Data analysis is an important part of evidence based planning

Chapter 4:
Basic Data Analysis

Estimated time needed for completion: 2 hours

Chapter Overview

In this chapter you learn some basic techniques for analyzing data. You will learn about different data types and some basic calculations that can be used to analyze that data. You will also learn how to use Excel to prepare graphical summaries of the data.

Learning Objectives

At the end of this chapter, you will be able to:

- define and distinguish between four data types;
- calculate numeric summaries of location (mean median, mode), spread (range, variance, and standard deviation), and shape (skewness); and
- use Excel to prepare graphical summaries that show relationships between two variables.

4.1. Introduction

Analysis transforms data into meaningful information that can be used by programme decision-makers. For a district level monitoring and evaluation (M&E) officer the process of transforming data into meaningful information will happen as follows: you will receive data from health programmes in your district and analyze and present these to programme officers and administrators to help them in their decision-making process. Based on the data, you may also identify further evaluation questions that need to be answered.

Previous chapters have also discussed how data is collected, managed, how one can ensure that the data received is of high quality to support decision making that is based on correct and timely information. In this chapter, we will focus on the different types of data and how they can be organized, summarized, and interpreted to effectively share information with programme officers and administrators.

NOTE: This chapter is more hands on. You will need a calculator and a computer with Microsoft Excel to follow the examples discussed in this chapter.

4.2 Data Types

Data may be available as numbers, text, symbols, and images. Usually for statistical analysis, you will be looking at numbers. The first step in analyzing data is to determine the type of data that you have. The type of data will help you decide how you will analyze and present the data. These two types of data are: categorical (qualitative) and numerical (quantitative).

4.2.1 Categorical Data

Categorical data are usually expressed as names or labels and can be assigned to a category. For example, food items can be sorted into distinct categories such as dairy, meat, fruits, vegetables, and grain. Each food item belongs to one and only one category. For example, milk can only be dairy; wheat can only be grain.

Categorical data can be further divided into *nominal* or *ordinal* data.

Nominal Data

Nominal data can be assigned a code which can be a name or a number. The number is only a label. Items with that number can be counted but cannot be put in any order or measured.

Examples:

- Gender (male or female can be given the code 1 & 2 respectively). 'Male' and 'Female' do not follow any particular order - you cannot say that one comes before the other. Nor can you add or subtract males from females; or females from males.
- Marital status (single, married, divorced, or widowed).
- Names of districts.
- Cadre of health care worker (HCW).
- Type of disease (tuberculosis (TB), malaria, and/or pneumonia).
- Reasons for HIV testing (clinical suspicion, prevention of mother-to-child transmission (PMTCT), medical exam, sexually transmitted infection (STI), voluntary counselling and testing (VCT), ape, needle stick injuries, TB, and/or other)

Ordinal Data

Ordinal data are numbers or values that have a natural order but do not contain a zero point. You can count the numbers and put them in order, or ranking. However, unlike true numbers, you cannot meaningfully calculate the distance between the values.

For example, in a race it is possible to ask worker to order 1-5 which things contribute to their satisfaction. Perhaps 1st ranked is a good salary, 2nd is subsidized housing, 3rd is good relationships with supervisors, 4th is good relationship with colleagues and 5th is reasonable working hours. What you would not know is the measurable difference between when the different reasons – was it a close decision between salary and housing or did salary far surpass housing? Do salary and housing combined make the other reasons insignificant or do they all contribute almost equally? All we know is that salary was ranked before housing which was ranked before supervisors etc. We do no know what the measured difference between these reasons would be. This example uses ordinal data of numbers. The following examples show some examples of ordinal data values.

Examples:

- Severity of infection can be categorized as mild, moderate or severe. 'Moderate' is clearly more than 'mild' and less than 'severe'. However, we cannot measure how much more 'medium' is than 'low'.
- Form in school (Form 1 through Form 5)
- Age groups of patients (0-14, 15 – 49, 50+)

4.2.2 Numerical Data

Numerical data usually represent and describe a quantity that can be measured. The data can have *discrete* or *continuous* values.

Discrete Data

Discrete data are measured with numbers that are distinct and separate. Discrete data are usually used in cases where we are counting something, such as number of people, equipment, services provided. Data is considered discrete if the thing you are counting, because of its nature, cannot be divided. For example, if we are counting the number of people in your immediate family, we would not say 5.5. People cannot be divided just as a vaccine cannot be divided if your are counting vaccine's delivered. The majority of data that M&E officers collect is discrete, as most data are counting the number of people who receive a given service.

Examples:

- Number of children seen at outpatient department
- Number of patients on ART
- Test scores
- Number of facilities that receive a certain piece of equipment

Continuous Data

Continuous data are numbers that can have any value, including infinity and have a set interval. Continuous data can be counted, ordered and measured. These numbers have a set zero point. For example, weight can be written based on rounding to the nearest kilo (75kg) or it can be written in a more exact manner with 2 decimal points (74.87kg). This data is continuous because it can be any number, or fraction of a number in relation to zero.

Examples:

- Weight
- Height
- Temperature

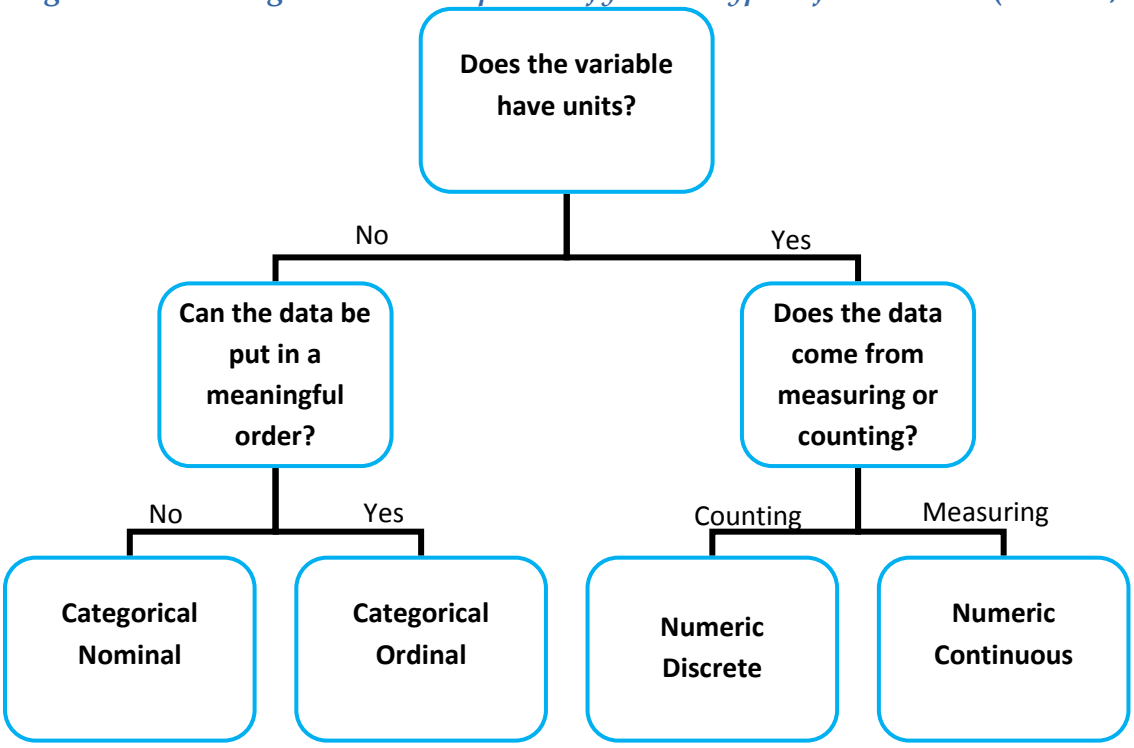
4.2.3 Variables

Variables are measurable bits or data. Variables are properties or characteristics of any entity (i.e. some event, object, or person) that can take on different values or amounts. They are called *variables* because the values of these characteristics can *vary*. For example, gender is a variable because it varies between two values: male and female. In fact, each of the examples of the different data types above can be considered a variable – marital status, form, test scores, age – because they all can take on multiple values. You can use the terminology of the data type coupled with the term variable to describe these characteristics. For example, form is a categorical ordinal variable.

When you receive your data, the first question you ask is whether these are numerical or categorical. If they are categorical (or labels), then you ask if they can be put in any order. If no, then they are categorical nominal. If yes, they are categorical ordinal.

If the data are numerical and not categorical, you ask if the numbers come from counting or measuring. If by counting, they are discrete data. If by measuring, they are continuous data. Figure 4.2.3 visually represents the algorithm of questions to ask.

Figure 4.2.3 An Algorithm to Help Identify Data Types of a Variable **(Bowers, D 2008)**



Once we have determined the type of data that we have, we can use this to determine that type of analysis that we can use for the data. For quantitative data analysis we use statistics to determine the relationships between the data and how this describes the population that are being analyzed. There are two main forms of statistics; descriptive statistics and inferential statistics. Below we will describe descriptive statistics. Workbook 3 will review inferential statistics in more detail.

Learning Activity 4.2.3

Differentiating Between Different Types of Variables

Directions: For the table below, identify each of the variables as either nominal, ordinal, continuous, or discrete.

Variable	Values
1. Height (cm)	158, 169.3, 170, 200.6
2. Weight (kg)	10.2, 50, 69.4, 84
3. Outcome of disease	Recovery, chronic illness, death
4. Age category	0-5, 4-25, 26-50, 51+
5. Parity *(# of children)*	0, 1, 6, 8, 10
6. Multiple pregnancy	Yes, no
7. Marital status	Single, married, widowed, separated, cohabiting
8. Stage of disease	I, II, III, IV
9. Hemoglobin (g/dl)	8.9, 14.2, 12.7
10. Number of AIDS cases	278, 301, 313, 350

Discussion 4.2.3

Differentiating Between Different Types of Variables

1. *Height is presented as a continuous data.*
2. *Weight is presented as a continuous data.*
3. *Outcome of disease is presented in a nominal category.*
4. *Age is a continuous data, but here it is presented in an ordinal category.*
5. *Parity is presented in an ordinal category.*
6. *Multiple pregnancies is presented in a nominal category.*
7. *Marital status is presented in a nominal category.*
8. *Stage of disease is presented in an ordinal category.*
9. *Hemoglobin is presented in as a continuous data.*
10. *Number of AIDS cases are presented in discrete measures.*

4.3. Descriptive Statistics and Graphics

The simplest form of statistics is descriptive statistics. You use descriptive statistics to summarize your data set. Descriptive statistics helps you to display this data and shows it in a simple way. Though you may lose some of the detail when you summarize data, you will still be able to perform some powerful comparisons across people or questions and therefore that data will be more useful than in its raw form. First you need to determine the type of data that you have. Does your data come from a survey with many questions; from only a few questions but from many people; or many questions of many people? Determining what type of data you have will let you theorize what the data should look like. This is important when analyzing the distribution of the data.

4.3.1 Distribution

Once you have identified your data type, you need to determine the distribution of your data. The distribution summarizes the number of times a response occurs. The *distribution* of a variable tells us what values the variable takes and how often it takes these values. For example, the *distribution* of a *categorical* variable lists the categories and gives either the count or percent of individuals that fall in each category.

To find out the distribution, you can begin by considering only one question at a time and listing all the possible responses and a count of how many people gave that response. For example, one question might be whether people are "male" or "female". You would then simply count how many were male and how many were female (If there are 124 males and 76 females, you can start by listing 124 or 62% of males and 76 or 38% for females).

Sometimes you may have to regroup (or re-categorize) the data after you see the distribution. This occurs when there are too many responses categories with relatively few responses in each category. This happens most often with numerical data. For example, if you look at the age of the people in your district you could consider each year of age as a separate unit. However, this will result in many units and some of the ages may have very few or no people. Instead, you can decide to create categories of ages by grouping ages. Keep in mind that these categories must be meaningful.

For example, in health care fields, the first 5 years of life, pre-pubescent, pubescent, reproductive age, and post-reproductive age are some important categories of age. Therefore you would create the following age categories:

- 0-5= infant/child,
- 6-11=pre-pubescent,
- 12-16=pubescent,
- 16-45=reproductive age,
- and 45+= post-reproductive age.

However, these age categories would not be as useful is you were considering HIV infection and you wanted to determine the testing rates for people in their 20's and 30's. In this case you may change the categories to:

- 0-12=child,
- 13-19=teenager,
- 20-27=youth,
- 27-34=young adult,
- 35-45=adult,
- 45-60= older adult, and
- 60+=senior.

After deciding categories that apply best to the analysis that you would like to do you would then count how many people belonged in each age category.

4.3.2 Frequency

Once you have created the categories that are meaningful in your data the first thing you do is to create a frequency distribution.

Frequency Count

A frequency count is a measure of the number of times that an event occurs. For example, we can count the number of STI patients who receive HIV counseling and testing at various clinics. A frequency count is useful for identifying any data that is missing and any extreme values. In fact, it will show which category has the highest concentration in terms of distribution, and does not indicate an error. *Extreme values* are those that are much higher or lower than the other values seen in the data set. Extreme values may be due to an error in data collection, recording or reporting. It may also be a genuine result. For example, if one clinic had no STI patients who came for HIV counseling, while other clinics in the area had an average of 150 patients, you would have to consider if this is a mistake in how this was noted either in the clinic or in the dataset. If this is not a mistake, then we would have to understand why no STI patients came for counseling – Was the counselor away on holiday? Were there no STI patients at the clinic during that time period? Are STI patients not being referred for HIV testing?

Relative Frequency

Relative frequency is the fraction of times a variable takes a specific value in your data. This will usually be represented by a percentage. Using the example above a relative frequency answers the question "Out of all the individuals who received HIV counseling and testing in a given month, what percentage were STI patients?" In this case the variable would be STI status and the value is "received counseling and testing".

To calculate relative frequency, you need the frequency count for the total population and the frequency count for the subgroup of the population you are considering. The relative frequency for the subgroup is then:

Relative frequency = Subgroup count / Total count x 100%

For the STI patients mentioned above you might have a table that looks at various STIs and their frequency.

Figure 4.3.2a STIs and their Frequencies

Status	Number of individuals	Relative Frequency
No STIs	54	37%
Herpes	43	29%
Syphilis	26	18%
Gonorrhea	23	17%
Total	146	100%

In this example you can see the frequency of each STI or no STIs to the total. You could also sum all of the counts of STIs and create a relative frequency of "No STIs" vs. "STIs". In this example "No STIs" would be 54/146 or 0.37. The relative frequency of STI patients who have been through counseling and testing would therefore be 92/146 or 0.63.

As another example let us say you have data on use of birth control methods by a sample of 290 individuals as follows:

Figure 4.3.2b

Method	Number of individuals
Abstinence	9
Oral contraceptive	93
Depo-provera	26
Loop	49
Spermicide	20
Condoms	75
Vasectomy	9
Hysterectomy	6
Norplant	3
Total	290

Presenting data as a relative frequency, or a proportion, can help readers to interpret your data more easily because they are on a standardized scale from 0 – 1. This can help make it easier to compare data.

For example, imagine that in addition to the birth control method data you have above, you have a similar data set that shows that 53 out of 170 use condoms. How does this data compare with the 75 out of 290 from our data above? Just using frequencies, it can be difficult to interpret because the total number of individuals is different, but with relative frequencies, it is easy to compare. In the second dataset 53/170 corresponds to a proportion of .31 who use condoms as compared to the 75/290 or 0.26 in the first dataset.

Sometimes, when a proportion is very small, it is useful to present it as a *rate.* For example, if our population size was 80,000, and our subpopulation of condom users is 75, then the proportion of condom users would be 0.0009. This number is more usefully expressed as a rate, that is, by multiplying it by 10,000 to get a rate of 9 per 10,000 contraceptive users.

The frequency of each variable can be summarized in a *frequency distribution* table or as a graph. In this next section we'll discuss how to present categorical data with tables and graphs. Later on in this chapter we'll discuss how to present numerical data with tables and graphs.

4.3.3 Presenting Frequencies of Categorical Data in Tabular and Graphical Form

Frequency Table

When a table is used to show frequency counts for a categorical variable, it is called a *frequency table*. There are several ways you can choose to present data in a frequency table. One way you might choose to present the data that you received on contraceptive use is with a *one-way frequency table* as in Figure 4.3.3.a below.

Figure 4.3.3.a One-Way Frequency Table

Birth control method	Frequency	Percent	Cumulative Percent
Abstinence	9	3	3
Oral contraceptive	93	32	35
Depo-Provera	26	9	44
Loop	49	17	61
Spermicides	20	7	68
Condoms	75	26	94
Vasectomy	9	3	97
Hysterectomy	6	2	99
Norplant	3	1	100
Total	290	100	100

Note that figure 4.3.3a also includes a cumulative percent which can be used to indicate frequencies of aggregate values. The cumulative percent is calculated by adding each percent with the previous cumulative percent. For example, the percentage for abstinence is 3% while the percentage for oral contraceptives is 32%. Add 32 + 3 which equals to 35% to obtain the cumulative percent of abstinence and oral contraceptive. Similarly, 44% is obtained by adding 9+35. Cumulative percentages are useful in calculating the middle point of your data (median score) if the items are place in ascending or descending order of frequency.

Bar Graphs

The same information about frequency of birth control methods can also be presented graphically, using a bar graph. Note that these graphs are created using either the frequency counts (Figure 4.3.3.b) or the percentage (Figure 4.3.3.c) of the total sample.

Figure 4.3.3.b Bar graph using frequency counts

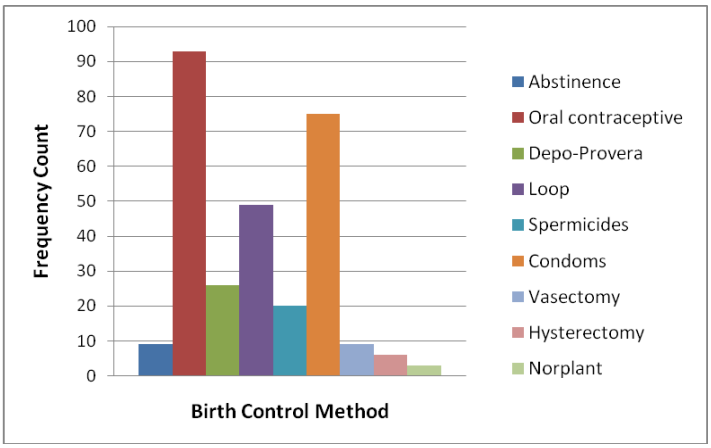
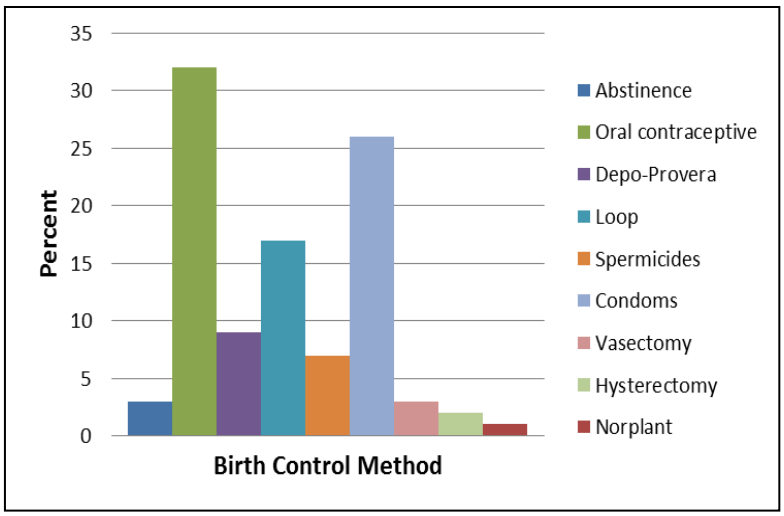


Figure 4.3.3.c Bar Graph for Birth control Method using Percentage



Histograms are used in a similar way for continuous data. Bar graphs, which use separated columns or columns that are differentiated in colours, are used for categorical data (such as birth control methods) while histograms are used to graph continuous data such as height, haemoglobin levels, etc.

