Federal Democratic Republic of Ethiopia

Ministry of Health

Data Quality Assurance Guideline

Federal Ministry of Health

January 2015

Data Quality Assurance tool and implementation Guideline

Federal Ministry of Health

January 2015

Table of Contents

List of Table

List of Figures

ACRONYMS

DQA	Data Quality Assessment
FF	Family Folder
FMOH	Federal Ministry of Health
HIV/AIDS	Human Immunodeficiency Virus/Acquired immunodeficiency syndrome
HIQIP	Health Information Quality Improvement Plan
ITN	Insecticide Treated bed nets
PRISM	Performance of Routine Information System Management
M&E	Monitoring and Evaluation
RDQA	Routine Data Quality Assessment
RHIS	Routine Health Information System
TB	Tuberculosis
VF	Verification Factor

1. INTRODUCTION

1.1 Background

The ability of health system stewardship and strategic decisions is impacted by the quality of their health data. Ethiopia is implementing the reformed HMIS since 2008 with a strong emphasis to improve data quality and information use at each level of health sector nationwide. There has been extensive improvement in quality of health data over time; however ensuring high quality of data collected through HMIS for sound information use and decision making still remains a challenge in the country.

Hence, the FMOH has prioritized and undertaken a major nationwide initiative with the aim of enhancing the quality of HMIS data radically. Accordingly, Health Information Quality Improvement Plan (HIQIP) has been developed and dedicated personnel were assigned to monitor and implement the HIQIP plan to meet the desired quality of data at each level of health system in the country. Moreover NAC/HMIS was revitalized and as result subsequent TWG has been established to accomplish different tasks like development of HMIS Mentorship guideline, and some other upcoming needs

In addition the ministry has been adopting and applying various data quality assurance tools to address the data quality dimensions taking in to account relevance to the context of the country. Currently as part of the HIQIP initiatives, the FMOH has identified and is utilizing RDQA and LQAS tools that are widely recommended by WHO and other international organizations. Therefore the purpose of this guideline is to provide guidance how, why, where and when to apply those different recommended data quality assurance tools.

1.2 Definitions and dimensions of data quality

There is no one definition of data quality that is used consistently across institutions. Data quality is a multi-dimensional construct. Overall data quality, then, becomes a function of each of its dimensions. Therefore data quality can be defined as the state of completeness, validity, consistency, timeliness, accuracy, integrity and confidentiality that makes the data appropriate for specific use. And in some cases it can be defined as the totality of features and characteristics of data that bears on their ability to satisfy a given purpose or the sum of the degrees of excellence for factors related to data.

1.3 Dimensions of Data Quality

Assessment of quality of data in any information system involves a comparison of data within the system against an agreed set of standards for the data which we usually call them as dimensions of data quality. Some of the most common and relevant sets of standards or dimensions of data are listed and defined as follows

- **Accuracy:** 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.
- **Timeliness:** data is collected, transmitted and processed according to the prescribed time and available for making timely decisions.
- **Completeness:**
  - At service delivery point, it refers to all the relevant data elements in a patient/client register are filled

- At Health Administrative unit – data completeness has two meanings:
  1. All the data elements in a database or report are filled
  2. The health administrative unit has reports from all the health facilities and/ or lower level health administrative units within its administrative boundary.

- **Precision:** Data collected and analyzed should be large enough and have sufficient detail to support to support the decision and to take action.
- **Integrity:** Data have integrity when the system used to generate them is protected from deliberate bias or manipulation for political or personal reason.
- **Reliability:** 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.
- **Confidentiality:** 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).

1.4 Types of Data quality Assurance tools

There are different data quality assurance tools that address the data quality dimensions fully or partially depending on the purpose and who uses the tools such as external audit team or teams for internal consumption.

1. **LQAS (Lots Quality assurance sampling):-** is a technique useful for assessing whether the desired level of data accuracy has been achieved by comparing data in relevant record forms (i.e. registers or tallies and FF in health post) and the HMIS reports. Based on a small Sample size, self-assessment is done to estimate the level of data quality using LQAS at health facility level.
2. **DQA (Data Quality Audit Tool)**:- provides guidelines to be used by an external audit team to assess a program/project's ability to report quality data
3. **RDQA (Routine Data Quality Assessment Tool)**-: is a simplified version of the DQA Tool for auditing that allows programs and projects to assess the quality of their data and strengthen their data management and reporting systems.
4. **PRISM Tool (Performance of routine information system management):** is a tool to identify specific technical, behavioral, and organizational factors that affect RHIS performance and provide the methods to objectively measure data quality and the degree to which information is used for evidence-based decision making

Table 1: Major Distinctions among LQA, DQA, RDQA and PRISM

LQAS	DQA	RDQA	PRISM
• Self assessment including producers of the reports • Simple and uses small sample size for continues quality assurance at facility level • can be used through data accuracy check lists • limited to few data quality components (mostly accuracy)	• Assessment by funding agency • Standard approach to implementation • Conducted by external audit team • Limited input in to recommendations by programs	• Self-assessment by program • Flexible use by programs for monitoring and supervision or to prepare for an external audit • Program makes and implements own action plan	• To assess whether technical, behavioral and organizational determinants have influence on RHIS performance • used by People involved in the collection, analysis and use of data in RHIS • provide structured way for assess the quality of data and use of information

Therefore, different tools can be used based on the context and relevance and countries can adopt and use their preferred ones based on their context and priorities to address and meet the most important data quality dimensions.

Among the many tools, currently the ministry has adopted RDQA and LQAS to assure data quality at all levels of its structure all the way to health facilities

1.5 Conceptual Framework

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. There are six functional components of data management and reporting as shown in the figure below that must be established for a health system to produce quality information for decision-making.

