MODULE 7:

DATA SOURCES FOR MALARIA SURVEILLANCE, MONITORING, AND EVALUATION

MODULE 7: DATA SOURCES FOR MALARIA SURVEILLANCE, MONITORING, AND EVALUATION

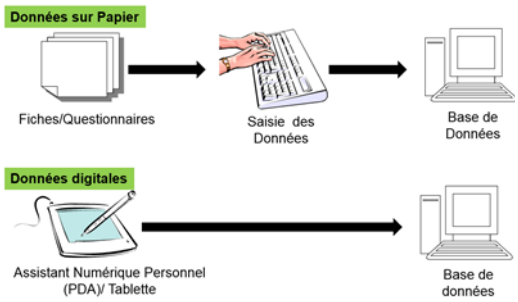
This module introduces the common data sources and systems for malaria programs and specifically discusses the different types and sources of data, the strengths and weaknesses of these data sources, and different issues affecting data quality for malaria programs.

Module Objectives

By the end of this module, you will be able to:

- Identify different data sources and systems for surveillance, monitoring, and evaluation (SME).
- Differentiate between routine and nonroutine data sources
- Identify strengths and weaknesses of common malaria data sources
- Identify different issues affecting data quality for malaria programs
- Identify strategies for linking malaria-related data sources

Figure 18. An example of data flow



Source: MEASURE Evaluation's anglophone SME regional course, Ghana, 2017

Introduction to Malaria-Related Data Sources

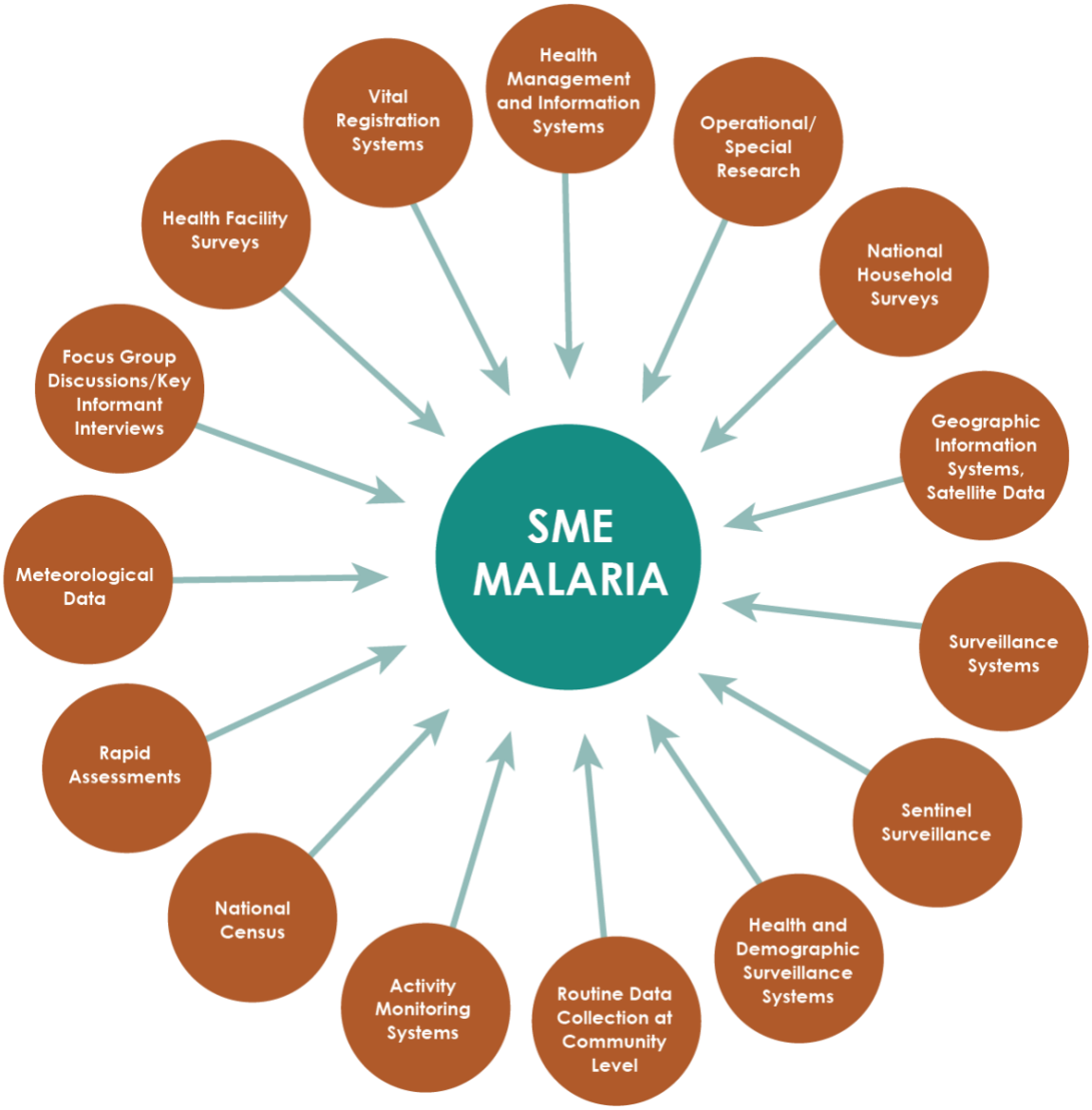
A program needs data to keep track of activities, follow the program's progress over time, make program management decisions, and use for program improvement. To collect these data, a program first requires a framework that outlines how the program is to work. Indicators should then be selected based on the framework, while also taking into account the different data sources that are available.