Two Way Table

Sometimes called a *contingency table*, a two way table is a useful tool for examining relationships between two categorical variables. Similar to one way tables, the entries in the cells of a two-way table can be frequency counts or relative frequencies.

Let us consider an example of data collected on results of pulmonary tuberculosis (PTB) among those with HIV and those without HIV. This type of data can be presented in a two way table as shown below in Figure. 4.3.3.d.

Figure 4.3.3.d A two way table showing the relationship of PTB test results and HIV test results using a row total and percent.

	PTB status		
HIV status	Positive	Negative	**Total**
Positive	875 (57.8%)	639 (42.2%)	1514 (100.0%)
Negative	57 (8.9%)	594 (91.1%)	651 (100.0%)
Total	932 (43.0%)	1233 (57.0%)	2165 (100.0%)

The above two-way table in Figure 4.3.3.d shows PTB and HIV status of 2165 individuals. Because entries in the table are frequency counts, the table is a frequency table.

Entries in the "Total" row and "Total" column (highlighted below) are called *relative frequencies*. It simply tells you, how many people have a condition. In our example below, the row totals (1514 and 651) tell you how many in the population have HIV and how many do not have HIV. The column totals (932 and 1233) on the other hand tell you how many in the population have PTB and how many do not have PTB.

Figure 4.3.3e Relative frequencies of HIV and PTB

	PTB status		
HIV	Positive	Negative	**Total**
Positive	875 (58%)	639 (42%)	1514 (100%)
Negative	57 (9%)	594 (91%)	651 (100%)
Total	932 (43%)	1233 (57%)	2165 (100%)

The relative frequencies in the body of the table are called *conditional frequencies* or the *conditional distribution*. Conditional frequencies refer to the number of people categorized by the presence or absence of two conditions. In our example below, we can see how many people with HIV had PTB and how many did not have PTB.

Figure 4.3.3.f Conditional Frequencies of HIV and PTB

	PTB status		
HIV	Positive	Negative	**Total**
Positive	875 (58%)	639 (42%)	1514 (100%)
Negative	57 (9%)	594 (91%)	651 (100%)
Total	932 (43%)	1233 (57%)	2165 (100%)

This might make us conclude that HIV positive people also testing positive for PTB (58%) had a roughly close status to people with HIV positive people testing negative for TB (42%). Yet, the conditional frequencies show that people with HIV have a higher proportion of PTB (58%) when compared to people without HIV (9%). This means that people with HIV have a higher burden of TB than people without HIV.

In two-way tables, you can show relative frequencies for the whole table, for rows, or for columns. The above table shows relative frequencies for rows. It can also be presented to show relative frequencies for columns, as shown below. In the table above each group (for example HIV positive and PTB positive) is shown a frequency of the total number of people. In the table below the frequency of each HIV status within each PTB status is shown. Therefore each PTB status has a total of 100%, of which some percentage will be HIV positive and some percentage HIV negative.

Figure 4.3.3.g A two way table showing the relationship of PTB test results and HIV test results using a column total.

	PTB status		
HIV	Positive	Negative	**Total**
Positive	875 (93.9%)	639 (51.8%)	1514 (69.9%)
Negative	57 (6.1%)	594 (48.2%)	651 (30.1%)
Total	932 (100.0%)	1233 (100.0%)	2165 (100.0%)

The decision to choose to show frequencies for rows or columns depends on the information you are presenting. In the case of tabulating relationships between HIV and PTB test results, whether you choose to show relative frequencies for columns of rows might depend on whether you are presenting data from an antiretroviral therapy (ART) clinic or a TB clinic. If you are presenting data from an ART clinic and are describing the PTB status of the HIV patients, you would want to choose to show frequencies for rows.

The relationship between variables is often easier to detect when they are displayed graphically in a bar chart. For example, it can be presented graphically using frequency counts or percentages. Figure 4.3.3.h-1, demonstrates that HIV+ persons are more likely to be PTB+. Similarly, Figure 4.3.3h-2 Percentage of PTB status of Individuals by HIV status.

Figure 4.3.3.h-1 Frequency Counts of HIV status for PTB+ and PTB - patients

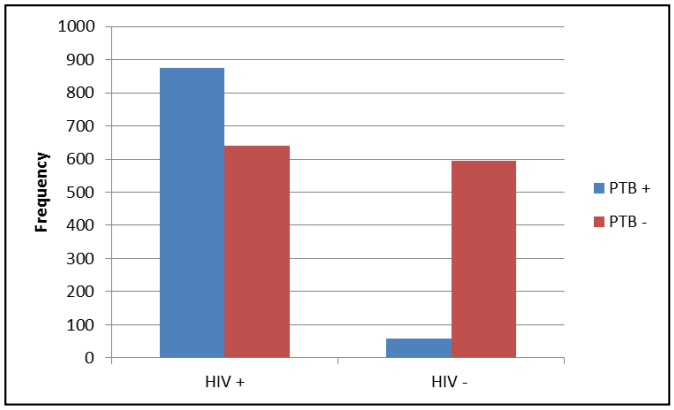
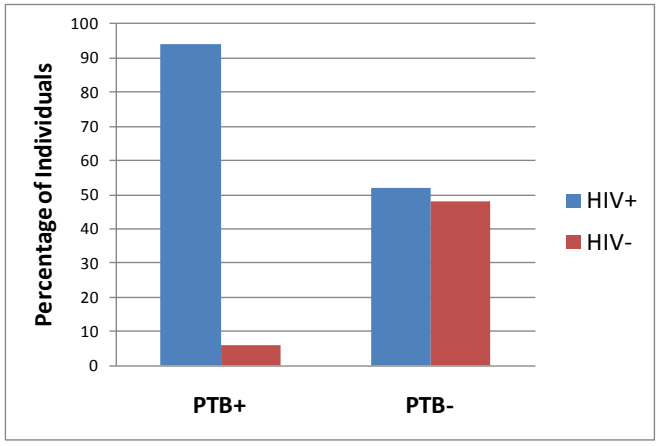
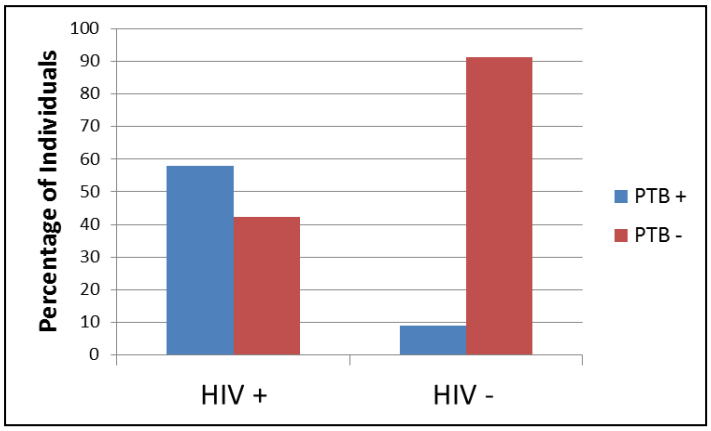


Figure 4.3.3.h-2 Percentage of HIV+ and HIV- individuals by PTB status



Inverting the relationship (with column frequencies) can also be presented graphically as shown in Figure 4.3.3.h-3.

Figure 4.3.3.h-3 Percentage of PTB+ and PTB- individuals by HIV status



What differs between these last two graphs (4.3.3h-2 and 4.3.3.h-3), is the relationship that they specify. While the graph based on the first table shows the proportion of HIV-infected and non-infected patients with and without TB, the second table shows the proportion of TB and non-TB patients with and without HIV.

Depending on who your audience is would dictate which graph (or relationship) that you would be most relevant to present. For example, if you were presenting the ART programme, you would likely want to present the graph based on Table 4.10, which focuses on HIV status. If you were presenting to the TB programme, you may want to focus on the second graph which focuses on TB status (Figure 4.3.3.h-3).

While two-way tables always have two categorical variables (as the name implies) these variables can have more than two labels. For example, one could have a two-way table that presents HIV infection and the various TB treatment outcomes, as shown below in Figure 4.3.3.i.

Figure 4.3.3.i HIV status by TB Treatment Outcome

HIV status	**TB Treatment Outcomes**						
	Cured	Completed	Failure	Died	Defaulted	Transferred Out	**Total**
Positive	66	18	3	8	2	3	875
Negative	43	4	1	3	2	4	57
Total	109	22	4	11	4	7	932

While there is no strict rules about how many labels the two variables can have in a two way table, a total of around 8 labels between the two variables (i.e. 4x4, 2x6, 3x5) would be the maximum you would want to present in this way.

Excel Tutorial 1

In this section, we will use the data presented earlier in this chapter to show you how to create bar charts for one-way and two-way data tables. Please refer to this when you have access to a computer equipped with Excel.

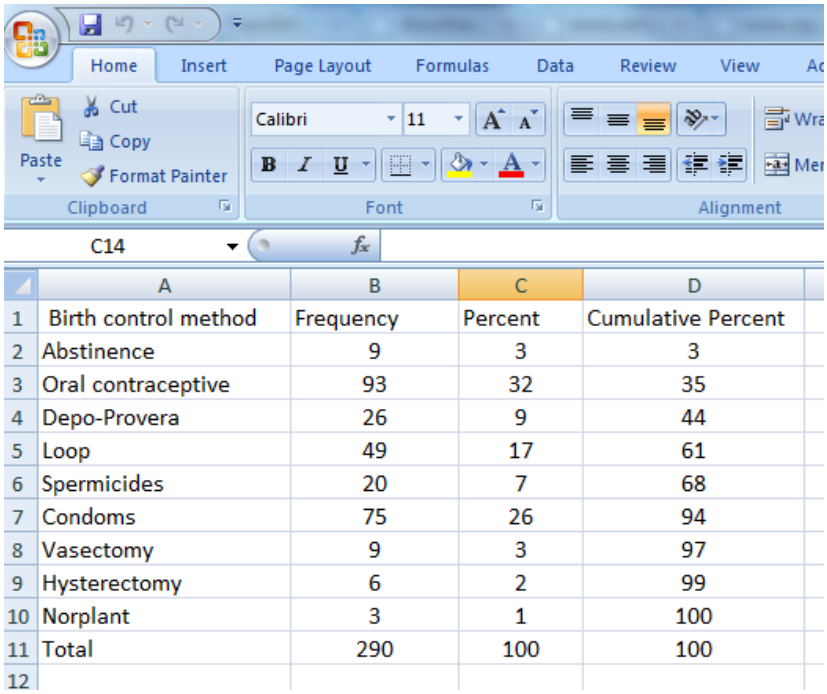
Creating Bar Charts for One-Way Frequency & Relative Frequency Tables

For this example, we will be using the birth control data set from the section on one-way frequency tables above.

Birth control method	Frequency	Percent	Cumulative Percent
Abstinence	9	3	3
Oral contraceptive	93	32	35
Depo-Provera	26	9	44
Loop	49	17	61
Spermicides	20	7	68
Condoms	75	26	94
Vasectomy	9	3	97
Hysterectomy	6	2	99
Norplant	3	1	100
Total	290	100	100

Part I: Enter the Data

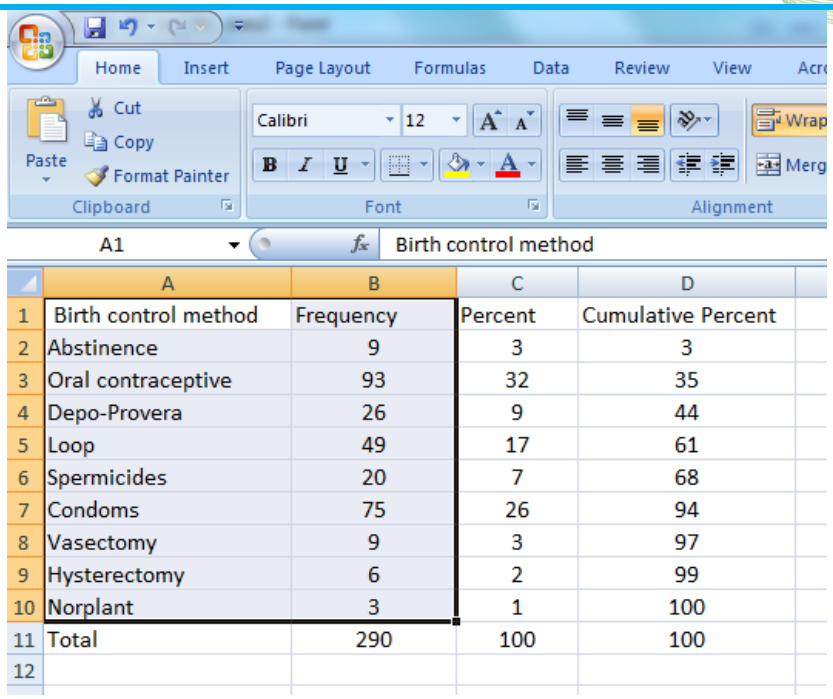
1. Open Excel and enter the data in the table into a worksheet.



	A	B	C	D
1	Birth control method	Frequency	Percent	Cumulative Percent
2	Abstinence	9	3	3
3	Oral contraceptive	93	32	35
4	Depo-Provera	26	9	44
5	Loop	49	17	61
6	Spermicides	20	7	68
7	Condoms	75	26	94
8	Vasectomy	9	3	97
9	Hysterectomy	6	2	99
10	Norplant	3	1	100
11	Total	290	100	100
12				

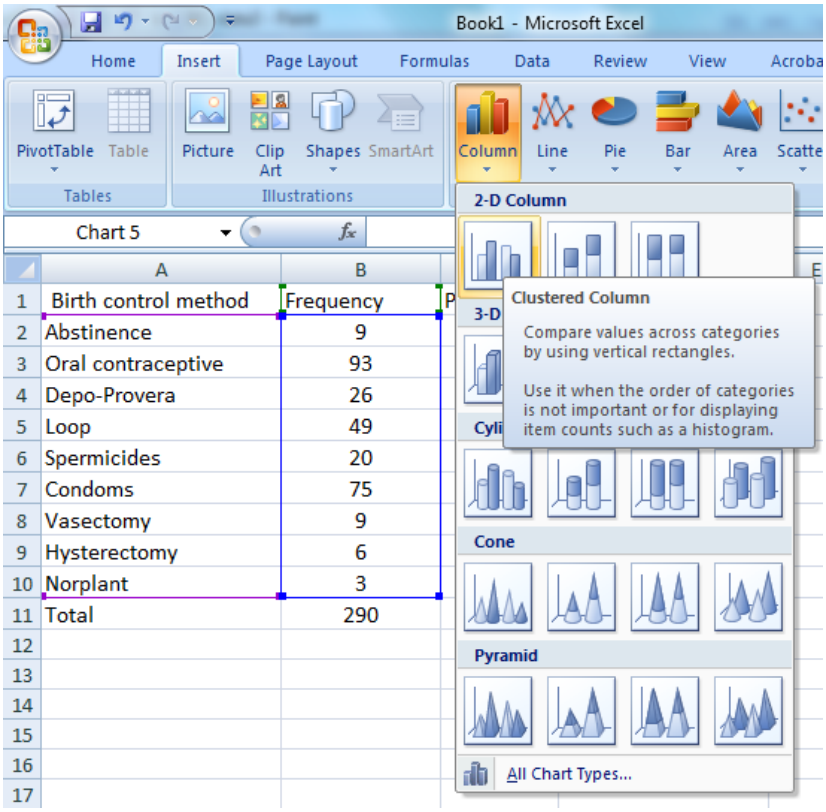
Part II: Create a Bar Graph for Frequency

1. Drag your cursor to highlight the first two columns only. Do not highlight the row of totals.

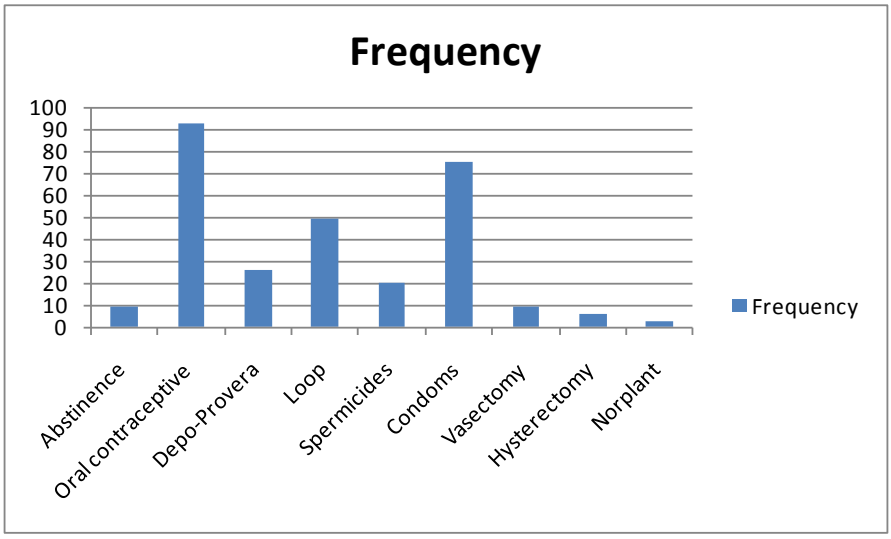


	A	B	C	D
1	Birth control method	Frequency	Percent	Cumulative Percent
2	Abstinence	9	3	3
3	Oral contraceptive	93	32	35
4	Depo-Provera	26	9	44
5	Loop	49	17	61
6	Spermicides	20	7	68
7	Condoms	75	26	94
8	Vasectomy	9	3	97
9	Hysterectomy	6	2	99
10	Norplant	3	1	100
11	Total	290	100	100
12				

2. Click on the **Insert** tab and then on the **Column** button. Choose the first Graph under the 2-D Column.

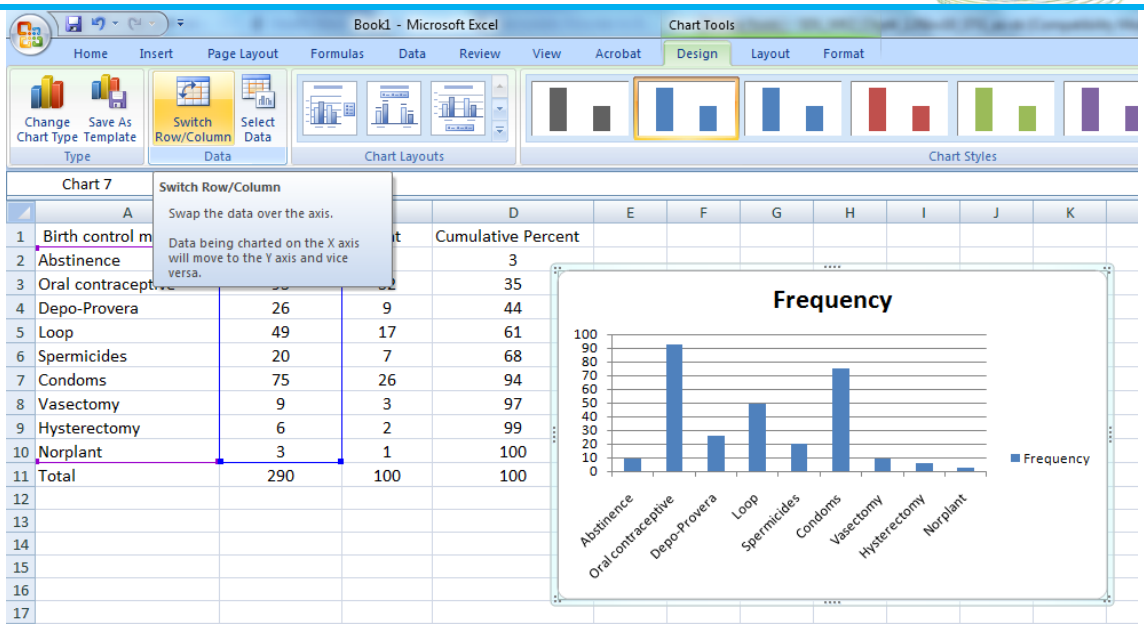


3. A graph will appear that looks like the one below. Congratulations you have made a graph of a one-way frequency table!