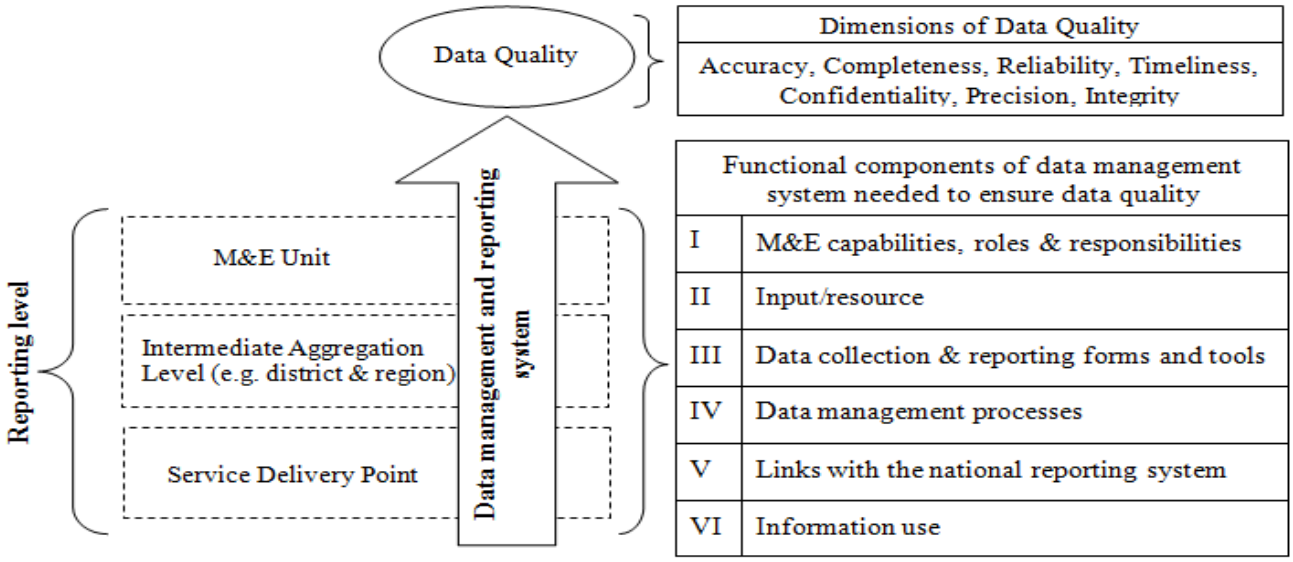

Figure 1: Conceptual Framework for the Data Quality Assurance: Data Management and Reporting Systems, Functional Areas, and Data Quality.

1.6. Objective and Scope

1.6.1. Objective

1.6.1.1 General Objective

- ✓ To set a standard and uniform approach for assessing and improving overall HMIS data quality

1.6.1.2 Specific Objective

- ✓ To provide guidance on data quality assurance practice to all stakeholders
- ✓ To maintain standards of data quality assurance practice across all levels of health system
- ✓ To promote data quality for effective decision making practice
- ✓ To enable all health institution to conduct data quality assessment

1.6.2. Scope

This user guideline is prepared to be used by all levels in the health system in order to make sure that all data quality dimensions are addressed and standardized data management system is maintained.

2. METHODOLOGY

2.1 Selection of study sites

2.1.1 Types of sampling methods for selecting sites for the RDQA

There are different sampling methods for selecting sites for the RDQA:

1. **Purposive selection:** The sites to be visited can purposely be selected according to size, geographical proximity or concerns regarding the quality of reported data. In this case, there is no need for a sampling plan. However, the data quality assessment findings produced from such a "purposive" or targeted sample cannot be used to make inferences or generalizations about all the sites, or a group of sites, in that area.
2. **Restricted site design**: Only one site can be selected for the RDQA. The benefit of this approach is that the team can maximize its efforts in one site and have a high degree of control over implementation of the assessment and knowledge of the site-specific systems from which the results are derived. This approach is ideal for measuring the change in data quality attributable to an intervention (e.g. Mentoring, data management training etc).
3. **Stratified random sampling**: This involves the drawing of a stratified random sample of a sub-national group of sites (Regions, Zones or Woredas) where a particular variable of interest is chosen as the basis of the sites to be visited. Such stratified random sampling allows making inferences from the sample findings to all the sites that belong to the stratification variable of interest (Region, Zone, and Woreda).
4. **Random sampling**: It is often desirable to make judgments about data quality for an entire program or larger area. Random sampling techniques allow selecting a relatively small number of sites from which conclusions can be drawn which are generalizable to all the sites in a program/project. This method involves the random selection of a number

of sites that together are representative of all the sites where activities supporting the indicator(s) under study are being implemented. The purpose of this approach is to produce quantitative estimates of data quality that can be viewed as indicative of the quality of data in the whole program/ project, and not simply the selected sites.

5. **Cluster Sampling Selection:** Cluster sampling is a variation on simple random sampling (where all sites would be chosen randomly) that permits a more manageable group of sites to be assessed. Cluster sampling allows for the selection of a few districts, thereby reducing the amount of travel required by the RDQA team. The primary sampling unit for Sampling is a cluster, which refers to the administrative or political or geographic unit in which Service Delivery Sites are located. Probability proportionate to size (PPS) is applied to derive the final set of sites that should be assessed. Clusters are selected in the first stage using systematic random sampling, where clusters with active programs reporting on the indicator of interest are listed in a sampling frame. In the second stage, Service Delivery Sites from selected clusters are chosen using stratified random sampling where sites are stratified on volume of service.

Most recommended method of sampling is random and the objective of the study is the base for selecting the type of sampling. Random selection of a number of sites creates representative of all the sites where activities supporting the indicator(s) under study are being implemented. Representative means that the selected sites are similar to the entire population of sites in terms of attributes that can affect data quality (e.g., size, volume of service, and location).

2.1.2 Determining the number of sites at National M&E unit

In Ethiopia the health organizations in the health system is divided into 11 Regional health bureaus. Each regional health bureaus is sub-divided into zonal health departments, each zonal department into Wareda health offices, under each Wareda health offices there are service delivery points (health posts, Health centers and hospitals). However in some regions Zonal health department is not functional and the woreda health offices directly report to Regional health bureau.

Study sites are widely distributed and the various administrative levels are not of equal size, hence the need to have a sampling frame that involves selection of clusters accordingly. All regions will be involved in the RDQA and the primary sampling unit for the sampling is cluster or districts which refer to the administrative or political or geographic unit in which Service Delivery Sites are located. A probability proportionate to size (PPS) will be used to derive the total set of cluster from each region that the assessment will include.

