

# Routine Data Quality Assessment Tool



## User Manual

January 2017





# Routine Data Quality Assessment Tool

## User Manual

January 2017

**MEASURE** Evaluation—Strategic Information for South Africa (MEval-SIFSA) Project

138 Muckleneuk Street

Nieuw Muckleneuk, Pretoria

South Africa

Tel: + 27 12 346 7490

<http://www.measureevaluation.org/sifsa>

This research has been supported by the President's Emergency Plan for AIDS Relief (PEPFAR) through the United States Agency for International Development (USAID) under the terms of MEASURE Evaluation—Strategic Information for South Africa associate award AID-674-LA-13-00005. MEASURE Evaluation—SIFSA is implemented by the Carolina Population Center at the University of North Carolina at Chapel Hill, in partnership with ICF International; John Snow, Inc.; Management Sciences for Health; Palladium; and Tulane University. Views expressed are not necessarily those of PEPFAR, USAID, or the United States government. MS-17-117

ISBN: 978-1-9433-6442-8



# CONTENTS

- Acknowledgments .....iv
- Abbreviations .....v
- Section 1: A Primer on the RDQA..... 1
  - About the RDQA ..... 1
- Section 2: How to Conduct an RDQA..... 5
  - Components of the RDQA ..... 5
  - About the RDQA Workbook ..... 6
  - Detailed Steps for Conducting an RDQA ..... 8
    - Data Collection Sheets ..... 11
    - Setting Up Your Workbook ..... 12
    - Data Collection Forms ..... 12
    - Data Verifications at the Service Delivery Sites ..... 12
    - Data Verifications at the District and Regional Aggregation Sites ..... 15
    - Combining Results from Multiple Workbooks ..... 16
    - Analyzing Your Results ..... 17
    - Verification Factor (VF) ..... 17
    - System Assessment Score..... 18
    - Cross-Check Results..... 19
    - Dashboards ..... 19
      - Service Delivery, District Aggregation, and Regional Aggregation Site Dashboards ..... 19
      - Global Dashboard ..... 23
    - Summary Tables ..... 24
      - Summary of System Assessment Comments ..... 25
      - Detail of System Assessment..... 26
    - Advanced Quantitative Metrics ..... 27
    - Action Plans for System Strengthening ..... 29
      - Templates for Action Plans ..... 29
    - Bonus: Extracting the Data for Further Analysis..... 32
- Section 3: Annexes..... 32
  - Appendix 1. The Link Between the Reporting System and Data Quality..... 32
  - Appendix 2. System Assessment Complete Questionnaire, by Level ..... 34
  - Appendix 3. Instructions for Sampling Sites Using Two-Stage Cluster Sampling ..... 41

## ACKNOWLEDGMENTS

The Data Quality Audit and Routine Data Quality Assessment tools have been developed, adapted, and improved through the contributions of many technical experts from various organizations. Those most directly involved in development of the original tool were Ronald Tran Ba Huy, of the Global Fund to Fight AIDS, Tuberculosis and Malaria; David Boone, of USAID-funded MEASURE Evaluation; Cyril Pervilhac, of the World Health Organization; and Annie La Tour Schwartz, of the Office of the Global AIDS Coordinator.

This tool benefited from feedback on the Data Quality Audit tool from a number of participants in workshops and meetings held in Ghana, Nigeria, Senegal, South Africa, Switzerland, Vietnam, and the United States. In 2015, the tool was updated to support more sophisticated analysis of RDQA findings. This manual was developed to accompany the updated RDQA tool, with additional calculations and dashboards to facilitate analysis of the results.

## **ABBREVIATIONS**

DQA	data quality audit
HMIS	health management information system(s)
M&E	monitoring & evaluation
RDQA	routine data quality assessment

# Section 1

# A Primer on the RDQA

## ABOUT THE RDQA

Strong, robust systems for capturing health program data are essential to tracking progress toward health objectives, such as the Millennium Development Goals, and will be central to supporting data-informed decisions as part of the new Sustainable Development Goals.

The data quality assessment tools were originally developed as part of global efforts to combat AIDS, malaria, and tuberculosis. Ambitious plans for national programs and donor-funded projects were in the works to reduce the burden of disease in countries around the world. Measuring the success and improving the management of these initiatives is predicated on strong monitoring and evaluation (M&E) systems that produce good-quality data related to program implementation.

In the spirit of the [“Three Ones,”](#) the [“Stop TB Strategy,”](#) and the [“Roll Back Malaria Global Strategic Plan,”](#) a number of multilateral and bilateral organizations collaborated to develop the [Data Quality Audit \(DQA\) Tool](#). This tool captures high-priority indicators from HIV and AIDS, tuberculosis, and malaria programs and offers a common approach to assessing and improving overall data quality. Having a single tool helps to ensure that standards are harmonized and allows for joint implementation by partners and national programs.

Implementing the DQA tool revealed the need for a capacity-building and self-assessment version. To that end, MEASURE Evaluation (funded by the U.S. Agency for International Development), the World Health Organization, the U.S. President’s Emergency Plan for AIDS Relief, and the Global Fund to Fight AIDS, Tuberculosis and Malaria worked together to develop the Routine Data Quality Assessment (RDQA) Tool. We designed it to build the capacity of health programs to assess and improve the quality of their data. The tool has subsequently been applied many times—both by individual health programs and by country health management information systems (HMIS).

The RDQA tool verifies the quality of reported data and assesses the underlying data management and reporting systems for standard program-level output indicators.

### In summary:

The two versions of the data quality tool are:

1. The **Data Quality Audit Tool**: a tool for formal data quality audits that includes indicator-specific audit templates and guidelines for use by an external audit team to assess a program’s or project’s ability to report good-quality data on a random sample of health facilities
2. The **Routine Data Quality Assessment Tool**: a simplified version of the DQA which allows programs and projects to assess the quality of their data and strengthen their data management and reporting systems (The RDQA is generic with regard to indicators and programs and intended for use with or without rigorous sampling methods.)

**Table 1. Distinctions between DQA and RDQA**

DQA	RDQA
<ul style="list-style-type: none"> <li>• Assessment by funding agency</li> <li>• Standard approach to implementation</li> <li>• Conducted by external audit team</li> <li>• Limited input into recommendations by programs</li> <li>• Program and indicator specific</li> <li>• Utilizes a modified two-stage cluster sampling technique for the selection of health facilities</li> <li>• Every several years for priority indicators</li> </ul>	<ul style="list-style-type: none"> <li>• Self-assessment by program</li> <li>• Flexible use by programs for monitoring and supervision or to prepare for an external audit</li> <li>• Program makes and implements its own action plan</li> <li>• Generic to program and indicator</li> <li>• Convenience sampling</li> <li>• Regular (repeated) data quality measurements during routine supervision</li> </ul>

## CONCEPTUAL FRAMEWORK FOR DATA QUALITY

The RDQA approach assesses the dimensions of data quality and the functional components of the data management system needed to ensure data quality.

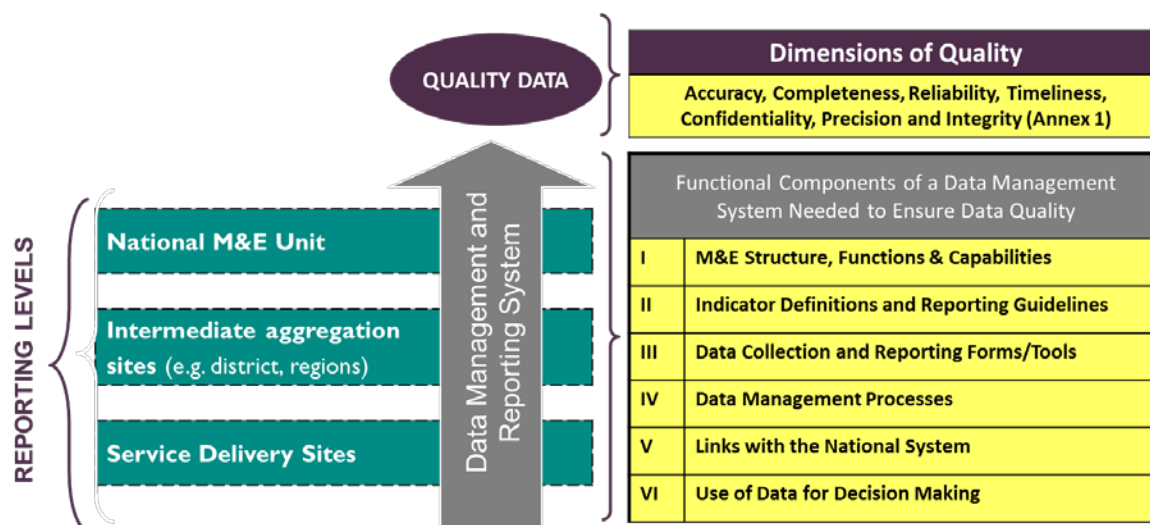
The conceptual framework for the DQA and RDQA is illustrated in Figure 1 (below). As you can see, quality data (measured through the dimensions of quality) are generated through a strong data management and reporting system (made up of the various functional components) that spans the different levels of the system.

### **Want to learn more about the RDQA Tool in action? Check out these resources.**

- MEASURE Evaluation’s Data Quality Assurance Tools  
<http://www.cpc.unc.edu/measure/our-work/data-quality>
- Botswana’s Integration of Data Quality into Standard Operating Procedures: an Adaptation of the RDQA  
<http://www.cpc.unc.edu/measure/publications/sr-13-79>



Figure 1. Conceptual framework for the DQA and RDQA



Generally, the quality of reported data is dependent on the underlying data management and reporting systems; stronger systems should produce better quality data.

In other words, for good quality data to be produced by and flow through a data management system, key functional components need to be in place at all levels of the system—the points of service delivery, the intermediate level(s) where the data are aggregated (e.g., districts, regions), and the M&E unit at the highest level to which data are reported.

The RDQA tool is designed to facilitate three key actions central to improving data quality:

1. Verify the quality of the data,
2. Assess the system that produces that data, and
3. Develop action plans to improve both.

The tool itself is designed for **data collection** and also includes a number of dashboards that summarize findings at various levels and facilitate **analysis** of the data. We'll review all of the sections of the application in detail in subsequent sections.

## OBJECTIVES OF THE RDQA

By using the RDQA tool, you can achieve three main objectives.



## USES OF THE RDQA TOOL

The RDQA is designed to be flexible in use and serve multiple purposes. Some potential uses of the tool are listed below, though it is most effective when used routinely.

### **RDQA Use Case**

*Routine data quality checks as part of ongoing supervision*

### **Example**

Routine data quality checks can be included in already planned supervision visits at the service delivery sites.

*Initial and follow-up assessments of data management and reporting systems*

Repeated assessments (e.g., biannually or annually) of a system's ability to collect and report quality data at all levels can be used to identify gaps and monitor necessary improvements.

*Strengthening program staff's capacity in data management and reporting*

M&E staff can be trained on the RDQA and be sensitized to the need to strengthen the key functional areas linked to data management and reporting in order to produce quality data.

*Preparation for a formal data quality audit*

The RDQA tool can help identify data quality issues and areas of weakness in the data management and reporting system that would need to be strengthened to increase readiness for a formal data quality audit.

*External assessment by partners of the quality of data*

Such use of the RDQA for external assessments could be more frequent, more streamlined, and less resource intensive than comprehensive data quality audits that use the DQA version for auditing.

The potential users of the RDQA are program managers, supervisors, and M&E staff at national and subnational levels, as well as donors and other stakeholders.

