Module 3
Site assessment of data quality: data verification and system assessment
IMPLEMENTATION GUIDE
Module 3
Site assessment of data quality: data verification and system assessment
IMPLEMENTATION GUIDE
VERSION UPDATE - DECEMBER 2020
Acknowledgements ....................................................... vii

Abbreviations .............................................................. viii

Chapter 1. Overview and methods .......................................... 1
  Background .................................................................. 1
  Objectives .................................................................... 2
  Methodology .................................................................. 2
    Site assessment of data quality and system capabilities: Discrete data verification and system assessment .................................... 3
  Indicators ..................................................................... 4
  Cross-cutting versus in-depth .......................................... 4

Chapter 2. Planning for survey implementation ......................... 6
  Roles and responsibilities ............................................... 6
    Ministry of Health ...................................................... 6
    Implementation agency ................................................ 6
    Agency providing quality assurance and technical support ............. 6
    Partners ..................................................................... 6
    DQR coordinating group .............................................. 7
  Indicator selection ....................................................... 7
  Survey steps .................................................................. 8
  Timeline ....................................................................... 9
  Requirements .................................................................. 10
    Budget ........................................................................ 12

Chapter 3. Implementation .................................................... 13
  Resource requirements .................................................... 13
    Scope of the assessment ............................................... 13
    Sampling methods ...................................................... 13
    Sample selection ......................................................... 17
    Master facility list ....................................................... 20
    Data requirements ...................................................... 21
    Data collection tools .................................................... 22
    Completing the DV/SA questionnaire ............................... 22
    Recording the responses ............................................... 22
    Instructions to the interviewer ....................................... 24
    “Skip” instructions ....................................................... 24
    Question types .......................................................... 25
    Numeric responses ...................................................... 26
    Ensuring quality ......................................................... 28
    Correcting mistakes ..................................................... 28
    Checklist ..................................................................... 28
    District-level .............................................................. 29
# Table of Contents

## Data quality review

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Data quality improvement plan materials</td>
<td>16</td>
</tr>
<tr>
<td>2</td>
<td>Data quality assurance plans</td>
<td>23</td>
</tr>
<tr>
<td>3</td>
<td>Data quality review: site assessment of data quality: data verification and system assessment – Implementation guide</td>
<td>29</td>
</tr>
<tr>
<td>4</td>
<td>Data analysis</td>
<td>80</td>
</tr>
<tr>
<td>5</td>
<td>Dissemination and use of DQR results</td>
<td>97</td>
</tr>
</tbody>
</table>

### Chapter 4. Data analysis

**Automated chartbooks in MS Excel**
- Calculate DQR indicators
- District-level data analysis
- Deriving the sample weights

**Chapter 5. Dissemination and use of DQR results**

- Data validation workshop
- Final report
- Develop a Data Quality Improvement Plan
  - Purpose
- Data Quality Improvement Plan process
  - Engage stakeholders
  - Review of DQR results
  - Action planning
  - Monitor and follow up
- Example data quality problems and potential solutions
  - Accuracy of reporting
  - Missing or incomplete data
  - Example outline of a Data Quality Improvement Plan

### Annex 1: Recommended indicators

- Additional indicators
List of tables
Table 1. Recommended core indicators for the DQR
Table 2. Survey steps
Table 3. Illustrative timeline for implementation
Table 4. Resources requirements for DV/SA
Table 5. Hardware and software specifications
Table 6. Budget template for DQR implementation
Table 7. Selective sample size calculations with a range of marginal prevalence values, percentage agreement and corresponding kappa values
Table 8. Intended audiences for each training session
Table 9. Suggested training lengths and emphases for different target audiences
Table 10. Illustrative agenda for 5-day data collector training
Table 11. Example table from DQR facility-level Data Analysis Excel chartbook – verification factors
Table 12. Example table from DQR facility-level Data Analysis Excel chartbook – System assessment
Table 13. Tabular summary of a representative sample survey of facilities (n = 310)
Table 14. Template: Agenda for data analysis and verification workshop
Table 15. Stakeholder engagement matrix
Table 16. Intervention prioritization matrix
Table 17. Example table for scheduling and budgeting activities

List of figures
Figure 1. Select the facilities to be included in the sample
Figure 2. Freeze the random numbers
Figure 3. Filter the random numbers
Figure 4. Recording responses to the questionnaire
Figure 5. Questionnaire instructions
Figure 6. Knowing which questions to skip
Figure 7. Recording answers to precoded questions
Figure 8. Using a response grid
Figure 9. Recording numeric responses
Figure 10. Correcting a mistake
Figure 11. Recording of facility reports at the district level
Figure 12. Aggregating and comparing data
Figure 13. Data collection form
Figure 14. Maintaining a list of facilities that have been assessed
Figure 15. Opening the DV/SA data entry application
Figure 16a. The Garmin eTrex GPS unit
Figure 16b. GPS instruction screens
Figure 17. Format of the DV/SA paper questionnaire
Figure 18. CSPro tools
Figure 19. Keeping track of facilities sampled
Figure 20. Opening the facility application
Figure 21. CSPro facility data
Figure 22. Facility batch results in CSPro
Figure 23. Specify the output data file and dictionary file
Figure 24. Run the compare tool
Figure 25. Typical output from data file comparison
Figure 26. Batch files and weighting
Figure 27. Running the indicator batch file
Figure 28. Checking the list of indicators
Figure 29. Exporting data to MS Excel
Figure 30. Inserting labels into the Excel chartbook
Figure 31. Entering survey information into the Excel chartbook
Acknowledgements

This toolkit is the result of collaboration between the World Health Organization (WHO), The Global Fund to Fight AIDS, Tuberculosis and Malaria (The Global Fund), Gavi, the Vaccine Alliance (Gavi) and the United States Agency for International Development (USAID)/MEASURE Evaluation.

The toolkit proposes a unified approach to data quality. It integrates and builds upon previous and current tools and methods designed to assess data quality at facility level, taking into account best practices and lessons learned from many countries.

Kavitha Viswanathan oversaw the technical development of the toolkit, with overall guidance provided by Kathryn O’Neill and with technical input from Ties Boerma, David Boone, Robert Pond, Olav Poppe, Claire Preaud, Ashley Sheffel, Amani Siyam, and Marina Takane.

All modules in the toolkit are based on technical contributions from WHO departments and programmes as well as other agencies with specific contributions from Isabelle Bergeri, Txema Callejas, Thomas Cherian, Richard Cibulski, Katherine Floyd, Marta Gacic-Dobo, Philippe Glaziou, Jan Van Grevendonk, Peter Hansen, Chika Hayashi, Ryuichi Komatsu, Chung-won Lee, Michael Lynch, John Puvimanasinghe, Annie Schwartz, Alka Singh, Hazim Timimi, Serge Xueref, Nathalie Zorzi.

Particular thanks are due to the country partners that tested and implemented different components of the approach and provided valuable feedback, namely: Benin, Burkina Faso, Cambodia, Democratic Republic of the Congo, Kenya, Mauritania, Sierra Leone, Togo, Uganda, Zambia, and Zanzibar in United Republic of Tanzania.