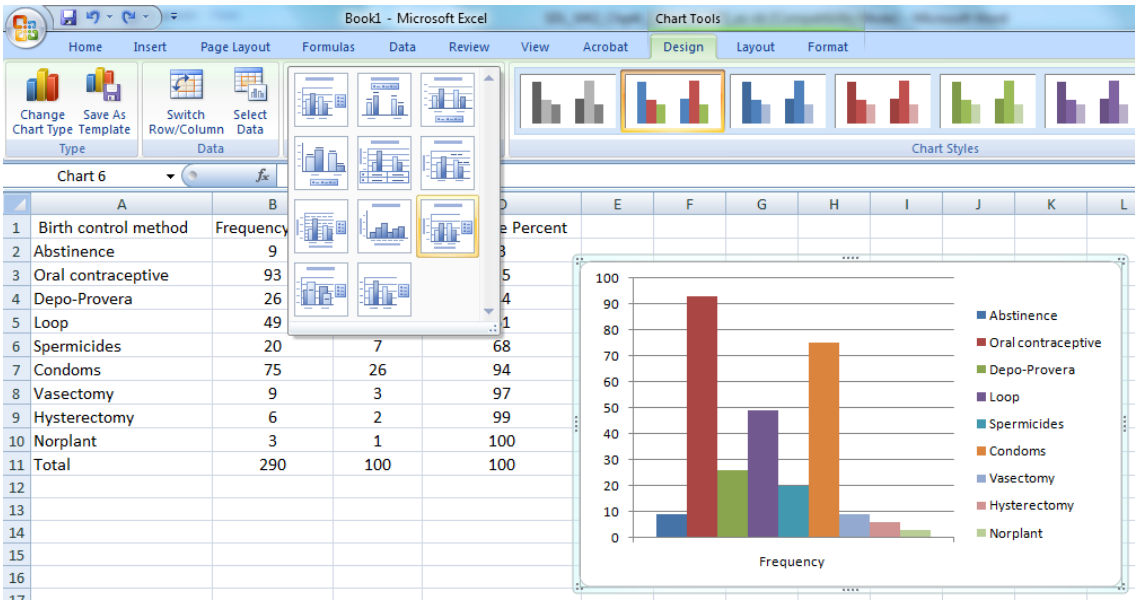
4. The final step is to format your graph. There are many options for formatting under the **Design** tab – **Chart Layouts** button. Choose from these to present the data the way you would like. To create a graph exactly like the one presented earlier in the chapter use the following steps:

    a. Click on the graph, then click on the **Design** tab of the **Chart Tools**. Click **Switch Row/Column**.



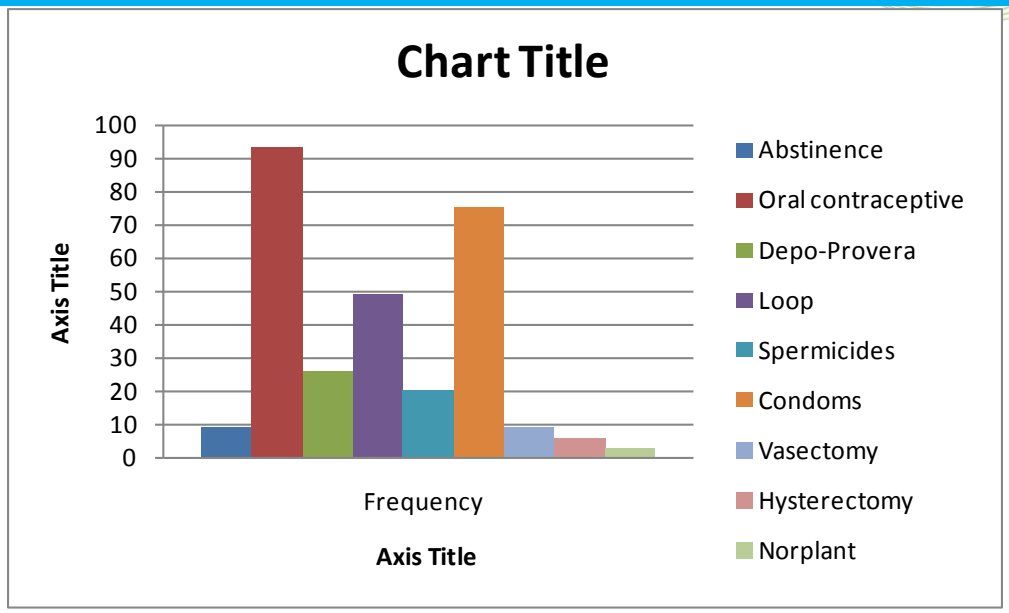
	A	B	C	D
1	Birth control m		t	Cumulative Percent
2	Abstinence			3
3	Oral contracept			35
4	Depo-Provera	26	9	44
5	Loop	49	17	61
6	Spermicides	20	7	68
7	Condoms	75	26	94
8	Vasectomy	9	3	97
9	Hysterectomy	6	2	99
10	Norplant	3	1	100
11	Total	290	100	100

b. A new graph will appear. Click on the **Chart Layout** button and choose **Layout 9**.

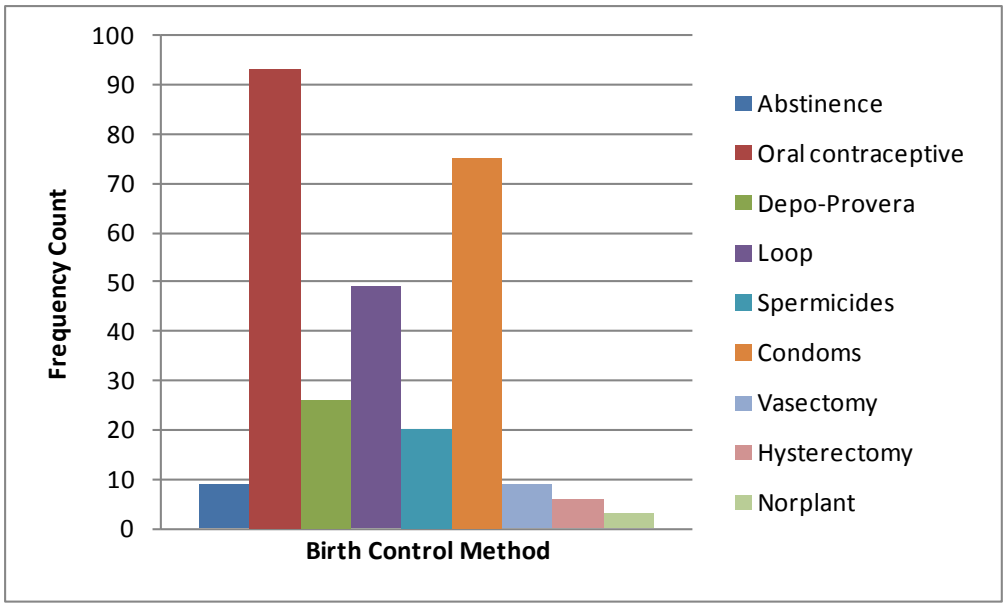


	A	B	C	D
1	Birth control method	Frequency		e Percent
2	Abstinence	9		3
3	Oral contraceptive	93		5
4	Depo-Provera	26		4
5	Loop	49		1
6	Spermicides	20	7	68
7	Condoms	75	26	94
8	Vasectomy	9	3	97
9	Hysterectomy	6	2	99
10	Norplant	3	1	100
11	Total	290	100	100

c. Once you have the correct layout, you can click on the axis title and chart title to add the correct headers.



d. Change the title on the x-axis (bottom axis) to read "Birth Control Method" and the y-axis (side axis) to read "Frequency Count". Delete the chart title and the box containing the word "Frequency" on the x-axis.

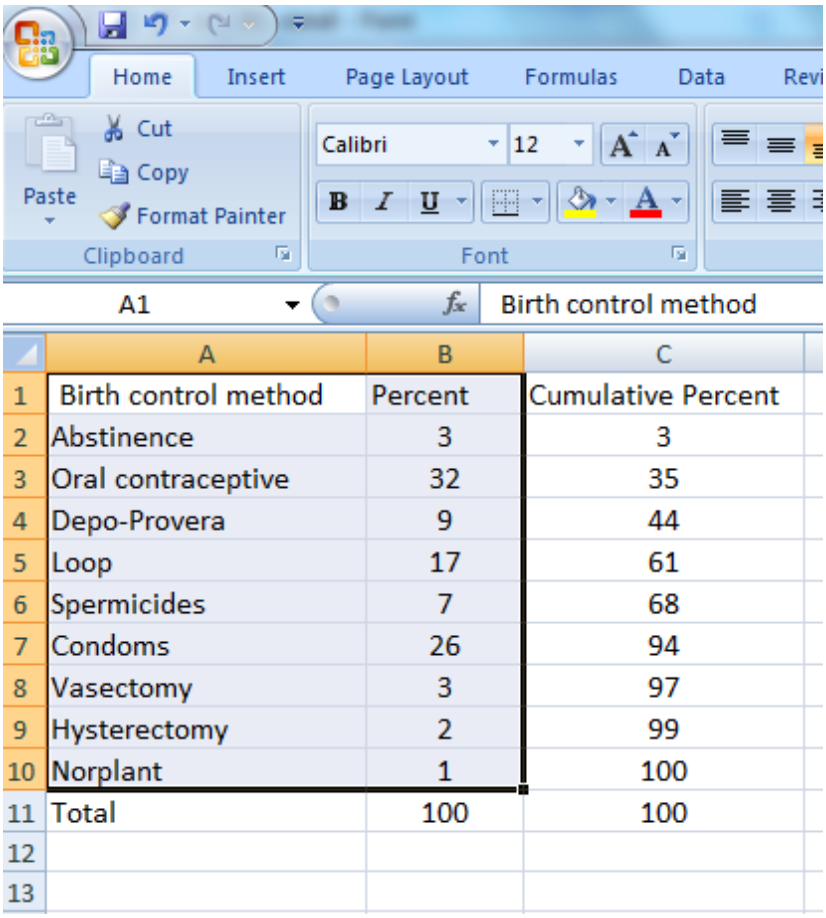


e. Congratulations! You have created a graph exactly the same as the one presented in the workbook.

Part III: Create a Bar Graph for Percent

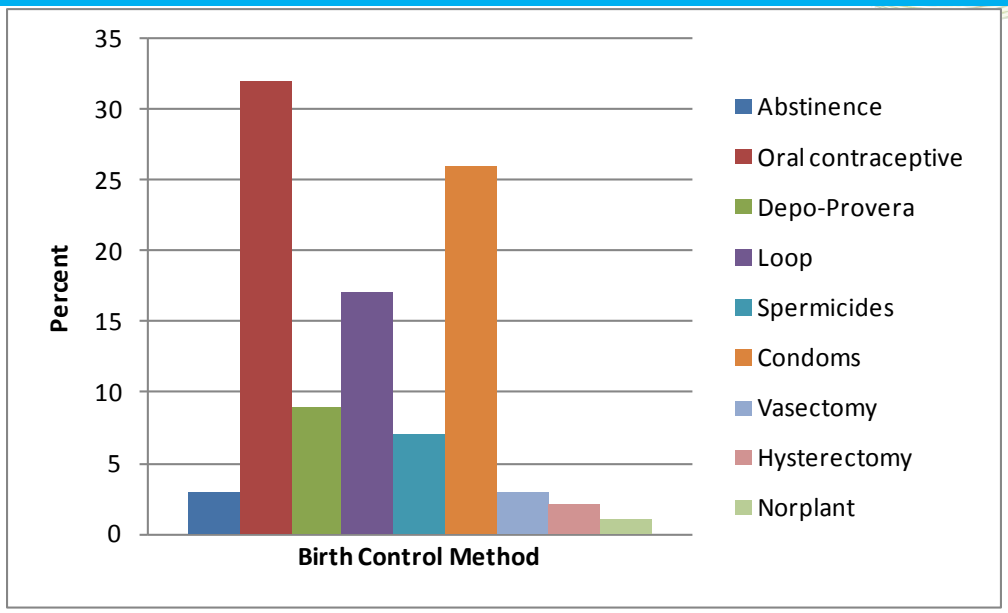
The directions for creating a bar graph for percent are very similar to those for creating one for frequency.

1. Rather than selecting the frequency column, select the percent column. You can do this by cutting and pasting the columns into a new worksheet and deleting the frequency column.



	A	B	C
1	Birth control method	Percent	Cumulative Percent
2	Abstinence	3	3
3	Oral contraceptive	32	35
4	Depo-Provera	9	44
5	Loop	17	61
6	Spermicides	7	68
7	Condoms	26	94
8	Vasectomy	3	97
9	Hysterectomy	2	99
10	Norplant	1	100
11	Total	100	100
12			
13			

2. Follow Part II steps 2-4 above to create your graph, changing the word "Frequency" in the graph to "Percent."



Bar Charts for Two-Way Frequency & Relative Frequency Tables

The instructions for creating a bar chart to for a two-way frequency and relative frequency tables are very similar to those above. The decision about whether to create a two-way frequency table versus a relative frequency table is based on the question you are trying to answer. For example, using the example of a two-way frequency table for HIV and pulmonary tuberculosis (PTB), do we want to know:

- Do more people with HIV have TB? or
- Do more people with TB have HIV?

Let us return to our two-way frequency table for HIV and pulmonary tuberculosis (PTB).

HIV	PTB status		
	Positive	Negative	Total
Positive	875 (57.8%)	639 (42.2%)	1514 (100.0%)
Negative	57 (8.9%)	594 (91.1%)	651 (100.0%)
Total	932 (43.0%)	1233 (57.0%)	2165 (100.0%)

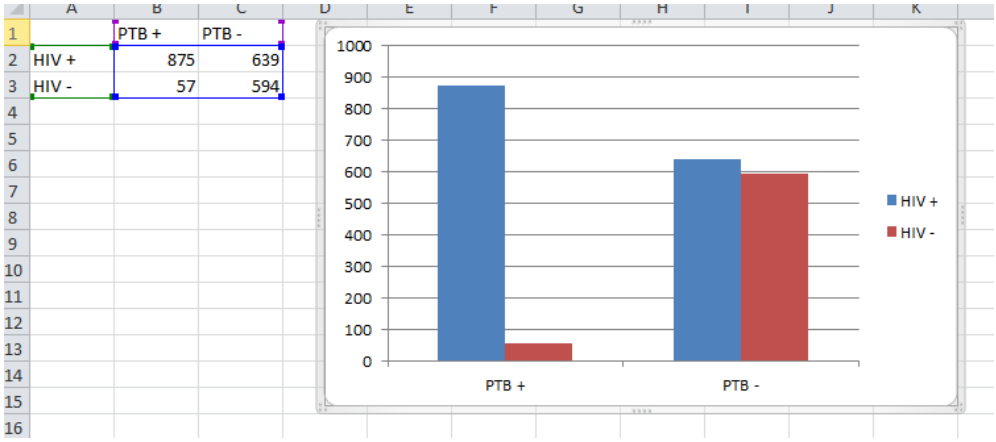
Part I: Enter the Data

1. Open Excel and enter the data in the table, including the column and row headings, into a worksheet.

	A	B	C
1		PTB +	PTB -
2	HIV +	875	639
3	HIV -	57	594
4			

Part II: Create a Bar Graph for Frequency

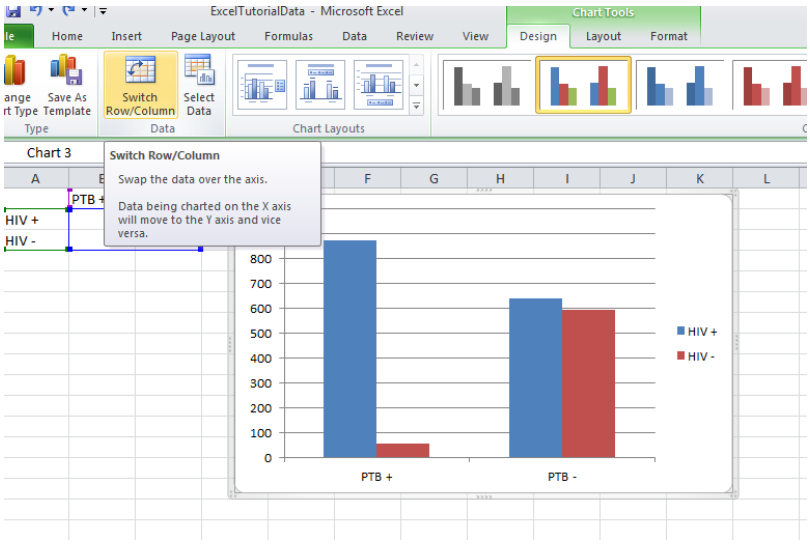
1. Drag your cursor to highlight the data, including the headings.
2. Click on the **Insert** tab and then on the **Column** button. Choose the first Graph under the 2-D Column. Excel will generate a bar chart that summarizes the information on both variables.



This bar chart shows us that the frequency of HIV(+) individuals is higher among PTB(+) individuals than among PTB(-) individuals, summarizing the data with the column totals as the table above.

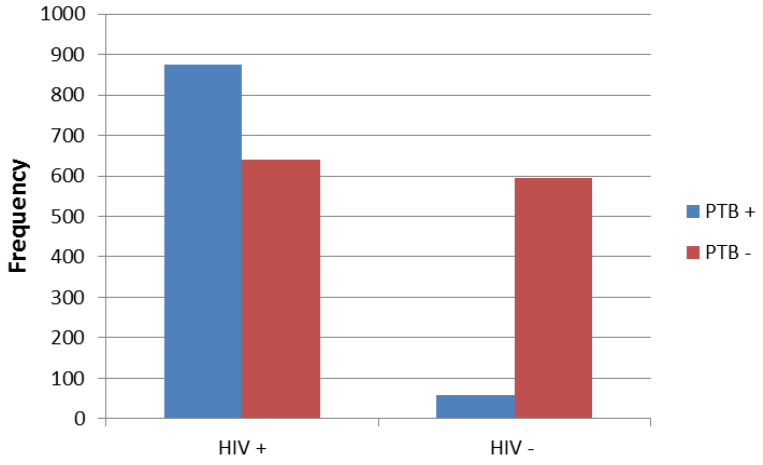
To change the graph so that it shows that the frequency of PTB(+) individuals is higher among HIV(+) individuals than among HIV(-) individuals, we'll need to take one more quick and easy step.

3. Click on the graph, then click on the **Design** tab of the **Chart Tools**. Click **Switch Row/Column**.



This approach gives us the plot that we want.

Notice how it is instantly apparent in the table below that the relative frequency of PTB(+) patients is much higher in HIV(+) individuals than in HIV(-) individuals.