A number of different malaria-related data sources are available for use, as shown in the graphic on the next page. A few of the most common data sources for malaria programs include health information systems, health facility and population-based surveys, and surveillance systems. It is important to remember that for data sources to be useful, they must be complete, accurate, relevant or representative, and timely.

Potential Data Sources

Figure 19 presents the potential data sources for malaria programs, which are described in more detail in this section.

Figure 19. Potential data sources for malaria programs



Health management and information system: A system that collects and aggregates all health-related information and data at the multiple administrative levels in a country

Operational/special research: The systematic and objective assessment of the availability, accessibility, quality, or sustainability of services designed to improve service delivery

National household survey: A large-scale, nationally representative survey carried out at the household level

Geographic information system (GIS), satellite data: A system that captures, stores, analyzes, manages, and presents data that are linked to a specific location

Surveillance system: A systematic ongoing process of assessing the health status of a population by using these four main activities:

- Collection of relevant data
- Aggregation and tabulation of data
- Analysis and interpretation of data
- Dissemination and use of data and results for a decision

Sentinel surveillance: Conducted in a small number of health facilities called sentinel sites, which are selected on the basis of well-defined criteria for the collection of routine data and malaria-specific data with varying frequency

Health and demographic surveillance system (HDSS): A set of field and computing operations to handle the longitudinal follow-up of well-defined entities or primary subjects (individuals, households, etc.) and all related demographic and health outcomes within a clearly circumscribed geographic area

Routine collection at community level: In the case of malaria routine data collection, community surveillance consists of malaria detection, reporting, providing adequate response, and ensuring SME at the community level.

Activity monitoring system: A system that collects data related to the progression or implementation of a program's activities

National census: A procedure for systematically acquiring and recording information about the members of a population at a national level

Rapid assessment: A smaller-scale survey that uses a small, reliable sample and is carried out over a short duration and typically examines only a small select set of variables

Meteorological data: Data related to weather conditions; for example, information on air temperature, winds, humidity, and precipitation

Focus group discussion or key informant interviews: A qualitative data collection method for obtaining in-depth information on concepts and perceptions about a certain topic through group discussion that is guided by a facilitator

Health facility survey: Survey of a representative sample of facilities. The aim of a facility survey is usually to assess the provision and quality of services provided within the health facility.

Vital registration system: A national system for registering all births and deaths of citizens and residents of a country, including the cause of death

Basic Data Types

Data are most frequently classified as either routine or nonroutine.

Routine and Nonroutine Data

Routine data are data that are continuously or regularly collected. These include data that are collected daily, weekly, or monthly. For example, routine data could include weekly reports on how many insecticide-treated nets (ITNs) were distributed in health facilities in a community, or how many children under five came into the health center and were diagnosed with malaria.