Then the actual Clusters (districts) are selected in the first stage using systematic random sampling, where clusters having active HMIS reporting system are listed in a sampling frame by region.

In the second stage, Service delivery Sites from selected clusters are chosen using stratified random sampling where the service delivery sites are stratified on volume of service (or OPD attendance per capita (<=0.5 and >0.5). And because of financial and logistic feasibility, two health centers from each stratum and one hospital will be selected randomly from each selected districts.

2**. Determine the number of clusters and sites**

To estimate the sample size of the clusters (districts) from the regions a single population proportion formula will be used:

$$n = \frac{p(1-p)z_{1-\propto/2}}{s^2}$$

Where:

- ✓ p= the estimated proportion of data quality (If a perevious study exists, p will be the accuracy level of the indicator wich provide the highest sample size or p will be 50% if no study exists)
- ✓ z1-α/2 = the z score corresponding to the probability with which it is desirable to be able to conclude that an observed change of size could not have occurred by chance (α= 0.05 (z1-α/2= 1.96) and from the precision or margin of error denoted by (s) found that 0.05.

If N (the total number of clusters or districts) < 10,000, a correction formula will be used.

$$nf = \frac{\mathrm{n}}{1+(\frac{\mathrm{n}}{\mathrm{N}})}$$

2.1.3 Determining the number of sites at Regional level

The above stated sampling methodologies can be employed to select the appropriate number of sites and clusters based 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 4 clusters (3 sites each) are sufficient to gain an understanding of the quality of the data and the corrective measures required. The Ethiopian MOH recommends the following sample size and methodology for RDQA:

1. **In regions with zones:**
    - Randomly select 4 zones
    - From each of the selected zones, randomly select three Woredas
    - From selected Woredas, select randomly one health centre or hospital
2. **In Regions without zones**
    - Randomly select 4 Woredas
    - From each selected Woredas, randomly select three health centers or hospitals
3. **For Zonal level**
    - Randomly select 4 Woredas
    - From selected each Woredas, select randomly three health centers or hospitals
4. **For Woreda level**
    - Use census of all health centers and hospitals in the Woreda

2.2 Selection of Indicators

Determination of indicators and reporting period that should be included in the assessment is also an important step in RDQA. It is recommended that up to two indicators be selected within a Disease/Health Area and that, if multiple Diseases/Health Areas are included in a Data Quality assessment, that a maximum of four indicators can be included. More than four indicators could lead to an excessive number of sites to be evaluated.

The criteria for selecting the indicators for the RDQA could be the following:

1. "Must Review" Indicators: Indicators that should be selected first depending on the indicator's national and global importance/ priority.
2. Relative Magnitude of the Indicators: The amount of budget and activity associated with the indicator(s).
3. "Case by Case" Purposive Selection: Indicators for which data quality questions exist and the government wants to be routinely verified. Those reasons should be documented as justification for inclusion.

2.3 Frequency

It is suggested that frequency of RDQA has to be based on the objective of the assessment and the level of the organization conducting it. Accordingly the data verification part has to be done quarterly integrating it with supportive supervision visits by organizations at all levels; whereas it is recommended that a comprehensive RDQA (Data verification and system assessment) should be done biannually by Federal or regional level coordinating bodies. It is also important to clearly identify the reporting period associated with the indicator(s) to be assessed. Ideally, the time period should correspond to the most recent relevant reporting period or schedule in HMIS.

2.4 The Tool and its Components

The RDQA allows programs and projects to rapidly self-assess the quality of their data and to strengthen their data management and reporting systems.

Data collection instrument was adapted from ***WHO DQA tool,*** which include Accordingly, the RDQA tool is comprised of two components:

(1) Verification of reported data for key indicators; and

(2) Assessment of data management and reporting systems at selected sites.

2.4.1 Verification of Reported Data for Key Indicators

The purpose is to assess, on a limited scale, if service delivery and intermediate aggregation sites are collecting and reporting data to measure the indicator(s) accurately and on time — and to cross-check the reported results with other data sources. To do this, the RDQA will determine if a sample of Service Delivery Sites have accurately recorded the activity related to the selected indicator(s) on source documents. It will then trace that data to see if it has been correctly aggregated and/or otherwise manipulated as it is submitted from the initial Service Delivery Sites through intermediary levels to the program/project M&E Unit.

The data verification exercise will take place in two stages:

1. In-depth verifications at the Service Delivery Sites; and
2. Follow-up verifications at the Intermediate Aggregation Levels and at the program/ project M&E Unit.