This toolkit has been produced with the support of grants from the Bloomberg Philanthropies Data for Health Initiative; Gavi; the Global Fund to Fight AIDS, Tuberculosis and Malaria; and the Norwegian Agency for Development Cooperation (Norad).
## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AHSR</td>
<td>Annual Health Sector Review</td>
</tr>
<tr>
<td>ANC</td>
<td>Antenatal care</td>
</tr>
<tr>
<td>ANC1</td>
<td>Antenatal care first visit</td>
</tr>
<tr>
<td>ART</td>
<td>Antiretroviral therapy</td>
</tr>
<tr>
<td>DHIS 2</td>
<td>District Health Information System Version 2</td>
</tr>
<tr>
<td>DHMT</td>
<td>District Health Management Team</td>
</tr>
<tr>
<td>DHS</td>
<td>Demographic and Health Surveys</td>
</tr>
<tr>
<td>DQR</td>
<td>Data Quality Review</td>
</tr>
<tr>
<td>DTP3</td>
<td>Diphtheria-tetanus-pertussis three-dose vaccine</td>
</tr>
<tr>
<td>DV</td>
<td>Data verification</td>
</tr>
<tr>
<td>EDC</td>
<td>Electronic data collection device</td>
</tr>
<tr>
<td>FBO</td>
<td>Faith-based organization</td>
</tr>
<tr>
<td>Gavi</td>
<td>Gavi, the Vaccine Alliance</td>
</tr>
<tr>
<td>GPS</td>
<td>Global positioning system</td>
</tr>
<tr>
<td>HIV</td>
<td>Human immunodeficiency virus</td>
</tr>
<tr>
<td>HMIS</td>
<td>Health management information system</td>
</tr>
<tr>
<td>M&amp;E</td>
<td>Monitoring and evaluation</td>
</tr>
<tr>
<td>MA4H</td>
<td>Measurement and Accountability for Results in Health</td>
</tr>
<tr>
<td>MCH</td>
<td>Maternal and Child Health</td>
</tr>
<tr>
<td>MCV</td>
<td>Measles-containing vaccine</td>
</tr>
<tr>
<td>MDR-TB</td>
<td>Multidrug-resistant tuberculosis</td>
</tr>
<tr>
<td>MFL</td>
<td>Master facility list</td>
</tr>
<tr>
<td>MICS</td>
<td>Multiple Indicator Cluster Survey</td>
</tr>
<tr>
<td>NGO</td>
<td>Nongovernmental organization</td>
</tr>
<tr>
<td>PCV</td>
<td>Pneumococcal conjugate vaccine</td>
</tr>
<tr>
<td>PDA</td>
<td>personal digital assistant</td>
</tr>
<tr>
<td>Penta</td>
<td>Pentavalent vaccine</td>
</tr>
<tr>
<td>PMTCT</td>
<td>Prevention of mother-to-child transmission</td>
</tr>
<tr>
<td>RDT</td>
<td>Rapid diagnostic test</td>
</tr>
<tr>
<td>RHIS</td>
<td>Routine health information system</td>
</tr>
<tr>
<td>RMNCH</td>
<td>Reproductive, maternal, newborn and child health</td>
</tr>
<tr>
<td>RR</td>
<td>Rifampicin-resistant</td>
</tr>
<tr>
<td>SA</td>
<td>System assessment</td>
</tr>
<tr>
<td>SARA</td>
<td>Service Availability and Readiness Assessment</td>
</tr>
<tr>
<td>SMART</td>
<td>Specific, measurable, achievable, relevant and time-bound</td>
</tr>
<tr>
<td>TB</td>
<td>Tuberculosis</td>
</tr>
<tr>
<td>The Global Fund</td>
<td>The Global Fund to Fight AIDS, Tuberculosis and Malaria</td>
</tr>
<tr>
<td>TT</td>
<td>Tetanus toxoid</td>
</tr>
<tr>
<td>USAID</td>
<td>United States Agency for International Development</td>
</tr>
<tr>
<td>VF</td>
<td>Verification factor</td>
</tr>
<tr>
<td>WHO</td>
<td>World Health Organization</td>
</tr>
</tbody>
</table>
Chapter 1. Overview and methods

Background

High-quality data is essential to monitor progress towards the Sustainable Development Goals (SDGs), WHO Triple Billion targets and national or subnational health priorities. It is also vital to strengthen country capacity to prevent, prepare for and respond to health emergencies such as COVID-19. Timely, reliable, actionable data is essential for delivering interventions to improve the health of populations.

The recently published global report on country capacity on health data and systems using the SCORE technical package has highlighted that improving data quality is essential for policy and planning. Data availability does not automatically translate into availability of the quality data needed for policy, planning and patient health care. Data quality is a critical issue for health facilities with about 40% of countries not showing clear evidence that data quality assurance processes have been followed for their published health facility data1.

1 https://cdn.who.int/media/docs/default-source/world-health-data-platform/score/who_2021-01-31_global-report-score_tb_v2.pdf?sfvrsn=cf86a4fb_3&download=true

High-quality data is essential to monitor progress towards the Sustainable Development Goals (SDGs), WHO Triple Billion targets and national or subnational health priorities. It is also vital to strengthen country capacity to prevent, prepare for and respond to health emergencies such as COVID-19. Timely, reliable, actionable data is essential for delivering interventions to improve the health of populations.

The recently published global report on country capacity on health data and systems using the SCORE technical package has highlighted that improving data quality is essential for policy and planning. Data availability does not automatically translate into availability of the quality data needed for policy, planning and patient health care. Data quality is a critical issue for health facilities with about 40% of countries not showing clear evidence that data quality assurance processes have been followed for their published health facility data1. The Data Quality Review (DQR) is a methodology for rapid evaluation of the quality and adequacy of health data used for planning. The DQR aims to institutionalize data quality assessment as a systematic and routine aspect of health sector and programme planning and provide a minimum standard of quality for routine health data. It is intended to be applied across programme areas to provide a holistic picture of country data quality from health facility-based information systems, and to identify areas in need of strengthening. The methodology and indicators for the DQR have been developed in consultation with international health programme experts from leading donor and technical assistance agencies such as the World Health Organization (WHO), Gavi, the Vaccine...
The DQR examines data within four domains: 1) completeness and timeliness of data, 2) internal consistency of reported data, 3) external comparison with other data sources, and 4) consistency of population data used in the calculation of rates for monitoring programme coverage. Within these domains priority indicators are examined to find anomalous or extreme values, quantify missing or zero values, and evaluate reporting consistency over time. The output of the DQR is intended to highlight potential data quality problems and stimulate discussion about their causes. Implementation of the DQR should result in action plans to fill gaps, correct errors, and strengthen health sector planning data.

Implementation of the DQR can help build confidence in the data for both national and external stakeholders. Knowing the data and their limitations can improve decision making during planning exercises and provides reassurance to donors and other key stakeholders that the evidence base for planning has undergone a known minimum level of scrutiny that adheres to international standards. The DQR is a cross-cutting tool for all health facility administrative data which can be supplemented by in-depth programme-specific assessment on a periodic basis. It is intended to harmonize with, or work alongside, existing programme-specific tools with similar aims. It should be implemented with an element of independence in order to promote transparency in the data and the health sector planning process.

The DQR is a suite of tools and guidelines, including electronic tools to facilitate data collection and analysis. These guidelines provide instructions for collecting the data, preparing the data for analysis, conducting data verifications, analysing and interpreting the results, as well as guidance on how and when to apply the methods. The electronic tools facilitate data analysis and presentation, as well as the identification of problematic data points and subnational reporting units.

Objectives

The DQR is designed to assess the quality of data generated by information system(s) based in health facilities. The objectives of the DQR are:

- to institutionalize a system for assessing the quality of data, including routine monitoring of data, discrete data quality reviews (conducted annually) and periodic in-depth assessments of priority health programmes;
- to identify weaknesses in the data management system and interventions for system strengthening; and
- to monitor the performance of data quality over time and the capacity to produce good-quality data.