Here are some examples of routine data sources:

- Health information systems (HIS)
- Surveillance
- Routine service reporting
- Administrative systems
- Vital registration systems

Nonroutine data are only collected periodically. For example, these data are collected quarterly, annually, or every few years. A good example of a nonroutine data source is a population-based survey that is conducted every three to five years. Because these types of surveys are large scale, they require a lot of resources and time, and therefore it is only possible to conduct them every few years.

Here are some examples of nonroutine data sources:

- Special program reporting systems
- Facility surveys
- Household surveys
- Censuses
- Interviews
- Focus groups
- Direct observations
- Research and special studies
- Rapid assessments

Qualitative and Quantitative Data

Qualitative data are descriptive and deal with aspects that cannot be measured numerically. They are most often used to help understand why something is happening. For example, to understand why ITN use is not common in a community, a researcher might conduct qualitative interviews with members of the community to better understand why they do not use ITNs.

Examples of qualitative data sources are as follows:

- In-depth interviews
- Key informant interviews
- Focus group discussions
- Direct observations

Quantitative data measure characteristics numerically; for example, by using a count or a scale. These data allow for statistical analysis that helps us understand different trends or the relationships between different factors.

Examples of quantitative data sources are as follows:

- HIS
- Surveillance
- Facility surveys
- Household surveys
- Censuses
- Routine service reporting
- Vital registration systems
- GIS
- Remote sensing

Data Source Types Quiz

Think back to the examples of data source types on the previous pages and answer the questions in the quiz. Your responses will not be graded.

Data Source Types Questions

1. Select all of the following data sources that are classified as **routine.**

    a. Surveillance
    b. Facility surveys
    c. Administrative systems
    d. Focus groups

2. Select all of the following data sources that are classified as **nonroutine.**

    a. Censuses
    b. Routine services reporting
    c. Direct observations
    d. Vital registration systems

3. The following is an example of which two data types?

    "Interviews conducted with household head members to understand why ITNs are not used in Community Y"

    a. Quantitative
    b. Routine
    c. Nonroutine
    d. Qualitative

4. The following is an example of which two data types?

    "Carrying out a nationally representative household survey to gather information on the country's population, health, and nutrition"

    a. Qualitative
    b. Quantitative
    c. Routine
    d. Nonroutine

5. The following is an example of which data type?

    "Monthly reports from health facilities on the total number of deaths from malaria are sent to the district health office where they are compiled and aggregated before being sent to the national level."

    a. Qualitative
    b. Quantitative

6. The following is an example of which data type?

   "Monthly reports from health facilities on the total number of deaths from malaria are sent to the district health office where they are compiled and aggregated before being sent to the national level."

   a. Routine
   b. Nonroutine

7. *True or False:* The following statement is an example of a **quantitative** data source.

   "Focus group discussions with caregivers about their perceptions of the quality of care at the local health facility are conducted to understand why use of health services in the community is so low."

   a. True
   b. False

8. *True or False:* The following statement is an example of a **nonroutine** data source:

   "Focus group discussions with caregivers about their perceptions of the quality of care at the local health facility are conducted to understand why use of health services in the community is so low."

   a. True
   b. False

Data Source Types Answers

1. Select all of the following data sources that are classified as **routine.**

   a. Surveillance
   c. Administrative systems

   In addition to surveillance and administrative systems, other examples of routine data sources include HIS, routine service reporting, and vital registration systems.

2. Select all of the following data sources that are classified as **nonroutine.**

   a. Censuses
   c. Direct observations

   In addition to censuses and direct observations, other examples of nonroutine data sources include special program reporting systems, facility surveys, household surveys, interviews, focus groups, research and special studies, and rapid assessments.

3. The following is an example of which two data types?

   "Interviews conducted with household head members to understand why ITNs are not used in Community Y"

   c. Nonroutine
   d. Qualitative

   The information collected is regarding perceptions and beliefs about malaria transmission and prevention; it is not information that can be numerically measured. This type of information would also not be collected on a regular basis, given that attitudes and beliefs do not change quickly or frequently. Therefore, this type of data source is considered a qualitative, nonroutine data source.

4. The following is an example of which two data types?

   "Carrying out a nationally representative household survey to gather information on the country's population, health, and nutrition"

   b. Quantitative
   d. Nonroutine

   The information collected is on characteristics (population, health, and nutrition) that can be measured numerically, and it is only collected every three to five years due to the amount of effort and resources needed to conduct the survey. Therefore, this type of data source is considered a quantitative and nonroutine data source.