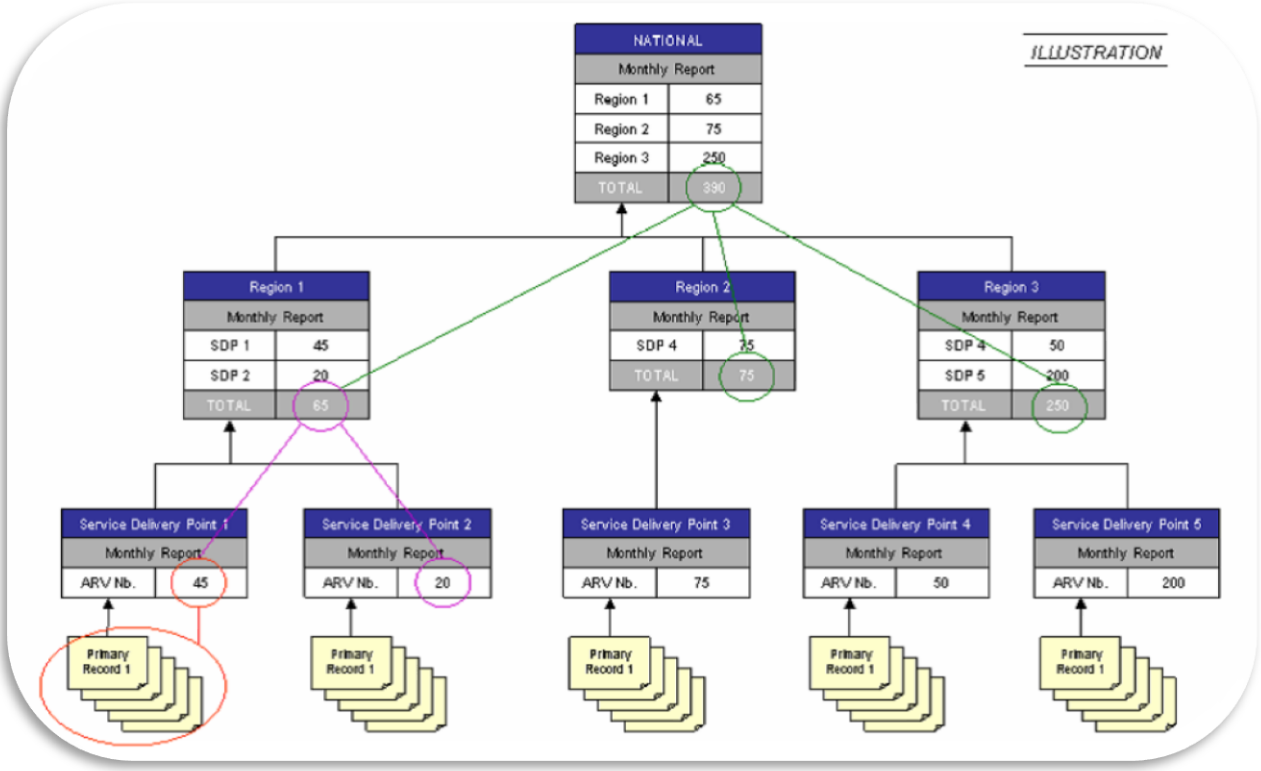

Figure 2: tracing and verifying Report Totals from the Service Delivery Site through Intermediate Reporting Levels to the Program/Project M&E Unit (National level).

The first stage of the data-verification occurs at the Service Delivery Sites. There are five types of standard data-verification steps that can be performed at this level.

Table 2: Service Delivery Site: Five Types of Data Verifications

Verifications	Description	Required
1. Description	Describe the connection between the delivery of services and/or commodities and completion of the source document to record that delivery	In all cases
2. Documentation of review	Review availability and completeness of all indicator source documents for the selected reporting period	In all cases
3. Trace the verification	Trace the verify reported numbers: (1) Recount the report numbers from available source documents; (2) Compare the verified numbers to the site reported number; (3) Identify reasons for any differences	In all cases
4. Cross-checks	Perform "cross-checks" of the verified report totals with other data-sources (e.g. inventory records, laboratory reports, registers, etc…	If feasible
5. Spot-checks	Perform "spot-checks" to verify the actual delivery of services and/or commodities to the target populations	If feasible

The second stage of the data-verification occurs at the Intermediate Aggregation Levels (e.g.
Woreda, Zone, Regions). The RDQA evaluates the ability at the intermediate level to accurately aggregate or otherwise process data submitted by Service Delivery Sites, and report these data to the next level in a timely fashion. The following verifications will therefore be performed at Intermediate Aggregation Levels and M&E unit.

Table 3: Intermediate Aggregation Levels: Two Types of Data Verifications

Verifications	Description	Required
1. Documentation Review	Review availability, timeliness, and completeness of expected reports from Service Delivery Sites for the selected reporting period.	In all cases
2. Trace and Verification	Trace and verify reported numbers: (1) Re-aggregate the numbers submitted by the Service Delivery Sites; (2) Compare the verified counts to the numbers submitted to the next level; (3) Identify reasons for any differences.	In all cases

2.4.1.1 The steps for completing the RDQA verification part

1. Select key data elements from the HMIS reports that will be studied
2. List the data items in the RDQA table
3. For each of the selected data elements recount the number of cases or events recorded during the reporting period by reviewing the relevant source documents available at the selected sites [A]
4. Copy the number of cases or events for the selected data elements reported by the site during the reporting period from the HMIS reports submitted by the selected sites [B]
5. Add up all the recounted figures for the corresponding data elements from the 12 sites [∑A]
6. Add up all the figures for the same data elements copied from the HMIS reports of all the 12 sites [∑B]
7. Calculate the ratio of recounted to reported numbers. [∑A / ∑B]
   This figure gives the Verification: Accuracy Ratio for the respective data element studied. The final output of the RDQA is an indicator of the program recording and reporting for monitoring and improving data quality. Verification factor (Recounted/Reported)

   - < 0.85 or 85% indicates over reporting,
   - 0.85 – 1.15 (85 – 115%) indicate acceptable accuracy level
   - > 1.15 (115%) signifies under reporting

A bar-chart will be used to illustrate the quantitative data generated from the data verifications

HMIS Data Element		Health Facility												Total	V.F
		1	2	3	4	5	6	7	8	9	10	11	12		
	Recounted figure (A)													∑A=	∑A/∑B=
	Reported figure (B)													∑B=	
	Recounted figure (A)													∑A=	∑A/∑B=
	Reported figure (B)													∑B=	
	Recounted figure (A)													∑A=	∑A/∑B=
	Reported figure (B)													∑B=	
	Recounted figure (A)													∑A=	∑A/∑B
	Reported figure (B)													∑B=	

2.4.2 Assessment of data management and reporting systems

The purpose is to identify potential challenges to data quality created by the data management and reporting systems at three levels: (1) the program/project M&E Unit, (2) the Service Delivery Sites, and (3) any Intermediary Aggregation Level (at which reports from Service Delivery Sites are aggregated prior to being sent to the M&E Unit).

The assessment of the data management and reporting systems will take place in two stages:

1. Off-site desk review of documentation provided by the program/project;
2. On-site follow-up assessments at the program/project M&E Unit and at selected Service Delivery Sites and Intermediate Aggregation Levels (e.g., Woredas, Zones, Regions).

The assessment will cover six functional areas:

1. M&E structures, functions and capabilities;
2. Input/resource
3. Data collection and reporting forms, tools and guidelines;
4. Data management process
5. Links with national reporting system and
6. Information use.