Alliance (GAVI) and The Global Fund to Fight AIDS, Tuberculosis and Malaria (the Global Fund) and constitutes consensus on a minimum standard for data quality.
Methodology

The DQR is envisioned as a suite of regularly implemented tools coordinated to provide an evidence base for data quality in advance of health sector planning. The suite of tools has a variety of components, each with recommended periodicity. The tools and methods include:

- **Site assessment of data quality and system capabilities**
  - Discrete site assessment at facilities and districts – data verification and system assessment (DV/SA) on a nationally representative sample of health facilities to provide information on the accuracy of reporting for priority indicators that are generalizable to all health facilities providing the service. Ideally conducted annually, the site assessment should be implemented as often as is feasible with country resources and should feature prominently in the health sector’s five-year planning cycle.
  - Routine data quality assurance checklists – a system of routine and regular (i.e. monthly) reviews of data quality of the health management information system (HMIS) or other programme reporting systems as part of a feedback cycle that identifies and rectifies errors in near real-time. The routine reviews are conducted as a part of regularly scheduled supervisory visits to health facilities using a standard data quality checklist. This routine system of data quality checks has two components: monthly self-assessment of HMIS data conducted by health facility staff, and a periodic (ideally quarterly) assessment of health facility data by district-level staff during supervisory visits to the health facility.

- **Data quality desk review** – an analysis of aggregate reported data in the HMIS to look for gaps, outliers and inconsistencies for priority indicators across health and disease programmes.
  - Discrete assessment – An ad hoc desk review of data quality usually conducted at national level and scheduled to coincide with a discrete site assessment of data quality and system capabilities, an annual planning event, or used to investigate suspicious reporting patterns.
  - Continuous desk review of data quality – routine (e.g. monthly analysis of standard metrics to determine completeness and consistency of reported data from health facilities. This should ideally be conducted at the district level for facilities in the district so that errors are found and corrected as they are reported. The review can also be applied at the national level and for specific health and disease programmes.

The DQR framework documents describe the methodology (how it is conducted) and metrics (what is assessed) used in the DQR and provides guidance on the use of all these tools. This Implementation guide is specific to the discrete site assessment of data quality and system capabilities and addresses requirements for conducting a health facility assessment on a sample of health facilities. A sister document – the DQR Implementation guide for discrete desk review – is also available. Guidance for routine site assessment and desk review can be found in

---

Chapter 1. Overview and methods
Site assessment of data quality and system capabilities: Discrete data verification and system assessment

The site assessment uses data verification (DV) to validate reported data for priority indicators. Data are verified by comparing a reported value for a given reporting period to a validated value – i.e. a value recounted by assessment team members in health facilities using available source documentation (e.g. registers and report forms). The system assessment (SA) is a qualitative evaluation of the capability, or readiness, of the information system to report accurate data. Together they are referred to in shorthand as “DV/SA”. DV/SA is typically implemented in a representative sample of health facilities. It can be implemented as a stand-alone data quality assessment or as a component of a larger health facility assessment – e.g. to measure service availability and readiness, and/or programme quality. The DV/SA is meant to be a feature of the planning cycle whereby data quality assessment is conducted in advance of planning so that planners have knowledge of the strengths and limitations in the data prior to planning events. Thus, the DV/SA should be scheduled several months before the health-sector planning process.

The DV/SA is conducted for up to five tracer indicators (one tracer indicator per health programme). Completeness of source documents and completeness and timeliness of reporting are also measured from the health-facility and district data.

Standard data collection tools have been prepared and a database application (CSPro 7) is available for both PC and tablet computers to facilitate and standardize data collection and entry.

Indicators

The DQR is designed to assess data quality for routine health information systems holistically. It uses tracer indicators from up to five programme areas to judge data quality for the whole system. Tracer indicators are those that are indicative of data quality for all indicators in the health programme. WHO recommends the indicators and programmes listed in Table 1.

Table 1. Recommended core indicators for the DQR

<table>
<thead>
<tr>
<th>Programme area</th>
<th>Abbreviated name</th>
<th>Indicator name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maternal health</td>
<td>Antenatal care 1st visit (ANC1) coverage</td>
<td>Number and % of pregnant women who received antenatal care at least once during their pregnancy</td>
</tr>
<tr>
<td>Immunization</td>
<td>DTP3/Penta3 coverage</td>
<td>Number and % of children &lt; 1 year receiving three doses of DTP/ Penta vaccine</td>
</tr>
<tr>
<td>HIV</td>
<td>New on ART</td>
<td>Number of people living with HIV who initiate ART</td>
</tr>
<tr>
<td>TB</td>
<td>TB notification rate</td>
<td>Number of new and relapse cases of TB that are notified per 100 000 population</td>
</tr>
<tr>
<td>Malaria</td>
<td>Confirmed malaria cases</td>
<td>Number of malaria cases positive by microscopy, RDT or molecular diagnostic test</td>
</tr>
</tbody>
</table>

Note: ANC = antenatal care; ART = antiretroviral therapy; DTP3 = diphtheria-tetanus-pertussis three-dose vaccine; Penta = pentavalent vaccine; RDT = rapid diagnostic test; TB = tuberculosis.
While it is recommended that countries should assess the indicators from the core list, they may select other indicators or expand the set of indicators depending on current needs and available resources. A full set of core and supplemental indicators is available in Annex 1 of the DQR framework document *Module 3 – Site assessment of data quality: data verification and system assessment*.

As “tracer indicators”, these indicators should be indicative of data quality for the entire health programme. As such, they should not be the most difficult to collect and compile, nor the easiest. The selection of priority indicators is also often determined by suspicions of data quality problems, or the level of investment made to collect and report the data. All these factors should be weighed when selecting the appropriate indicator for each programme area.

**Cross-cutting versus in-depth**

The DQR provides information on up to five programme areas to give an overall view of data quality for the health system. For the DQR to remain practical as a facility assessment, the information requirements need kept to a manageable minimum for each health programme. Not all information on data quality can be collected for all health programmes. In reality, health programmes often need more detail on data quality for programme management and planning than can be obtained by the cross-cutting DQR. In such cases, the DQR can be adapted to focus periodically on the broader information needs of a particular health programme. Such application of the DQR is referred to as “in-depth DQR” and is anticipated by the DQR framework. An in-depth DQR would be likely to feature 4–5 indicators from a given health programme, such as vaccinations for priority antigens with data on commodities tracking for the immunization programme, or the testing and treatment cascade for HIV/AIDS. In-depth assessments can be included every few years for a given health programme, depending on in-country needs.

Please see the DQR toolkit *Module 3 – Site assessment of data quality: data verification and system assessment* for more information on in-depth application of the DQR, and the table “Additional indicators” in Annex 1 of this document for a list of suggested additional indicators by programme area.

Standard DQR data collection tools (paper and electronic) and analysis tools would require adaptation for the in-depth DQR. See the section on Tool adaptation below for details on the adaptation of DQR tools for country use.
Chapter 2. Planning for survey implementation

Roles and responsibilities

The survey is usually undertaken under the overall leadership of the Ministry of Health. The following section briefly outlines the roles and responsibilities of the key parties involved in the planning of the DQR.

Ministry of Health

The Ministry of Health will have overall responsibility for the coordination of this process. It will coordinate and provide support to obtain permission to conduct data collection activities, and will help with the coordination of analysis and results dissemination meetings by inviting all appropriate governmental departments and key nongovernmental and development partners. The Ministry of Health will also promote the use of these data for policy and planning.

Implementation agency

The implementation agency will be responsible for conducting field data collection for the DQR and the data verification component of the Data Quality Review. The implementation agency is often a unit within the Ministry of Health (e.g. Health Information Management Unit, Statistics Bureau, etc.) or a nongovernmental organization (NGO) with survey research experience.