5. The following is an example of which data type?

   "Monthly reports from health facilities on the total number of deaths from malaria are sent to the district health office where they are compiled and aggregated before being sent to the national level."

   b. Quantitative

   Reported deaths from malaria can be measured numerically; therefore, this is considered to be a quantitative data source.

6. The following is an example of which data type?

   "Monthly reports from health facilities on the total number of deaths from malaria are sent to the district health office where they are compiled and aggregated before being sent to the national level."

   a. Routine

   In this case, because the reports are sent on a regular basis from the health facility to the district health office, the data are considered to be a routine data source.

7. *True or False:* The following statement is an example of a **quantitative** data source.

   "Focus group discussions with caregivers about their perceptions of the quality of care at the local health facility are conducted to understand why use of health services in the community is so low."

   b. False

   The information gathered during the focus group discussions is on caregivers' perceptions, attitudes, and beliefs about the quality of care, and thus is considered descriptive and cannot be measured numerically. Therefore, this would be considered a qualitative data source.

8. *True or False:* The following statement is an example of a **nonroutine** data source.

   "Focus group discussions with caregivers about their perceptions of the quality of care at the local health facility are conducted in order to understand why use of health services in the community is so low."

   a. True

   These discussions most likely would occur once and not on a regular basis; thus, this would be considered a nonroutine data source.

Choosing an Appropriate Data Source

When you are developing your SME plan and deciding on the appropriate data sources for your program, the following questions can help you determine whether a data source is an appropriate or feasible source to use:

- Do the data exist for the specified or required time period?
- Do the data exist for the specified population? For example, are data available for most-at-risk or special populations or disaggregated by gender or specific age groups?
- Do the data exist for the specified geographic area? For example, are data available at the subnational or program area levels?
- Do the data exist for the appropriate administrative or functional level? For example, are there data available at the following levels:
    - Policy and program
    - Service environment
    - Client
    - Population
    - Spatial/geographic
- If the data do not currently exist, will it be feasible and do you have the resources to set up the necessary structures to collect the data?

Data Sources: Health Information Systems

An **HIS** is a data system that collects and aggregates all health-related information and data at the multiple administrative levels in a country from a number of different sources. The information varies by country, but it generally includes information about the delivery, cost, and use of health services, patient demographics, and health status.

Table 7. Main strengths and weaknesses of HIS in developing countries

Strengths	Weaknesses

• Ideally reflective of and integrated within health systems activities
• Collected continuously and suitable for frequent reporting
• System already exists:
  • Need fewer resources for new infrastructure or systems
  • Helps build local capacity and is sustainable
• Typically available at lowest administrative levels
	• Data not representative of population
	• Difficult to determine population at risk or denominators for coverage estimates
	• Indicators determined centrally by Ministry of Health and may not be easily altered to answer new questions
	• Quality and completeness of reporting frequently varies
	• May only cover government facilities
	• Potential for double counting, both within and between facilities

Data Sources: Heath Facility Surveys

Health facility surveys collect data specifically on the type and quality of services, as well as on outcomes and impact of services provided at health facilities. They are typically conducted on a nonroutine basis and can be conducted in both public and private health facilities.

Table 8. Main strengths and weaknesses of health facility surveys in developing countries

Strengths	Weaknesses

• Can be nationally or regionally representative
• Can be tailored to specific program needs
• Quality control may be easier than in routine systems
• Provide more detailed data than is typically available in routine systems
• Timing can coincide with program implementation
• Can cover both public and private health facilities
• Can combine with a population survey for outcome monitoring and impact evaluation
	• Overall less sustainable and not carried out routinely:
	  • Data collection is periodic and less connected to ongoing program decision making
	  • Information can become rapidly outdated
	  • Requires devoted personnel, resources, and time
	• Survey sampling design and analysis can be complex
	• Coverage and sample size constraints exist:
	  • National vs. subnational coverage
	  • May not have enough of a specific type of facility to be completely representative
	• May have small client sample sizes for some services

Data Sources: Population-Based Surveys

Population-based surveys are large, nationally representative surveys conducted typically every three to five years. They provide important population and health data at the outcome and impact levels. Examples of common population-based surveys include the DHS, the MIS, and the MICS.