## ETHICAL CONSIDERATIONS

Data quality assessments must be conducted with the utmost adherence to the ethical standards of the country. Though assessment personnel may require access to personal information (for example, medical records), this should not be shared with nonassessment staff or disclosed in any way during the conduct of the assessment.

## Section 2

# How to Conduct an RDQA

## COMPONENTS OF THE RDQA

We'll explain the process in greater detail, but first it's important to understand that the RDQA has two main components. The RDQA tool facilitates assessment of the quality of selected indicator data (**data verifications**) and the strength of the overall data management and reporting system (**system assessment**).

Let's break down what that means in more detail.

### What are data verifications?

Part 1 of the RDQA Tool facilitates a quantitative comparison of recounted to reported data and a review of the timeliness, completeness, and availability of reports. The purpose of this part of the RDQA is to assess if 1) service delivery and intermediate aggregation sites are collecting and reporting data accurately, completely, and on time; and 2) whether the data agrees with reported results from other data sources.

### What is the system assessment?

Part 2 of the RDQA Tool enables qualitative assessment of the relative strengths and weaknesses of functional areas of a data management and reporting system. The purpose of assessing the data management and reporting system is to identify potential threats to data quality posed by the design and implementation of data management and reporting systems.

The six functional areas of a data management and reporting system (shown in the earlier data quality framework) are as follows:

- M&E functions, roles, and capabilities
- Indicator definitions and reporting guidelines
- Data collection and reporting forms and tools
- Data management processes
- Links with national reporting system
- Use of data for decision making

Appendix 2 lists the questions posed for the systems assessment and the levels to which the questions pertain.

The RDQA tool facilitates assessment of the **quality of selected indicator data** (data verifications) and the **strength of the overall data management and reporting system** (system assessment).

## Do I have to complete the full tool, including both the data verifications and the system assessment?

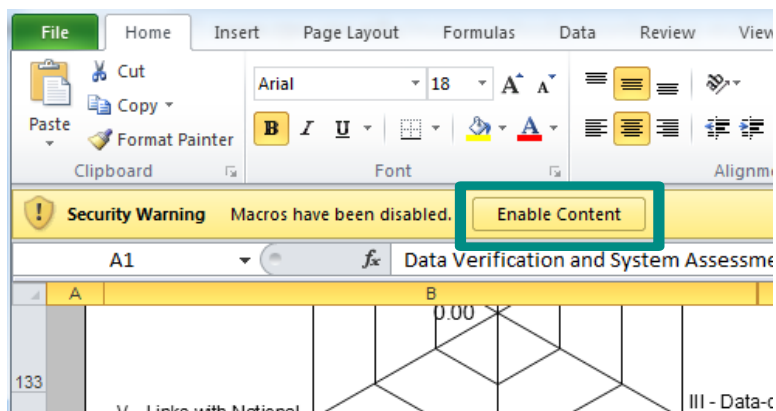
While it is recommended that both parts of the RDQA Tool—data verifications and system assessment—be used to fully assess data quality, depending on the assessment objectives, one or both of these protocols can be applied and adapted to local contexts.

Parts 1 and 2 of the RDQA Tool can be implemented, at any or all levels of the data management and reporting system: M&E unit; intermediate aggregation levels (e.g. region and district); and/or service delivery points. However, **the data verification aspect of the tool is vital** and should be conducted regularly. The system assessment protocol, which verifies the presence and adequacy of program inputs, could be applied less often.

## ABOUT THE RDQA WORKBOOK

Let's explore the RDQA tool itself in more detail.

When you open the RDQA tool in Microsoft Excel, you'll see a pop-up bar in the upper left of the workbook. Click on "Enable Content" as illustrated in the figure below to use the tool.



Across the bottom of the workbook, you'll see a series of **19 worksheets**. Each sheet fills one of the following functions:

- Gives you **information** about how to use the tool,
- Facilitates **data collection**, or
- Provides an **output in the form of a graph or a table** to use to develop your action plan.

### Don't be overwhelmed!

Many of these sheets have tables and graphs automatically generated to facilitate your analysis of the results. We've color coded the sheets in the table on the next page **to highlight which sheets require data entry and which are automatically calculated or aggregated by the tool**. We'll talk more about the specific use of these in the subsequent sections of this manual.

Gives information	Data collection	Analysis (automatically generated)
-------------------	-----------------	------------------------------------

1	Start	To select the number of service sites, district aggregation sites, and regional aggregation sites to be included in the RDQA
2	Instructions	Details on how to use the Excel-based tool
3	Information page	To record the country, program/project, indicator reviewed, reporting period reviewed, and the assessment team
4	Service delivery site 1	<ul style="list-style-type: none"> <li>Record results of the assessment on data verifications, systems assessment and cross-checks at the service delivery level</li> <li>Record recommendations for the service site</li> </ul>
5	Service site summary	Displays a dashboard of results of the data verification and systems assessment for the service site (more detail provided below)
6	District site 1	<ul style="list-style-type: none"> <li>Record results of the assessment on data verifications and systems assessment at the <b>regional</b> aggregation site</li> <li>Record recommendations for the intermediate aggregation level site</li> <li>Display a dashboard of results of the data verification and systems assessment for the <b>regional</b> aggregation site</li> </ul>
7	District-level summary	Displays a dashboard of results of the data verification and systems assessment for the intermediate aggregation level sites
8	Regional site 1	<ul style="list-style-type: none"> <li>Record results of the assessment on data verifications and systems assessment at the <b>regional</b> aggregation site</li> <li>Record recommendations for the intermediate aggregation level site</li> <li>Display a dashboard of results of the data verification and systems assessment for the <b>regional</b> aggregation site</li> </ul>
9	Regional-level summary	To display a dashboard of results of the data verification and systems assessment for the regional aggregation sites
10	National-level M&E unit	<ul style="list-style-type: none"> <li>Record results of the assessment on data verifications and systems assessment at the M&amp;E unit</li> <li>Record follow-up recommendations and an action plan based on the RDQA</li> <li>Display a dashboard of results of the data verification and systems assessment for the M&amp;E unit</li> </ul>
11	System assessment summary	Displays a summary table of system assessment findings (scores by data quality dimension) for each service delivery site, district aggregation site, regional aggregation site, and the National M&E unit
12	Global dashboard	To display in graphic form aggregated results from all levels of the assessment (more detail provided below)
13	Overall action plan	To consolidate recommendations from each level into an overall action plan based on the RDQA (more detail provided below)

14	Quantitative comments: district	Dashboard display to group comments on <b>district level</b> quantitative findings so that the analyst can scan a column on one page to know what the assessors are saying about why things are as they are
15	Quantitative comments: facility	Dashboard display to group comments on <b>facility level</b> quantitative findings so that the analyst can scan a column on one page to know what the assessors are saying about why things are as they are
16	Summary of SA comments	Dashboard display to group comments on <b>system assessment</b> quantitative findings so that the analyst can scan a column on one page to know what the assessors are saying about why things are as they are
17	Detail of system assessment	Detailed table display with system assessment scores by question from each data collection site across levels
18	Advanced quantitative metrics	Detailed analysis of data accuracy for up to four indicators Provides insight on accuracy for unique situations (e.g., wide range in service volume among sites, or combination of over and underreporting)
19	All SA questions	Table showing the questions included in the system assessment by data quality dimension and level  Can be used as a key or reference when conducting the assessment or reviewing findings

## DETAILED STEPS FOR CONDUCTING AN RDQA

Now let’s walk through the detailed steps for conducting an RDQA.

This process can be customized to your unique needs, depending on what your purpose is for conducting an RDQA.



## Step 1. Determine the purpose of the RDQA.

You can find a more detailed explanation of the various reasons for conducting an RDQA in an earlier section, but as a reminder, you could conduct an RDQA for:

- Routine data quality checks as part of ongoing supervision
- Initial and follow-up assessments of data management and reporting systems
- Strengthening program staff's capacity in data management and reporting
- Preparation for a formal data quality audit
- External assessment by partners of the quality of data



## Step 2. Select levels and sites to be included (depends on resources and time)

Once the purpose has been determined, the second step in the RDQA is to decide what levels of the data management and reporting system will be included in the assessment: service delivery sites, intermediate aggregation levels (e.g., regions, districts), and/or the central M&E unit.

The levels should be determined once the appropriate reporting levels have been identified and “mapped” (e.g., there are 100 sites providing the services in 10 districts—reports from sites are sent to districts, which then send aggregated reports to the M&E Unit). In some cases, the data flow will include more than one intermediate level (e.g., regions, provinces or states, or multiple levels of program organizations).

It is not necessary to visit all the reporting sites in a given program to determine the quality of the data. Random sampling techniques can be used to select a representative group of sites whose data quality is indicative of data quality for the whole program.



Depending on the volume of service of the program—e.g., number of people treated with anti-retroviral treatment (ART)—the number of service delivery sites, and the quality of the data, as few as a dozen sites can be assessed to obtain a reasonable estimate of data quality for the program. Please see Appendix 3 for instructions on how to sample sites using two-stage cluster sampling.

Depending on the volume of service of the program, the number of service delivery sites and the quality of the data, as few as a dozen sites can be assessed to obtain a reasonable estimate of data quality for the program.

Precise measures of data accuracy are difficult to obtain for an entire program using these methods. “Reasonable estimates” of data accuracy are generally sufficient for the purposes of strengthening data quality, capacity building, or preparing for external auditing.

For a more rigorous sampling method leading to more precise estimates of accuracy, please see **the Data Quality Audit Tool and Guidelines** on the MEASURE Evaluation website.<sup>1</sup>



### Step 3. Identify indicators, data sources, and reporting period.

The RDQA is designed to assess the quality of data and underlying systems related to indicators that are reported to programs or donors to measure success in program areas related to specific diseases during specific reporting periods.

For each program area, a number of indicators are measured through various data sources. Here are two examples:

For tuberculosis, in the **program area** *treatment*, the international community has agreed to the harmonized **indicator**: *Number of new smear positive TB cases that successfully complete treatment*. The **data source** for this indicator is facility-based and the **source documents** are the district TB register along with the facility register and patient treatment cards.

<sup>1</sup> You can find the Data Quality Audit Tool guidelines at: <http://www.cpc.unc.edu/measure/tools/monitoring-evaluation-systems/data-quality-assurance-tools/dqa-auditing-tool-implentation-guidelines.pdf>. Additional MEASURE Evaluation data quality assurance resources are available at: <http://www.cpc.unc.edu/measure/our-work/data-quality>



For HIV/AIDS, under the U.S. President’s Initiative for AIDS Relief, a **program area** is *Orphans and Vulnerable Children*, and an **indicator** is: *Number of OVC served by OVC programs* (disaggregated by male, female, primary direct, and supplemental direct). The **data source** for this indicator will be at community-based organizations that serve OVC and the **source documents** will be client records (intake forms, daily logs, registers, etc).

When planning the RDQA, it is important to determine the data sources for the indicator(s) selected, and to determine the time period for assessing the reported data. For example, if data are reported every six months, the reporting period for the RDQA could be January–June, 2007. Using a specified reporting period gives a reference from which to compare the “recounted” data.



## Step 4. Conduct Site Visits.

Sites should be notified prior to the visit for the data quality assessment. This notification is important in order for appropriate staff to be available to answer the questions in the checklist and to facilitate the data verification by providing access to relevant source documents.

During site visits, the relevant sections of the appropriate checklists in the Excel file are filled out (e.g., the service site checklist at service sites, etc.). These checklists are completed following interviews of relevant staff and reviews of site documentation.

### Data Collection Sheets

At each site, you’ll need to collect data using the RDQA tool. If you have a team that is visiting multiple sites simultaneously, you can use multiple workbooks to collect your data and compile the data in one workbook when the site visits are complete.