2.4.2.1 The steps for completing the RDQA Assessment of data management and reporting systems part

1. Write the answer for the detailed system questions which are categorized under the six functional areas from the information obtained from both Off-site desk review of documentation provided by the program/project and On-site follow-up assessments. (Questions are found on annex Y)
2. The system will analyze the outcome of this assessment as strengths and weaknesses for each functional area. The values range from 0 to 3.0 and are classified in to three dimensions.

Color key code		
Green	2.5-3.0	Yes, completely
Yellow	1.5-2.5	Partly
Red	<1.5	No, Not at all

2.5 Contents of RDQA tool worksheets in General

In order to do the verification and system assessment RDQA tool employs an excel sheet which has the following components:

Sheet 1- Header: to select the number of service sites and intermediate aggregation level sites to be included in the RDQA.

Sheet 2- Instructions: to inform users how to use the Excel spreadsheet.

Sheet 3- Information: to record the country, program/project, indicator reviewed, reporting period reviewed, and the assessment team.

Sheet 4- Service Delivery Point: to record results of the assessment on data verifications, systems assessment and cross-checks at the service delivery level and to record recommendations for the service site and a dashboard of results of the data verification and

Systems assessment for the service site

Sheet 5- Intermediate Aggregation Site: to record results of the assessment on data verifications and systems assessment at the intermediate aggregation level site and to record recommendations for the intermediate aggregation level site and a dashboard of results of the data verification and systems assessment for the intermediate aggregation level site.

Sheet 6- M&E Unit: to record results of the assessment on data verifications and systems assessment at the M&E Unit, to record follow up recommendations and an action plan based on the RDQA, and to show a dashboard of results of the data verification and systems assessment for the M&E Unit.

Sheet 7- National Dashboard: to present in graphic form aggregated results from all levels of the assessment

Sheet 8- RDQA Final Action Plan: to consolidate recommendations from each level into an overall action plan based on the RDQA

Sheet 9- List of survey questions: a reference page to map the functional areas assessed in the systems assessment part of the RDQA with components of data quality.

Sheet 10- Feedback Form: For users of the RDQA to provide feedback to the developers of the RDQA tool.

2.6 Implementation Steps

Step 1 – PREPARATION

A. Levels of the M&E System to be Included

Part of determining the scope of the RDQA is to decide what levels of the M&E system will be included in the assessment – service sites, intermediate aggregation sites (woreda, zone and regions) and a national M&E unit.

B. Indicator(s), Data Sources and Reporting Period

The RDQA is designed to assess the M&E systems and to verify data related to indicators that are reported to the next level in the health system. Therefore, it is important to select one or more indicators – or at least program areas – to serve as the subject of the RDQA. This choice will be based on the list of indicators reported in the health system.

For each program area, a number of indicators are measured, through various data sources. For example, under the Disease TB, in the program area Treatment, 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 report along with the facility register and patient treatment cards.

The RDQA can be implemented on one or more indicators, but it is important to keep in mind that data come from various sources, most notably facility-based, community-based, and commodity distribution-based data sources. When planning the RDQA, it is important to determine the data sources that will need to be assessed related to the indicator(s) and program area(s).

The RDQA designed to assess data related to indicators during a specific (selected) time period, generally a reporting period. Using a specified reporting period gives a reference from which to compare the "recounted" data, and is the recommended method for the RDQA.

C. Determine and Notify Sites to be visited

The types of sites included will depend on the scope of the RDQA. Once the scope and levels have been set, the actual sites to be visited should be selected. The list of sites may include those:

- Due for routine M&E monitoring during a given supervision cycle.
- Selected based on the indicator(s) under review
- Selected sites of newly implementing programs.

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 M&E systems questions in the checklist and to facilitate the data verification by providing access to relevant source documents.

Step 2 – SITE VISITS

Survey teams will include interviewers, and supervisors. The number of data collectors and supervisors in each region will depend on the number of sample facilities in the region. The team members will be participants from the FMOH, RHBs, Zones, Districts and Development partners depending on the initiator of the study. The number of days the study to occur depends on the number of sample facilities included

Data collection approaches will include interviews using closed and open ended questionnaires, observations, and Verifications of reports and registries

Assess the Program/project's data management and reporting system at all or selected level of the M&E Unit, any intermediate aggregation level and at selected service sites. Verify data for selected indicator(s) from source documents and compare with reported results.

Part 1 of the RDQA checklist is the systems assessment should be administered at each of the levels of the M&E system that is included in the RDQA. There is a specific worksheet or part of the checklist for each level included in the RDQA (e.g. worksheets for service sites, intermediate

aggregation sites and the M&E Unit). The questions should be asked of the staff member(s) at each level who are most familiar with the data management and reporting system. There are six functional areas of the M&E system which are included in the assessment of the data management and reporting system. Relevant questions are asked at each level of the M&E system.

Part 2 of the RDQA, data verifications, should also be filled out for each level included. At service sites, recounting includes recounting the number of people, cases or events recorded during the reporting period as contained in the relevant source documents. This number is compared to the number reported by the site during the reporting period (on a summary report sent to the next level). At the Intermediate Aggregation Level, recounting includes re-aggregation of the numbers from reports received from all service delivery points and a comparison of that number with the aggregated result that was contained in the summary report prepared by the intermediate aggregation level and submitted to the next level. At the national M&E Unit, recounting includes re-aggregating reported numbers from all reporting regions and comparing them to the summary report that was prepared by the national M&E Unit for the reporting period. At this level, the reports should also be reviewed to count the number that are on available, on time, and complete – all measures of data quality.

Alternatively, the recount can be simplified by comparing the recounted results in the relevant register to the summary report. Using this method, a sample of names from the register can be selected using the methodology for the "cross checks" in Part 2 of the checklist. Those entries on the register can be compared to the same information on the source documents (e.g. name, age, sex, diagnosis, treatment date(s), treatment regimen, etc. as relevant for the indicator). If errors are found, the methodology of recounting from source documents (described in the paragraph above) should be used to thoroughly check the data.

Step 3 – ACTION PLAN

Based on the findings from the RDQA, prepare an action plan for strengthening the data management system and for improving the quality of data.

Part 3 of the RDQA has a template for an action plan that is based on the findings from the RDQA and includes follow up actions, responsibilities, timelines and resource needs.