Table 9. Main strengths and weaknesses of population-based surveys

Strengths	Weaknesses
• Representative of the general population, which helps eliminate selection bias if the sample is truly random • Can collect a wide range of outcome-level indicators, such as program coverage • Questionnaires can be adapted to cover specific issues and topics • Involve well-tested instruments with good quality control	• Very expensive and time-consuming to conduct; thus, are typically carried out only every three to five years • Not suitable for some types of data; for example, if collecting retrospective data, the data will be subject to recall bias • Do not provide input/process-level data • May not be adequately powered for subnational or district-level estimates • Cannot detect small changes or changes over short periods of time without large sample sizes

Data Sources: Surveillance

Surveillance refers to the ongoing, systematic collection, analysis, and interpretation of health data. It aims to provide accurate and timely information for decision-making purposes to facilitate rapid medical and programmatic response. It also serves to track outbreaks, monitor progress toward malaria elimination, and evaluate control and prevention activities.

Table 10. Main strengths and weaknesses of a surveillance system

Strengths	Weaknesses

• Very flexible and can be adapted to cover specific topics to collect data that are otherwise hard to obtain
• Can collect a wide range of data, from input data to impact data
• Especially useful and necessary when the events being monitored are rare and when a rapid response is required
	• Expensive and resource-intensive because the following are necessary:
	  • Identifying sites and providing adequate resources for them
	  • Training staff at sites
	  • Creating a system to monitor and transfer data to central authorities
	• Active surveillance even more resource-intensive

Data Sources: Sentinel Surveillance

Sentinel surveillance is the ongoing, systematic collection, analysis, and interpretation of health data carried out in a limited number of health facilities. Sentinel surveillance is useful for when:

- The routine information system is inadequate.
- There is a need for high-quality data to monitor trends.
- Disease outbreaks need to be rapidly identified.
- Geographical distribution of malaria varies greatly.

There are a number of limitations of sentinel surveillance to consider:

- It can be very costly to equip and operate.
- It requires frequent supervision.
- The limited number of sites is not representative of all health facilities, and patients are not necessarily representative of the community; thus, the data are not generalizable.
- Record keeping can be burdensome to facility staff.
- Changes in use of health services can bias trend data.

Data Sources: Health and Demographic Surveillance Systems

An **HDSS** is a longitudinal follow-up of individuals or households and all related demographic and health outcomes within a clearly defined geographic area. It allows you to assess demographic dynamics with the defined geographic region and provides risk sets and outcome measures for evaluating interventions as well as up-to-date sampling frames for identifying target populations for appraisal, intervention, and monitoring. The International Network for the Demographic Evaluation of Populations and Their Health is an example of a network of HDSS, and it currently has 56 HDSS field sites in 20 countries (map of HDSS sites).

An HDSS offers the following features:

- Conduct a baseline or initial census
- Regularly update data on households and individuals on the following events: marriages, pregnancies, births, deaths, in- and out-migration, and education and vaccination records
- Carry out verbal autopsies on all deaths to determine cause of death profiles
- Conduct morbidity surveys involving biomarkers

Table 11. Main strengths and weaknesses of health and demographic surveillance surveys

Strengths	Weaknesses
• Can monitor vital events in the demographic surveillance area: births, deaths, migrations, morbidity, socioeconomic development (poverty) • Can assess progress and impact of intervention • Can define population denominator • Could be linked with the health management information system • Could serve as sentinel sites • Could serve as operational research sites • Most field sites include malaria in their research agenda • Multidisciplinary team • Provides an ideal environment for training	• High maintenance cost • Community fatigue • Covers only a small area; thus, is unrepresentative of the national population • Has either a weak link or no link at all to the health management information system • Potential bias from over-study of the population • Data not easily accessible • Set up to address specific research questions, not necessarily set up for malaria SME

Putting It into Practice: Scenario 1

Read the following scenario and decide which data source would be the most appropriate. Remember to consider the type of data, the timing, and the level or scale at which the information is needed when trying to determine the most appropriate data source.