## Setting Up Your Workbook

Using the drop down lists on the *START* page of the RDQA workbook, select the appropriate number of regions, districts, and service delivery sites to be reviewed.



The image shows a blue rectangular panel with three rows of controls. Each row consists of a label on the left and a dropdown menu on the right. The first row is labeled 'Number of Regional Aggregation Sites' and has a dropdown menu with the number '2' selected. The second row is labeled 'Number of District Aggregation Sites' and has a dropdown menu with the number '4' selected. The third row is labeled 'Number of Service Delivery Sites' and has a dropdown menu with the number '8' selected.

The appropriate number of worksheets will automatically appear in the RDQA workbook (up to four regions, eight districts, and 24 service delivery sites).

## Data Collection Forms

The main “data collection” sheets of the RDQA Tool are the service delivery site, district and regional aggregation sites, and National M&E Unit sheets. Each of these sheets contains two parts for data collection: (1) data verifications, and (2) system assessment.

When you’re conducting an RDQA, you may be completing the full tool **OR** only the data verifications if the tool is being used for routine monitoring of data quality.

## Data Verifications at the Service Delivery Sites

At the service delivery level, Part 1, **Data Verification** of the RDQA Excel protocol has three parts:

- Documentation review
- Recounting reported results
- Cross-check reported results with other data sources

The sheets for regional and district aggregation sites and the M&E unit are found in the MS Excel spreadsheet.

For each of the stages of the **Data Verification** at the **Service Delivery Site**, let’s look at why the exercise is completed, how to complete the form, and a sample form.

### 1. Documentation review

For each of the indicators selected, review the availability and completeness of the indicator source documents for the selected reporting period. Select *Yes* (available and complete) or *No* (not available and/or complete) for each indicator, following the data collection form.

Where data is not available and/or complete, note in the comments what was unavailable or incomplete in the cells at right on the form (not pictured here).

<b>A - Documentation Review:</b>		Number Counseled and tested	Number testing positive	Number on Prophylaxis	Number of infants PCR tested
	<i>Review availability and completeness of all indicator source documents for the selected reporting period.</i>				
1	Review available data sources for the reporting period being verified. Are all necessary data sources available for review?	Yes	Yes	Yes	Yes
	<u>If no</u> , determine how this might have affected reported numbers.				
2	Are all available data sources complete?	No	Yes	Yes	No
	<u>If no</u> , determine how this might have affected reported numbers.				
3	Review the dates on the data sources. Do all dates fall within the reporting period?	No	Yes	No	No
	<u>If no</u> , determine how this might have affected reported numbers.				

## 2. Recounting reported results

Recount results from source documents, compare the verified numbers to the site reported numbers and explain discrepancies (if any) in the comments at right on the form (not pictured here).

<b>B - Recounting reported Results:</b>					
<i>Recount results from source documents, compare the verified numbers to the site reported numbers and explain discrepancies (if any).</i>					
4	Recount the number of people, cases or events during the reporting period by reviewing the data source. <b>[A]</b>	100	90	60	15
5	Enter the number of people, cases or events reported by the site during the reporting period from the site summary report. <b>[B]</b>	120	40	40	10
6	<b>Calculate the ratio of recounted to reported results. [A/B]</b>	<b>83%</b>	<b>225%</b>	<b>150%</b>	<b>150%</b>
7	What are the reasons for the discrepancy (if any) observed (i.e., data entry errors, arithmetic errors, missing data source, other)?				

## 3. Cross-check reported results

When collecting your data at the service delivery site level, you will be asked to cross-check your results with other available data. This should include the following three cross-checks:

- Primary source to secondary source
- Secondary source to primary source (or a different primary to secondary)
- Commodities management systems

Each of the three cross-checks can be captured on the data collection form for individual service delivery sites. Cross-checks can be performed by examining separate inventory records documenting the quantities of treatment drugs, test-kits, or ITNs purchased and delivered during the reporting period to see if these numbers corroborate the reported results.

You can include additional cross-checks at your discretion. These could include, for example, randomly selecting 20 patient cards and verifying if these patients were recorded in the unit, laboratory, or pharmacy registers.

To the extent relevant, the **cross-checks should be performed in both directions** (for example, from patient treatment cards to the register and from the register to patient treatment cards).

You can find more detail on the link between a data management and reporting system and the components of data quality in **Appendix 1**. An example of completed cross-checks is included on the subsequent page.

<b>Cross-Check 1:</b> Verify the <b>primary source of data</b> against the <b>secondary source of data</b> . (If the cross-check conducted is different than the one that was planned, specify the cross-check performed in the cells to the right.)		ANC Register to Testing Register	ANC Register to Testing Register	ANC Register to Patient Card	EID Register to PCR test register
1.1	If feasible, select 5% of units being counted (or at least 20 units) in the secondary data source. How many units were selected?	20	20	20	10
1.2	For how many units does the information for the indicator in the secondary data source match the information in the the primary data source?	8	11	10	4
1.3	<b>Calculate % difference for cross check 1:</b> If difference is below 90%, select an additional 5% of individual client records (or at least an extra 10 records) and redo the calculation (ADD the numbers to the existing numbers in the above cells).	40.0%	55.0%	50.0%	40.0%
<b>CROSS-CHECK 2:</b> Cross-check <b>secondary data source</b> with the <b>primary data source</b> . (If cross-checks are different from the planned cross-check, i.e. the cross-checks entered on the Information Page, specify the cross-checks performed in the comment cells to the right.)		Testing Register to ANC Register	Testing Register to ANC Register	Patient Card to ANC Register	PCR Test register to EID Register
2.1	If feasible, select 5% of units being counted (or at least 20 units) in the secondary data source. How many units were selected?	10	10	10	10
2.2	For how many units does the information for the indicator in the secondary data source match the information in the the primary data source?	5	3	4	6
2.3	<b>Calculate % difference for cross check 2:</b> If difference is below 90%, select an additional 5% of individual client records (or at least an extra 10 records) and redo the calculation (ADD the numbers to the existing numbers in the above cells).	50.0%	30.0%	40.0%	60.0%
<b>CROSS-CHECK 3:</b> Between <b>stock movement</b> and <b>commodities distributed</b> by the site.		Number of HIV test kits	Confirmation Notifications	Number of units prophylaxis prescribed	PCR Tests used
3.1	Enter the number of commodities in stock at the site at the beginning of the reporting period (initial in stock).	110	110	250	40
3.2	Enter the number of commodities received by the site during the reporting period.	50	50	200	100
3.3	Enter the number of commodities in stock at the site at the end of the reporting period (closing in stock).	60	40	250	120
3.4	Enter the number of commodities distributed by the site during the reporting period.	90	100	200	15
3.5	<b>Calculate % difference in cross check 3.</b> (i.e., Distributed / (Beginning stock + Stock received - End stock)) If there is a discrepancy between in stock and distributed commodities during the reporting period, determine why, and if and how the store or site addressed this discrepancy.	90.0%	83.3%	100.0%	75.0%

## Data Verifications at the District and Regional Aggregation Sites

At the district and regional levels, your focus is on the accuracy of the reporting. Your source documents at these aggregation levels for data verifications will be the reports from the service delivery sites within that district or the districts within the region.

At the aggregation sites, Part 1, **Data Verification** of the RQDA Excel protocol has two parts (below). Each of these parts requires review of the source documents, and percent calculations for scores are created automatically by the tool.

### 1. Recounting reported results

Recount results from the periodic reports sent from service sites to the district and compare to the value reported by the district. Explain discrepancies (if any).

<i>Recount results from the periodic reports sent from service sites to the District and compare to the value reported by the District. Explain discrepancies (if any).</i>		<b>Number Counseled and tested</b>	<b>Number testing positive</b>	<b>Number on Prophylaxis</b>	<b>Number of infants PCR tested</b>
1	Re-aggregate the numbers from the reports received from all Service Delivery Sites. What is the re-aggregated number? <b>[A]</b>	60	50	29	10
2	What aggregated result was contained in the summary report prepared by the District (and submitted to the next reporting level)? <b>[B]</b>	75	45	22	8
3	<b>Calculate the ratio of recounted to reported results. [A/B]</b>	<b>80%</b>	<b>111%</b>	<b>132%</b>	<b>125%</b>
4	What are the reasons for the discrepancy (if any) observed (i.e., data entry errors, arithmetic errors, missing data source, other)?				

### 2. Reporting performance

Review availability, completeness, and timeliness of reports from all service delivery sites. How many reports should there have been from all sites? How many are there? Were they received on time? Are they complete?

5	How many reports should there have been from all service delivery sites? <b>[A]</b>		12
6	How many reports are there? <b>[B]</b>		9
7	Calculate % <b>Available Reports [B/A]</b>		75%
8	Check the dates on the reports received. How many reports were received on time? (i.e., received by the due date). <b>[C]</b>		2
9	Calculate % <b>On time Reports [C/B]</b>		22%
10	How many reports were complete? (i.e., complete means that the report contained all the required indicator data, the date of reception of the report by the District, and the authentication by the responsible staff). <b>[D]</b>		7
11	Calculate % <b>Complete Reports [D/B]</b>		78%

## Combining Results from Multiple Workbooks

Often it is necessary to combine the results from separate workbooks. In this case, the most efficient way to transfer data between workbooks is to copy and paste an entire worksheet from one workbook to another.

To copy an entire worksheet, select the entire sheet by clicking in the upper left-hand corner of the worksheet (there is a tiny square to the left of the column heading “A” and above the row heading “1” — click this square to select the whole sheet) then copy the sheet into memory (control-c). Then, navigate to the destination workbook, select the entire worksheet where you want to paste the copied results and paste (control-v).

When copying sheets from one workbook to another containing formulas in the copied sheet that point to other sheets in the workbook (i.e., sheets other than those being copied), the formula will retain the original workbook identifiers in the formula. This pertains mostly to the “header” information for each site reviewed. These formulas can be updated in a batch once you have copied the sheets. Correct the formulas that draw the names of the sites from the “Information Page” onto the site-specific pages by running the macro “fixformulas” to fix the formulas automatically (go to “macros” on the “view” ribbon, select “view macros,” then select the macro and click “run”).

Also, the graphics for site-level pages will not replace, or copy over, the existing graphics. The graphics from the copied page will be placed on top of the graphics from the destination page. The graphics from the copied page will still point to (i.e., have as data source) the originating sheet. You will need to delete the “new” graphics from the destination pages to reveal the appropriate graphics. The graphics can also be deleted from the originating sheet before the copy and paste step.

If, for whatever reason, the graphics do not point to the appropriate data source they can be updated in a batch to correct the data source by running a couple simple macros. Run the macro “fixcharts\_x\_attrib\_SDP” to revise the charts’ sources to the appropriate site-level pages, and “fixcharts\_x\_attrib\_Dist” to revise the charts’ sources for district and region page graphics.

**Conducting assessments at multiple sites may require that you combine workbooks to review your results together. The RDQA tool makes that simple and straightforward.**



## Step 5. Review outputs and findings.

The outputs from the RDQA described above should be reviewed for each site visited. Site-specific summary findings in the form of recommendations are noted at each site visited.

The RDQA tool is designed to facilitate review and analysis of results – the tool **automatically generates multiple dashboards and tables** based on the data collected during the assessment.

### Analyzing Your Results

As you’ve seen, the RDQA is an Excel-based tool. This allows for flexibility: you can choose to fill the form on the computer as you conduct the assessment or print the sheets and fill them by hand, with data entered at a later point. Excel also facilitates the generation of graphs and summary tables once the data collection pages are completed.