Agency providing quality assurance and technical support

It is recommended that an independent party should be involved in the implementation process. This support can be provided by a separate national institute or independent consultant. He/she will be responsible for providing support to the implementation team on planning and implementing DQR; providing a quality assurance role to ensure due processes are followed during training, data collection, cleaning and analyses (including validation visits in 5–10% of the facilities); and providing assistance and oversight to the implementing team on the production of the DQR and data quality assessment report.

Partners

The Measurement and Accountability for Results in Health (MA4H) 5-point call to action recommends that partner investments in health information be fully aligned with a single country platform for information and accountability. Thus, development partners will probably
be stakeholders in the DQR implementation and results. It is important to ensure that in-country partners are included in the decision making process for planning and implementing the DQR. Additionally, partners can be a valuable source of technical assistance and other resources for survey implementation.

**DQR coordinating group**

Bringing country stakeholders together is a critical first step towards successful implementation of the DQR. One of the first activities is to identify and establish a group of core stakeholders at country level to oversee, coordinate and facilitate the planning and implementation of the DQR and the dissemination and use of the DQR findings.

The group should comprise technical focal points among health-sector stakeholders from government (including the different programme stakeholders), development partners and multinational organizations such as WHO, GAVI and the Global Fund. Monitoring and evaluation (M&E) technical working groups or health information system governance boards, which already exist in many countries, can serve as the DQR coordinating group. Development and technical partners can greatly contribute to the success of efforts to improve data quality and should agree on a standardized set of data quality indicators.

The role of the DQR coordinating group is:

- to develop a harmonized plan for data quality assessments;
- to identify technical support requirements for implementation and quality assurance;
- to identify funding sources;
- to oversee the selection of core indicators and the establishment of benchmarks;
- to monitor implementation of the DQR;
- to ensure promotion and dissemination of the findings.

**Indicator selection**

Indicators should be selected with care. Each programme indicator should be indicative of data quality for the whole programme since we are judging data quality for the programme on the basis of the results of the selected tracer indicator. As such the indicator selected should not be the most difficult to compile and report monthly, or the easiest. Suspicions of data quality problems, or the level of investment in terms of time and resources for certain indicators, will often ultimately determine the selection of priority indicators for the assessment. Ensure that all stakeholders have had a chance to give their views on the selection of indicators and that consensus is reached before finalizing the selection.
Survey steps

A DV/SA should be planned to coincide with, and generate data to feed into, the national health planning cycle. The time needed to complete a DV/SA depends on the size of the country and whether or not there is a need for a full facility census. From the initial country-adaptation of the assessment tool to the dissemination of data and production of country reports, the entire process generally takes from three to six months.

Table 2 provides an overview of the survey’s steps and the activities to be undertaken at each step.

Table 2. Survey steps

<table>
<thead>
<tr>
<th>Steps</th>
<th>Survey activities</th>
</tr>
</thead>
</table>
| 1. Survey planning and preparation | 1. Establish a survey coordinating group of country stakeholders to oversee and facilitate the objectives, scope, design, implementation and analysis.  
2. Obtain a list of all health-facility sites (public, private, nongovernmental organizations [NGOs] and faith-based organizations [FBOs]), including country facility registry codes.  
3. Determine an appropriate design methodology (census or sample), develop an implementation plan and budget, and secure funding.  
4. Review and adapt questionnaires to meet country-specific needs.  
5. Recruit survey personnel (survey manager, field supervisors, data collectors, data entry/processing personnel, data analysts).  
6. Prepare a survey schedule.  
7. Identify the survey sites (sampling frame). Select the sample size and a sample of health facilities (if a sampling methodology is chosen).  
8. Procure logistics, including equipment and transport, taking into consideration the number of sites to be visited, the number of data collection teams, drivers, vehicles, fuel, etc.  
9. Plan and conduct training courses for interviewers and field supervisors.  
10. Pilot-test the survey in a selected number of health facilities, evaluate results and make amendments if necessary. |
| 2. Data collection in the field | 11. Plan the data collection visits (prepare a letter of introduction, contact each site, prepare a schedule of visits).  
12. Prepare materials and tools for data collectors.  
13. Arrange for transport and regular communications during fieldwork.  
15. Confirm appointments with health facilities.  
16. Visit health facilities and collect DQR data in teams (usually two interviewers and a driver).  
17. At the end of the interview, check the questionnaire and resolve missing/unreliable information.  
18. Return completed forms and/or transfer electronic files to the field supervisor at the conclusion of each day.  
19. Return forms (paper and/or electronic) to the survey manager when data collection is complete.  
20. Conduct validation visits in surveyed sites (5-10%) to ensure quality of the collected data. |
| 3. Data processing, analysis and interpretation | 21. Enter data using the CSPro application1 (on site or at the end of the day).  
22. Edit, validate and clean the data set. Check for consistency and accuracy.  
23. Export the data set for analysis (DQR indicators).  
24. Conduct analyses of DQR core indicators using the DQR Chartbook automated tables and graphs as well as any country-specific indicators of interest.  
25. Conduct Desk Review analyses of routine data available at national level. |
| 4. Results dissemination | 26. Meet with the survey coordinating group to analyse and interpret the survey results and to finalize recommendations.  
27. Conduct a validation workshop on the data and results to check the results for accuracy.  
28. Prepare the final report.  
29. Plan and implement dissemination activities. The results should be used to support annual health reviews and feed into the M&E platform for the national health plan.  
30. Document and archive the survey using metadata standards. |

Note: ANC = antenatal care; ART = antiretroviral therapy; DTP3 = diphtheria-tetanus-pertussis three-dose vaccine; Penta = pentavalent vaccine; RDT = rapid diagnostic test; TB = tuberculosis.
Timeline

The DQR is ideally conducted in advance of health-sector planning so that the results are available prior to the planning event. From planning to results dissemination the total time required could be as long as six months. Surveys with regional-level or district-level domains of estimation could take longer due to the larger sample size requirements. Ample time should be allowed to ensure adequate planning and preparation for the survey implementation. If tools need to be obtained (e.g. tablets for electronic data entry), these should be ordered well in advance to ensure they arrive in the country well before the period reserved for survey implementation. If technical assistance is required, consultants should be identified and the contractual details arranged well in advance. Finally, large surveys are rarely completed exactly as planned or on schedule. Anticipate delays and have plans, staff and resources in place to address problems quickly as they arise and resolve them. See Table 2 for a list of steps for survey implementation. See Table 3 for an illustrative timeline for survey implementation.

Table 3. Illustrative timeline for implementation
Requirements

Planning for data collection requires consideration both of the logistical needs of data collection teams (Table 4) and of the hardware and software needed for data collection (Table 5).