Step 4 – FOLLOW UP

Implement activities included in the RDQA action plan and follow up with relevant sites/locations to insure implementation. The purpose of the RDQA is to strengthen M&E systems related to collecting, managing and reporting data related to indicators. It will be important to follow-up to ensure that strengthening measures identified in the action plan are actually carried out.

2.7 DATA PROCESSING AND ANALYSIS

WHO_ DQA Excel spread sheet will be used to calculate verification factor and system level performance. It will also used to display the status using spider diagram and graphs. In addition SPSS Version 17 will be used to further study association between independent and dependent variables. Bivariate and multivariate analysis can be done to identify determinant factors and to model their effect on data accuracy or use.

2.8 OUTPUT OF THE RESULT

When the Excel RDQA checklists completed electronically, a summary of each attribute by level and bar chart presentation shows, a number of dashboards produce graphics of summary statistics for each site or level of the reporting system and a "national" dashboard that aggregates the results from all levels and sites included in the assessment (Figure 2).

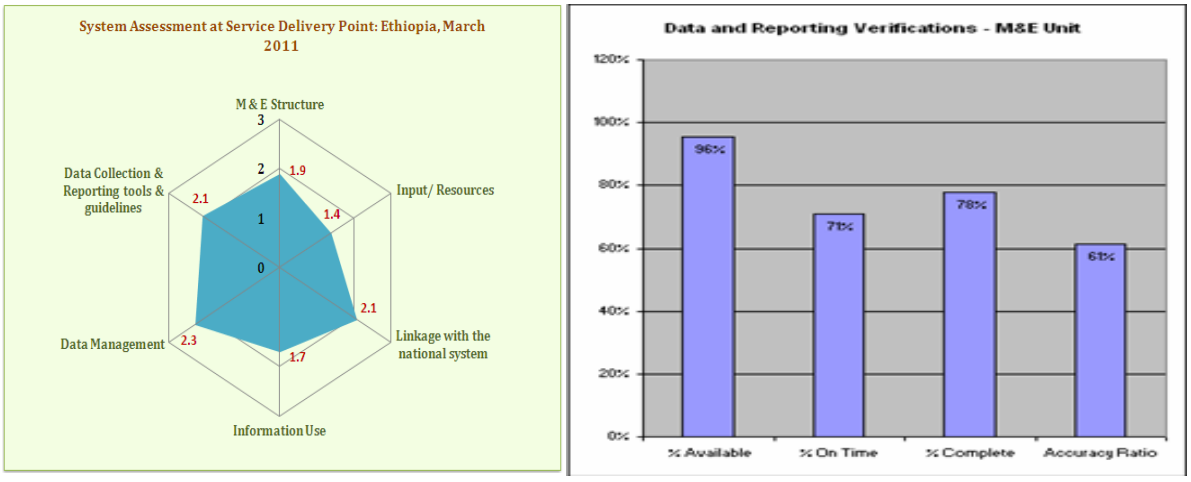

Figure 3: Dashboard displays two graphs for each site visited

- The spider-graph on the left 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 on the right shows the quantitative data generated from the data verifications; these can be used to plan for data quality improvement.

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.

2.9 LQAS AT SERVICE DELIVERY POINT

Lot Quality Assurance Sampling (LQAS) - is a technique useful for assessing whether the desired level of data accuracy has been achieved by comparing data in relevant record forms (i.e. registers or tallies) and the HMIS reports at each Service Delivery Point.

This is a method for testing hypothesis related with the level of HMIS data quality whether it is achieved or not. It uses a sample size of 12 data elements and tries to check the accuracy of reports.

If the number of sampled data elements not meeting the standard exceeds a pre-determined criterion (decision rule), then the lot is rejected or considered not achieving the desired level of pre-set standard. "Decision rule" table is used for determining whether the pre-set criterion is met or not. Comparison of LQAS results over time can indicate the level of change.

2.9.1 STEPS IN LQAS AT SERVICE DELIVERY POINT

The following steps show how the quality of HMIS data can be estimated using a sample of 12 data elements and comparing the results with a standard LQAS table. Selected data elements from the monthly report submitted to the Woreda are compared with the tallies and register sums that are the sources of these data elements. If a high proportion of the numbers are the same, then the quality of the data can be assumed to be high; if a low proportion is the same, then the quality of the data is low.

1. Selection of data elements is random, which means data elements are selected without any preference. A broad representation of the data elements from different sections of the monthly report form is required to assure all data elements are given equal opportunity for selection. A sample of 12 data elements is required based on LQAS table.
2. Select randomly one data element from each section of the previous monthly report. Write the selected data element in the first column of the data accuracy check sheet given below. Repeat the procedure until all data elements from different sections are entered in first column.
3. Copy the figures of the selected data elements as reported on the monthly report form in second column of data quality check sheet, under the heading of "figures from monthly report form".
4. Pick the register or tally sheet which has the selected data element. Sometimes there may be several registers or tally sheets. Count the actual entries in the register or tally related to a specific selected data element. Put the figure you counted in third column of check sheet, under the heading "figure from register". Repeat this procedure for all data elements.
5. If the figures in column 2 and 3 are same, tick under YES in column four. If they are not the same (do not match), put a tick under NO in column four. Repeat this procedure for all data elements.
6. Count the total ticks under "YES" and write in row of total for "YES". Repeat the procedure for "NO" column. The sum of YES and NO totals should be equal to the sample size of 12.

Table 4: Data Accuracy Check Sheet

Month for which data accuracy is checked_____				
Randomly Selected Data Elements from the monthly reporting form	Figures from the Monthly report form (2)	Figures counted from registers & tallies (3)	Do figures from columns 2 & 3 Match? (4)	
			YES	NO
1.				
2.				
3.				
4.				
5.				
6.				
7.				
8.				
9.				
10.				
11.				
12.				
Total				

The total in number in the "Yes" column corresponds to the percentage of data accuracy in the following LQAS table. For example, if total "yes" number is 2, the accuracy level is between 30-35%; if total number in the "yes" column is 7, the accuracy level is between 65-70%.