Across the levels of the system, there are two key metrics you should know how to interpret and use as you analyze your results and use them to create action plans for system strengthening.

### Verification Factor (VF)

**What it is**

The VF is the key metric for assessing the quality of the reported data, by comparing the reported data to the source data (i.e., the register or other HMIS record at the service delivery point)

**Scoring scale**

**Scale:** 0-200%

**What the scores mean**

**Values >100%:** Under-reporting (i.e., recounted data from the primary source document is higher than the reported value)

→ This means the report says there were fewer services rendered than your source document shows.

**100%:** Perfect data quality (exact match of recounted to reported), which is rare.

**Values <100%:** Over-reporting (i.e., recounted data from the primary source document is lower than the reported value)

→ This means the report says there were more services rendered than your source document shows.

**Acceptable values:** For the purposes of the RDQA, 90-110% is considered acceptable (within a 10% range of a perfect match).

**Where you'll see it in the results**

Each of the dashboards for the individual sites and the summary dashboard will have a bar chart of the verification factors for each indicator on the chart titled "Data Verifications." You'll see a band that shows the acceptable range of 90-110%. Bars that fall outside of this band indicate the site is over or underreporting.

## System Assessment Score

**What it is**

For each of the six dimensions of data quality, the RDQA tool has a series of questions. The system assessment score for each dimension is the **average** of the scores across the questions for that dimension.

This tells us the strength of the system for the individual dimensions, which can help with identifying what the site is doing well and where there are opportunities for improvements.

**Scoring scale**

**Scale:** 1-3

The scores correspond to each of the responses in the system assessment as follows:

1 = No, not at all

2 = Yes, partly

3 = Yes, completely

Then, for each component, the scores for each individual question are averaged to create an aggregate score. The lowest possible aggregate score is 1, meaning all questions had a "no" response for that component; the highest possible aggregate score is 3, meaning all questions had a "yes" response for that component.

**What the scores mean**

The closer an aggregate score is to 3, the stronger the site or level of the system is functioning for that component. The lower the score, the poorer the performance.

**Where you'll see it in the results**

Each of the dashboards for the individual sites and the summary dashboard will have a spider graph that shows the results of the assessment for each of the M&E system components. Read on to learn more about how to interpret this chart type.



In addition, your cross-check results further validate your results by comparing different data sources.

## Cross-Check Results

<b>What it is</b>	Cross-checks compare a subset of units in your source data to a secondary source. The value reported for your cross-check indicates the percent of the source records you selected that were also reported in the comparison document.
<b>Scoring scale</b>	0-100%
<b>What the scores mean</b>	<p>The lower the value, the fewer of your source records also appeared in a second data source.</p> <p>If you conduct the cross-checks with ~5% of your source records and the cross-check value is &lt;90% (more than 1 in 10 records was missing in your secondary document), select another ~5% or 10 records (whichever is greater) to add to your sample.</p>
<b>Where you'll see it in the results</b>	The cross-checks are an additional means of assessing data quality at the service delivery point and are included in the individual and aggregate dashboards for the service delivery sites.

## Dashboards

The RDQA tool is designed to produce outputs that facilitate your analysis and use of the data to understand the current status of the data quality for your selected indicators and develop a targeted action plan.

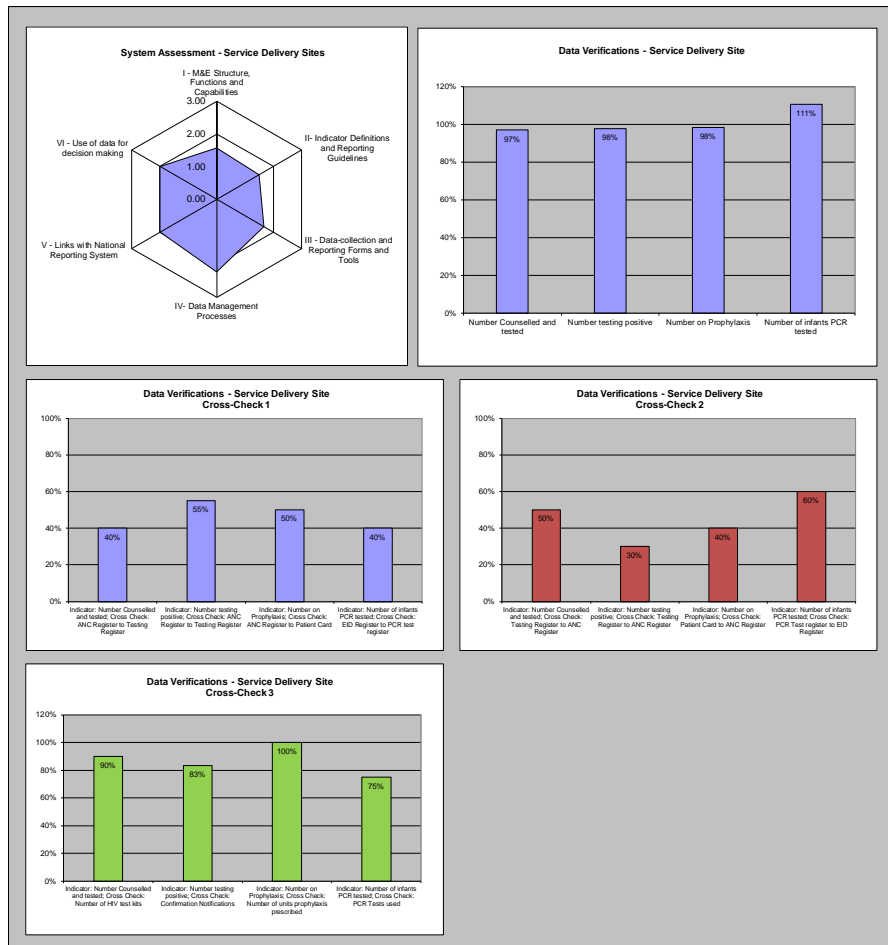
When completed electronically, a number of dashboards produce graphics of summary statistics for each site or level of the reporting system and a “global” dashboard that aggregates the results from all levels and sites included in the assessment (Figure 4).

Across the different dashboards, you'll find similar graphs for each level. There are two kinds of graphs you need to understand how to interpret in order to review the data in the summary dashboards.

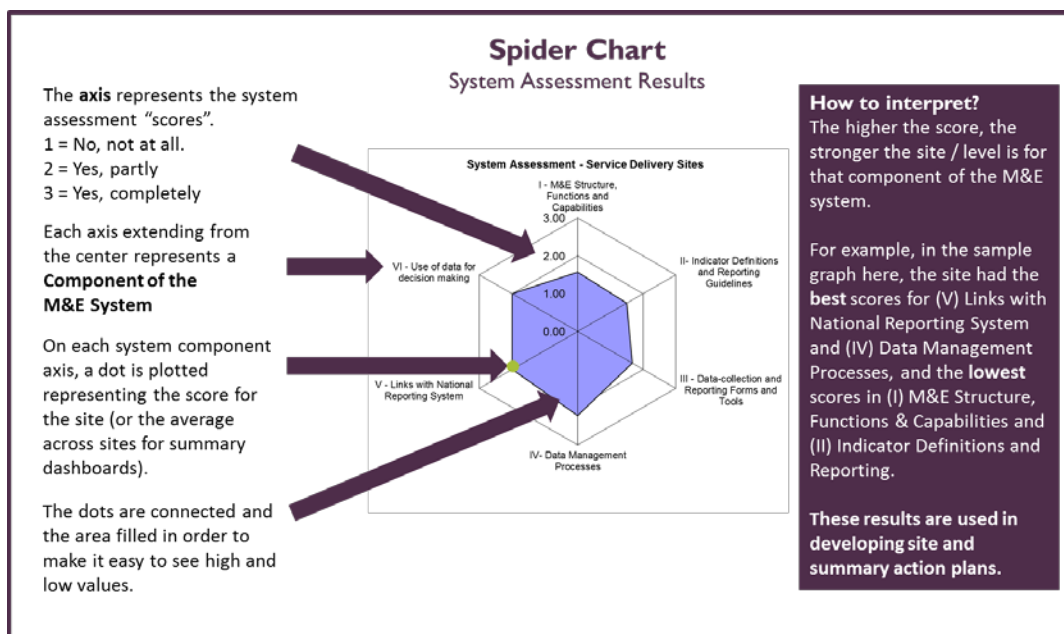
## Service Delivery, District Aggregation, and Regional Aggregation Site Dashboards

There are two types of dashboards for each of these levels: a small dashboard at the bottom of the sheet for each individual site, and a summary dashboard for each level.

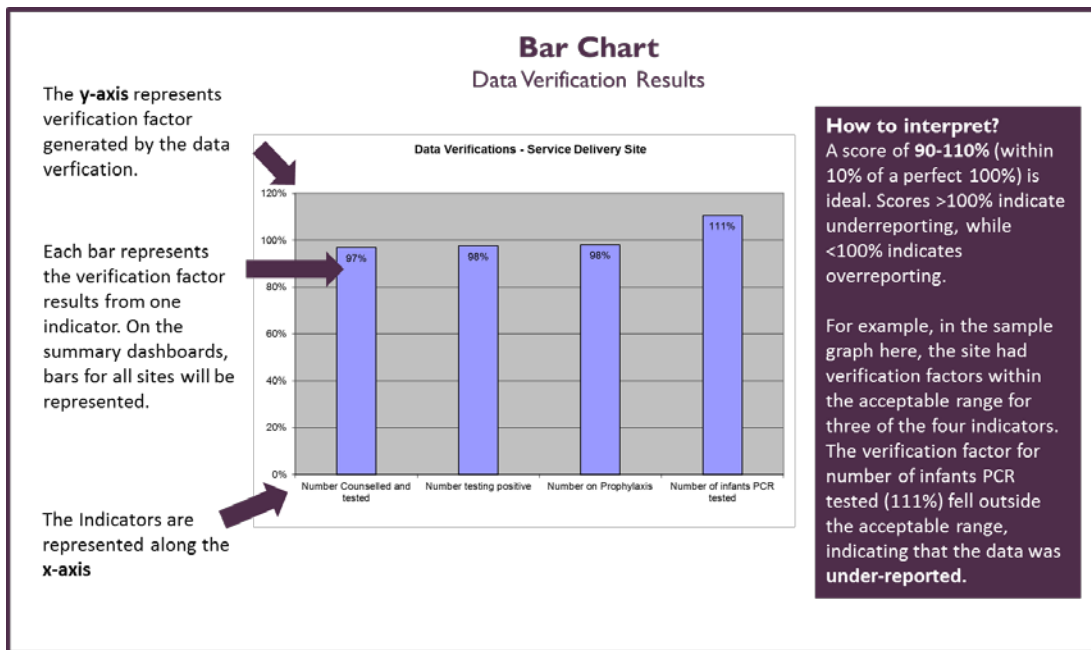
The **site-specific dashboard** displays five graphs for each site visited.



The **spider chart** below displays qualitative data generated from the assessment of the data management and reporting system and can be used to prioritize areas for improvement.