Table 4. Resources requirements for DV/SA

<table>
<thead>
<tr>
<th>I. Data resources</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Master facility list (MFL)</td>
<td>List of all health facilities in the country (including private sector)</td>
</tr>
<tr>
<td>HMIS tools</td>
<td>Standard registers, tally sheet and reporting forms</td>
</tr>
<tr>
<td>Country adapted data verification module</td>
<td>Selection of 4 or 5 indicators from the proposed list of indicators</td>
</tr>
<tr>
<td>Desk Review data requirements</td>
<td>In order to conduct the DQR the following HMIS data should be made available: monthly data for the selected RMNCH, immunization, HIV, TB and malaria indicators</td>
</tr>
<tr>
<td></td>
<td>Annual totals at district level for the above indicators for the preceding 3 years</td>
</tr>
<tr>
<td></td>
<td>Other country-specific indicators (depending on the data verification selection)</td>
</tr>
<tr>
<td></td>
<td>Estimated denominators for the above indicators (for the year of the Annual Health Sector Review (AHSR) and preceding years if possible)</td>
</tr>
<tr>
<td></td>
<td>Data on the number of facilities reporting and number of facilities in the district (for each district)</td>
</tr>
<tr>
<td></td>
<td>Data on the number of districts reporting every month/trimester</td>
</tr>
<tr>
<td></td>
<td>Latest household survey data aligned to indicators selected for review</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>II. Human resources</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Survey manager</td>
<td></td>
</tr>
<tr>
<td>Field supervisors</td>
<td>Number depends on the sample size and defined timeframe</td>
</tr>
<tr>
<td>Data collectors</td>
<td>Number depends on the sample size and defined timeframe</td>
</tr>
<tr>
<td>Data entry personnel</td>
<td>Depends on sample size; however, the following guideline can be used for planning purposes: for double data entry – two teams of two data entry clerks can enter data for 100 facilities in about 5 days</td>
</tr>
<tr>
<td>Data analysts</td>
<td>Key resource persons from the survey team, technical units and partners to be involved</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>III. Logistics</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Transport for data collection fieldwork activities</td>
<td>Vehicles and drivers</td>
</tr>
<tr>
<td>Field accommodation for data collectors</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>IV. Training resources</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Training venue</td>
<td></td>
</tr>
<tr>
<td>Daily accommodation for participants</td>
<td></td>
</tr>
<tr>
<td>Transport</td>
<td>Vehicles and drivers</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>V. Electronic equipment</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Computer</td>
<td>Computer (PC) for data entry and processing</td>
</tr>
<tr>
<td></td>
<td>Charger</td>
</tr>
<tr>
<td></td>
<td>Extra battery</td>
</tr>
</tbody>
</table>
Mobile electronic data collection devices (EDC)

Mobile data collection unit (such as PDA, tablet computer, laptop computer): for training purposes, one per participant; for survey purposes, one per data collection team plus two backup units
Charger/battery: two chargers per unit, two sets of batteries per unit
Mobile unit carry-case: one per unit
PC/mobile unit connector cable: one per unit
Memory card (if applicable): one per unit

GPS devices (if running adjunct to primary EDC)

GPS device: for training purposes, one unit per participant; for survey purposes, one unit per data collection team plus two backup units
Charger/battery: two chargers per unit, two sets of batteries per unit
GPS carry-case: one per unit
PC/GPS connector cable: one per unit

Communication equipment

Cell phones: one per data collection team
Chargers: one per unit
Mobile phone credit

Software for data collection/entry

The Census and Survey Processing System (CSPro) is recommended unless the country already uses other software. CSPro is a public domain software package for entering, editing, tabulating and disseminating data from censuses and surveys. It can be downloaded from www.census.gov/ipc/www/cspro/index.html. More information on hardware and software specifications can be found in Table 2.
Software manual: one per data collection team

Data analysis software

VI. Supplies

Printing

Printer/copier (two in one or as separate machines)
PC/printer connector cable: one per pair
Ink cartridge
Printing paper: standard size is A4

Pens, pencils

Projector and projector screen

Multi-port extension cable

International power adapters

USB keys (1GB)

Table 5. Hardware and software specifications

<table>
<thead>
<tr>
<th>Computer and software specifications</th>
<th>PDA hardware and software specifications</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Desktop or laptop computer</td>
<td>9. Pocket PC</td>
</tr>
<tr>
<td>2. Pentium processor</td>
<td>10. Windows Mobile 5.0 or 6.0</td>
</tr>
<tr>
<td>3. 512 MB of RAM</td>
<td>11. USB cable or cradle to connect the Pocket PC to the desktop or laptop computer.</td>
</tr>
<tr>
<td>4. SVGA monitor</td>
<td>12. Microsoft ActiveSync 4.2 or later. This should come with your Pocket PC and is also available for download at <a href="http://www.microsoft.com/windowsmobile/en-us/help/synchronize/device-synch.mspx">http://www.microsoft.com/windowsmobile/en-us/help/synchronize/device-synch.mspx</a>.</td>
</tr>
<tr>
<td>5. Mouse</td>
<td>13. CSProMobile. This is the installer for the Pocket PC. It is separate from the installer for the “standard version” of CSPro on a desktop computer. It can be found on the CSPro CD or downloaded from the CSPro website.</td>
</tr>
<tr>
<td>6. 100 MB of free hard drive space</td>
<td></td>
</tr>
<tr>
<td>7. Microsoft Windows 98se, Me, NT, 4.0, 2000, XP, Vista or Windows 7.0</td>
<td></td>
</tr>
<tr>
<td>8. CSPro Version 5.2 (or 4.1 if using PDAs)</td>
<td></td>
</tr>
</tbody>
</table>
Budget

A detailed budget should be developed well in advance of implementing the survey. Funding sources should be identified early in the planning process in order to determine which aspects of the survey will be funded by which organization so that problems do not arise during implementation. Budgets should be developed jointly with partners through a transparent process. Compliance with local policies must be ensured regarding the payment of stipends and/or per diem for survey implementers. Work out ahead of time how expenses should be reconciled against the budget. Involve and budget for finance personnel so that adequate accounting procedures are in place and adhered to. A sample budget template is shown in Table 6.

Table 6. Budget template for DQR implementation

<table>
<thead>
<tr>
<th>Area of work</th>
<th>Activities</th>
<th>No./Quantity</th>
<th>Frequency</th>
<th>Cost/Unit</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Preparation</td>
<td>Planning meetings</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Survey adaptation</td>
<td>Adaptation of the questionnaire(s) and data entry application</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Translation of the questionnaire (if applicable)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Training of data collectors</td>
<td>Training workshop for field supervisors and data collectors (xx data collectors &amp; xx supervisors):</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1. per diem xx USD * number of persons * number of days (8–10 days)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2. travel costs of participants (if applicable)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3. venue, lunch</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Printing of documents for training</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Technical assistance (travel, fee and per diem of facilitators)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pilot-testing</td>
<td>Pilot-testing in at least 3 facilities</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1. USD xx per diem * number of people * 1 day</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2. USD xx transportation * 3 facilities * 1 day</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subtotal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Field survey</td>
<td>Data collectors per diem (USD xx per diem * number of people * number of days)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Field supervisors per diem (USD xx per diem * number of people * number of days)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Drivers, vehicles and petrol @ USD xx * number of days</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Equipment: data collection devices * number needed</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Supplies (e.g. paper forms, mobile phone + units etc.)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subtotal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Data processing, analysis</td>
<td>Data processing and analysis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>and dissemination</td>
<td>1. manager/analyst * 6 weeks</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2. statistician/analyst * 6 weeks</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Analytical workshop</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1. per diem xx USD * number of people * 1 day travel</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2. cost of participants (if applicable)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Presentation of analytical report</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1. venue, lunch</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Validation workshop</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1. per diem xx USD * number of persons * number of days</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dissemination of results (printing report, web posting etc.)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subtotal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contingency/unpredictable costs (around 10%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grand total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Chapter 3. Implementation

WHO recommends that the health facility survey component of the DQR should be conducted in conjunction with a larger health-facility survey to maximize efficiency and conserve resources. The DQR is often conducted as part of a Service Availability and Readiness Assessment (DV/SA). Combination with an existing survey is an efficient way to obtain information from health facilities and may allow for a greater sample size, thereby improving the precision of survey estimates. However, the health-facility survey of the DQR may also be administered as a stand-alone survey.