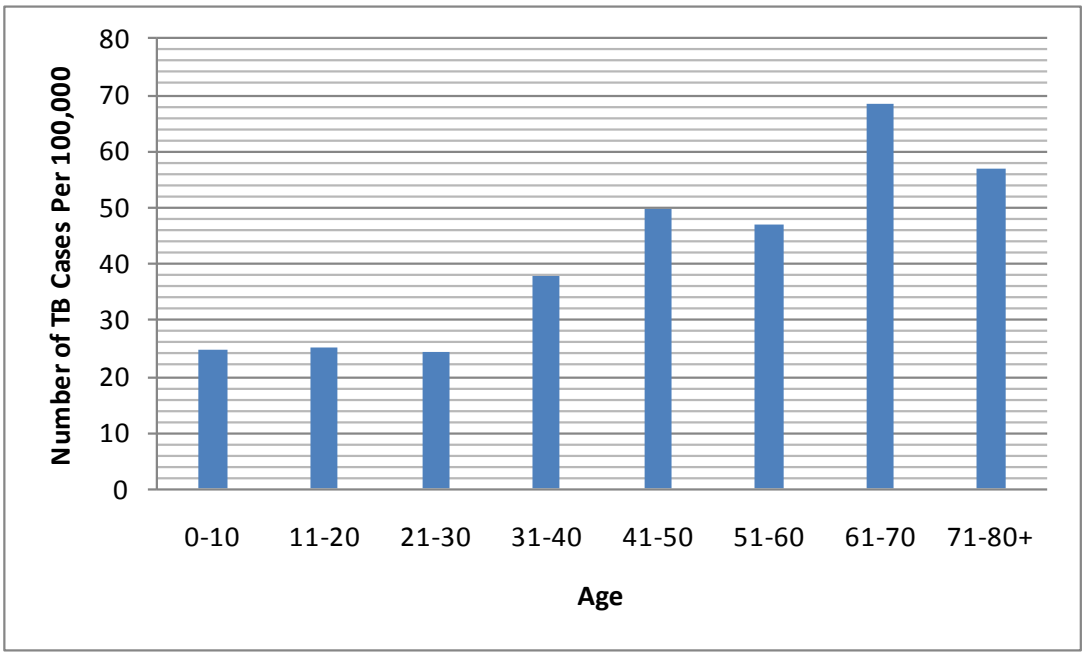


4.4. Shape

When you generate a frequency, you can see how your data is distributed. You can assess how many times something is observed; what are the observations most seen, and the halfway point of the observations. Another way to look at how the data is distributed is by looking at its shape, how many peaks it has, and whether it is symmetrical around a certain point.

The shape shows us if our data is thickly distributed, with all of the data points stacked closely together or thinly distributed, with all the data points spread out. Being aware of this can influence how we would categorize and present our data, as well as prompt us to ask the right questions. For example, seeing the shape of the data in the graph below might prompt us to ask the question: Why do older people test positive for TB more than younger people, because the peaks of the graphs are at 61-70 and 71-80?

Figure 4.4.a Number of TB Cases by Age

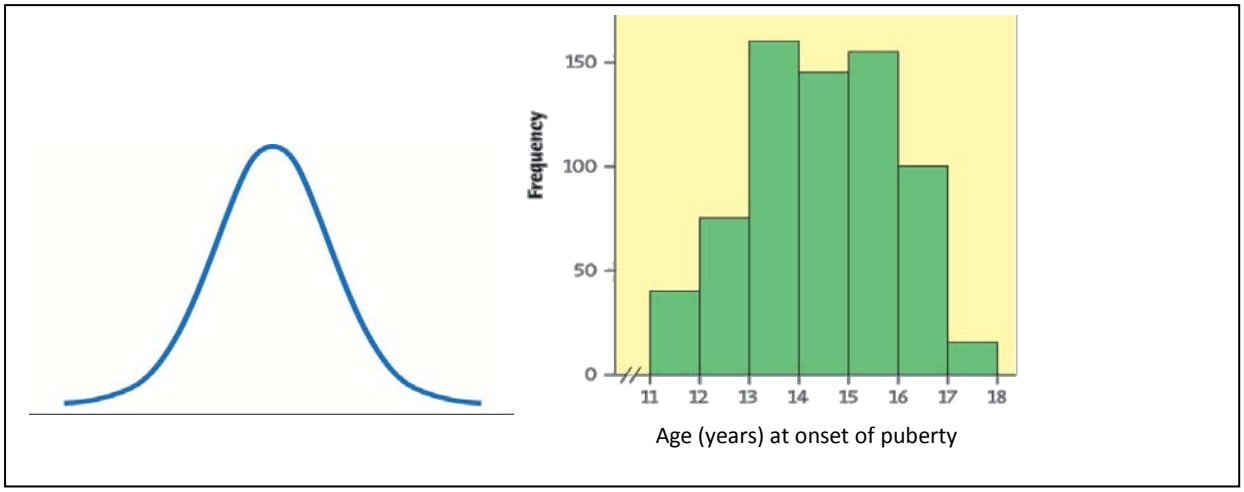


The *shape* of a distribution can be symmetric or skewed. Figures 4.4.b-4.4.e below show representations of symmetric, left-skewed, and right-skewed distributions. The left and right sides of the distribution are also called left and right *tails.*

In *symmetric distribution,* the right and left sides are approximately mirror images of each other, so that the values are more or less equally distributed from the middle of the sample. When our data has symmetric distribution, then the data is easily interpreted.

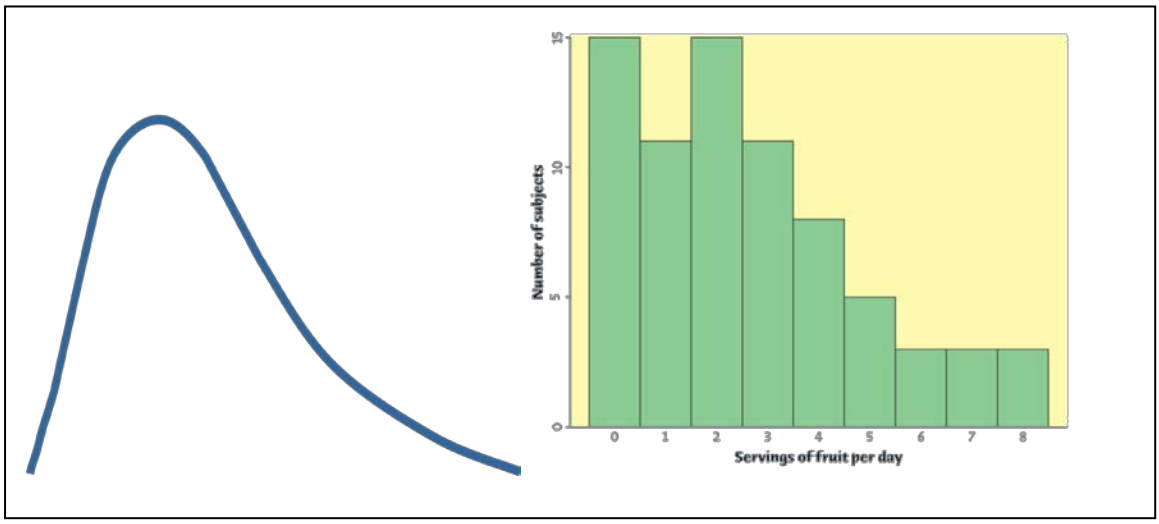
You can treat the extreme values on both sides of the middle equally, and, you can compare the distribution with other data sets. Symmetry is an important criteria for many statistical tests. You will learn about statistical tests in Chapter 4 of workbook 3.

Figure 4.4.b Symmetric Distribution **(Baldi 2009)**



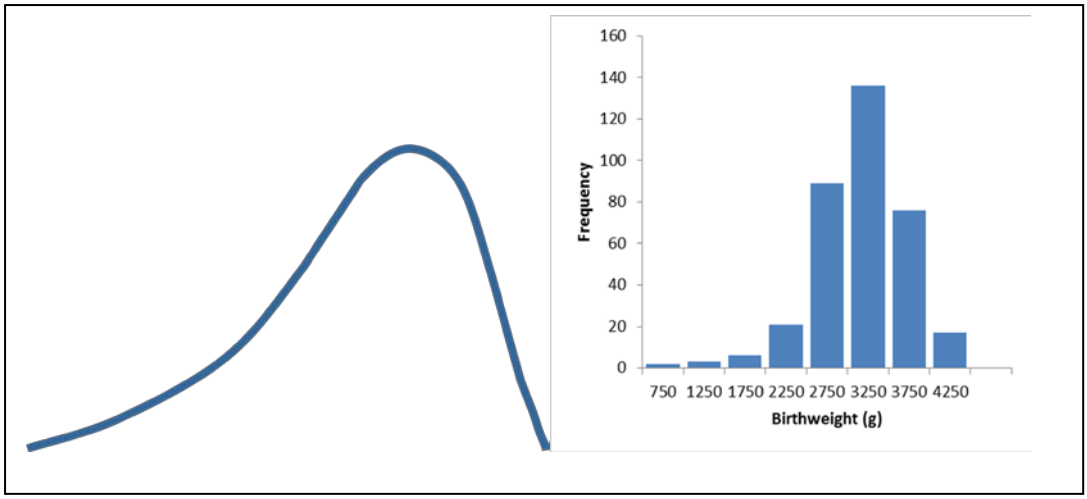
When a distribution is skewed, the data are asymmetric from the middle and are heavier on one side. When the right side extends much farther out than the left side, the data are said to be *right skewed* (or *positively skew).* This tells you that your data has a longer tail at the right and more subjects have the indicated on the left hand side of the x-axis. For example, in figure 4.4.c, more number of subjects have 0-2 servings of fruits per day.

Figure 4.4.c Right Skewed Distribution **(Baldi 2009)**



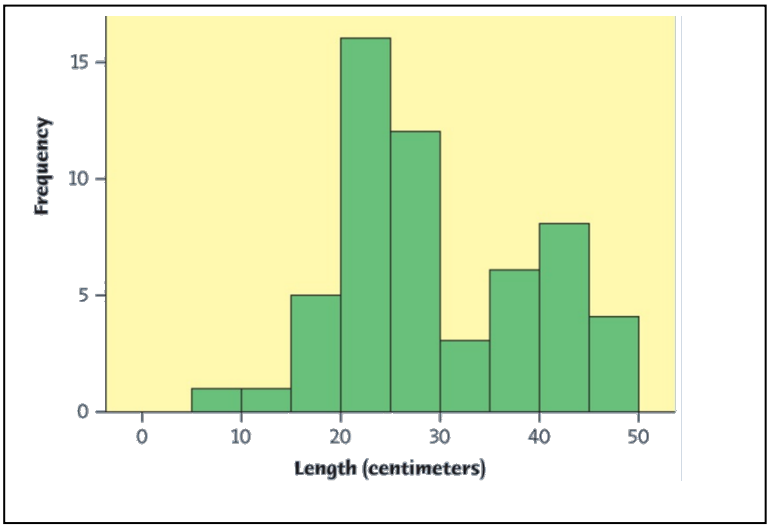
When a distribution is *left skewed (or negatively skew)*, the left side extends much farther out than the right side. As shown in figure 4.4.d, this means that more subjects have the value indicated in the right side of the x-axis (in this case, birth weights of 2750, 3250 and 3750 grams).

Figure 4.4.d Left Skewed Distribution **(Baldi 2009)**



Note that all distributions might not be simply described. For example, a graph can have a bimodal distribution if it has two peaks, called modes.

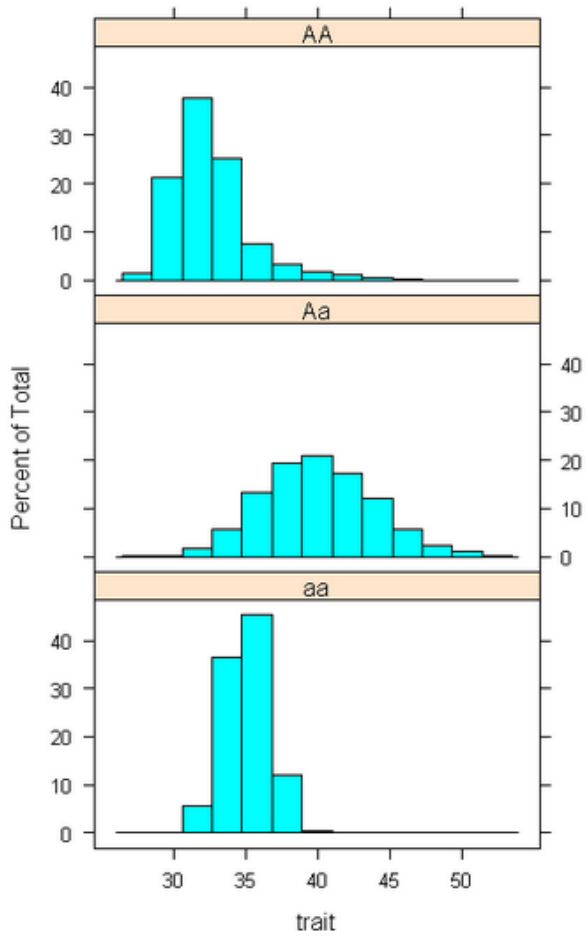
Figure 4.4e Bimodal Distribution **(Baldi 2009)**



Learning Activity 4.4

Describing Shape

Directions: Study the features of the 3 graphs below. What characteristics do you notice about the shape of these three histograms?



Source: (Getting Genetics Done: Software, Tips, & Productivity Hacks for Getting things Done in Genetics Research 2010)

Discussion 4.4

Describing Shape

All 3 histograms show unimodal data; in other words, they all have one peak.

The bottom two are approximately symmetric.

The top one is not symmetric; rather, it is right-skewed.

4.5. Mean, Mode, and Median

Another way of looking at the distribution of a given dataset is looking for the location of the middle or the centre of a distribution. We do this because numbers tend to cluster in the middle. If you think of the height of people in Botswana, there are some who are very tall, some who are very short but most are medium. This phenomena of data points clustering around a centre is known as central tendency. Mean, mode, and median can be used to measure central tendency.

4.5.1 Mean

Mean (commonly referred to as the 'average'): can be calculated by adding up all the numbers and dividing by the total number of data points.

The mathematical calculation for mean is as follows:

$$\text{Mean} = \mu = \Sigma X / N$$

Where:

Σ denotes sum of
X = all values

N = total number of data points

For example, you may have the ages of 12 patients presenting with hypertension (high blood pressure). You can calculate the average as follows in Figure 4.5.1.a.

Figure 4.5.1.a Mean

Patient #	Age
1	35
2	45
3	48
4	60
5	60
6	61
7	70
8	72
9	75
10	80
11	85
12	90
Sum=	781
Average =	781/12 = 65.1

4.5.2 Mode

Mode is the most frequently occurring number in a dataset. It is the variable with the most frequent counts in your data.

In the example above, the most common age is 60. Of the patients who presented with hypertension, two were aged 60. All other ages were only represented once. Thus, 60 is the most frequently occurring variable and is the mode.

4.5.3 Median

Median: can easily be remembered as the "middle" number. Half of the data points are below the median and half are above it. Determining the median in your data set is useful when there are outliers (values that are much less than or greater than the other numbers) in your data. You can then compare the median to the mean to determine if the outliers are significant. To find the median, sort the data in rank order (smallest to largest), then count in from one end until you find the middle. If the sample size is an even number, take the average of the two middle numbers.

In the case above, there is no clear middle. So, we will average the age between the sixth and seventh patient ([61+70]/2) to say that half the patients are below 65.5 years old. That is the median age is 65.5.

The mean, median, and mode are the same when the data has a symmetric distribution. Usually, you will be using averages like the mean. However, when your data has outliers, you should use the median to describe your sample. When your data is heavily skewed, it is better to describe your data using the mode as this is where the centre of your data will be contained.

Learning Activity 4.5.3

Calculating Mean, Median, and Mode

Directions: Based on the definitions presented in this chapter, answer the questions below.

1. Calculate the mean, median, and mode for the dataset on birth weights below.

Patient number	Birth weight (kg)
1001	3.4
1002	2.6
1003	3.0
1004	3.0
1005	3.6
1006	3.4
1007	3.0
1008	4.0
1009	6.2
1010	2.8

2. Which of these three is the best to use in describing your data and why?

Discussion 4.5.3

Calculating Mean, Median, and Mode

1. Mean (recall that the mean is the average of all the numbers) : 3.5

Median (recall that the median is the 'middle' number): 3.2

Mode (recall that the mode is the most frequently occurring number in the dataset): 3.0

2. The median is the best number to use to summarize the distribution because there is an outlier of 6.2.

4.6 Spread: Range, Variance, and Standard Deviation

How your data falls around the central tendency is called the *spread* or *dispersion* of the data. Looking at the spread of your data will assist you in identifying outliers in your dataset. Outliers are data points that are way out of the normal range. Outliers can exist due to a mistake or abnormally high or low values. In such cases, you might choose to include or exclude the outliers you have identified.

In order to study the spread of your data, you can start by asking: is the data clustered around a couple values or does it vary between very small and very large values? This spread can be measured as the range and variation from the mean. As an M&E officer, analyzing routine data, this can be very important. If you determine that a service, for example, should be provided to everyone but is clustered around a certain age group, or if the patient load at many facilities is similar but there is one facility that either sees many more patients or a lot fewer patients. The range, variance, and standard deviation can help you better understand this data.

4.6.1 Range

Range is the difference between the largest and smallest values in a given dataset. You can calculate the range of your dataset by subtracting the lowest score from the highest. In our example of patients with hypertension above, the range is 90-35 = 55. Note that the range can be highly affected by outliers. For example, if the lifespan of the average person is 80 and someone lives to be 110 they would be an outlier. Therefore, if some of the values in your dataset are really high or really low, the range may be a little misleading. In this case it might be better to use the standard deviation, which we will explain below.

4.6.2 Variance

Variance is defined as the average squared deviation from the mean. This simply means that we want to look at how far each data point is from the mean on average, how much they vary. We begin by calculating the difference between each data point and the mean. However, when the mean is subtracted from values below the mean, we get a negative number. So we square each difference between the number and the mean to avoid negative numbers. Let us assume that our mean age from the earlier example, of patients presenting with hypertension, is 65 years old. We then take each value and subtract 65.

Figure 4.6.2a-1 Variance

Deviation from the mean
35 – 65 = - 30
45 – 65 = - 20
48 – 65 = -17
60 – 65 = -5
60 – 65 = -5
61 – 65 = -4
70 – 65 = 5
72 – 65 = 7
75 – 65 = 10
80 – 65 = 15
85 – 65 = 20
90 – 65 = 25

Because we do not want some numbers to be negative numbers, we square this difference.

Figure 4.6.2a-2

(Deviation from the mean)^2	Is equal to
$(-30)^2$	900
$(-20)^2$	400
$(-17)^2$	289
$(-5)^2$	25
$(-5)^2$	25
$(-4)^2$	16
$(5)^2$	25
$(7)^2$	49
$(10)^2$	100
$(15)^2$	225
$(20)^2$	400
$(25)^2$	625

Remember, we want the *average* deviation of values from the mean, so we need to sum up the differences and divide them by the number of observations.

Figure 4.6.2a-2

(Deviation from the mean)^2	Is equal to
$(-30)^2$	900
$(-20)^2$	400
$(-17)^2$	289
$(-5)^2$	25
$(-5)^2$	25
$(-4)^2$	16
$(5)^2$	25
$(7)^2$	49
$(10)^2$	100
$(15)^2$	225
$(20)^2$	400
$(25)^2$	625
TOTAL	3079
Average	3079/12 = 256.6

The mathematical formula for calculating variance is as follows:

$$\sigma^2 = \Sigma (X_i - \mu)^2 / N$$

Where:

σ^2 = variance

Σ denotes the sum of

μ = mean

X_i = the i^{th} element of the given variable, refers to each data point

N = the number of elements in the data

Understanding the variance of the data set is important as it tells you what the expected response was (central tendency) and what was actually observed. The full importance of understanding this is explained in Chapter 4 of Workbook 3.

4.6.3 Standard Deviation

Standard deviation (SD) is a slightly more sophisticated way to provide an indication of the spread or variation in the data. Similar to variance, it measures how other values deviate from the mean. In calculation, it is the square root of the variance.

You might remember, we squared each difference of each value subtracted by the mean to get positive numbers that we can add together for variance? We are simply getting the square root to reverse this step which is necessary to calculate the sum of the squares to get a simpler number. Thus, the standard deviation of a data set is:

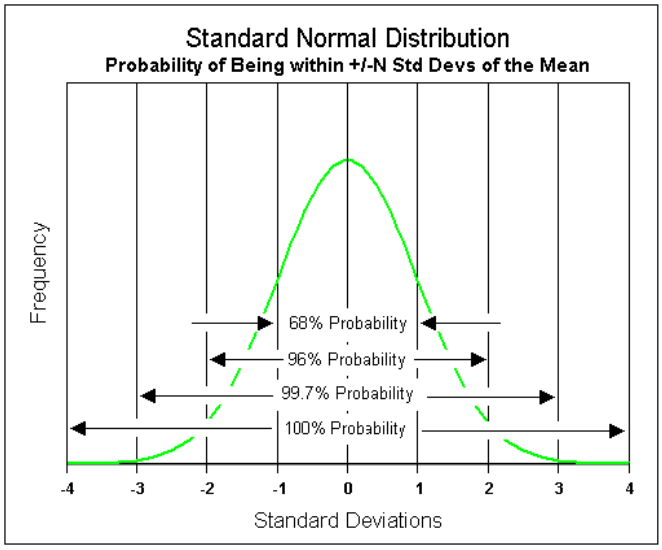
$$\sigma = \text{sqrt} [\sigma^2] = \text{sqrt} [\Sigma (Xi - \mu)^2 / N]$$

In our example, this is equal to sqrt 256.6 = 16.02.