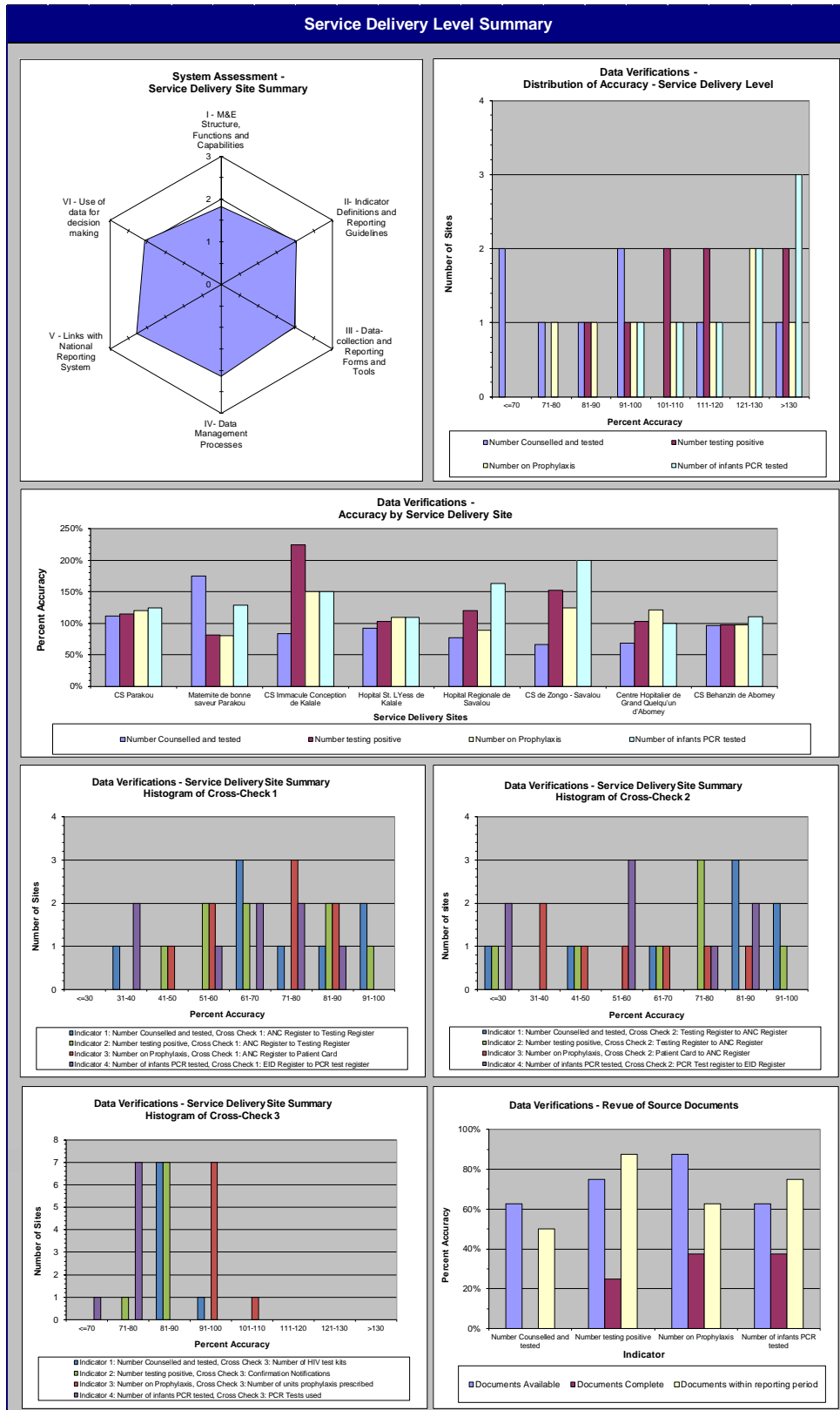


The **bar chart** below shows the quantitative data generated from the data verifications; these can be used to plan for data quality improvement.



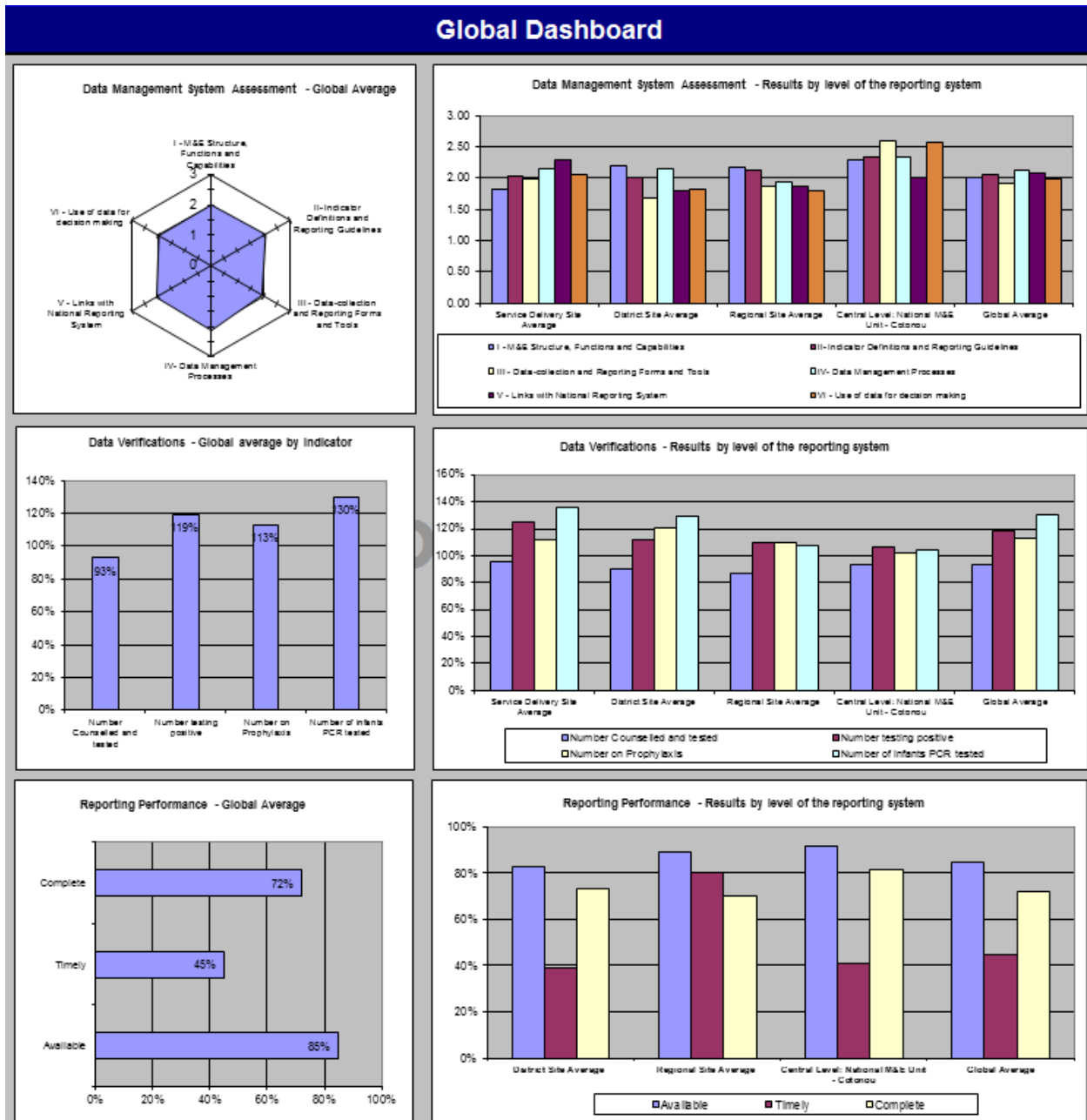
The three bar charts in the center and lower rows of the dashboard display the results from the three cross-checks. These can be used to determine the extent to which alternative sources of data agree with the results compiled from primary source documents and reported on the periodic reports.

The **summary dashboard** for each level provides additional details and facilitates comparison (below).



# Global Dashboard

A “Global” summary dashboard is produced to show the aggregate results from the data verification and data management system assessment. In addition, a dashboard is produced to show findings from the systems assessment by the components of data quality. A sample global dashboard is below.



## Summary Tables

To simplify the process of reviewing feedback from various sites or at various levels, the latest version of the RDQA tool has been updated to include worksheets with tables that automatically populate with the comments and remarks about the responses to the RDQA questions.

**Data Verification** results are summarized on the *Quantitative Comments for the District and Facility Levels* sheets. Here, you can view the results of the documentation review, recounting reported results, and the three cross-checks. Each row provides data for one service delivery site or district (depending on which sheet you're reviewing). By reviewing results and comments across service delivery or aggregation sites (i.e., in columns), trends and systematic data quality problems can be identified which aide in the formulation of data quality improvement plans.

The Service Site # corresponds with the worksheet where the data was captured

The Unit Name gives the actual name of the service site / district.

Results are automatically drawn from the data collection worksheets

Headers indicate the section of the data verifications

Service Site #	Service Site	B - Recounting reported Results:					
	Unit Name	Recount the number of people, cases or events during the reporting period by reviewing the data source. [A]		Enter the number of people, cases or events reported by the site during the reporting period from the site summary report. [B]		Calculate the ratio or recount What are the reasons for the discrepancy (if any) observed. <small>data source</small>	
						Number Counseled and tested	Number testing positive
Service Delivery Site 1	CS Parakou	1) 100 2) 75 3) 30 4) 10	0%	1) 30 2) 85 3) 25 4) 8	0%	110% no reported results; no verified results	115% 0
Service Delivery Site 2	Maternite de bonne saveur Parakou	1) 140 2) 45 3) 20 4) 90	0	1) 80 2) 55 3) 25 4) 70	0	175% no reported results; no verified results	82% 0
Service Delivery Site 3	CS Immacule Conception de Kalale	1) 100 2) 30 3) 60 4) 15	0	1) 120 2) 40 3) 40 4) 10	0	83% no reported results; no verified results	225% 0
Service Delivery Site 4	Hopital St. LYess de Kalale	1) 180 2) 150 3) 101 4) 23	0	1) 185 2) 145 3) 82 4) 21	0	92% no reported results; no verified results	103% 0
Service Delivery Site 5	Hopital Regionale de Savalou	1) 210 2) 180 3) 80 4) 31	0	1) 270 2) 150 3) 90 4) 19	0	78% no reported results; no verified results	120% 0

*System Assessment* results are summarized in two summary tables.

## Summary of System Assessment Comments

Once you've completed your assessment, you'll be able to review all of the results for the different levels in the summary table instead of maneuvering between the worksheets in the tool. As you can see in the example below, the summary provides both the response and the detailed comments. Where you see "*Please provide a comment*" the response was either "*No*" or "*Partly*" and the data collector did not include a comment.

Again, by reviewing results and comments across service delivery or aggregation sites (i.e., in columns), trends and systematic data quality problems can be identified which aide in the formulation of data quality improvement plans.

The Service Site # corresponds with the worksheet where the data was captured

The Unit Name gives the actual name of the service site / district.

The system assessment question and response are listed.

Any explanatory comments will populate automatically

	Service Site						
Service Site	Unit Name	All relevant staff have received appropriate training on the data management processes and tools.	The responsibility for recording the service delivery of the source document is clearly assigned to the relevant staff.	There are designated staff responsible for reviewing periodic reports prior to submission to the next level (e.g. sub-district, district or national levels).			
Site Delivery Site 1	CS Parakou	Yes - completely		Yes - completely	Partly	Please Provide a Comment.	
Site Delivery Site 2	Maternite de bonne saveur Parakou	No - not at all	Please Provide a Comment.	No - not at all	Please Provide a Comment.	Partly	Please Provide a Comment.
Site Delivery Site 3	CS Immacule Conception de Kalale	Yes - completely		Yes - completely	Partly	Please Provide a Comment.	
Site Delivery Site 4	Hopital St. LYess de Kalale	No - not at all	Please Provide a Comment.	No - not at all	Please Provide a Comment.	Partly	Please Provide a Comment.