Resource requirements

The level of effort required for data verification depends on the number of facilities to be assessed (i.e. the sample size), the number of indicators included in the data verification exercise, the volume and organization of the data at the health facilities, and the complexity of the reporting system. To ensure quality in the verification of the data it is recommended that data verifiers work in pairs.

Data verification and system assessment at small facilities generally requires 3–4 hours for an assessment of 4–5 indicators. Larger facilities or hospitals will require more time because the volume of service provision (and of records of service provision) is greater. In general, for a sample of 100 health facilities, 10 data collection teams (with two persons in each) will take 8–10 working days, depending on the factors noted above. This amounts to 160–200 person-days. Depending on whether the data collection is conducted with paper or electronic versions of the questionnaire (or both), several days may be required for data entry and checking prior to analysis.

Scope of the assessment

The DQR coordinating group will determine the scope of the assessment on the basis of the needs of the health system stakeholders and planners, and the resources available. While regional-level or district-level survey estimates are more valuable for planning purposes, the sample sizes required, and therefore the costs for implementation, are much higher. The DQR coordinating group, with donors, partners and other stakeholders, must weigh the relative value of increased granularity of survey estimates against the increased cost to obtain them, and must determine the appropriate scope of the assessment.

Sampling methods

The sample size will depend on the desired precision of the key estimates of interest of the health-facility survey (including data accuracy) and the acceptable margin of error. Other
considerations include the availability of resources and the desired level of application of the estimates (Note: provincial-level estimates require a greater sample size than estimates at the national level). The DQR coordinating group will need to work with a survey statistician and the health-facility survey organizers to determine the appropriate sample size for the survey on the basis of the country’s priorities with regard to the level of application of the estimates, available resources and the precision desired for the estimates.

Below is some brief guidance on key considerations that are necessary when calculating sample sizes for either a stand-alone data verification exercise or for conducting data verification with another health-facility survey. The aim is to determine the sample size that can achieve statistical power or precision of estimation – which means deciding on the minimum number of facilities necessary to obtain a statistically significant result or a confidence interval with a fine enough width to judge the level of agreement.

Most of the estimates described in this guidance involve “agreement” between recounts from source documents and those found in monthly reports. Agreement is a product of 1) a marginal prevalence (i.e. the chance of finding both the source document and monthly report), and 2) the expected proportion of agreement in the counts for the key service outputs being verified (e.g. Penta3, ANC1, confirmed malaria cases, etc.) from the source document and monthly reports. Consequently, it is imperative to ensure a minimum sample size to support a robust measure of agreement (in this instance it is termed “kappa”) beyond what is expected by chance alone. Kappa (ranging from 0 to 1) is a measure of the chance-corrected agreement calculated from the overall percentage agreement and expected agreement by chance.\(^1\) Table 7 provides a selection of sample sizes calculated relative to three scenarios of the marginal prevalence and the permissible range of the necessary two levels of percentage agreement (minimum acceptable agreement \([P_0]\) versus the expected agreement by the study \([P_A]\)) and their corresponding adjusted kappa values.

In scenario A, the DQR coordinating group may not have enough knowledge of the situation regarding the availability of both source documents and monthly reports, so it is appropriate to consider the marginal prevalence value of 0.3 (i.e. 30% chance of finding both documents). Similarly, the team requires an indication of the minimum acceptable agreement level between the two document counts, which needs to be at least 70%. Therefore, with 70% minimum agreement (i.e. \(P_0 = 0.70\)) and a conservative better-than-expected agreement level of 80% (i.e. \(P_A = 0.80\)), a minimum national sample size of \(n = 144\) facilities is needed to provide 80% power and a 95% Confidence Interval (CI) for all key estimates based on the sample, as necessary. In addition, the sample provides inter-observer reliability (given recounts using source documents versus counts reported in monthly reports) and a fair measure of agreement (kappa is between 0.29 and 0.52) that is beyond chance alone.

In **scenario B**, the DQR coordinating group may have a fair knowledge of the chances to find both source documents and monthly reports. If so, the marginal prevalence value of 0.5 is appropriate to consider (i.e. a 50% chance of both documents being available). The team needs to discuss and then choose the minimum acceptable agreement level between the two counts presented in the documents – e.g. at 80% (i.e. $P_0 = 0.80$) and a better than expected agreement level of 90% (i.e. $P_A = 0.90$). With these considerations, a minimum national sample size of $n = 126$ facilities which also provides inter-observer reliability and a substantial measure of agreement (kappa is between 0.60 and 0.80) that is beyond chance alone. If the DQR coordinating group lacks enough knowledge to assert the minimum acceptable agreement level, then the lowest advisable value to consider is 70% (i.e. $P_0 = 0.70$), as indicated in Table 7, with a conservative “better than expected” agreement level of 80% (i.e. $P_A = 0.80$) and therefore a minimum national sample size of $n = 165$ facilities that also guarantee a moderate kappa estimate between 0.4 and 0.6.

In **scenario C**, the DQR coordinating group may have substantial knowledge of the possibility to find both source documents and monthly reports, in which case the marginal prevalence value of 0.80 is appropriate to consider. However, if the DQR coordinating group anticipates a high degree of agreement between counts in source and monthly documents then the minimum acceptable agreement level can be 80% (i.e. $P_0 = 0.80$) with a “better than expected” agreement level of 90% (i.e. $P_A = 0.90$). With these considerations, a minimum national sample size of $n = 100$ facilities is sufficient (with a close-to-moderate estimate of kappa between 0.38 and 0.53).

Finally, taking a closer view of Table 7, two extra points are worth mentioning:

- The sample size increases when the difference between the minimum acceptable level of agreement and that expected from the study is smaller (e.g. when the marginal prevalence is 50%, or 0.5). Choosing $P_0 = 0.80$ and $P_A = 0.85$, the difference is 5% and requires a sample size of $n = 502$, compared to $P_A = 0.90$ when the difference is 10% and requires a sample size of $n = 126$.

- The sample size calculation can also be applied in settings where a subnational representation of the DQR sample is necessary. For instance, in a country with considerable interregional variability in the expected availability of source documents and monthly reports, the DQR coordinating group can choose a conservative marginal prevalence of 30%, a minimum acceptable level of agreement of 75% ($P_0 = 0.75$) to a wider expected agreement level ($P_A = 0.95$), in which case a minimum sample size of $n = 37$ facilities per region is suitable.

For all scenarios, the following sample size formula was used to generate sample size estimates.
If the facility DV/SA is being implemented in conjunction with another health facility survey such as a SARA and a separate sample size calculation has been calculated for each survey, evaluate both sample sizes and apply the larger sample size to both survey modules.

