A Menu of Tools for Data Quality Assessment and Review

Name	Representative sample survey		Diagnostic data quality assessment of specific health program or indicators			Intervention- based DQA
	Data Quality Review (DQR)	**Data Quality Audit (DQA)**	**Lot Quality Assurance Sampling (LQAS) Triage System**	**Routine Data Quality Assessment (RDQA)**	**Mini-DQA**	**Expedited DQA (EDQA)**
Description	National assessment across health program areas to assess quality of reported data and adequacy of the information system	Assessment of program health data to assess the impact of data quality on performance	Site-level assessment/ supervision tool to assess completeness and consistency of records and investigate suspected data quality problems	Versatile capacity building and self-assessment tool for a program in a single health area as part of monitoring and supervision	Emphasis on verified counts from source documents and reported data within a single program. This may happen as part of monitoring of a site or at district/ national level	Focused on validating the "active file" (i.e., the current roster of patients on treatment at a service delivery point [SDP]). Includes identifying and following up with patients classified as lost to follow-up (LTFU) in a program.
Dimensions of quality assessed	Accuracy, completeness, timeliness, internal consistency, external consistency, evaluation of denominator values	Accuracy, completeness, reliability, precision, completeness, timeliness, integrity, confidentiality	Accuracy, completeness, timeliness, reliability, internal consistency	Accuracy, completeness, availability, completeness, timeliness	Accuracy, internal validity	Accuracy, internal validity
Source document for validation	Registers, patient files, (and other sources as necessary)	Data from source documents sampled from SDPs (client intake forms, facility or community registers)	Patient medical records at SDPs	Recount within select SDPs from patient files and/or registers	Patient files/ cards, registers, pharmacy records	Active patient files
Verified against	Reported values from SDPs (i.e., previously reported aggregates in the health information system)	Reported values from SDPs (including intermediate and national aggregation levels)	Alternate data sources within the SDP	Aggregate numbers reported by SDPs	Aggregate numbers reported by SDPs	Aggregate numbers reported by SDPs
Suggested sampling (sampling will depend on assessment scope and available resources)	List sampling; sample size determined by desired level of estimation and level of stratification	Cluster sampling with probability proportional to size (PPS); stratified random sample within clusters recommended to achieve a nationally representative sample	Systematic random sampling of patient records within SDPs, with LQAS classification to determine quality	Targeted group of SDPs; purposive or convenience sample. Cluster-based sampling can also be used with PPS.	Purposive sampling to target high-volume facilities/SDPs to obtain an accurate count for a specific indicator	Facilities identified to have significant discrepancies following standard data quality assessment should be targeted for this exercise.
Number of indicators*	Up to 5 indicators for up to 5 priority health programs, or 5 program-level indicators for in-depth program assessment	Program- level indicator or group of related indicators	Up to 4 indicators within the same health or disease program	Up to 4 indicators within the same health or disease program	Single (or up to 3 related indicators)	Single (or up to 3 related indicators)
Time to implement	Allow 3–6 months for health facility surveys. Desk review component requires 2–4 weeks	6 phases of implementation over 6–11 weeks, depending on sample	1–2 weeks, depending on sample	Requires roughly 1 week to plan, gather data, implement, and report results	Depends on a number of SDPs; a single program indicator may require 2–3 days	5–6 weeks to plan, organize, train, and conduct a verified count
Frequency of implementation	Recommended annually or at minimum at the start and midpoint of a 5-year planning cycle	As needed, every several years	Not specified; as needed	Regular and repeated as part of routine supervision	Not specified; as needed	Not specified; as needed
Cost**	$$$	$$	$-$$	$-$$	$-$$	$-$$$

*Indicators are constructed of data elements, so you may be collecting several data elements to construct an indicator.

**Costs provided here are estimates based on previous assessment. Costs will vary greatly, depending on scope, sample size, location, personnel, and training requirements

$: $50,000–100,000 at the time of implementation (2018); **$$:** $100,000–$200,000; **$$$:** $200,000+

Introduction

Robust systems are essential to track progress toward health objectives, such as the United Nations Sustainable Development Goals, and to support evidence-based decision making. Different approaches may be followed, to assess and improve data quality and data management and to make informed decisions for planning to improve quality and to achieve expected health outcomes.

This document presents a menu of options for data quality assessment and is meant to provide guidance on which approach would be the most suitable for the data and system to be assessed. The scope of the assessment and the depth of data to be collected will depend on the purpose of the assessment. Data quality assessments will focus on one or more dimensions of data quality, such as accuracy, completeness, reliability, timeliness, confidentiality, precision, and integrity.

For details on methods for data quality assessment including recommended sampling, please refer to the Comparative Analysis of Data Quality Assessment Tools:
https://www.measureevaluation.org/resources/publications/tr-18-300?searchterm=comparative+analysis+of+data+

▶ Data Quality Review (DQR) Toolkit:

https://www.measureevaluation.org/our-work/data-quality/data-quality-review

What is it? Backed by the World Health Organization and the Global Fund, the DQR is a comprehensive approach for harmonized data quality assurance across health programs. It can be holistic or program-specific in its approach. DQR uses a master facility list for sampling. Sample size is determined by the desired level of estimation (e.g., national or regional) and the level of stratification required.