## Detail of System Assessment

The results are further summarized into the worksheet that has the detail of the system assessment. Here, you can see the various service points and the responses for each of the questions. This layout allows you to quickly skim for comparisons for each question attached to each dimension (lots of red indicates an overall weakness, while lots of green indicates an overall strength).

Headers indicate which component of the M&E system the questions link to

The system assessment questions are listed.

The Service Site # corresponds with the worksheet where the data was captured

The actual name of the site.

		I - M&E Structure, Functions and Capabilities						
		All relevant staff have received appropriate training on the data management processes and tools.	The responsibility for recording the service delivery on the source document is clearly assigned to the relevant staff.	There are designated staff responsible for reviewing periodic reports prior to submission to the next level (e.g. sub-district or national levels).	There is a process in place to ensure that data compilation and reporting is completed in the event that the responsible staff is not available to do the job (e.g. shared duties, a team approach etc.)	The health facility receives regular feedback on the quality of their submitted reports according to the guidelines.	The health facility receives regular supportive supervisory visits from district and/or national level staff according to the guidelines.	...if yes, the last visit was within the past three months.
Service Point #	Service Delivery Site	Yes - completely	Yes - partly	Partly	No - not at all	Yes - completely	No - not at all	Yes - completely
Service Point 1	CS Parakou	Yes - completely	Yes - completely	Partly	No - not at all	Yes - completely	No - not at all	Yes - completely
Service Point 2	Matemite de bonne saveur Parakou	No - not at all	No - not at all	Partly	No - not at all	Yes - completely	No - not at all	No - not at all
Service Point 3	CS Immacule Conception de Kalale	Yes - completely	Yes - completely	Partly	No - not at all	Yes - completely	No - not at all	Yes - completely
Service Point 4	Hopital St. LYess de Kalale	No - not at all	No - not at all	Partly	No - not at all	Yes - completely	No - not at all	Yes - completely
Service Point 5	Hopital Regionale de Savalou	Partly	No - not at all	Partly	Partly	Yes - completely	No - not at all	Partly
Service Point 6	CS de Zongo - Savalou	Partly	No - not at all	Partly	No - not at all	Partly	Partly	No - not at all
Service Point 7	Centre Hospitalier de Grand Quelqu'un d'Abomey	Partly	Partly	Partly	No - not at all	Yes - completely	Partly	No - not at all
Service Point 8	CS Behanzin de Abomey	No - not at all	No - not at all	Partly	No - not at all	Yes - completely	No - not at all	Partly
<b>yes - completely</b>		2	2	0	0	7	0	3
<b>Partly</b>		3	1	8	1	1	2	2
<b>No - not at all</b>		3	5	0	7	0	6	3
<b>N/A</b>		0	0	0	0	0	0	0
<b>Total</b>		8	8	8	8	8	8	8
<b>% yes</b>		25%	25%	0%	0%	88%	0%	38%
<b>% partly</b>		38%	13%	100%	13%	13%	25%	25%
<b>% no</b>		38%	63%	0%	88%	0%	75%	38%
<b>%N/A</b>		0%	0%	0%	0%	0%	0%	0%

Responses are summarized in a simple quick reference table to facilitate speedy comparisons between sites and overall strengths and weaknesses.

The number and percent of each response category (Yes, Partly, No) is summarized by question.



## Advanced Quantitative Metrics

The latest version of the RDQA tool has a worksheet to calculate advanced metrics around data quality. The calculations for these metrics are conducted using the data you've already entered into the data collection forms, and allow for a more in-depth analysis of your results within the tool.

In the table below, we list each of the advanced metrics, what the metric means and/or how it was calculated, and how you could use it in your analysis.

Advanced metric	What it means	When it's useful
Average verification factor (VF) across sites	The extent to which reported results could be verified for the selected reporting period and indicators.	Always useful in a general sense, but less useful (or misleading) when you have diversity among sites with regard to service volume, or a mix of over/underreporting.
Average across sites (excluding zero reporting sites)	The extent to which reported results could be verified for all sites with results to report (i.e., excluding the results from sites that scored zero on validation due to missing source documents).	This is useful when there are sites that score zero on data verification. The zero value brings down the average VF for the rest of the sites leading to an artificially low VF. The interpretation of the VF would then be: of all sites with available source documents, the average percent of verified data is...
Weighted average VF	This measure weights the VF on volume of service (the indicator value) so that the influence a site's VF has on the average is proportional to the volume.	When you have high volume sites which have a large influence on the VF, or to limit the influence of smaller volume sites. Use this when you have a mix of high volume and low volume sites.

Advanced metric	What it means	When it's useful
Average of 1-Abs(VF) across sites	Over and underreporting together can cancel each other out, which makes the VF tend toward 1.0, or perfect agreement between recounted and reported, especially for small samples. The Average of the 1-Abs(VF) across sites looks at the magnitude of the aggregate deviation from perfect agreement between recounted and reported and forces the error in the same direction. It is an absolute measure of discrepancy between recounted and reported.	When you have a lot of over or underreporting.
Average of 1-Abs(VF) across sites (excluding zero reporting sites)	Same as above but excluding sites for which the VF is 0 as a result of missing source documents.	When there is both over and under reporting and there are sites with missing source documents.
Weighted Average of 1-Abs(VF) across sites	The Average of 1-Abs(VF) weighted on volume of service.	When you have both over and underreporting, and both high volume and low volume sites.
Percent underreporting	Gives the percentage of sites in the sample of sites verified that are underreporting on the indicator.	Permits an appreciation of the extent of underreporting in the population of sites verified.
Percent overreporting	Same as above but for overreporting.	Same as above but for overreporting.
Percent missing	Gives the percentage of sites in the sample of sites verified that have missing data for data verification	Permits an appreciation of the extent of missing data in the population of sites verified.
Percent of sites with accurate data	Tells you the percent of your sites with VF in the ideal range between 90-110%.	Provides a quick metric on the overall performance of the sites you sampled.

## Step 6. Develop a system strengthening plan, including follow-up actions.

Given the findings and recommendations for each site, an overall action plan is developed. We'll talk about the template in the subsequent section.

### Action Plans for System Strengthening

At each site and after reviewing the overall results, you will create **action plans** to improve data quality. These recommendations should be created based on the findings of the assessment in consultation with the team members who are involved in the reporting process. By engaging the team members, you create ownership of the plan and get the direct insights from the people in the field.

Decisions on where to invest resources for system strengthening should be based on the relative strengths and weakness of the different functional areas of the reporting system identified via the RDQA, as well as consideration of practicality and feasibility.

#### **How you develop your action plan will depend on if you conducted data verifications only or the full assessment.**

##### **If...**

You are only conducting the data verifications

You have conducted the system assessment and the data verifications

##### **Then...**

Identify which dimension of data quality is the most pressing (e.g., accuracy, timeliness) and consider action items that could improve that dimension.

Review the results and identify which component(s) of the M&E system need to be addressed. The conditional formatting in the tool makes it simple to scan through the system assessment responses.

Look for components with multiple red ("no – not at all") responses, which indicate a weakness in the system. If none of the components have multiple "no – not at all" responses, start to look for the yellow "partly" responses.

### Templates for Action Plans

To facilitate this process, the RDQA Tool provides templates for recommendations for service delivery sites and intermediate aggregation level sites included in the assessment. You will find the template for recommendations on the individual worksheets for each site (service delivery site, district, and regional), and can add additional rows as needed.

The following table shows the recommendations template for service delivery sites. The same template is used at the intermediate aggregation level. A similar template is used at the M&E Unit, with directions to summarize key issues that the program should follow up at various levels of the system (e.g., issues found at site level and/or at intermediate aggregation site level).

## Recommendations for the service delivery site (template)

Part 3 : Recommendations for the service delivery site				
Based on the findings of the systems' review and data verification at the service site, please describe any challenges to data quality identified and recommended strengthening measures, with an estimate of the length of time the improvement measure could take. These will be discussed with the Program.				
	Identified Weaknesses	Description of Action Point	Responsible(s)	Time Line
1				
2				
3				
4				

The final output of the RDQA is an action plan for improving data quality which describes the identified strengthening measures, the staff responsible, the timeline for completion, resources required, and follow up. The template for the action plan is shown in Figure 7.

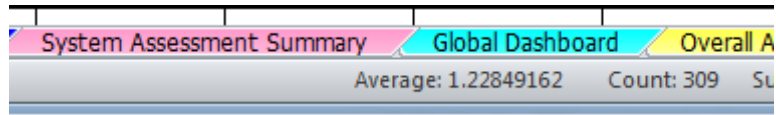
RDQA Final Action Plan				
Country:				
Program/Projet				
Date of Evaluation :				
Date of Follow-up				
Summary of Identified Data Quality Challenges	System Strengthening Measures	Responsibles	Deadline	Comments

## Bonus: Extracting the Data for Further Analysis

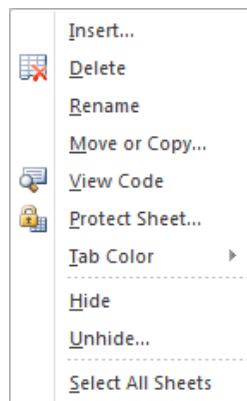
All data are extracted from the worksheets and collated on a single hidden sheet called “Data Export.” This sheet, a “flat file” of data where indicators are in columns and sites are in rows, can be extracted from the workbook and imported into a database for further analysis. Unhide the “Data Export” sheet by right clicking on the worksheet tabs at the bottom of the screen and selecting the tab “Data Export” from the list of hidden worksheets. (Or, type *Control-Format-Sheet-Unhide, Ctrl-o-h-u*).

Not sure how to unhide the sheet? Here’s a step-by-step guide:

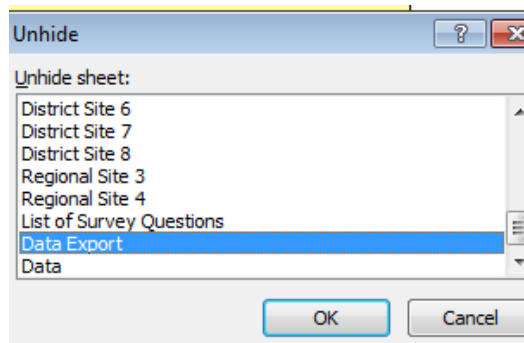
1. Right click on any sheet in the workbook.