### Table 6. Budget template for DQR implementation

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Marginal prevalence</th>
<th>Percentage agreement</th>
<th>Kappa*</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P₀</td>
<td>P₀</td>
<td>P₀</td>
<td>P₀</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Under the minimum agreement</td>
<td>Under the expected agreement</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>0.3</td>
<td>0.90</td>
<td>0.95</td>
<td>0.76</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.85</td>
<td>0.95</td>
<td>0.64</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.80</td>
<td>0.90</td>
<td>0.52</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.80</td>
<td>0.85</td>
<td>0.52</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.75</td>
<td>0.95</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.75</td>
<td>0.85</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.75</td>
<td>0.80</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.70</td>
<td>0.80</td>
<td>0.29</td>
</tr>
<tr>
<td>B</td>
<td>0.5</td>
<td>0.90</td>
<td>0.95</td>
<td>0.80</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.80</td>
<td>0.90</td>
<td>0.60</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.80</td>
<td>0.85</td>
<td>0.60</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.75</td>
<td>0.95</td>
<td>0.50</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.75</td>
<td>0.85</td>
<td>0.50</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.75</td>
<td>0.80</td>
<td>0.50</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.70</td>
<td>0.80</td>
<td>0.40</td>
</tr>
<tr>
<td>C</td>
<td>0.8</td>
<td>0.90</td>
<td>0.95</td>
<td>0.69</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>0.85</td>
<td>0.90</td>
<td>0.53</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>0.80</td>
<td>0.90</td>
<td>0.38</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>0.80</td>
<td>0.85</td>
<td>0.38</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>0.75</td>
<td>0.80</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>0.75</td>
<td>0.85</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>0.75</td>
<td>0.95</td>
<td>0.22</td>
</tr>
</tbody>
</table>

* Kappa statistic: 0.21–0.40, fair; 0.41–0.60, moderate; 0.61–0.80, substantial.
** Sample size calculated for positive kappa value (type 1 error = 5%, power = 80%).
Sample selection

Facility

Once the sampling frame has been established, probability sampling principles are used to draw a selection of facilities for inclusion in the assessment. Usually, a multistage or stratified sampling plan is followed in order to ensure representation across various domains of the eligible facilities. In stratified random sampling, the sampling frame (or the population) is partitioned into strata (or subpopulations) which are then independently sampled (usually a simple or systematic random sample within each stratum). The results from the different strata are then combined to form estimates for the entire population.

There are several reasons why it is better to use a stratified sample rather than a simple random sample of all facilities. First, a stratified sample guarantees that a prescribed number of facilities from each stratum (or subpopulation) will be assessed, whereas taking a simple random sample of all facilities might result in under-representation of certain types of facilities. In addition, the number of hospitals in a country is generally small compared with the number of primary care facilities, and therefore a simple random sample of all facilities in a country is likely to include only a very small number of hospitals – or it might miss them altogether. By stratifying the sample by facility type, the number of hospitals and primary care facilities can be controlled to ensure that a sufficient number of hospitals are included in the sample. Second, more precise estimates can be obtained where facilities within each stratum are relatively homogeneous and the variation between strata is relatively large. The recommended sampling methodology for Service Availability and Readiness Assessment (SARA) is to select all tertiary-level facilities or hospitals in a country plus a simple random sample of the lower-level facilities stratified by a combination of region, facility type, managing authority and urban–rural distribution. If the facility DV/SA is being implemented in conjunction with a SARA survey, the sample sampling methodology can be applied. If disproportionate allocation is used, sample weights must be applied when analysing the data to calibrate for national representation.

It is often desirable to have separate estimates by region, facility type or other groupings of facilities called domains. Domains are the analytical groupings, whether geographical or categorical, for which separate estimates are required when analysing the results (e.g. primary care facilities versus hospitals; urban areas versus rural areas; public sector facilities versus private sector facilities; different regions). Domains and strata are often synonymous, but this is not always the case because the former is determined by analytical considerations while the latter serves to improve sampling efficiency. The greater the number of domains, the larger the sample size that is required to obtain good estimates.

In general, when equal reliability is wanted for each domain, this necessitates multiplying the calculated sample size needed for a domain by the number of categories in the domain. Thus, if equally reliable estimates were wanted for, for instance, five regions, the sample size would be about five times the value calculated using the equation above. The survey budget would
probably preclude such a large sample, so certain compromises would have to be made. One such compromise is to relax the confidence interval criterion for the domain estimates. Another possibility is to select the most important domains for the stricter reliability and allow the others to be measured with whatever reliability a proportionately allocated sample would yield.

The sample size calculation presented above assumes that all facilities visited will offer all services that are being assessed. However, some services such as those for HIV and TB are offered only at designated facilities. If a complete MFL is available which includes information about the distribution of facilities by services offered, facilities offering HIV and TB should be oversampled in order to ensure they are adequately represented in your sample. However, if information is not available on which services are offered at each facility, it will not be possible to deliberately oversample facilities. In this case, assess the levels of coverage and non-response once the survey is complete. If either is at or above 1.3, it is recommended to present an unweighted analysis.

**District**

Ideally, all districts in a country would be included in the district DV/SA assessment. However, if resources do not permit a census of district offices, an alternative is to select district offices on the basis of the facilities sampled in the facility DV/SA. In this approach, district offices would be included in the district DV/SA only if a health facility within that district was selected for the facility DV/SA (i.e. a team would already be travelling to that district). With this approach it is important to note that the comparison will be restricted to districts that have been implicitly selected by the facility sample.

**Probability sampling using MS-Excel**

Once the sampling fractions for each stratum have been determined, the facilities from each stratum should be selected using a probability sampling method. An example on how to do sample selection is shown in MS-Excel; however, a user can choose any software package and use the appropriate software for the selection. The list frame should be partitioned according to the chosen stratification; and within each stratum (e.g. a list of hospitals in Region 1) the facilities to be included in the sample should be selected by simple random sampling or systematic sampling. Replacement facilities for those facilities that are closed or otherwise cannot be accessed can be selected using the same method. Alternatively, to facilitate logistics, the closest facility of the same type in the same geographical area can be selected.

First select the facilities to be included in the sample from the MFL. The MFL should be divided according to the categories selected to determine the sample. If the MFL is in a Microsoft Excel workbook, copy and paste each stratum of facilities into a new worksheet within the workbook. On each sheet, add a column called Random. Type “Random” into the first cell. In the column to the right of the column called Random, type the word “TRUE” in the first cell, as illustrated by the yellow fields in the Figure 1.
Use the following formula to assign a random unique number to each facility:

\[ \text{IF}(\text{SB}$1, \text{TRUNC}(\text{RAND()} \times (1000000-1)+1), \text{A2}) \].

Copy and paste the formula into the first cell of the column called Random. Place the cursor at the lower right corner of the cell with the formula and pull it downwards. If the columns named “Random” and “TRUE” are not the first two columns (A and B), please change the letter of the “Random” column to A and the letter of the “TRUE” column to B in the formula. A random number will be assigned to each of the facilities.

Then change the word TRUE to FALSE. This will freeze the random numbers so that they do not generate new random numbers (Figure 2).
A warning box may appear similar to the following:

Click on OK. Then filter the data so that the numbers in the Random column are in descending order, from the largest to the smallest (Figure 3).

**Figure 3. Filter the random numbers**

Determine how many facilities in the stratum should be selected on the basis of your sample size calculation. Highlight starting from the first facility in the list through to the total number of facilities needed for the sample in that stratum. These facilities will be included in the survey sample. Repeat for each of the strata identified above. Then select the next ten facilities in each worksheet as replacement facilities.

**Master facility list**

A comprehensive list of facilities with unique identifiers for facilities and with attribute data – i.e. information on region/district, facility type, managing authority, urban/rural designation – is also known as a Master Facility List (MFL). If an MFL exists for the country it can serve as the sampling frame.
A list frame that is complete, accurate and up to date, covering both public and private sectors, may often not exist. If that is the case, it will need to be constructed before a sample can be selected. Unless the country maintains a comprehensive MFL, authorities do not always have up-to-date records of health facilities functioning in the country. Coverage of private facilities is often incomplete and out-of-date; they may have closed or moved, and there is often no standard definition for facility type in the private sector.

If the MLF is not up to date it should be complemented with information other sources, such as private-sector coordinating bodies, social ministries where NGOs register their activities, or directly from faith-based, private and parastatal organizations. District Health Management Teams (DHMTs) are another good source for information on health facilities in the country. District Health Management Officers should be consulted on the accuracy of the MFL as it pertains to their respective districts, and revisions made as necessary. In situations where it is not possible to obtain a reliable sampling frame list of facilities, a dual-frame sampling methodology may be used. This method combines a simple random sample of hospitals and large facilities with a sample of geographically defined areas of the country.