A donor would like to determine whether its program has been able to improve coverage of prompt and effective treatment for children under five with malaria. Thus, the donor wants to know how many children under five in the program's intervention areas received antimalarial treatment within 24 hours of the onset of malaria. What would be the most appropriate data source to use to provide the answer? *Correct answer is provided on the next page.*

1. HIS
2. Health facility surveys
3. Population-based surveys
4. National census

Correct answer in bold:

1. **HIS: The HIS is the most appropriate data source because the donor wants to know how many children received antimalarial treatment only (rather than a percentage, which would require data on all children who had malaria). The HIS also has information on confirmed cases, and these data are accessible for the specific intervened areas in the HIS. Lastly, HIS will not require extra resources to obtain the data.**

2. Health facility surveys: A health facility survey could answer the donor's question; however, it is very resource-intensive to carry out. In this scenario, the HIS is the most appropriate data source because the donor wants to know how many children received antimalarial treatment only (rather than a percentage, which would require data on all children who had malaria). Further, it would be specific to the intervened areas.

3. Population-based surveys: A population-based survey would not be an appropriate data source in this scenario for many reasons. First, these types of surveys are conducted every three to five years; it would be difficult in terms of timing. Second, these surveys provide nationally representative data and percentages, not the absolute number of children who had received antimalarial treatment. Further, it would be specific to the intervened areas. In this scenario, the HIS is the most appropriate data source because the donor wants to know how many children received antimalarial treatment only.

4. National census: A national census would not provide information on the number of children who had received prompt and effective treatment with antimalarial drugs. In this scenario, the HIS is the most appropriate data source because the donor wants to know how many children received antimalarial treatment only (rather than a percentage, which would require data on all children who had malaria). Further, it would be specific to the intervened areas.

Putting It into Practice: Scenario 2

Read the following scenario and decide which data source would be the most appropriate. Remember to consider the type of data, the timing, and the level or scale at which the information is needed when trying to determine the most appropriate data source.

The National Malaria Control Programme wants information on the percentage of homes that own at least one ITN. What would be the most appropriate data source to use to provide the answer? *Correct answer is provided on the next page.*

1. HIS
2. Health facility surveys
3. Population-based surveys
4. National census

Correct answer in bold:

1. HIS: An HIS generally includes information about the delivery, cost, and use of health services and patient demographics and health status. It would not provide information regarding household ownership of ITNs. In this scenario, population-based surveys, like the DHS and MIS, would be the most appropriate source for providing the answer. In most cases, these surveys will provide a national and regional estimate of the percentage of household ITN ownership.

2. Health facility surveys: A health facility survey would not be an appropriate data source, because it does not capture information on household ITN ownership. In addition, it would not provide an accurate reflection of all households, because it collects data only on those who attend health facilities. In this scenario, population-based surveys, like the DHS and MIS, would be the most appropriate source for providing the answer. In most cases, these surveys will provide a national and regional estimate of the percentage of household ITN ownership.

3. **Population-based surveys: In this scenario, population-based surveys, like the DHS and MIS, would be the most appropriate source for providing the answer. In most cases, these surveys will provide a national and regional estimate of the percentage of household ITN ownership.**

4. National census: A national census gathers data on the members of a population, and thus would not provide information regarding household ownership of ITNs. In this scenario, population-based surveys, like the DHS and MIS, would be the most appropriate source for providing the answer. In most cases, these surveys will provide a national and regional estimate of the percentage of household ITN ownership.

Data Quality: Why Is It Important?

Data quality refers to the worth and accuracy of the data collected. The quality of the data determines how useful the data are. Ensuring that data quality is maintained throughout the data collection, analysis, interpretation, and dissemination process is critically important. When data quality is high, the data reflect true performance and can provide vital information for decision making and program improvement. If data quality is poor, it can cause the following:

- Erroneous program management decisions and the use of additional program resources to take corrective actions
- Missed opportunities for identifying program strengths and weaknesses
- Reduced stakeholder confidence and support

© Bonnie Gillespie, Voices for a Malaria-Free Future

Dimensions of Data Quality

There are six main data quality criteria that we need to consider and ensure that we are meeting throughout the entire data collection process:

- **Validity:** Data clearly, directly, and adequately represent what was intended to be measured.
- **Reliability:** Data are collected regularly using the same methodology, and if we repeat the same procedure over and over, we end up with the same results or findings.
- **Integrity:** Data are truthful. In other words, they are free from willful or unconscious error due to manipulation or through the use of technology.
- **Precision:** Data can be used to reproduce measurements consistently and to minimize random error.
- **Timeliness:** Data are regularly collected, and up-to-date data are available when needed.
- **Completeness:** Data collected and reported are complete.