2. Select “Unhide” from the navigation menu.



3. Select “Data Export” from the list of sheets that can be unhidden.



4. Copy the rows of data for the sites assessed, and paste the results into your master workbook.

	A	B	C	D	E	F	G	H	I	J	K	L
	Facility	Level	fac_code	district	dis_code	date	vf1	recount1	report1	vf2	recount2	report2
1	Overall	1					0.93			1.18		
2	Central Level: National M&E Unit - Cotonou	2	0	0	0	0	0.94	1660	1775	1.06	1400	1320
3	Borgu	3	0	0	0	0	0.91	200	220	1.09	180	165
4	Collines	3	0	0	0	0	0.83	250	300	1.11	200	180
5		3	0	0	0	0	-	0	0	-	0	0
6		3	0	0	0	0	-	0	0	-	0	0
7		3	0	0	0	0	-	0	0	-	0	0
8		3	0	0	0	0	-	0	0	-	0	0
9		3	0	0	0	0	-	0	0	-	0	0
10		3	0	0	0	0	-	0	0	-	0	0
11	Parakou	4	0	0	0	0	0.80	60	75	1.11	50	45
12	Kalale	4	0	0	0	0	1.16	104	90	1.15	75	65
13	Savakou	4	0	0	0	0	0.67	80	120	1.09	70	64
14	Abomey	4	0	0	0	0	0.96	130	133	1.14	115	101
15		4	0	0	0	0	-	0	0	-	0	0
16		4	0	0	0	0	-	0	0	-	0	0
17		4	0	0	0	0	-	0	0	-	0	0
18		4	0	0	0	0	-	0	0	-	0	0
19		4	0	0	0	0	-	0	0	-	0	0
20	CS Pansou	5	1234	Pansou	Borgu	40160	1.11	100	90	1.15	75	65
21	Maternité de bonne saviour Parakou	5	1224	Parakou	Borgu	40180	1.75	140	80	0.82	45	55
22	CS Immacule Conception de Kalale	5	1123	Kalale	Borgu	40181	0.83	100	120	2.25	90	40
23	Hopital St. LYess de Kalale	5	1427	Kalale	Borgu	40161	0.92	100	195	1.03	150	145
24	Hopital Regionale de Savakou	5	2216	Savakou	Collines	40164	0.78	210	270	1.20	100	150
25	CS de Zongo - Savakou	5	2365	Savakou	Collines	40165	0.67	50	75	1.52	32	21
26	Centre Hospitalier de Grand Quelqu'un d'Abomey	5	3567	Abomey	Collines	40167	0.68	75	110	1.03	69	67
27	CS Bèhanzin de Abomey	5	3657	Abomey	Collines	40168	0.97	165	160	0.98	121	124
28		5	0	0	0	0	-	0	0	-	0	0

### APPENDIX 1. THE LINK BETWEEN THE REPORTING SYSTEM AND DATA QUALITY

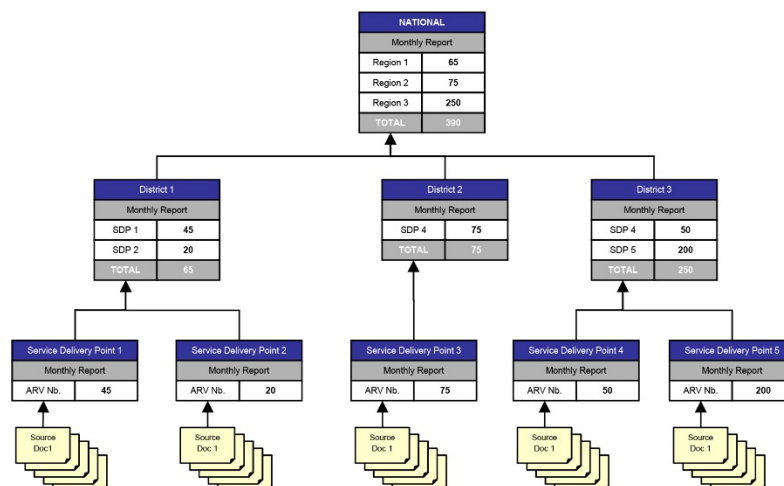
The RDQA has been developed based on a multidimensional concept of **data flows through a program/project data management and reporting system** that operates at three (or more) levels and the **seven dimensions of data quality** that can pose challenges at each level of the system. Furthermore, the RDQA identifies **seven functional areas** that should be assessed to strengthen the data management and reporting system and improve the quality of data the system produces.

#### A. Levels of the Data Management and Reporting System

Data collected, aggregated, and reported to measure indicators flow through a data management and reporting system that begins with the recording of an encounter between a client and a program staff member, a commodity distributed, or a person trained. Data are collected on source documents (for example, patient records, client intake sheets, registers, training registers, and commodity distribution logs). Through the data management and reporting system, the data from source documents are aggregated and sent to a higher level (for example, a district, a partner or principal recipient, or a subpartner or a subrecipient) for further aggregation before being sent to the next level, culminating in aggregation at the highest level of a program (for example, the M&E unit of a national program, the principle recipient of a Global Fund grant, or the SI unit of a USG program). The data from countries is frequently sent to international offices for global aggregation to show progress in meeting goals related to health initiatives.

Figure 8 illustrates this data flow of data through the data management and reporting system that includes service sites, districts, and the national M&E unit. Each country and program/project may have a different data flow. Challenges to data quality can occur at each of these levels.

**Figure 8. Illustration of Data Flow**



## B. Data Quality Dimensions

The RDQA is grounded in the components of data quality, namely, that programs/projects need accurate and reliable data that are complete, timely, precise, credible, and maintained under conditions of confidentiality, when appropriate (see Table 1).

Appendix 1 - Table 1. Data Quality Dimensions	
Dimensions of data quality	Operational Definition
<b>Main dimensions of data quality</b>	
<b>Accuracy</b>	Also known as validity. Accurate data are considered correct: the data measure what they are intended to measure. Accurate data minimize errors (e.g., recording or interviewer bias, transcription error, sampling error) to a point of being negligible.
<b>Reliability</b>	The data generated by a program's information system are based on protocols and procedures that do not change according to who is using them and when or how often they are used. The data are reliable because they are measured and collected consistently.
<b>Subdimensions of data quality</b>	
<b>Precision</b>	This means that the data have sufficient detail. For example, an indicator requires the number of individuals who received HIV counseling and testing and received their test results by sex of the individual. An information system lacks precision if it is not designed to record the sex of the individual who received counseling and testing.
<b>Completeness</b>	Completeness means that an information system from which the results are derived is appropriately inclusive: it represents the <i>complete</i> list of eligible persons or units and not just a fraction of the list.
<b>Timeliness</b>	Data are timely when they are up-to-date (current), and when the information is available on time. Timeliness is affected by: (1) the rate at which the program's information system is updated; (2) the rate of change of actual program activities; and (3) when the information is actually used or required.
<b>Integrity</b>	Data have integrity when the system used to generate them are protected from deliberate bias or manipulation for political or personal reasons.
<b>Confidentiality</b>	Confidentiality means that clients are assured that their data will be maintained according to national and/or international standards for data. This means that personal data are not disclosed inappropriately, and that data in hard copy and electronic form are treated with appropriate levels of security (e.g., kept in locked cabinets and in password-protected files).

## APPENDIX 2. SYSTEM ASSESSMENT COMPLETE QUESTIONNAIRE, BY LEVEL

	Service Delivery Level	Subnational Unit Level	National Level	
<b>I. M&amp;E Structure, Functions, and Capabilities</b>	Statement 1.1		There is a documented organizational structure/chart that clearly identifies positions that have data management responsibilities at the M&E unit (to specify which unit: e.g., MOH, NAP, GF, World Bank).	
	Statement 1.2	All relevant staff has received appropriate training on the data management processes and tools.	All relevant staff has received appropriate training on the data management processes and tools.	
	Statement 1.3		There is a training plan which includes staff involved in data-collection and reporting at all levels in the reporting process.	
	Statement 1.4	The responsibility for recording the service delivery on the source document is clearly assigned to the relevant staff.		
	Statement 1.5	There is designated staff responsible for reviewing periodic reports prior to submission to the next level (e.g., subdistrict, district, or national levels).	There is designated staff responsible for reviewing and approving the district monthly report prior to submission to the national level.	A senior staff member (e.g., the program manager) is responsible for reviewing the aggregated numbers prior to the submission/release of reports from the M&E unit.
	Statement 1.6		There is designated staff responsible for reviewing the quality of data (i.e., accuracy, completeness, and timeliness) received from health facilities.	There is designated staff responsible for reviewing the quality of data (i.e., accuracy, completeness, timeliness, and confidentiality) received from sub-reporting levels (e.g., regions, districts, service points).



<b>II. Indicator Definitions and Reporting Guidelines</b>	Statement 1.7	There is a process in place to ensure that data compilation and reporting is completed in the event that the responsible staff is not available to do the job (e.g., shared duties, a team approach etc.).	There is a procedure in place to ensure the monthly facility reports are compiled and the monthly district report is completed and submitted in the event the responsible staff is unavailable (e.g., shared duties, a team approach etc.).	
	Statement 1.8		Feedback is systematically provided to all health facilities on the quality of their reporting (i.e., accuracy, completeness, and timeliness).	Feedback is systematically provided to all subnational units on the quality of their reporting (i.e., accuracy, completeness, and timeliness).
	Statement 1.9	The health facility receives regular feedback on the quality of their submitted reports according to the guidelines.	The subnational unit receives regular feedback from the national program on the quality of their submitted reports.	
	Statement 1.10		The subnational unit conducts regular supervisory visits to the health facilities in the district according to the guidelines.	The central level conducts regular supervisory visits to subnational units according to the guidelines.
	Statement 1.11	The health facility receives regular supportive supervisory visits from district and/or national level staff according to the guidelines.	The subnational unit receives regular supervisory visits from the national program according to the guidelines on supervision.	
	Statement 1.12	...If yes, the last visit was within the past three months.	...If yes, the last visit was within the past three months.	
	Statement 2.1			The national program/M&E unit has documented and shared the definition of the indicator(s) with all relevant levels of the reporting system (e.g., regions, districts, service points).
	Statement 2.2	The health facility has an up-to-date copy of written guidance on indicator definitions provided by the national program/M&E unit.	The subnational unit has an up-to-date copy of written guidance on indicator definitions provided by the national program/M&E unit.	

<b>III. Data Collection and Reporting Forms and Tools</b>	Statement 2.3			There is a description of the services that are related to each indicator measured by the program/project.
	Statement 2.4	The health facility has an up-to-date copy of written guidance provided by the national program/M&E unit on the content of services that are related to each indicator measured by the program/project.	The subnational unit has an up-to-date copy of written guidance provided by the national program/M&E unit on the content of services that are related to each indicator measured by the program/project.	
	Statement 2.5			The national program/M&E unit has provided written guidelines to all reporting entities (e.g., regions, districts, service points) on reporting requirements and deadlines.
	Statement 2.6	The health facility has a copy of written guidelines provided by the national program/M&E unit on reporting requirements and deadlines.	The subnational unit has a copy of written guidelines provided by the national program/M&E unit on reporting requirements and deadlines.	
	Statement 2.7	The written instructions provided by the program are adequate to ensure standardized recording and reporting of program data.	The written instructions provided by the program are adequate to ensure standardized recording and reporting of program data.	
	Statement 3.1	The national program has identified standard reporting forms/tools to be used by all reporting levels.	The national program has identified standard reporting forms/tools to be used by all reporting levels.	The M&E unit has identified standard reporting forms/tools to be used by all reporting levels.
	Statement 3.2	...If yes, the standard forms/tools are consistently used by the service site.	...If yes, the standard forms/tools are consistently used by the subnational unit.	
	Statement 3.3	If multiple organizations are implementing activities under the program/project, they all use the same reporting forms and report according to the same reporting timelines.	If multiple organizations are implementing activities under the program/project, they all use the same reporting forms and report according to the same reporting timelines.	If multiple organizations are implementing activities under the program/project, they all use the same reporting forms and report according to the same reporting timelines.

Statement 3.4			The national program/M&E unit has developed and shared with all reporting units clear instructions on how to complete the data collection and reporting forms/tools.
Statement 3.5	The health facility has a copy of clear instructions on how to complete the data collection and reporting forms/tools provided by the national program/project.	The subnational unit has a copy of clear instructions on how to complete the data collection and reporting forms/tools provided by the national program/project.	
Statement 3.6			The M&E unit has clearly documented data aggregation, analysis, and/or manipulation steps performed at each level of the reporting system.
Statement 3.7		The subnational unit has a copy of clear instructions provided by the national program/M&E unit on data aggregation, analysis, and/or manipulation steps to be performed at the subnational level.	
Statement 3.8			The data collected by the M&E system has sufficient precision to measure the indicator(s) (i.e., relevant data are collected by sex, age, etc., if the indicator specifies disaggregation by these characteristics).
Statement 3.9	All source documents and reporting forms relevant for measuring the indicator(s) are available for auditing purposes (including dated print-outs in case of computerized system).	All reporting forms relevant for measuring the indicator(s) are available for auditing purposes (including dated print-outs in case of computerized system).	
Statement 3.10	There are sufficient stocks of blank data collection and reporting forms at the service site.	There are sufficient stocks of blank reporting forms at the subnational unit.	