Who should use it? Ministries and their partners looking to do a comprehensive, independent review of data quality to inform health sector planning. The DQR is frequently implemented as one module of a larger health facility assessment (e.g., a Services Availability and Readiness Assessment [SARA]) to take advantage of typically larger sample sizes.

▶ Data Quality Audit (DQA):

https://www.measureevaluation.org/resources/publications/ms-08-29

What is it? The DQA tool comprises 16 indicator-specific quantitative assessment templates for HIV/AIDS, malaria, and Tuberculosis in Microsoft Excel to evaluate data quality, and a generic qualitative "System Assessment" module to assess the reporting system for gaps and weaknesses. The DQA employs cluster-sampling and traces reported results for a selected reporting period from sampled health facilities through intermediate aggregation levels to the national level, from which country-level estimates of reporting accuracy are derived. Sample size requirements will depend on desired precision of the estimates and domain of estimation (i.e. national or regional).

Who should use it? Those looking to a do a rigorous, independent (often external), program-specific data quality audit, e.g. for performance-based financing models. Sampling is typically smaller than for a DQR. The DQA can be used as needed to identify issues of data quality for specific health programs.

▶ Routine Data Quality Audit (RDQA):

https://www.measureevaluation.org/resources/publications/ms-17-117

What is it? The RDQA is a self-assessment and capacity building version of the DQA. Its principle purpose is to conduct routine data quality checks on a targeted group of health facilities and the aggregation levels through which they report. Identified data quality problems are grouped by the level of the health system, to better tailor interventions to improve data quality. The Gender-Integrated RDQA (RDQA+G)—https://www.measureevaluation.org/our-work/gender/gender-integrated-routine-data-quality-assessment-rdqa-g-tool/gender-integrated-routine-data-quality-assessment-rdqa-g-tool—builds on this tool, enabling national programs or projects to evaluate their own data quality with a special focus on gender data (including sex and age disaggregation). The RDQA can be used with up to four indicators at one time, or modified to look at one indicator over four time periods.

> **Read about Botswana's adaption of the RDQA Tool:** https://www.measureevaluation.org/resources/publications/sr-13-79—to develop national procedures for data quality assurance.

Who should use it? Stakeholders who need to assess their own data quality and prepare for audits by donors. RDQA is versatile and can be easily integrated in program SOPs. Because it can assess up to four program indicators in a given health area, RDQA is particularly well suited for cascade-type indicators.

▶ Lot Quality Assurance Sampling (LQAS) Triage System

What is it? A data quality assessment that uses LQAS to classify source documents in health facilities as meeting or not meeting a predetermined standard of quality. The so-called "LQAS triage system" allows managers to quickly assess the completeness and consistency of data in source documents and diagnose specific data quality issues. The LQAS triage system can be applied as a part of routine supervision, to catch data quality problems before they have an impact on reported numbers.

> **Read about how MEASURE Evaluation-supported LQAS sampling in Burundi:** https://www.measureevaluation.org/resources/publications/tr-18-316/

Who should use it? Stakeholders wishing to address weaknesses in data collection at facilities. The LQAS triage system is particularly well suited to indicators with extensive record keeping from multiple sources, such as longitudinal treatment-based indicators (e.g., for HIV and tuberculosis).

▶ Mini Data Quality Assessment (Mini-DQA)

What is it? The mini-DQA's goal is to obtain an accurate count for a specific indicator. This can act as a targeted diagnostic of source documents. A mini-DQA may identify SDPs that require further query into quality of the data. Then an expedited data quality assessment (EDQA) can be used to follow up on issues and current forms.

> **Read about MEASURE Evaluation's use of Mini-DQA and EDQA together in Zambia:** https://www.measureevaluation.org/resources/publications/tr-17-228

Who should use it? Stakeholders who need a targeted data quality assessment of a specific indicator for high-volume clinics and who may be unable to conduct a DQA of larger sample size.

▶ Expedited Data Quality Assessment (EDQA)

What is it? This approach is a novel method for classifying cases by treatment status, with a view to organizing patient records and cleaning data. It is used to clean up patient files, by clarifying, identifying, and following up with patients classified as LTFU.

> **Read about MEASURE Evaluation's use of Mini-DQA and EDQA together in Zambia:** https://www.measureevaluation.org/resources/publications/tr-17-228

Who should use it? Stakeholders needing to clean up files and improve the quality of data for the HIV patients currently on treatment. The EDQA can be used as an intervention when source documents are found to be incomplete or inconsistent after assessment with the LQAS triage system, or following a more comprehensive data quality assessment using one of the above tools.

USAID
FROM THE AMERICAN PEOPLE

MEASURE
Evaluation