If the distribution of our data is symmetric (also referred to a normally distributed) and does not contain large outliers, then we can make some assumptions about our dataset as follows:

- 68% of our data falls within + 1 and -1 stand deviation of the mean.
- 96% of our data falls within +2 and -2 stand deviation of the mean.
- 2% of our data falls within +3 and -3 stand deviation of the mean.

Figure 4.6.3 Standard Normal Distribution



In other words, most of our data lies within 2 standard deviations of the mean (95%) and very little lies outside of 3 standard deviations (1%). This concept underlies our assumptions when we want to test for statistical significance. You will learn more about this later in this chapter.

Learning Activity 4.6.3
Describing Spread

Directions:
Calculate the variance and standard deviation for the following sample data that has 8 values.

23 33 37 45 46 52 52 60

Discussion 4.6.3

Describing Spread

Step	Calculation
1. First, find the mean of the dataset by finding the sum of all 8 values and diving by the number of values (8).	$\bar{x} = \frac{23+33+37+45+46+52+52+60}{8}$ $= \frac{348}{8} = 43.5$
2. Calculate the variance using the formula below. The formula finds the difference of each value from the mean (in step 1), squares it, and sums all of the squared differences. It then divides it by N-1 which is 7 in this case. $\sigma^2 = [1/(N-1)]^* \Sigma (X_i - \mu)^2$	$s^2 = \frac{1}{(8-1)} \times [(23-43.5)^2 + (33-43.5)^2 + (37-43.5)^2 +$ $(45-43.5)^2 + (46-43.5)^2 + 2(52-43.5)^2 + (60-43.5)^2]$ $s^2 = \frac{1}{7} \times [420.25 + 110.25 + 42.25 + 2.25 + 6.25 +$ $2(72.25) + 272.25]$ $s^2 = \frac{1}{7} \times 998$ $s^2 = 142.57$
3. Calculate the standard deviation. $\sigma = \text{sqrt}\ [1/(N-1)^* \Sigma (X_i - \mu)^2]$	$s = \sqrt{142.57}$ $s = 11.94$

4.7 Presenting Frequencies of Numerical Data

4.7.1 Frequency & Relative Frequency Tables

One-way Frequency Tables

Frequency tables, used for categorical data, are usually not an appropriate method for displaying numerical data. For example, say you have 150 clients who came to health facility during one week and you are trying to present their age by frequency. As demonstrated by the abbreviated Figure 4.7.1a-1 below, this could get quite lengthy.

Figure 47.1a-1 Frequency table for numerical data (ages)

Age	Frequency	Percent	Cumulative Percent
0	1	.66	.66
1	2	1.33	2
2	1	.66	2.66
...			
12	4	2.66	10
13	3	2	12
...			
44	2	1.33	55.33
45	1	.66	56
...			
67	2	1.33	100
Total	150	100	100

One way to address this is to transform, or change, the numerical variables into categorical variables by grouping them into meaningful categories. As mentioned above, in health, the important categories of age are the first 5 years of life, pre-pubescent, pubescent, reproductive age, and above. Thus, for the example above, we might recreate this *one-way frequency* table to look like the one presented below in Figure 4.7.1a-2.

Figure 4.7.1a-2: One-way Frequency table for numerical data once it has been transformed to categories

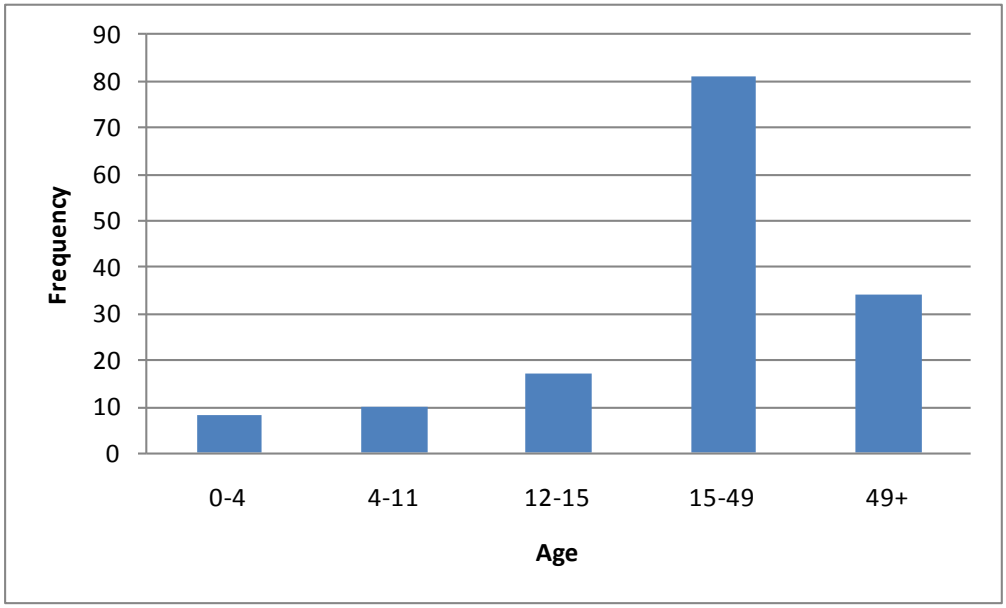
Age	Frequency	Percent	Cumulative Percent
0-4	8	5	5
5-11	10	7	12
11-15	17	11	23
15-49	81	54	77
49+	34	23	100
Total	150	100	100

Presenting the data in this way, by age groups, gives a quick visual indication of what age group are affected. In this example it is clear that those 15-49 are most represented.

Bar Graphs

Similar to categorical variables, you can also create bar graphs to represent this data. You can follow the instructions in Excel Tutorial #1 above to create a graph similar to the one below.

Figure 4.7.1b: Frequency of Age Categories



Two-Way Frequency Tables

When numerical data is changed into categorical variables you can create two way frequency tables.

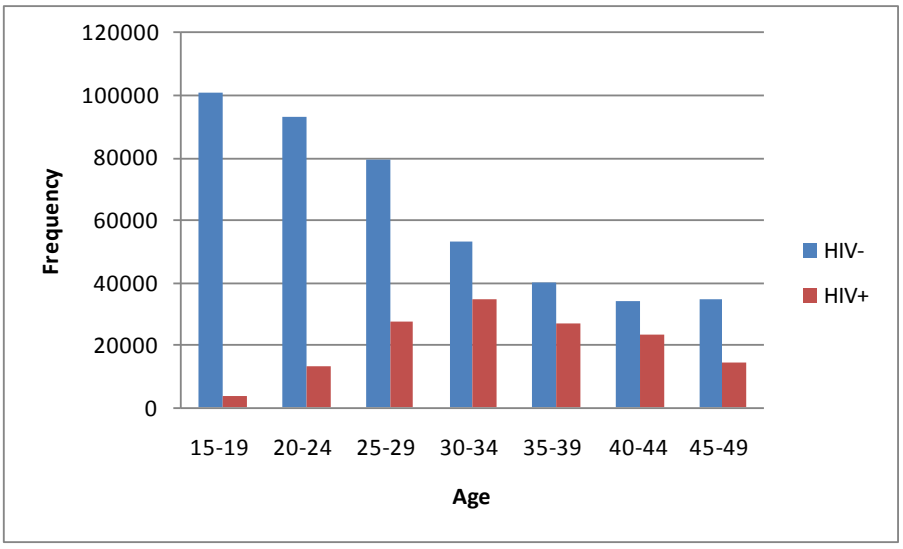
Figure 4.7.1c1 HIV Status by Age **(Central Statistics Office 2009b)**

Age	HIV Status		
	Positive	Negative	Total
15-19	101111 (23.2%)	3907 (2.7%)	105018 (25.9%)
20-24	93388 (21.4%)	13063 (9.0%)	106451 (30.4%)
25-29	79258 (18.2%)	27732 (19.1%)	106990 (37.3%)
30-34	53214 (12.2%)	35018 (24.2%)	88232 (36.4%)
35-39	40041 (9.2%)	27290 (18.8%)	67331 (28.0%)
40-44	33922 (7.8%)	23204 (16.0%)	57126 (23.8%)
45-49	34784 (8.0%)	14781 (10.2%)	49565 (18.2%)
Total	435718 (100.0%)	144995(100.0%)	580713 (100.0%)

Bar Graphs

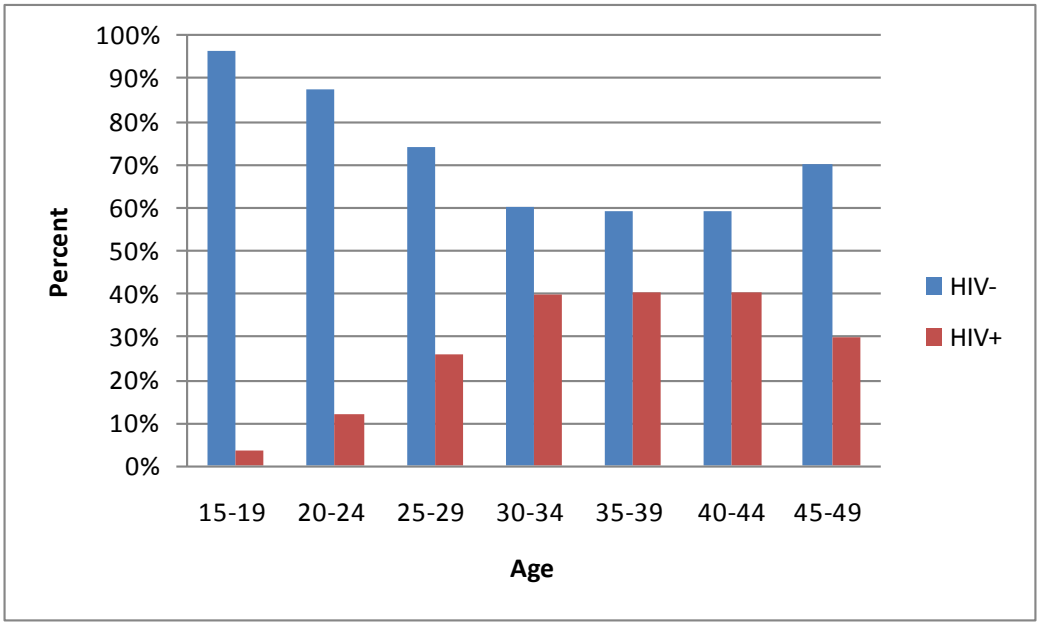
These tables can also be turned into graphs in the same way that two-way tables for categorical variables are made. The bar graph below is an example using frequency, presenting the number of individuals who are infected with HIV by age.

Figure 4.7.1d-1 HIV Status by Age - Frequency **(Central Statistics Office 2009b)**



Using frequencies with this data is difficult because the total number of individuals in each age category varies, from 49,565 to 106,990. Therefore, the frequency in one category as percentage of the total would be very different than in another category. A better way to present this data would be using percentages. This helps to make the data comparable across age groups. Figure 4.7.1d-2 below clearly demonstrates that the age groups 30-44 have the highest rates of HIV, at roughly 40%.

Figure 4.7.1d-2 HIV Status by Age – Relative Frequency **(Central Statistics Office 2009b)**



4.7.2 Other Summary Tables Using Mean and Standard Deviation

Another way to present numerical data is to use the mean and standard deviation. These descriptive statistics help give the reader an idea not only of the middle point of the data but also an idea of the spread or distribution.

Figure 4.7.2a Demographics of HIV Patients at Two Facilities

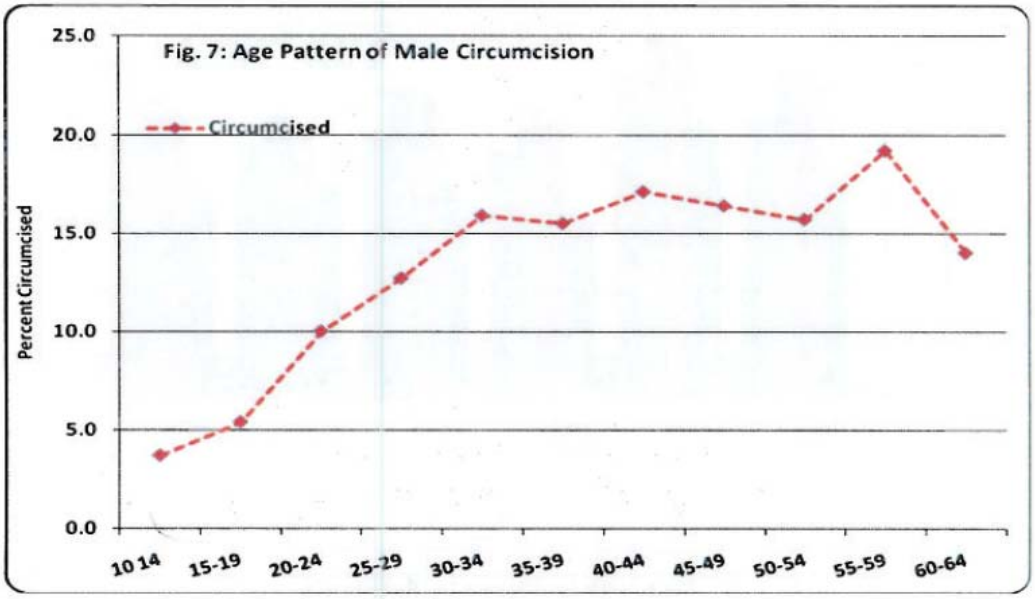
	Facility 1 (n=745)	**Facility 2 (n=1142)**	
	M (SD)	**M (SD)**	**Total**
Age (years)	42 (3.2)	44 (3.0)	43 (3.1)
CD4 Count (mm^3)	330 (60)	455 (72)	397 (65)
Weight (kg)	77.1 (22)	83.9 (26)	82 (24)
Duration of Treatment (months)	46 (17)	55 (23)	52 (20)

In this table you can compare the demographics between two facilities. While the ages of the patients in both facilities are similar and have a similar SD their CD4 counts are different. CD4 counts, patient weight, and duration of treatment are better at the second facility.

Line Graphs

Another way to present numerical data is to use a line graph. Line graphs can help present a clearer picture of your data, particularly when you have 10 or more points on the X-axis. Below is one example of a line graph which maps the percentage of men who are circumcised within each age group.

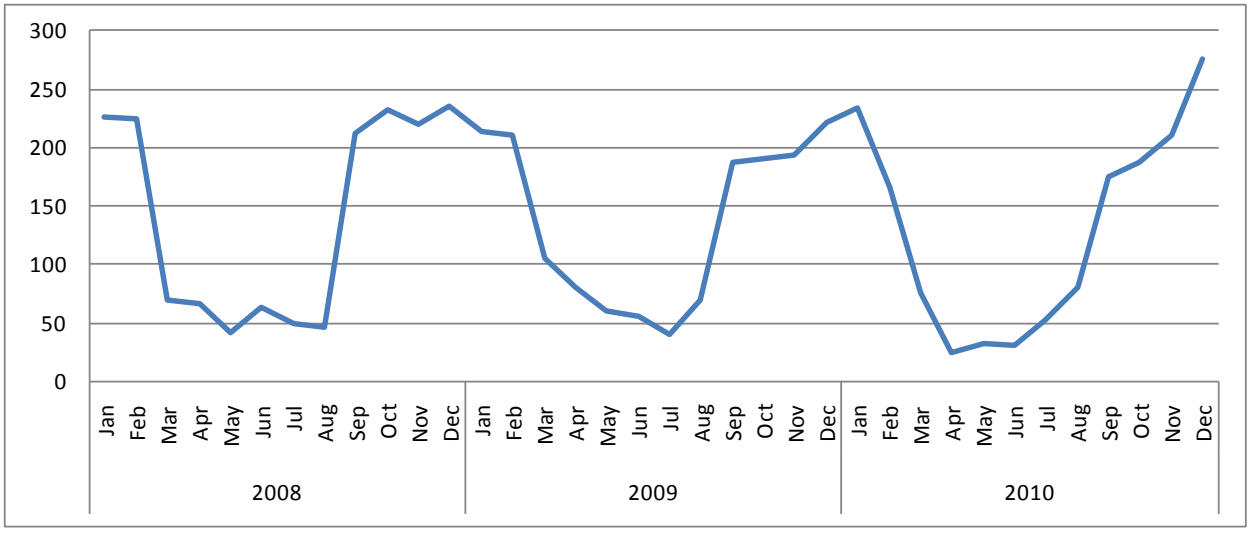
Figure 4.7.2b Age Pattern of Male Circumcision **(Central Statistics Office 2009b)**



Time Series Plots

One type of line graph that will be particularly useful is time series plots. A time plot tracks time on the horizontal (x) axis and a numerical, or quantitative, variable on the vertical (y) axis. Time series charts are especially useful in tracking the performance of a process over time and document the story of improvement work. You can look at charts (such as bar graphs) that give you a snapshot pictures as providing still photos, while charts such as time series charts provide information of a process as a video would.

Figure 4.7.2.c Number of Patients Diagnosed with Malaria in Ngamiland



As shown in Figure 4.7.2.c above, a time plot can be used to assess:

- Overall trends: the number of patients diagnosed with malaria is relatively stable year to year.
- Cyclical patterns: the number of patients diagnosed with malaria goes up and down within each year. This repeated pattern suggests a seasonal cycle of malaria, where there are more cases during the rainy season than in other times of the year.

Excel Tutorial #2

In this section, we will practice creating a time series line graph like the one above. For this example, we will be using the malaria diagnosis data.

Part I: Enter the data

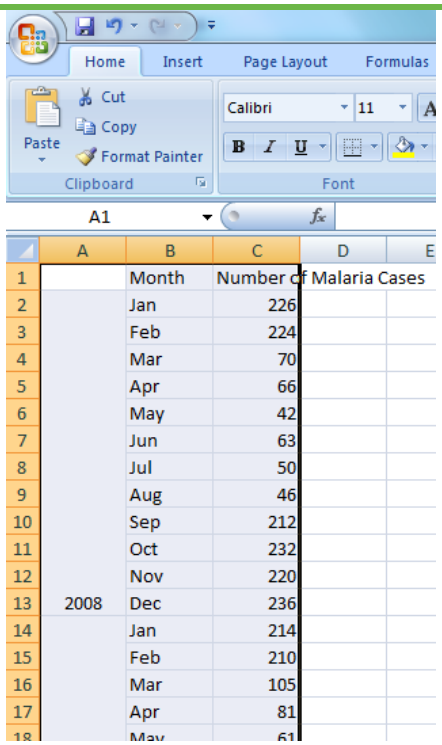
1. Open Excel and enter the data in the table into a worksheet. Tip: On the **Home** tab, highlight the year in column A and click **Merge & Center.** This will ensure that there is only listed once in the graph and will be centered under the months.