<b>IV. Data Management Processes</b>	Statement 4.1	If applicable, there are quality controls in place for when data from paper-based forms are entered into a computer to ensure the accuracy of data entry (e.g., edit and/or logic checks, post-data entry verification, etc.).	If applicable, there are quality controls in place for when data from paper-based forms are entered into a computer to ensure the accuracy of data entry (e.g., edit and/or logic checks, post-data entry verification, etc.).	(If applicable) There are quality controls in place for when data from paper-based forms are entered into a computer (e.g., double-entry, post-data entry verification, etc.).
	Statement 4.2	There are quality controls in place for compiling data for the monthly facility report to ensure accuracy (e.g., detection of transcription errors).	There are quality controls in place for compiling data for the monthly district report to ensure accuracy (e.g., detection of transcription errors).	There are quality controls in place for compiling data for the monthly district report to ensure accuracy (e.g., detection of transcription errors).
	Statement 4.3	The service delivery site routinely creates back-up files of program data.	The subnational unit routinely creates back-up files of program data.	(If applicable) There is a written back-up procedure for when data entry or data processing is computerized.
	Statement 4.4	If yes, the latest date of back-up is appropriate given the frequency of update of the computerized system (e.g., back-ups are weekly or monthly).	If yes, the latest date of back-up is appropriate given the frequency of update of the computerized system (e.g., back-ups are weekly or monthly).	If yes, the latest date of back-up is appropriate given the frequency of update of the computerized system (e.g., back-ups are weekly or monthly).
	Statement 4.5	The recording and reporting system avoids double-counting people within and across service delivery sites (e.g., a person receiving the same service twice in a reporting period, a person registered as receiving the same service in two different locations, etc.).		
	Statement 4.6	The reporting system enables the identification and recording of a "drop out," a person "lost to follow-up," and a person who died.		
	Statement 4.7	Relevant personal data are maintained according to national or international confidentiality guidelines.		

<b>V - Links with National Reporting System</b>	Statement 4.8		There is a written procedure to address late, incomplete, inaccurate, and missing reports; including following-up with service points on data quality issues.	There is a written procedure to address late, incomplete, inaccurate, and missing reports; including following-up with subreporting levels on data quality issues.
	Statement 4.9		If data discrepancies have been uncovered in reports from service points, the district has documented how these inconsistencies have been resolved.	If data discrepancies have been uncovered in reports from subreporting levels, the M&E unit (e.g., districts or regions) has documented how these inconsistencies have been resolved.
	Statement 4.10		The subnational unit can demonstrate that regular supervisory site visits have taken place and that data quality has been reviewed.	The M&E unit can demonstrate that regular supervisory site visits have taken place and that data quality has been reviewed.
	Statement 4.11	There is a written policy that states for how long source documents and reporting forms need to be retained.	There is a written policy that states for how long source documents and reporting forms need to be retained.	There is a written policy that states for how long source documents and reporting forms need to be retained.
	Statement 4.12	The health facility maintains an adequate archive of program documents (i.e., clean, dry, with sufficient space, etc.).	There is a written policy that describes how program documents (e.g., source documents and reporting forms) should be archived (e.g., filing cabinets, storage rooms, etc.).	There is a written policy that describes how program documents (e.g., source documents and reporting forms) should be archived (e.g., filing cabinets, storage rooms, etc.).
	Statement 5.1	When available, the relevant national forms/tools are used for data collection and reporting.	When available, the relevant national forms/tools are used for data collection and reporting.	When available, the relevant national forms/tools are used for data collection and reporting.
	Statement 5.2	When applicable, data are reported through a single channel of the national information systems.	When applicable, data are reported through a single channel of the national information systems.	When applicable, data are reported through a single channel of the national information systems.
	Statement 5.3	The system records information about where the service is delivered (i.e., region, district, ward, etc.)	The system records information about where the service is delivered (i.e., region, district, ward, etc.)	The system records information about where the service is delivered (i.e., region, district, ward, etc.).
	Statement 5.4	If yes, place names are recorded using standardized naming conventions.	if yes, place names are recorded using standardized naming conventions.	...if yes, place names are recorded using standardized naming conventions.

<b>VI - Use of Data for Decision Making</b>	Statement 6.1	The service delivery site develops charts, graphs, maps, etc. (If yes, ask to see them.)	The district develops charts, graphs, maps, etc. (If yes, ask to see them.)	The national program develops charts, graphs, maps, etc. to depict analyzed data. (If yes, ask to see them.)
	Statement 6.2	If yes, there are assigned staff to develop them regularly.	...If yes, there are assigned staff to develop them regularly,	...If yes, there are designated staff to develop them.
	Statement 6.3	There are assigned staff to interpret and analyze the data/results.	There are assigned staff to interpret and analyze the data/results.	There are designated staff to interpret and analyze the data/results.
	Statement 6.4	Staff at the health facility have access to guidance/technical assistance on data use (e.g., peer review meetings, or during supervisory visits).	The subnational unit provides guidance/technical assistance on data use to health facility level (beyond routine reporting).	The national program provides guidance on data use to the reporting levels (beyond routine reporting).
	Statement 6.5	The analyzed data/results are presented/disseminated to other information system stakeholders in the community in a timely manner so that the information can be used to inform decisions. (Ask to see examples.)	The analyzed data/results are presented/disseminated to other information system stakeholders in the subnational administrative unit in a timely manner so that the information can be used to inform decisions. (Ask to see an example.)	The analyzed data/results are presented/disseminated to stakeholders and other levels of the information system in a timely manner so that the information can be used to inform decisions (for example, a bulletin, planning document or other compilation of analyzed data). (If yes, ask to see an example.)
	Statement 6.6	When data are disseminated, they include indications of any limitations that may exist in the data. (If yes, ask to see examples.)	When data are disseminated, they include indications of any limitations that may exist in the data. (If yes, ask to see examples.)	When data are disseminated, they include indications of any limitations that may exist in the data. (If yes, ask to see examples.)
	Statement 6.7	There are programmatic decisions taken by the service delivery site based on analyzed data/results. (If yes, ask for examples.)	There are programmatic decisions taken by the subnational unit based on analyzed data/results. (If yes, ask for examples.)	There are programmatic decisions taken by the national program based on analyzed data/results. (If yes, ask for examples.)

## APPENDIX 3. INSTRUCTIONS FOR SAMPLING SITES USING TWO-STAGE CLUSTER SAMPLING

- 1. Determine the number of clusters and sites.** The assessment team should work with the relevant stakeholders (NACA, MOH, SI Team, CCM, etc.) to determine the number of clusters and sites within clusters. The appropriate number of sites and clusters depends on the objectives of the assessment; precise estimates of data quality require a large number of clusters and sites. Often, it isn't necessary to have a statistically robust estimate of accuracy. That is, it is sufficient to have a reasonable estimate of the accuracy of reporting to direct system strengthening measures and build capacity. A reasonable estimate requires far fewer sites and is more practical in terms of resources. Generally, 12 sites sampled from within four clusters (three sites each) is sufficient to gain an understanding of the quality of the data and the corrective measures required.
- 2. More than one intermediate level.** In the event there is more than one intermediate aggregation level (i.e., the data flows from district to region before going to national level,) a three-stage cluster sample should be drawn. That is, two regions should be sampled and then two districts sampled from each region (four total districts).
- 3. No intermediate level.** If the data are reported directly from service delivery point to the national level (i.e., no intermediate aggregation sites) the site selection will be conducted as above (cluster sampling with the district as the primary sampling unit), but the data will not be reviewed for the intermediate level and results from service delivery sites will be aggregated to derive the national total.
- 4. Prepare the sampling frame.** The first step in the selection of clusters for the assessment will be to prepare a sampling frame, or a listing of all districts (or clusters) where the activity is being conducted (e.g., districts with ART treatment sites). The method calls for selecting clusters proportionate to size, i.e., the volume of service. Often, it is helpful to expand the sampling frame so that each cluster is listed proportionate to the size of the program in the cluster. For example, if a given cluster is responsible for 15% of the clients served, that cluster should comprise 15% of the elements in the sampling frame. See the Illustrative Example Sampling Strategy D (Appendix 4, Table 3) from the Data Quality Audit Guidelines<sup>2</sup> for more details. Be careful not to order the sampling frame in a way that will bias the selection of the clusters. Ordering the clusters can introduce periodicity; for example, every third district is rural. Ordering alphabetically is generally a harmless way of ordering the clusters.

---

<sup>2</sup> <http://www.cpc.unc.edu/measure/resources/tools/health-information-systems/data-quality-assurance-tools/data-quality-assurance-tools>

5. **Calculate the sampling interval.** The sampling interval is obtained by dividing the number of elements in the sampling frame by the number of elements to be sampled. Using a random number table or similar method, randomly choose a starting point on the sampling frame. This is the first sampled district. Then proceed through the sampling frame selecting districts which coincide with multiples of the sample interval. The *starting number + sampling interval = 2<sup>nd</sup> cluster*. The *starting number + 2(sampling interval) = 3<sup>rd</sup> cluster*, etc.
6. **Stratify service delivery points.** Order the service delivery points within each of the sampled districts by volume of service—i.e., the value of the indicator for the reporting period being assessed. Divide the list into strata according to the number of sites to be selected. If possible, select an equal number of sites from each strata. For example, if you are selecting three sites, create three strata (small, medium, and large). If selecting two sites, create two strata. For six sites, create three strata and select two sites per stratum and so on. Divide the range (subtract the smallest value from the largest) by the number of strata to establish the cut points of the strata. If the sites are not equally distributed among the strata use your judgment to assign sites to strata.
7. **Select service delivery points.** For a large number of sites per district you can use a random number table and select sites systematically as above. For a small number of sites, simple random sampling can be used to select sites within clusters.
8. **Select backup sites.** If possible, select a backup site for each stratum. Use this site only if you are unable to visit the originally selected sites due to security concerns or other factors. Start over with a fresh sampling frame to select this site (excluding the sites already selected). Do not replace sites based on convenience. The replacement of sites should be discussed with the funding organization and other relevant stakeholders, if possible.
9. **Know your sampling method.** The sites are intended to be selected for the assessment as randomly (and equitably) as possible while benefiting from the convenience and economy associated with cluster sampling. You may be asked to explain why a given site has been selected. Be prepared to describe the sampling methods and explain the equitable selection of sites.



**MEASURE** Evaluation–Strategic Information for  
South Africa (MEval-SIFSA) Project  
138 Muckleneuk Street  
Nieuw Muckleneuk, Pretoria  
South Africa  
Tel: + 27 12 346 7490  
<http://www.measureevaluation.org/sifsa>

This research has been supported by the President's Emergency Plan for AIDS Relief (PEPFAR) through the United States Agency for International Development (USAID) under the terms of MEASURE Evaluation–Strategic Information for South Africa associate award AID-674-LA-13-00005. MEASURE Evaluation–SIFSA is implemented by the Carolina Population Center at the University of North Carolina at Chapel Hill, in partnership with ICF International; John Snow, Inc.; Management Sciences for Health; Palladium; and Tulane University. Views expressed are not necessarily those of PEPFAR, USAID, or the United States government. MS-17-117

ISBN: 978-1-9433-6442-8