LQAS: Decisions Rules for Sample Sizes of 12 and Coverage Targets/Average of 20-95%																	
Sample Size	Average Coverage (Baselines)/ Annual Coverage Targets (Monitoring and Evaluation)																
	Less than 20%	20%	25%	30%	35%	40%	45%	50%	55%	60%	65%	70%	75%	80%	85%	90%	95%
12	N/A	1	1	2	2	3	4	5	5	6	7	7	8	8	9	10	11

7. Circle the data accuracy percentage and write it in the monthly report and submit to the Woreda office.
    - You could set a target for achievement in a specified period and use it for monitoring progress. The target can be broken down on monthly basis. For example, if data accuracy is improving by 5% on monthly basis, the correct match number should increase accordingly as shown in the LQAS table. As the correct match number increases compared to previous months, it reflects improvement in level of data accuracy.
    - Achievement of data accuracy level at 95% means a high level of accuracy and needs to be maintained at that level.

Note: Please note that with sample size of 12 data elements, the data accuracy ranges +15%. That means if the data accuracy is 30%, the range is between 15% and 45%.

3. ETHICAL CONSIDERATIONS

Data quality assessments must be conducted with the utmost adherence to the ethical standards of the country. While those undertaking the RDQA may require access to personal information (e.g. medical records) under no circumstances, information of any personal should be disclosed in relation to the conduct of the assessment.

ANNEXES

Table 1: Data verification protocol

Part 1: Data Verifications			
A - Documentation Review:			
	Review availability and completeness of all indicator source documents for the selected reporting period.	***(Yes completely, partly, No not at all)***	***Reviewer Comments***
1	Review available source documents for the reporting period being verified. Is there any indication that source documents are missing?		
	If yes, determine how this might have affected reported numbers.		
2	Are all available source documents complete?		
	If no, determine how this might have affected reported numbers.		
3	Review the dates on the source documents. Do all dates fall within the reporting period?		
	If no, determine how this might have affected reported numbers.		
B - Recounting reported Results:			
	Recount results from source documents, compare the verified numbers to the site reported numbers and explain discrepancies (if any).		
4	Recount the number of people, cases or events recorded during the reporting period by reviewing the *source document*s. **[A]**		
5	Copy the number of people, cases or events reported by the site during the reporting period from the site *summary report*. **[B]**		
6	**Calculate the ratio of recounted to reported numbers. [A/B]** -		
7	What are the reasons for the discrepancy (if any) observed (i.e., data entry errors, arithmetic errors, missing source documents, other)?		

C - Cross-check reported results with other data sources:			
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. Other cross-checks 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 crosschecks should be performed in both directions (for example, from Patient Treatment Cards to the Register and from Register to Patient Treatment Cards).			
	List the documents used for performing the cross-checks.		
	Describe the cross-checks performed?		
	What are the reasons for the discrepancy (if any) observed?		

Table 2: **System assessment protocol**

Part 2. Systems Assessment			
I - M&E Structure, Functions and Capabilities		Answer	Comments
1	There are designated staff (HIT or HMIS focal person) responsible for aggregating data prior to submission to the next level (e.g., to woredas, to regional offices, to the central M&E Unit). [Y/P/N]		
	If yes for Q.No 1, answer questions 2-4		
2	What is his/her qualification?		
3	Does he/she take supervisory level training (above 10 days)? [Y/N]		
4	Does he/she is responsible to provide refresher training or to train new staffs? [Y/N] if no why?		
5	The responsibility for recording the delivery of services on source documents is clearly assigned to the relevant staff. [Y/N]		
6	Do all staffs need to be trained on HMIS got training? (Y/No) If yes skip Q No 7		
7	The number of staffs trained on HMIS?		
8	Total number of staffs needs to be		

	trained?		
9	Is there one central MRU (All programs are integrated including ART) [Y/P/N]		
10	Does the MRU has fast tracking window [Y/N]		
11	Does the MRU provide service for 24 hours [Y/N]		
12	Does the service delivery point established performance Review Team (as per standard) [Y/N]		
II- Input/Resource			
1	Card room Size (meter square):		
2	Is the card room adequate [Y/N]		
3	Number of Card room Workers		
4	Number of runners		
5	Number of standard shelves (Refer shelf standard from data collectors guide)		
6	Availability of standard MPI box (Refer shelf standard from data collectors guide)		
7	Does the facility has HMIS unit ? [Y/N] (Observe availability of table and chair for HMIS purpose)		
8	Is there computer for HMIS? [Y/N]		
9	Does your facility allocate budget for HMIS activities? [Y/N] *(for printing and supervision*		
10	Is there any concerned outside organ that provide supportive supervision on HMIS (WoHO, Development Partner)? [Y/N] Name:		
11	Are frequent written feedback are given on HMIS supervision finding (WoHO, Development Partner)? [Y/N] Name:		
III - Data - Collection and Reporting Forms, Tools and Guidelines			
1	Cards including tracer cards are available in adequate amounts (stock level of at least of 1 month). [Y/P/N]		
2	Data collection tools (registers and tally sheets) are available in adequate amounts		

	(stock level of at least of 1 month). [Y/P/N]		
3	The standard forms/tools are consistently used by the health facility. [Y/P/N]		
4	The service delivery point uses additional "unofficial" forms, registers, tally or reports.[Y/N]		
5	The service point hasHMIS recording, reporting,indicator and information use guidelines showing what and when it is supposed to report on (including the annual report) [Y/N]		
6	HIT or HMIS focal person knows the recording and reporting procedures (e.g. PMTCT data elements, ANC 1st visit coverage, FP data elements) [Y/P/N]		
7	Do you know how to calculate indicators (for HIT or HMS focal or respective process officers)? If yes, go to Q # 9		
8	HIT or HMIS focal person able to calculate indicators (e.g. PMTCT completion rate, CAR, ANC 1st visit coverage, CAR, etc) [Y/P/N]		
IV- Data Management Processes			
1	Individual folder ordered numerically [Y/N]		
2	Use MPI card for indexing [Y/N]		
3	Check the procedure how the providers provide MRN for new clients (Using sequential MRN or not) [Y/N]		
4	Number of complete Medical records (Take 10 Individual folder and check in all visits)		
5	MRs returned to MRU on daily basis after clients receive their service [Y/N]		
6	Registers are promptly used upon service delivery [Y/N]		
7	All Tally sheets are available in appropriate place and filled regularly [Y/P/ N]		
8	How frequent the HMIS focal person collects tally sheets from service delivery points? Why?		