Data Quality Assurance

Data quality assurance mechanisms should be implemented at every stage of the data management process. This includes during data collection, management, analysis, interpretation, and dissemination. For more information on data quality assurance methods and different data quality assessment tools, visit the MEASURE Evaluation website.

Here are some steps your program can follow to improve the quality of the data that you collect:

- Provide written instructions for how to use data collection instruments and tools. Include these instructions on each of the instruments and tools. This will help ensure that no matter who is collecting the data, they will be collected in the same way.
- Document processes for data entry, cleaning, and management.
- Provide continuous monitoring of data collection activities and perform routine checks to ensure that instructions are being followed properly.
- Randomly sample data and verify that they are accurate.
- Take proactive steps to report, document, correct, and communicate problems that compromise the quality of the data.
- Be transparent in the data analysis techniques used and the assumptions upon which the data are based.

Linking Data Sources

Linking data refers to connecting two or more data types or sources. Linking data offers the following benefits:

- Provides context by increasing understanding and informing analyses
- Helps corroborate data quality, trends, and associations within your data. In essence, linking data helps provide further evidence on a specific program or project output, outcome, or impact. This is also often referred to as triangulating data.
- Attributes causality by linking process-level data with impact and outcome-level data to establish causality

Data can be linked from different sources, across different levels, over time, across geography, and across different sectors. Examples include the following:

- **Sources:** Linking health facility survey or HIS data with household survey data to establish change in impact or outcome
- **Levels:** Linking HIS data from district, regional, and national levels to check data quality
- **Time:** Linking data on service provision for antenatal care with birth outcomes
- **Geography:** Linking malaria cases with GIS data to assess foci of transmission
- **Sectors:** Linking malaria cases from HIS with agricultural data on rainfall levels

It is important to remember that linking data appropriately requires advanced planning, preferably prior to data collection. It should be done when sufficient, good-quality data that are plausibly connected exist. It should NOT be done if there is no logical connection between the data.

Module 7 Assessment

Questions

Correct answers are provided on the next page.

1. Data and information serve which main purpose for programs? *(Select all that apply.)*

    a. To keep track of program activities
    b. To make program management decisions
    c. To provide evidence to improve programs
    d. To demonstrate a program's progress and achievements

2. The ongoing, systematic collection, analysis, and interpretation of health data carried out in a limited number of health facilities refers to which common malaria-related data source?

    a. Health information systems
    b. Surveillance
    c. Sentinel surveillance
    d. Population-based surveys

3. A routine data source refers to data that are continuously or regularly collected. Which of the following data sources is NOT an example of a routine data source?

    a. Health information systems
    b. Facility surveys
    c. Surveillance
    d. Vital registration systems

4. Linking different data sources serves all of the following purposes, except:

    a. Helps determine whether your data are of poor quality
    b. Provides context by increasing understanding and informing analyses
    c. Helps attribute causality by linking process-level data with impact and outcome-level data to establish causality
    d. Helps corroborate data quality, trends, and associations observed within your data

Correct Answers

Correct answers are noted in bold.

1. Data and information serve which main purpose for programs? *(Select all that apply.)*

   a. To keep track of program activities
   b. To make program management decisions
   c. To provide evidence to improve programs
   d. To demonstrate a program's progress and achievements

2. The ongoing, systematic collection, analysis, and interpretation of health data carried out in a limited number of health facilities refers to which common malaria-related data source?

   c. Sentinel surveillance

   Sentinel surveillance refers to the ongoing, systematic collection, analysis, and interpretation of health data within a limited number of health facilities.

3. A routine data source refers to data that are continuously or regularly collected. Which of the following data sources is not an example of a routine data source?

   b. Facility surveys

   Facility surveys are not carried out continuously or on a regular basis; therefore, they are considered a nonroutine data source.

4. Linking different data sources serves all of the following purposes, except:

   a. Helps determine whether your data are of poor quality

   Linking data sources does not help you necessarily to determine whether your data are of poor quality. Data quality checks and audits can help you to determine whether your data are of poor quality. Linking different data sources serves the following purposes: provides context by increasing understanding and informing analyses; helps attribute causality; and helps corroborate data quality, trends, and associations observed within your data.