Table X. Number of Patients Diagnosed with Malaria in Ngamiland

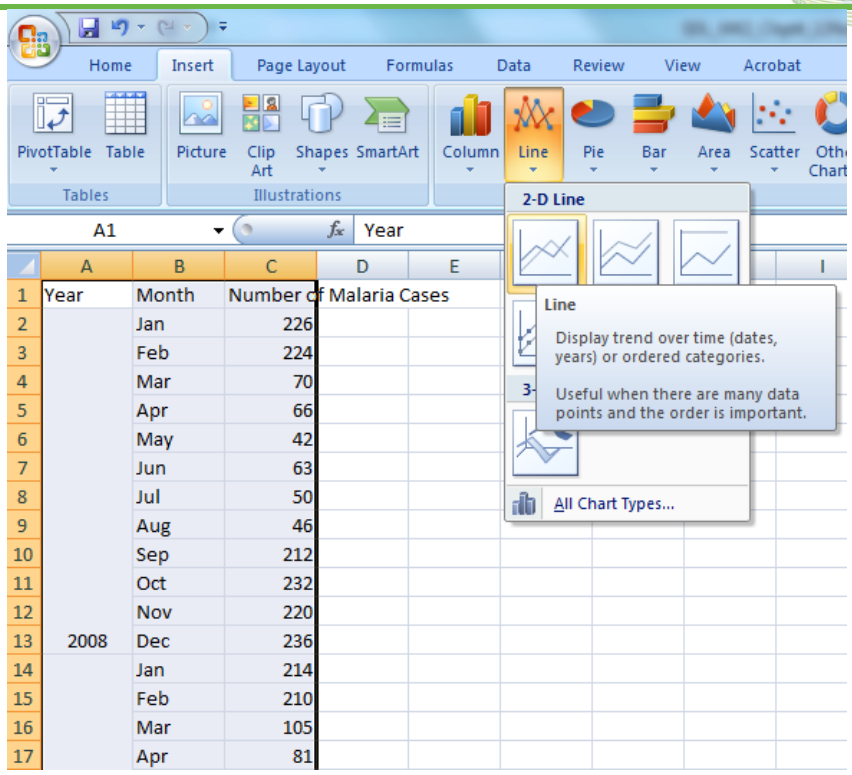
2008			2009			2010		
	Month	Malaria Cases		Month	Malaria Cases		Month	Malaria Cases
	Jan	226		Jan	214		Jan	234
	Feb	224		Feb	210		Feb	166
	Mar	70		Mar	105		Mar	76
	Apr	66		Apr	81		Apr	25
	May	42		May	61		May	33
	Jun	63		Jun	55		Jun	31
	Jul	50		Jul	41		Jul	53
	Aug	46		Aug	69		Aug	80
	Sep	212		Sep	187		Sep	175
	Oct	232		Oct	190		Oct	188
	Nov	220		Nov	194		Nov	210
	Dec	236		Dec	222		Dec	275

Part II: Create a Time Series Line Graph

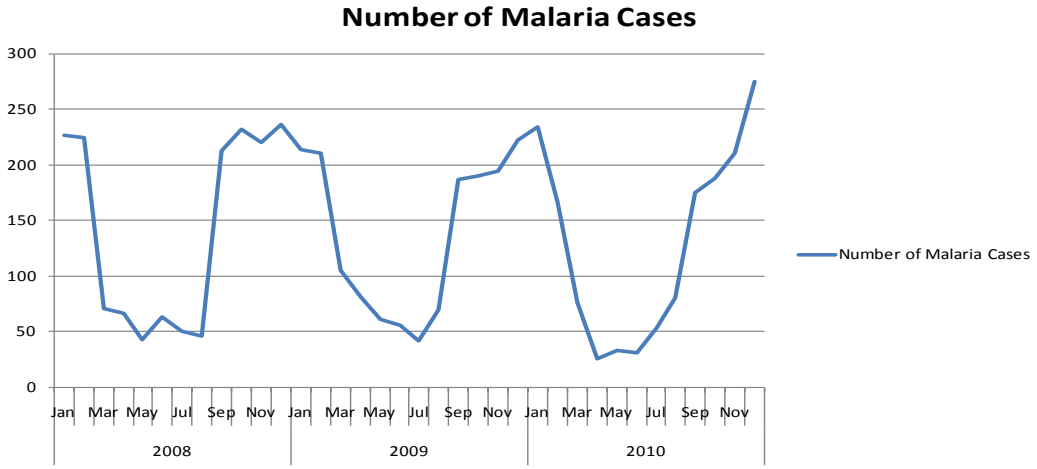
1. Drag your cursor to highlight the data and headers. Do not highlight the row of totals.



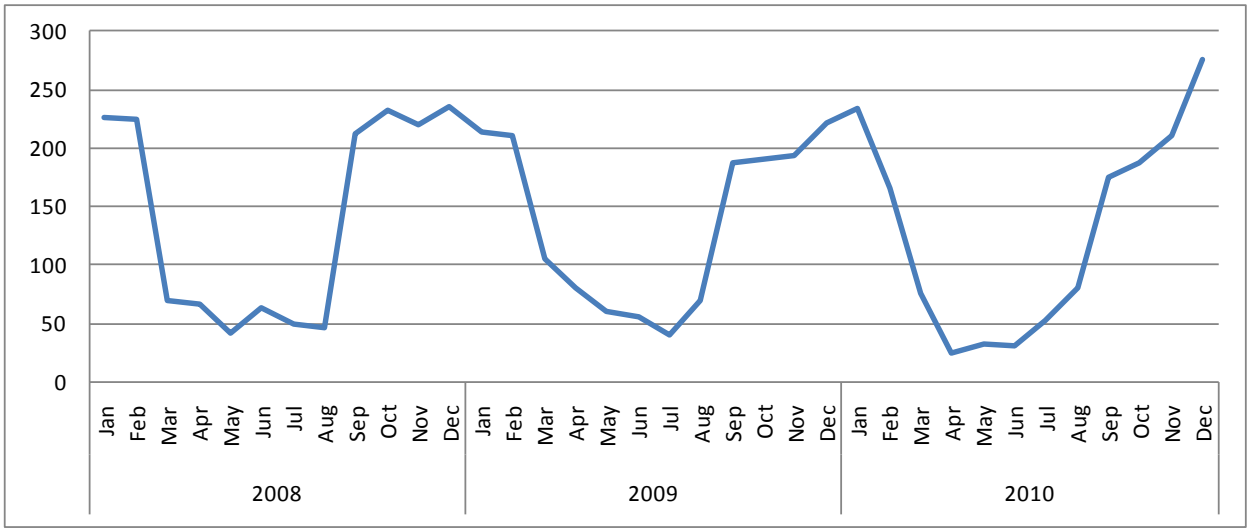
2. Click on the **Insert** tab and then on the **Line** button. Choose the first Graph under the 2-D Line.



3. A graph will appear that looks like the one below. Congratulations you have made a time series graph!



4. The final step is to format your graph so that all the months are clearly visible and the axis are labeled. There are many options for formatting under the **Design** tab – **Chart Layouts** button.



When using time series charts, you can add more details and the point estimate values you learned about in this chapter to tell a story of what changed, why it changed, and if a trend or a shift happened in the story you are telling. You can do this story telling by hand, or using Excel.

After constructing a time series chart, you can detect if a trend or a shift occurred in your data points.

- A *trend* happens when there is a continued movement in a single direction, either up or down.

- A *shift* indicates a movement to a different level in the pattern of a process or outcome you are measuring.

Let us take the example of the graph shown in Figure 4.7.2.c and discussed in Excel tutorial 2. You can use the following simple rules to detect a trend or shift that may be statistically significant (a concept which will be discussed in detail in Workbook 3, Chapter 4).

Detecting a Trend

To detect a trend, we look for movement in our time series chart. If there are at least five consecutive data points moving in the same direction, you may have a trend. Depending on the number of data points you have available in your graph, you may need more number of consecutive data points to detect a trend. If more consecutive data points are equal, only one counts towards the trend.

The following Figure, 4.7.2.d provides you with a minimum number of consecutive points you will need to have to detect a trend.

Figure 4.7.2.d Number of consecutive points required to detect a trend

Number of data points available	Number of consecutively increasing or decreasing data points required to detect a trend
≤8 data points	5 data points
9-20 data points	6+ data points
21+ data points	7+ data points

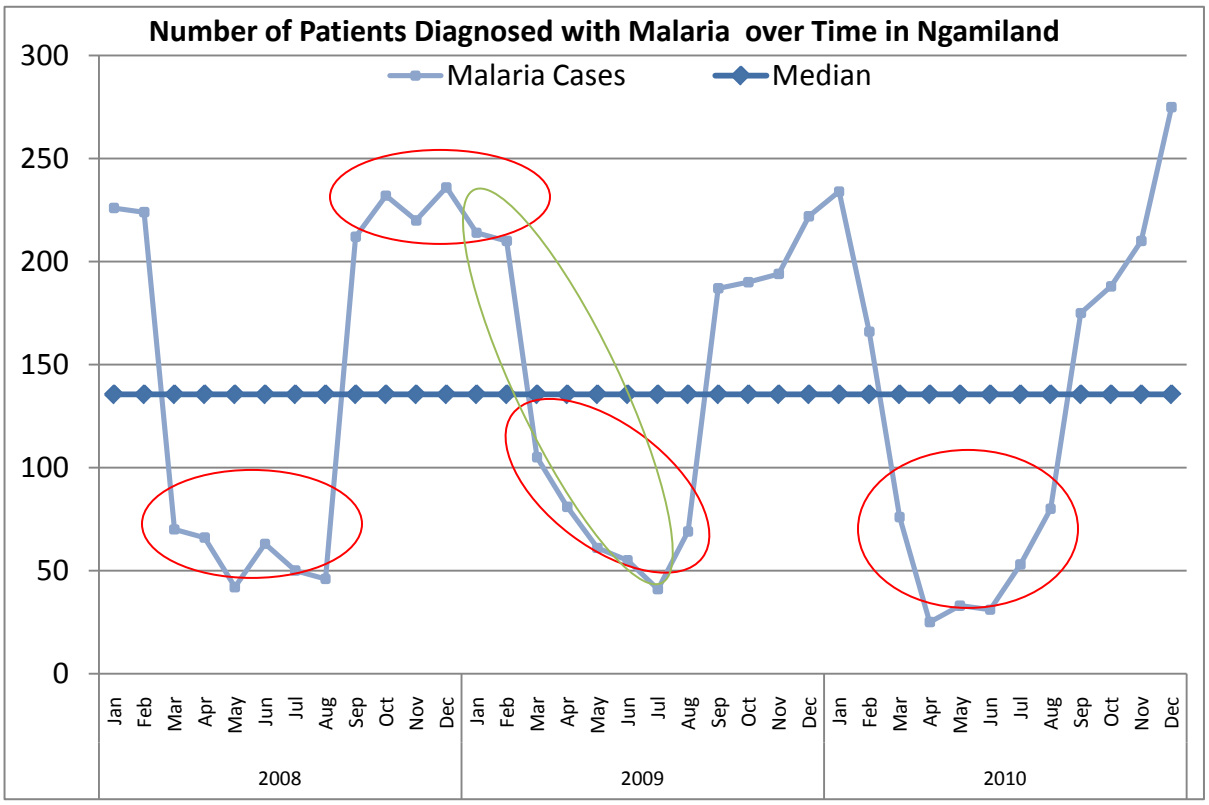
NOTE: You do not need a median to detect a trend.

Detecting a Shift

To detect a shift, you will need to calculate the median and include it in your graph. A shift requires at least six consecutive points on one side (above or below) the median line. Data points that fall on the median are not counted towards a shift.

Let us go back to the malaria diagnosis data from Ngamiland (discussed in Figure 4.7.2b and discussed in Excel tutorial 2).

Figure 4.7.2.e



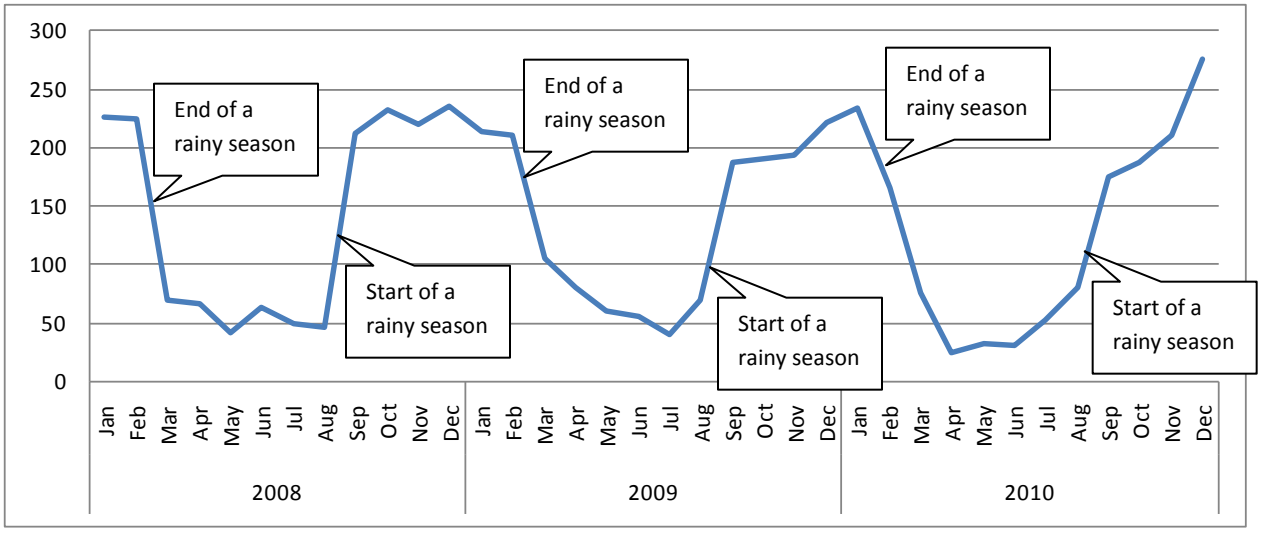
You will note that there are 36 data points in the graph (12x3). Therefore, we will be looking for seven data points to detect a trend. Given that rule, we only detect one trend (shown in green), between January and July of 2009. However, we have multiple times where a shift has occurred (depicted in red).

Annotations

So far you may have noted that the graphs we have discussed, including time series charts have been telling us what has changed but not why. Reasons for change can be added to graphs using annotations. Annotation is the process of adding commentary or explanatory notes to a chart. With annotations, we can denote when intervention was implemented and other explanation that might help explain trends and shifts. Doing this allows you to tell a more complete story by adding qualitative information that explains the variation shown in your charts.

The time series chart from Ngamiland, for example, can be annotated as the following to provide a more complete story that also explains the shifts in the time series chart.

Figure 4.7.2.f



Chapter Summary

Using descriptive statistics to analyze routine data helps to reveal the main features and characteristics of our data. It can tell us what the spread of the data is, how grouped data compares to other data, and if there are large outliers in the data. It can also reveal surprising trends. Thinking back to some of our examples above, we have been able to identify trends in our data, for example: older adults in one district seem to get TB more often, STI patients are not being tested for HIV in one clinic. This information can be presented in a table or graph, in order to help others quickly and easily absorb it. In the process of describing our dataset, we are able to understand basic aspects of the data, including: the data type, how best to present each data type, including relationships between key variables (such as age and HIV status), how to calculate key descriptive statistics, such as relative frequency, mean, median, mode, variance, standard deviation, and how to interpret the findings of our data.

At the same time, presenting descriptive statistics often brings up more questions that still need to be answered. For example, *why* do older adults seem to get TB more often? Why does it seem that STI patients are not being tested for HIV? To confirm that these trends are indicative of a true difference between groups we may need to use statistical tests. To understand the systems and processes behind *why* we are seeing what we see in our data, we may need to collect more information that is outside of our routine reporting systems. Both of these topics will be covered in Workbook 3, which builds on this workbook to strengthen evaluation skills.

Data analysis is an important part of evidence based planning

Self-Assessment Quiz

1. Classify the following variables as categorical or numerical. Once you have determined if they are categorical or numerical, determine if they are nominal, ordinal, discrete, or continuous.
   a. _____ Routine HIV Testing (RHT) Monthly Report - Types of HIV Tests (ELISA, Rapid, PCR) -
   b. _____ Botswana HIV and AIDS Response Information Management System (BHRIMS) 003B Report - Number of clients receiving IPT this quarter
   c. _____ BHRIMS 003B Report - Total cost of HIV and AIDS components of qualifying development programmes this quarter
   d. _____ ARV Site Manager's Report – Time on HAART Before Death (< 1month, 1-2 months, …, 6 months – 1 year, > 1 year)
   e. _____ MH1049 – Type of Notifiable Disease
   f. _____ MH1049 – Number of consultations

2. Using the Ngamiland malaria diagnosis data from the following data (also presented in Excel Tutorial #2); calculate the mean, variance, and the standard deviation. Interpret what the mean and standard deviation tell you about the data.

2008	Month	Malaria Cases	2009	Month	Malaria Cases	2010	Month	Malaria Cases
	Jan	226		Jan	214		Jan	234
	Feb	224		Feb	210		Feb	166
	Mar	70		Mar	105		Mar	76
	Apr	66		Apr	81		Apr	25
	May	42		May	61		May	33
	Jun	63		Jun	55		Jun	31
	Jul	50		Jul	41		Jul	53
	Aug	46		Aug	69		Aug	80
	Sep	212		Sep	187		Sep	175
	Oct	232		Oct	190		Oct	188
	Nov	220		Nov	194		Nov	210
	Dec	236		Dec	222		Dec	275

a. Mean:

b. Variance:

c. Standard deviation:

d. Interpretation:

3. The PMTCT Programme Coordinator has used the antenatal clinic (ANC) Register and the PMTCT Male Partner Testing Register to create a table of all the ANC clients registered in September 2010 and all their partners that got tested for the next three months within the district. She wants you to use this table to help her figure out what percentage of ANC clients' partners are getting tested. She also wants this data presented by the HIV status of the ANC client because she wants to know whether partners of HIV-infected ANC clients are more or less likely to get tested. Finally, she wants to know what proportion of the couples that tested are discordant, where one partner is HIV-infected and the other partner is not infected with HIV.

   Using this table, calculate the necessary relative frequencies and create the graph that best represents the data and answers the PMTCT Coordinator's questions.

Table X. ANC Clients by HIV Status and Partner HIV Testing Status

	ANC Clients			
Partners HIV Status	HIV+	HIV-	Declined	**Total**
HIV+	47	44	2	93
HIV-	57	277	4	338
Declined	25	10	0	35
Not Tested	82	190	7	279
Total	211	521	13	745

References / Bibliography

Baldi, B, Moore, DS 2009. *Practice of Statistics in the Life Sciences* [CD-ROM]. New York: W.H. Freeman and Company.

Bowers, D. 2008. *Medical Statistics from Scratch: An Introduction for Health Professionals.* Chapter 1 First things first- the nature of data. 2nd Edition. JohnWiley & Sons, Ltd.

Central Statistics Office November 2009. *2008 Botswana AIDS Impact Survey III*.

Central Statistics Office May 2009. *Preliminary Botswana AIDS Impact Survey III Results, Stats Brief.*

Getting Genetics Done: Software, Tips, & Productivity Hacks for Getting Things Done in Genetics Research. http://gettinggeneticsdone.blogspot.com/ [Accessed: November 2010]

MASA 2009. *Strategic Information: Epidemiology and Indicators.* Presentation at the 2009 ARV Assessment Orientation, Gaborone International Conference Center 26 Oct 2009.

Massoud, R., K. Askov, J. Reinke, LM. Franco, T. Bornstein, E. Knebel, and C. MacAulay. 2001. A Modern Paradigm for Improving Healthcare Quality. *QA Monograph Series* 1(1) Bethesda, MD: Published for the U.S. Agency for International Development (USAID) by the Quality Assurance Project.

Measurement Scales in Social Science Research. http://simon.cs.vt.edu/sosci/converted/Measurement/activity.html [Accessed December 2010].

StatTrek: Teach yourself statistics. http://stattrek.com/AP-Statistics-1/AP-Statistics-Intro.aspx?Tutorial=AP [Accessed: November 2010].

Zeribi KA, & LM. Franco. September 2001. Guidance for Analyzing Quality Improvement Data Using Time Series Charts. Bethesda, MD: Published for the U.S. Agency for International Development (USAID) by the Quality Assurance Project.