9	All required reportable data elements are aggregated and filled monthly by HMIS focal person [Y/N]		
10	Data and medical charts are kept confidential [Y/N]		
11	LQAS is performed as per standard monthly and documented [Y/P/N]		
12	The service delivery point keep copies of reports sent to WorHO/ZHD/ (check availability of at least one year report) [Y/N]		
13	The recording and reporting system avoids double counting of new and repeat family planning users within and across Service Delivery Points (e.g., a person receiving the same service twice in the same quarter, a person registered as receiving the same service in two different health facilities, etc). (explain)		
V - Links with National Reporting System			
1	The relevant national forms/tools are used for data-collection and reporting. [Y/N]		
2	Data are reported through a single channel of the national information systems. [Y/N]		
VI - Information Use			
1	The Service Delivery point use demographic data from sources such as survey, census, etc for planning? [Y/N]		
2	Performance Review Ream analyze report (plan vs. achievement) on monthly basis [Y/P/N]		
3	Does the facility develop action plan for recommended activities and disseminate to responsible bodies?[Y/N]		
4	Dos the facility document and follow execution of decisions?[Y/N]		
5	Service Delivery point has discussion about RHIS, findings such as patient utilization, disease data, service coverage, or medicine stock out [Y/P/N]		

6	Does the service Delivery point display information? Please indicate the types of data displayed and whether the data are updated for the last reporting period [Y/P/N]		indicate type of display (graph, table, chart or map)
7	The Service Delivery point has identification and tracing mechanism for "drop out", "lost to follow-up" and died cases in TB, ART and Immunization programs. [Y/P/N]		

Annex 2.Components of the Spider Diagram

1. M&E Structure

- Availability of designated staff (HMIS focal Person) responsible for reviewing aggregated numbers prior to submission to higher levels
- Availability of designated staff (HMIS focal Person) responsible for reviewing the quality of data (i.e., accuracy, completeness and timeliness) received from sub-reporting levels (e.g., service points). **(IAL)**
- Integration level of MRU **(SDP)**
- 24 hours service provision of the MRU **(SDP)**
- Proportion of trained staff on HMIS or data management processes and tools
- Availability of functional performance monitoring team (as per standard)
- Having regular supportive supervision to the lower levels**.(IAL)**

2. Input/Resource

- Adequacy of MRU **(SDP)**
- Number of Card room workers **(SDP)**
- Number of runners **(SDP)**
- Availability of standard shelves **(SDP)**
- Availability of MPI box **(SDP)**
- Availability of HMIS unit
- Availability of computers for HMIS
- Number of staffs work on HMIS ***(RHB & FMOH)***
- Number of data clerks ***(ZHD, RHB & FMOH)***
- Allocation of budget for HMIS activities

3. Data Collection and Reporting Forms, Tools and Guidelines

- Use of standard reporting tools and formats **(SDP)**
- Use of HMIS disease classification list **(SDP)**
- Provision of clear instructions to lower levels on how to complete the data collection and reporting forms/tools **.(IAL)**
- Identification of the standard reporting forms/tools to be used by lower levels **(IAL)**
- Use of the standard forms/tools consistently
- Availability of source documents and reporting forms relevant for measuring the indicator(s) for auditing purposes
- Use of additional "unofficial" forms, registers, tally or reports
- Availability of reporting guideline showing what and when it is supposed to report

4. *Data Management*

- **For Service Delivery Point**
  - ✓ Numeric order of Individual folder
  - ✓ Use of MPI card for indexing
  - ✓ Use of sequential MRN to avoid duplication
  - ✓ Completeness of MRs
  - ✓ Returning time of MRS to MRU
  - ✓ Use of registers at service delivery point
  - ✓ Regular use of tally sheets
  - ✓ Keeping MRs confidential
  - ✓ Performing and documenting LQAS as per standard
  - ✓ Availability of report copies sent to higher levels (one year report)
  - ✓ Aggregation of all reportable data elements on monthly basis

- **For Intermediate Aggregation Levels**
  - ✓ Availability of reports copies received from lower levels (one year report)
  - ✓ Availability of report copies sent to higher levels (one year report)
  - ✓ Compile RHIS data submitted by lower levels
  - ✓ Availability of quality controls when data entered from paper-based forms are into a computer (e.g., double entry, post-data entry verification, etc).
  - ✓ Systematic provision of feedback to lower levels on the quality of their reporting (i.e., accuracy, completeness and timeliness).
  - ✓ Performing and documenting LQAS as per standard
  - ✓ Availability of written procedure to address late, incomplete, inaccurate and missing reports; including following-up with service points on data quality issues.
  - ✓ Availability of documentation on how data discrepancies or inconsistencies have been resolved.
  - ✓ Availability of written back-up procedure for when data entry or data processing is computerized.

5. *Linkage with the National System*

- The relevant national forms/tools are used for data-collection and reporting
- Data are reported through a single channel of the national information systems.
- The health Institution disseminate report to the outside other than the next higher level (WorHO, ZHD, RHB, FMOH)
- The system records information about where the service is delivered (i.e. region, district, ward, etc.) ***(IAL)***

6. Information Use

- Issue reports containing RHIS information
- Performance monitoring team analyze report (plan vs achievement) on monthly basis
- Action Plan is developed and documented for further follow up
- Availability of updated display information/charts/graphs
- Use of demographic data for planning **(SDP)**
- Discussion based on RHIS finding such as patient utilization, service coverage, medicine stock out, etc… **(SDP)**
- Availability of display of summary of demographic information (population by target group, …) **(IAL)**
- Having and providing feedback report (using RHIS information) to lower levels **(IAL)**
- Availability of documents showing districts/senior management directives were based on use of information **(IAL)**
- Availability of feedback, quarterly, yearly or any other report on RHIS data available, which provides guidelines/recommendations for actions? **(IAL)**
- Publish a newsletter or report in last three months that included use of information success stories? **(RHB & FMOH)**