Self-Assessment Quiz Answer Key

1. Classify the following variables once you have determined if they are categorical or numerical, determine if they are nominal, ordinal, discrete or continuous.

   a. RHT Testing Monthly Report - Types of HIV Tests (ELISA, Rapid, PCR) - *Nominal*
   b. BHRIMS 003B Report - Number of clients receiving IPT this quarter – *Discrete*
   c. BHRIMS 003B Report - Total cost of HIV and AIDS components of qualifying development programmes this quarter – *Continuous*
   d. ARV Site Manager's Report – Time on HAART Before Death (< 1month, 1-2 months, …, 6 months – 1 year, > 1 year) – *Ordinal*
   e. MH1049 – Type of Notifiable Disease - *Nominal*
   f. MH1049 – Number of consultations - *Discrete*

2) Using the Ngamiland malaria diagnosis data from Excel Tutorial #2, calculate the mean, variance, and the standard deviation. Interpret what the mean and standard deviation tell you about the data.

   a. Mean: *135.06*
   b. Variance: *6611.60*
   c. Standard deviation: *81.32*
   d. Interpretation*: On average 135 individuals are diagnosed with malaria every month. Given the wide standard deviation of 81.32, the number of malaria cases each month varies widely.*

3) The PMTCT Programme Coordinator has used the ANC Register and the PMTCT Male Partner Testing Register to create a table of all the ANC clients registered in September 2010 and all their partners that got tested for the next three months within the district. She wants you to use this table to help her figure out what percentage of ANC clients' partners are getting tested. She also wants this data presented by the HIV status of the ANC client because she wants to know whether partners of HIV-infected ANC clients are more or less likely to get tested. Finally, she wants to know what proportion of the couples that tested are discordant, where one partner is HIV-infected and the other partner is not infected with HIV.

Using this table, calculate the necessary relative frequencies and create the graph that best represents the data and answers the PMTCT Coordinator's questions.

Table X. ANC Clients by HIV Status and Partner HIV Testing Status

	ANC Clients			
Partners HIV Status	HIV+	HIV-	Declined	**Total**
HIV+	47	44	2	93
HIV-	57	277	4	338
Declined	25	10	0	35
Not Tested	82	190	7	279
Total	211	521	13	745

Q1. To determine what percentage of ANC clients' partners are getting tested.

Step 1. Calculate relative frequencies for the ANC Client column totals. We use the columns because the PMTCT Programme Coordinators' first question is of all the ANC clients, what percentage of their partners are getting tested.

$$93+338/745 = 57.9\%$$

Q2. To determine whether partners of HIV-infected ANC clients are more or less likely to get tested than others.

Step 1. Simplify the table. Since we only care about whether or not partners get tested with can reduce this to two groups.

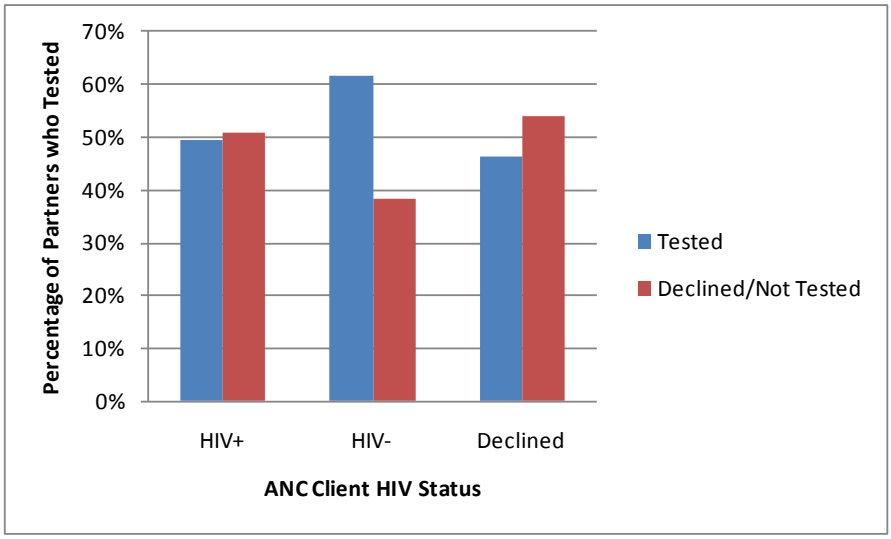
	ANC Clients			
Partners HIV Status	HIV+	HIV-	Declined	**Total**
Tested	104	321	6	431
Declined/Not Tested	107	200	7	314
Total	211	521	13	745

Step 2. Calculate the relative frequencies by column. Again because the question is, of the ANC clients, are the partners of ANC clients who are HIV-infected more or less likely to get tested.

	ANC Clients			
Partners' HIV Status	HIV+	HIV-	Declined	**Total**
Tested	104 (49.3%)	321 (61.6%)	6 (46.2%)	431
Declined/Not Tested	107 (50.7%)	200 (38.4)%	7 (53.9%)	314
Total	211 (100%)	486 (100%)	13 (100%)	745

Step 3. Create a graph. Using the instructions for a two-way frequency table, create a bar graph with the relative frequencies.

Figure X. Percent of ANC Partners who Tested by ANC Clients' HIV Status



Q3. Determine the proportion of the couples that tested who are discordant, where one partner is HIV-infected and the other partner is not infected with HIV.

Step 1. Simplify the original table to only include those where both partners tested.

	ANC Clients		
Partners HIV Status	*HIV+*	*HIV-*	***Total***
HIV+	*47*	*44*	*91*
HIV-	*57*	*277*	*334*
Total	*104*	*321*	*425*

Step 2. Using the table calculate the relative frequency of discordant couples.

Discordant: HIV+/HIV- and HIV-/HIV+ = 44+57 = 101; 101/425 = 23.8%

Self-Directed Learning Workbook2:
Chapter Five
Presentation Skills

Strong presentation skills are an important professional skill

Chapter 5:
Presentation Skills

Estimated time needed for completion: 2 hours

Chapter Overview

Throughout Workbook 2 we have addressed the role of the district-level M&E officer in collecting, managing, and maintaining high quality data. Further, we have discussed various ways in which the categorical and quantitative data you collect can be analyzed and summarized. Once the data have been analyzed and summarized, you may be asked to present it at District Multi-sectoral AIDS Committe (DMSAC), Technical Advisory Committee (TAC), District Development Committee (DDC), and District Health Management Team (DHMT) meetings. This chapter will help you to develop strategies and techniques for giving a successful presentation. We will also present techniques for using PowerPoint and guidelines for presenting data and graphs using PowerPoint. You will then have an opportunity to practice these skills as you complete this chapter and Workbook 2 by giving a presentation at one or more of the facilities where you work. In your presentation, you will be describing your role and responsibilities as a district level M&E officers.

Learning Objectives

At the end of this unit, participants will be able to:

- identify ten techniques for giving a strong presentation,
- identify five techniques for using PowerPoint effectively, and
- list guidelines for using PowerPoint for presenting graphs and data.

Materials Needed

- Computer equipped with Microsoft Office PowerPoint

5.1 Giving a Strong Presentation

As an M&E officer, sharing information is an important part of your job. One effective way to share information is by giving oral presentations. While public speaking can be a little intimidating at first, the skills involved in giving a good oral presentation are skills that can easily be learned and improved upon! There are several techniques that you can use to help ensure you are giving a strong and successful presentation.

Listed throughout this chapter are the top 10 tips for giving a good presentation. At the end of the chapter, you will be asked to deliver an oral presentation to the audience of your choice. These tips will help you during the planning and development phases of your presentation.

Reflection

Take a moment and consider the following questions:

- Have you ever had to give a presentation?
- If so, how did you feel when you gave the presentation?
- What part of giving a presentation did you find the most challenging? Developing the presentation? Or actually delivering the presentation?
- What did you enjoy about it?

5.1.1 Ten Tips for Giving a Strong Presentation

Tip 1: Have a clear purpose

You should be able to clearly articulate the purpose of your presentation before beginning work on developing the presentation. The purpose describes what you as the presenter wants to achieve by giving the presentation. There are many reasons why you might give a presentation. The purpose of the presentation may be to impart knowledge, change attitudes or behaviours, or to stimulate interest in a subject. Knowing your purpose for giving an oral presentation is the first step in planning and developing a strong presentation.

The purpose of a presentation will guide what messages are delivered and how you deliver the message. For example, when the purpose is to impart knowledge the presentation should include essential facts and an explanation of how the audience will be able to use the information. When the purpose of a presentation is to stimulate interest, the presentation should include specific ideas about how the information being presented will be able to directly benefit the audience. Finally, a presentation that is intended to persuade or change behaviour will often include such things as compelling statistics or a strong emotional appeal.

Learning Activity 5.1.1

Purpose of a Presentation

Directions: Determine what the purpose of the presentation might be for each example below. Behaviour change? Stimulate interest? Impart knowledge?

Scenario	Purpose
You are giving a presentation to the hospital matrons emphasizing the importance of HIV testing for pregnant women during their antenatal visits. Your presentation includes a comparison of incidence of HIV in infants whose mother had an antenatal HIV test and the incidence of HIV in infants whose mother did not.	
You are giving a presentation to newly-hired M&E officers describing your roles and responsibilities as a district level M&E officer. Your presentation includes information about your daily activities and how the data you collect is used.	
You area giving a presentation to DHMT staff about how to use Excel in order to analyze health data. You want to show how using Excel can be a more effective way of presenting some types of data. You would like the audience to use Excel more frequently when presenting at district level meetings. You begin your presentation by showing a slide that says: "*From 2001 – 2007 the rate of MTCT transmission has decreased from 21-40% to 4%*". You then show a slide with a histogram graphically demonstrating the decrease in the rate of mother-to-child transmission (MTCT) of HIV.	

Discussion 5.1.1

Purpose of a Presentation

Example	Purpose
You are giving a presentation to the hospital matrons emphasizing the importance of HIV testing for pregnant women during their antenatal visits. Your presentation includes a comparison of incidence of HIV in infants whose mother had an antenatal HIV test and the incidence of HIV in infants whose mother did not.	*Behaviour change.* *You are using persuasive and compelling statistics to show the difference that antenatal visits make in incidence of HIV.*
You are giving a presentation to newly-hired M&E officers describing your roles and responsibilities as a district-level M&E officer. Your presentation includes information about your daily activities and how the data you collect is used.	*Impart knowledge.* *You are informing you colleagues about the activities of a district level M&E officer.*
You area giving a presentation to DHMT staff about how to use Excel to analyze health data. You want to show how using Excel can be a more effective way of presenting some types of data. You would like the audience to use Excel more frequently when presenting at district level meetings. You begin your presentation by showing a slide that says: *"From 2001 – 2007 the rate of MTCT transmission has decreased from 21-40% to 4%"*. You then show a slide with a histogram graphically demonstrating the decrease in the rate of MTCT of HIV.	*Stimulate interest.* *By showing how much more dramatic and stimulating the excel chart is you are trying to motivate the new officers to learn and use Excel.*

Tip 2: Answer the question "WIIFM?"

While your purpose for giving the presentation might be to impart knowledge, the purpose of listening to the presentation must also be clear to the audience members. The primary question for audience members at the start of your presentation is; "What is in it for me?" otherwise known as WIIFM? It should be made clear to the audience at the outset of the presentation what they will get from listening to your presentation. Make it clear to the audience how some aspect of their work will be improved, enhanced, or made easier by watching the presentation.

In the context of M&E, answers to the question WIIFM? might include:

- clarification of roles and responsibilities,
- ways to contribute to better data flow,
- data to use for decision making, and
- understanding current trends in disease indicators.

Tip 3: Organize the presentation

Make sure the presentation has a clear beginning, middle, and end. The beginning should set the stage for the presentation and identify the topics that will be presented. The middle is where the bulk of the information is presented. This is also the longest part of the presentation. At the end of the presentation, key points and concepts should be summarized to help your audience remember the most important aspects of the presentation.

Tip 4: Eliminate unnecessary information

Include only the information that you most want the audience to know in order to meet the goal of the presentation. Do not overload the presentation with extra details and facts. This will only dilute your intended message. After developing your presentation, you can review it with the purpose of identifying and removing unnecessary information. If you determine that there is a large quantity of background or additional information that the audience may want to have you can offer it as a handout or provide it to them upon request.

Tip 5: Stay within time limits

Keeping your presentation within the set time limitations not only shows respect for your audience, but also helps to keep the presentation focused and eliminates unnecessary information. One way of making sure that your presentation does not go over the allotted time is to practice giving the presentation aloud. Another technique is to identify specific slides that will indicate how much material there is left to present. It can be helpful to use specific slides to identify when there is 10 minutes, 5 minutes, or 1 minute of material left to present. Alternatively, you can demarcate the amount of material you have left in your speaker's notes.

Tip 6: Concentrate on delivery

It is important that you face the audience and make eye-contact throughout the presentation. Print out a copy of the presentation and notes so that you do not need to read from the screen. Do not read the slides word for word as you will lose the interest of the audience. Know your presentation well so that just glancing at the screen for a key word should serve as reminder enough. Occasionally using hand gestures will also keep the attention of the audience on the presenter.

A number of techniques can help engage the audience and facilitate a successful presentation. Vary your tone of voice and avoid speaking in monotone in order to keep the audience interested and to highlight important points you want them to remember. Make sure to include pauses throughout the presentation. This allows the audience to consider what you are presenting and identify any questions they might have. Some presenters will actually write the word <<pause>> at various points on their slide notes to make sure that they include a pause.

Tip 7: Make it a performance

Making a presentation entertaining will not only help you engage the audience but will also help them recall the information you have presented. Rather than simply stating facts, attempt to tell a story. Show enthusiasm for the topic and information being presented.

Tip 8: Use skills to minimize nervousness

It is not uncommon to be nervous prior to and during a presentation. It is helpful to identify one or two techniques to help you take the edge off of your nervousness. Some people find that making direct eye contact with audience members can be calming. Looking at and speaking to a familiar person in the audience may also help you to reduce nervousness. However, the most reliable technique for reducing nervousness is to be prepared. Be very familiar with the material being presented and the organization of the presentation. It is also helpful to be familiar with the environment where the presentation will take place.

Tip 9: Use appropriate humour

Humour can help both the presenter and the audience relax. Humour can also be used to emphasize a point. However, it is essential that any humour you incorporate is appropriate and not offensive in any way. A good technique is for you direct humour at yourself, rather than someone else.

Tip 10: Practice, Practice, PRACTICE!!

Practicing the presentation is the number one way to make sure your presentation is successful. Practicing in front of a mirror can be useful, but it may be even more useful to practice in front of your peers. This not only provides you with an opportunity to receive feedback but also gives you an opportunity to make sure the timing of the presentation is appropriate. One can never be too familiar with the material being presented.

5.2 Using PowerPoint Effectively

PowerPoint is presentation software that is used widely throughout the world. It is relatively simple to use and can help make presentations professional and engaging. Like any tool though, PowerPoint can be misused and detract from your presentation. This section contains tips for designing an effective PowerPoint presentation.

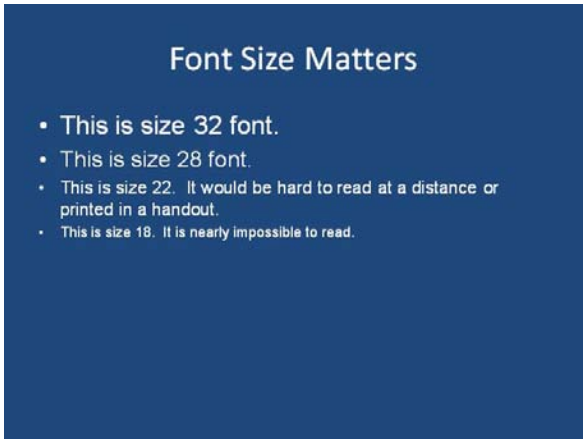
Reflection	*Think about presentations you have attended in which PowerPoint was not used effectively:* • What are some of the characteristics that you observed? For example, slides that contained too much information and thus were illegible? • Did the colour of either the background or text make them difficult to see? • What about font type and size?

5.2.1 Five Tips for Using PowerPoint Effectively

Tip 1: Choose an appropriate font

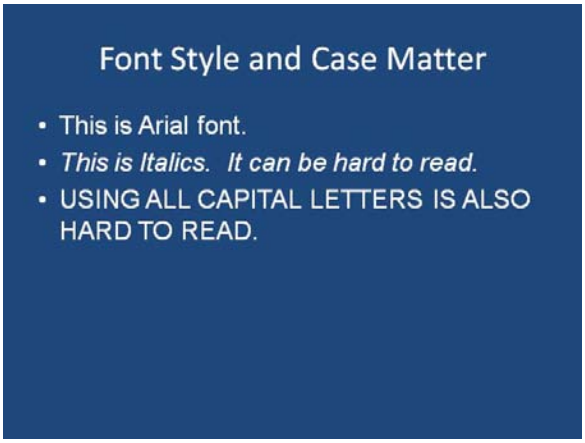
When choosing a font the size, style, colour, and case matter.

- Font Size: It is important to use a font size that is large enough for the audience to read. Usually size 32 font is used for headings and text. Size 28 can also be used for text, but nothing less than size 24 font should be included.



- Font Style: The best font styles are the Sans serif fonts such as Arial. Avoid using Italics for text; it can be hard to read.

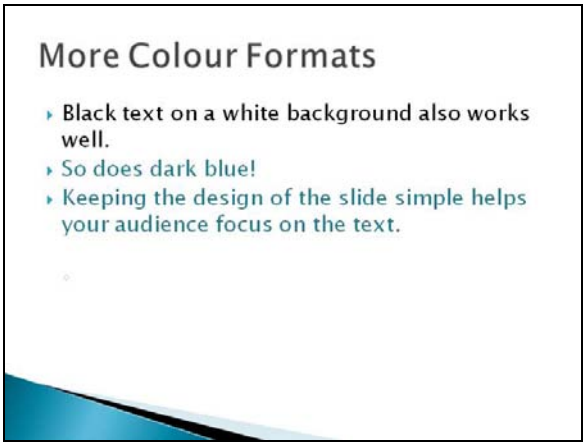
- Font Case: It is also best to use both upper-case and lower-case rather than all capitals. Using all capitals makes your slide hard to read and difficult to emphasize key points.



- Colour: There are two standard colour formats that project best when working with PowerPoint.

    - A blue background (either dark or light) with white or yellow text, or

- A white background with either dark blue or black text



If the intent is to print handouts from the slides, the second format using white background with dark text is the best option. Using coloured text, particularly, yellow, grey, orange and purple, can make the words difficult to see. If it is important to emphasize a word or concept, using **bold** or <u>underlining</u> the text is effective.

Tip 2: Do not use full sentences

Use bullets to list key words and phrases, but do not write complete sentences. The bullets should reflect the key points being made by the presenter. You can then expand on the key point identified in the bullet. Keep each bullet to a single line as much as possible and do not include more than 4-5 bullets on a single slide. Again, it is best if the audience is focussed on the presenter rather than trying to read the slides.

Tip 3: Use animation cautiously

Animating text, bullets, and phrases can be distracting and time consuming if not done sparingly. Often there is limited time scheduled for a presentation. Using animation is time consuming and valuable time can be wasted waiting for text to appear. Sound effects can also distract from the purpose of the presentation and diminish the professional nature of the presentation. A good rule to follow is the simpler, the better.

Tip 4: Incorporate graphics to enhance content

Images and graphics can be an excellent method of conveying information, especially in the context of M&E. Examples of possible images and graphics include: clip art, tables, figures, and photos.

By including an image or a graphic, you should be enhancing the audience's understanding of the content. It should help you achieve the purpose of your presentation. Some basic guidelines for using graphics in a PowerPoint presentation include:

- avoid getting too fancy,
- use images that are the proper size for the slide,
- remember Tip 1, related to slide's background, text size, font and colour as those rules apply to images as well,
- keep in mind that you may be printing handouts in black and white rather than colour, and
- consider placing detailed graphics and data sets into a handout.

We will discuss how to effectively incorporate tables and figures in the next section.

Tip 5: Print handouts

When possible, provide a handout of the presentation so that the audience can focus on you and what you are saying rather than furiously taking notes during your presentation. Providing handouts will also give the audience an opportunity to review the key points after the presentation has been given.

5.3 Incorporating Figures and Tables

It is highly probable that as a district level M&E officer, you will be asked to present programme specific data as part of a presentation. Tables and figures such as charts and graphs can be powerful methods of presenting that data. When incorporated into PowerPoint properly, they can be highly effective. However, when incorporated improperly, graphs, charts, and tables can be distracting and confusing to the audience. In this section we will discuss how to incorporate figures and tables into your presentation effectively and in a way that supports achieving the purpose of your presentation.

There are also several techniques that can be used to effectively incorporate graphs, charts and tables in PowerPoint. Refer to *Resource 5.1: Presenting Excel Charts and Graphs in PowerPoint* at the end of this chapter for a step-by-step guide on how to incorporate Excel charts and graphs into your presentation.

5.3.1 Five Tips for Incorporating Figures and Tables

Tip 1: Always label the x and y axis

When preparing the table and graph in PowerPoint it is essential that the *x* and *y* axes are clearly labelled (University of Newcastle n.d.). Make sure that the font used to label the axis is large enough for the audience to read. It is also critical to identify the units being represented in the graphic. For example, state whether the data represents the number of people with new infections, the number of women with new infections, the number of reports generated, etc. See Figure 5.3.1 for an example of a properly labelled slide.

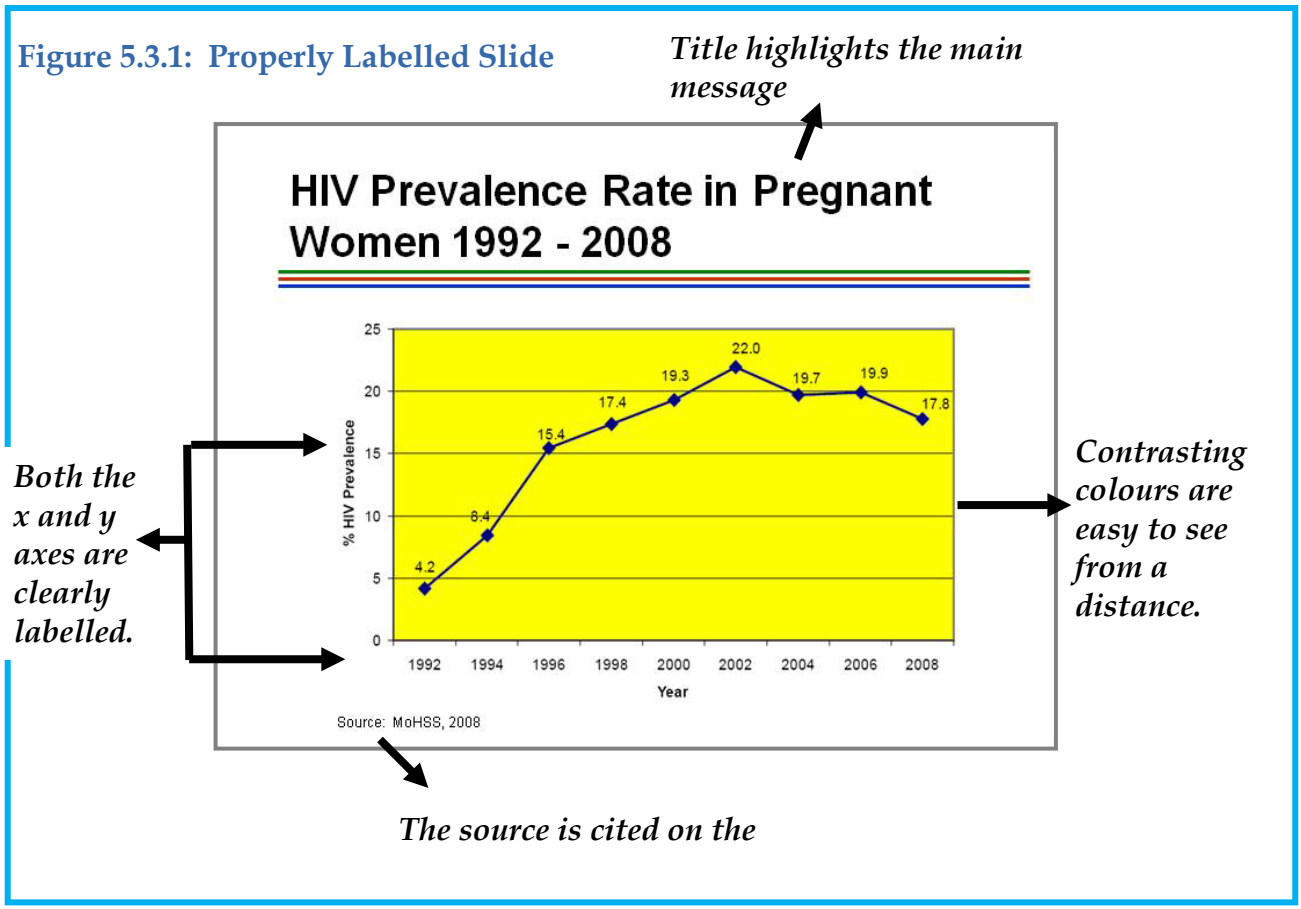
Tip 2: Orient your audience

Before delving into the details of the data being presented, take a moment to orient your audience to the image. Identify the x and y axis, the units of the data displayed, and the main message of the graphic. Where appropriate, you should consider titling the slide in such a way that it highlights the main message of the graphic.

Tip 3: Use colors that are easy to see at a distance

Use bold and basic colors such as red, green, and blue for text and a bright background that contrasts with the text. This is especially important when presenting data in line graphs, histograms, or other graphics making comparisons.

Figure 5.3.1: Properly Labelled Slide



Tip 4: Cite the sources of the data and graphics on the slide

It is important to cite the source of the data and/or the graphic presented on the slide. If you used data you collected to develop the graphic, then it is only necessary to cite the source of the data. However, if you are incorporating graphics that were created by somebody else, it is necessary to cite the source of the graphic as well.

Tip 5: Keep your audience in mind

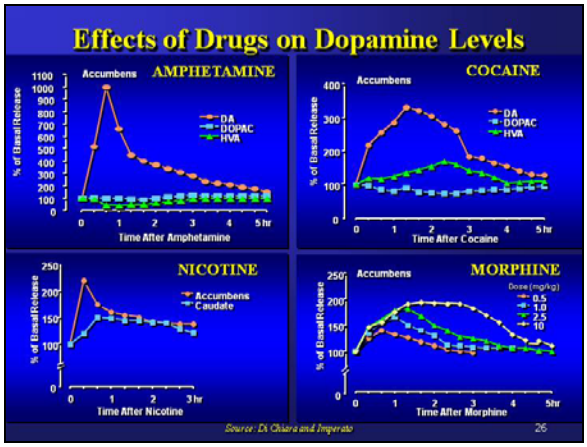
Create and include graphics that are appropriate for the audience you are presenting to. If your audience is not highly specialized in the subject of your presentation, do not include extremely detailed information. Present data and comparisons of data that are both relevant to them and to the purpose of your presentation.

Learning Activity 5.3.1

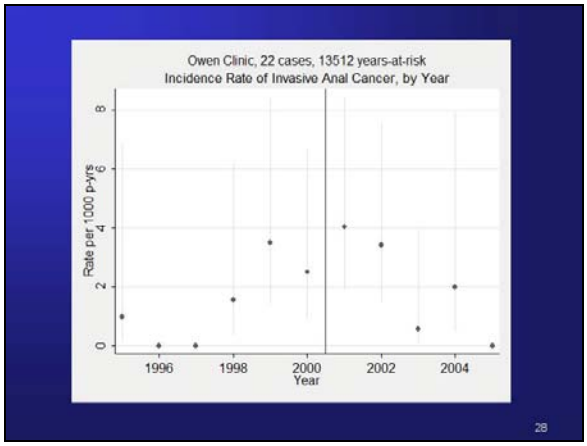
PowerPoint Slide Critique

Directions: Study each of the slides labelled A to E below. Then determine which critique below, numbered 1 to 5, best fits which slide. Write the letter of the slide next to the best fitting description.

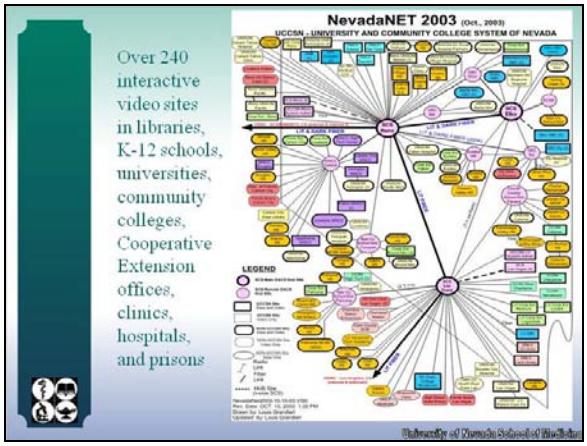
A



B



C



D

E

Public Health Tool

Unrecognized HIV Infection Among 1,767 MSM
(Baltimore, LA, Miami, NYC, San Francisco)

Age Group (yrs)	Total Tested	HIV Prevalence No.	%	Unrecognized HIV Infection No.	%
18-24	410	57	(14)	45	(79)
25-29	303	53	(17)	37	(70)
30-39	585	171	(29)	83	(49)
40-49	367	137	(37)	41	(30)
≥ 50	102	32	(31)	11	(34)
Race/Ethnicity					
White	616	127	(21)	23	(18)
Black	444	206	(46)	139	(67)
Hispanic	466	80	(17)	38	(48)
Multiracial	86	16	(19)	8	(50)
Other	139	18	(13)	9	(50)
Total	**1,767**	**450**	**(25)**	**217**	**(48)**

MMWR June 24, 2005

30

CDC/Janssen

1. ____Colours are faint. Lines are difficult to see. Difficult to identify where the main data points lie on the *x* and *y* axis.

2. ____Improper use of images. Unclear of the relevance of the image.

3. ____Overall, good slide. Shows nice comparisons of different variables. Also uses different shapes along the lines to distinguish between the lines which can be helpful, especially if printed in black and white. The contrast between the lines is not easy to read, especially with the small size of the graphics. There also needs to be more dramatic contrast between the colours of the lines as they blur together in places, especially if there is any problems with the light in the room or focus of the machine.

4. ____There is a lot of information of this slide, but it does a good job of highlighting the important information, has a relevant and clear slide title and is the source of the graphic appears on the slide.

5. ____The image is far too complex and there is too much text. It would be better to remove the image and place it into a handout if appropriate.

Discussion 5.3.1

PowerPoint Slide Critique

Review the answers to Exercise 5.3.1 below. Note further explanation for the correct answer is written in italics.

1. __B__Colours are faint. Lines are difficult to see. Difficult to identify where the main data points lie on the *x* and *y* axis. *Connecting the data points (dots) will demonstrate a trend better as well as make it easier to identify the data points.*

2. __D__Improper use of images. Unclear of the relevance of the image. *This slide came from an actual professional training for physicians. Although unclear, the point of the slide may have been to show that AIDS is a global epidemic an affects many people worldwide. Maybe the bird and cute puppy were aimed at decreasing stigma? The point of the images is not clear and does not directly support the purpose of the presentation.*

3. __A__Overall, good slide. Shows nice comparisons of different variables. Also uses different shapes along the lines to distinguish between the lines which can be helpful, especially if printed in black and white. The contrast between the lines is not easy to read, especially with the small size of the graphics. *There needs to be more dramatic contrast between the colours of the lines as they blur together in places, especially if there is any problems with the light in the room or focus of the machine. It also may be easier to see the lines and data points if the colour format of the slide is changed to include a white background.*

4. __E__There is a lot of information of this slide, but it does a good job of highlighting the important information, has a relevant and clear slide title and is the source of the graphic appears on the slide. *In general this slide embodies the concept of a good use of graphics on a slide. The only possible improvements include removing unnecessary information including the pink text box in the upper left hand corner, and the blue 'CDC/Janssen' text box in the lower right hand corner. These boxes are distracting and make the slide look busier than it actually is.*

5. __C__The image is far too complex and there is too much text. It would be better to remove the image and place it into a handout if appropriate. *Even if the point of this slide is to point out that there are many sites, the image is too complicated. The text could remain as a single bulleted point and the image put in a handout.*

Learning Activity 5

Develop and Deliver a Presentation

Directions: Using the techniques described in this chapter, develop a 5-10 minute PowerPoint presentation to be presented in the facility where you work. The purpose of this presentation is to orient your co-workers to district-level M&E activities and your role as an M&E officer within the district. Prior to giving your presentation in front of your colleagues, identify 2-3 colleagues who you can practice giving your presentation. Ask them for feedback following your practice. Remember to time your presentation during this practice session.

You can use the checklist below to help you develop your presentation. You may also ask a colleague to use it as a guide to evaluate your presentation. Does your presentation:

	• Answer the question, "WIIFM?" from the perspective of your audience. What will they get out of listening to your presentation? How will they benefit?
	• Identify between two and four main points you would like to make. Potential main points may include:
	▪ overview of the M&E activities in your district,
	▪ a bit of personal information about yourself,
	▪ describe some of your activities,
	▪ identify ways you and your audience may work together,
	▪ describe how the M&E activities you conduct could potentially impact the health population of the population in your district, and/or
	▪ describe how the M&E activities you conduct can potentially affect the national health programmes.
	• Organize points into a logical order.
	• Identify a beginning, middle, and end.
	• Eliminate any unnecessary information o Make handouts if necessary
	• Use appropriate humour

	Next develop PowerPoint slides to accompany your presentation. Do your slides:
	• Use a font size that is large enough for the audience to read. Usually size 32 font is used for headings and text. Size 28 can also be used for text, but nothing less than size 24 font should be included.
	• Incoporate appropriate background: o a dark blue background with white or yellow text, or o a white background with either dark blue or black text.
	• Use key words and phrases to bullet text on the slide
	• Incorporate graphics to enhance the slide. Remember the guidelines for including graphics:
	o the graphic or figure should help you to achieve the purpose of your presentation,
	o make sure the axes are labeled,
	o use colours that are easy to see at a distance,
	o if the graphics are taken from another source ensure you have cited the source on the slide, and
	o make sure the graphics chosen are clear and appropriate to your audience.
	Practice giving your presentation in front of peers
	o Make sure you can deliver the presentation in 5-10 minutes.
	Deliver your presentation to a health facility where you work.
	o Show enthusiasm,
	o use skills and techniques to manage your nervousness,
	o concentrate on delivery, and
	o face your audience, pace yourself, breath, and vary your tone of voice.

Discussion 5

Develop and Deliver a Presentation

After giving your presentation, ask yourself the following questions. Ask your colleagues questions 1 and 3.

1. What was successful about the presentation?

2. How did it feel to give the presentation?
    a. Do you think you got your message across?
    b. Did you feel prepared?
    c. Do you think your PowerPoint slides enhanced your presentation?

3. What were some areas for improvement?

Chapter Summary

As a district level M&E officer your role is to collect data and then present that data in order to inform programmatic decisions, leverage funding, and share lessons learned. You may be asked to present that data orally in a public forum. There are several tips that you can use to help you develop, design and deliver your presentation. Begin with having a clear purpose for giving your presentation. Remember that less is more as you develop your PowerPoint. The most important thing you can do to ensure a successful presentation is to prepare thoroughly and practice, practice, practice!

Strong presentation skills are an important professional skill

Self Assessment Quiz

1. What two things should you identify at the onset of developing your presentation?

2. Identify 3 guidelines when formatting the font of your slide.

3. True or False

   You are giving a presentation on the prevalence of HIV in Botswana. You find a bar graph on the DHAPC website that illustrates HIV prevalence nicely. It is ok to copy the graph from the website as long as you source where the data in the graph came from.

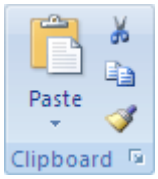
Resource 5.1:

Presenting Excel Charts and Graphs in PowerPoint

(Adapted from Microsoft Office website (MS Office Support 2010))

How to Copy a chart to a PowerPoint presentation

1. In Excel, select the embedded chart or chart sheet that you want to copy to a PowerPoint presentation.
2. On the **Home** tab, in the **Clipboard** group, click **Copy**  .



Keyboard shortcut You can also press CTRL+C.

3. Click in the PowerPoint presentation where you want to paste the copied chart.
4. On the **Home** tab, in the **Clipboard** group, click **Paste**.

Keyboard shortcut You can also press CTRL+V.

5. Click **Paste Options**  next to the chart, and then do one of the following:
   - To paste the chart with a link to its source data, click **Chart (linked to Excel data)**.
   - To paste the chart and to include access to the entire workbook in the presentation, click **Excel Chart (entire workbook)**.
   - To paste the chart as a static picture, click **Paste as Picture**.
   - To paste the chart in its original format, click **Keep Source Formatting**.
   - To paste the chart and format it by using the document theme that is applied to the presentation, click **Use Destination Theme**.

- To paste a chart in another format, click the arrow on the **Paste** button, and then click **Paste Special**. In the **As** list, select the format that you want to use.
  - To paste a chart as a static picture, click any **Picture** or **Bitmap** option. Picture format generally provides better display quality than Bitmap format.
  - Click **Microsoft Office Graphic Object** to paste a chart as a graphic object that can be edited by using the chart tools in PowerPoint. This option is not available when you are using Compatibility Mode in PowerPoint.
  - Click **Microsoft Office Excel Chart Object**, and then click **Paste link** to paste a chart as an embedded object that is linked to the source data of the original Excel chart so that it is updated when the chart data changes. **Microsoft Office Excel Chart Object** is not available when you are using Compatibility Mode in PowerPoint.

- To paste the chart as an Object Linking and Embedding (OLE) PowerPoint application icon (rather than pasting the actual chart) into a presentation, click **Microsoft Office Excel Chart Object**, and then select the **Display as icon** check box. You can click the icon to open the application, and then view the chart. **Microsoft Office Excel Chart Object** is not available when you are using Compatibility Mode in PowerPoint.

You can also create a Microsoft Office Excel 2007 chart directly in Microsoft Office PowerPoint 2007 by clicking the **Chart** button on the Ribbon (**Insert** tab, **Illustrations** group), and then by using the chart tools to modify or format the chart. Charts that you create in PowerPoint will be embedded in the PowerPoint presentation, and their source data is stored in an Excel worksheet that is incorporated in the PowerPoint file.

NOTE The Ribbon is a component of the Microsoft Office Fluent user interface

Bibliography

University of Newcastle upon Tyne, no date. *Presenting Data.* [PowerPoint presentation on internet]. Department of Chemical and Process Engineering Newcastle upon Tyne, UK. Available from: <http://lorien.ncl.ac.uk/ming/dept/tips/writing/data.pdf> [Accessed Aug. 2010]

Microsoft Office Support 2010. *Copy Excel data or Charts to PowerPoint.* Available at: http://office.microsoft.com/en-us/powerpoint-help/copy-excel-data-or-charts-to-powerpoint-HA010204555.aspx [Accessed 27 September 2010]

Self Assessment Answer Key

1. What two things should you identify at the onset of developing your presentation?

   - *Your purpose for giving the presentation.*
   - *How the audience benefit from listening to your presentation.*

2. Identify 3 guidelines when formatting the font of your slide.

   Font should be:

   - Arial or other Sans Serif font,
   - size 28 or larger,
   - use underlining and bold for emphasis,
   - use upper and lower case letters, and
   - use either a dark background with yellow or white text or a light background with black or blue text.

3. True or **False**

   You are giving a presentation on the prevalence of HIV in Botswana. You find a bar graph on the DHAPC website that illustrates the prevalence nicely. It is ok to copy the graph from the website as long as you source where the data in the graph came from.

   False- if the graph has already been created and you copied it directly into your presentation, you must cite the source where you got the graph.