Data requirements
The health facility assessment component of the DQR requires the following information from the sampled sites:

Health facility:
- validated monthly indicator values for three consecutive reporting periods (one quarter year) for selected indicators;
- reported values for the same indicators and periods from the same facilities;
- information on the completeness and availability of source documents and reports;
- causes of discrepancies between data recounted and data reported;
- causes of missing source documents and reports.

District level: From all health facilities in the district:
- indicator values for facilities in the district for the selected reporting periods;
- information on the availability, completeness and timeliness of reports from facilities in the district.

System assessment:
The DQR includes a qualitative survey conducted as an interview with the data manager or person in charge (i.e. whoever compiles the monthly report at the facility). This information helps
in identifying causes for weaknesses in the reporting system and interventions in order to help improve data quality. The system assessment includes questions on the following aspects of the reporting system:

- reporting practices;
- staff training,
- supervision;
- availability of data compilation and reporting guidelines;
- availability of data collection tools and reports;
- analysis and use of data.

**Data collection tools**

The DQR has standard data collection tools for the collection of data on either paper forms or electronically with CSPro. For CSPro, two methods of electronic data entry are available: 1) on a tablet in the field (with remote data transmission to a central server); and 2) on a PC once the data have been collected on paper forms. Separate data collection tools are available for the health facility and district levels for the DV/SA.

The paper-based data collection tools can be found here: [URL](#)
The CSPro data entry applications for facility and district levels can be accessed here: [URL](#)

**Health-facility level**

**Completing the DV/SA questionnaire**

The interviewer’s main responsibility is to use the questionnaire to collect information that is as accurate as possible by asking questions of the appropriate respondents and recording the responses accurately.

The instructions and examples below explain the questionnaire form, the types of questions and instructions and procedures for recording information correctly.

**Recording the responses**

When completing a paper version of the questionnaire, all responses are to be recorded using pens with blue ink. Blue ink is used because it can be distinguished from the black ink in which the questionnaires are printed. Red or green ink should never be used in recording responses since these colours are reserved for the survey manager and field supervisor in correcting the questionnaires in the office.
The information recorded in the fields of the questionnaire form will eventually be entered into an electronic database. At that point, it is very difficult to correct for errors or omissions in the questionnaires. Consequently, it is important that all answers are correctly recorded and that special instructions in the questionnaire are followed.

The procedures for recording responses will vary according to the type of question being asked. There are some basic questions in the questionnaire, such as pre-coded questions and those requiring a numeric response. Samples of all types of questions, and combinations of them, are reviewed below with examples.

NEVER LEAVE A RESPONSE BLANK! A blank is recorded as “missing information” because it is not known if the question was asked or not. If a response is negative, the negative response must be circled. Likewise, if a response is “don’t know”, the number corresponding to the “don’t know” response must be circled.

This questionnaire is typically divided into four columns, as shown in Figure 4. The first column contains the question number with each question numbered separately within each section. The second column contains the questions plus instructions to the interviewer asking the questions. The third column contains the response categories, and the fourth column contains “Skip” and other instructions if necessary.

Figure 4. Recording responses to the questionnaire
Instructions to the interviewer

It is important to ask the questions exactly as they are written in the questionnaire and in the order in which they appear. Questions are often accompanied by a set of instructions for the interviewer. Instructions are usually located in the question column and appear as CAPITAL LETTERS in bold type. Instructions will help you to remember important directions for asking questions, making correct observations, and recording information. The instructions should not be read to respondents.

Figure 5. Questionnaire instructions

“Skip” instructions

The questionnaire is designed to avoid as much redundancy as possible and to ask only questions that are appropriate to a given a situation. This is accomplished through the use of skip patterns. It is very important to follow the “skip” instructions as they will make the questionnaire more concise and relevant and thus increase the cooperation of respondents. In the sample question in Figure 6, if the answer to question DV_100 is “No” (they do not provide the antenatal care services), the providers in the health facility do not have to answer the following questions about antenatal care. The interviewer will skip questions DV101 to DV106 and go directly to question DV200. If the answer is “Yes” the interviewer will ask question DV_101.
Question types

Precoded questions

For some questions, we can predict the types of responses a respondent will give. The responses to precoded questions are listed in the questionnaire (Figure 7). To record a respondent’s answer, the number (code) that corresponds to the response should be circled. When numbers indicate coding categories, the interviewer records only one response for each question. The interviewer should make sure that each circle surrounds only a single number.

Figure 7. Recording answers to precoded questions
In some cases, a precoded question will include a category marked “other”. The “other” code should be circled when the answer provided is different from any of the precoded responses. The “other” response should then be specified and written in the space provided. If more room is needed, the margins can be used. The interviewer should also pay attention to any skip linked to a precoded answer.

Sometimes responses to questions must be entered in response grid/table (Figure 8). When recording a response in one of these grids, the interviewer has to be sure that the answer is entered in the proper row and column.

**Figure 8. Using a response grid**

<table>
<thead>
<tr>
<th>DV_609</th>
<th>Which of the following types of information is captured in the data visuals? PLEASE OBSERVE VISUALS FOR EACH ITEM BELOW.</th>
<th>OBSERVED</th>
<th>REPORTED NOT SEEN</th>
<th>NOT AVAILABLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>Maternal health care</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>02</td>
<td>Neonate and child health care (other than immunization)</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>03</td>
<td>Immunization</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>04</td>
<td>Top causes of morbidity and mortality</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>05</td>
<td>Other (specify)</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

**Numeric responses**

Several questions require a numeric response. These should be recorded in the appropriate boxes in the right column of the table (Figure 9).
If a respondent does not know the answer to a numeric question, the interviewer must circle the “Don’t know” response option. If “Don’t know” is not one of the responses, the interviewer must probe to obtain a numeric response to fill in the boxes. All boxes should have a number recorded in them. If a respondent’s answer requires fewer digits than are provided for in the response column, the interviewer must record zeros (0) in the left-hand box and the respondent’s answer in the right-hand box (for example: 0032).

---

**Figure 9. Recording numeric responses**

<table>
<thead>
<tr>
<th>DV_203</th>
<th>Please confirm the availability of the main source document used for reporting of DTP3/Penta3 for Month1 to Month3. If available and information on DTP3/Penta3 for children under 1 is recorded, please recount the number of DTP3/Penta3 visits for children under 1 for Month1 to Month3.</th>
<th>(A) SOURCE DOCUMENT AVAILABLE</th>
<th>(B) RECOUNT NUMBER OF DTP3/Penta3 FOR CHILDREN UNDER 1 IN SOURCE DOCUMENT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>YES, SOURCE DOCUMENT AVAILABLE WITH INFORMATIONRecorded FOR DTP3/Penta3*</td>
<td>NO, SOURCE DOCUMENT NOT AVAILABLE OR INFORMATION ON DTP3/Penta3 NOTRecorded</td>
<td></td>
</tr>
<tr>
<td>01</td>
<td>Month1</td>
<td>1 → 8</td>
<td>2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DV_403</th>
<th>Please confirm the availability of the main source document used for reporting of notified cases of TB for the quarter (Month1 to Month3). If available and information on notified cases of TB is recorded, please recount and record the number of notified cases of TB for the quarter (Month1 to Month3).</th>
<th>(A) SOURCE DOCUMENT AVAILABLE</th>
<th>(B) RECOUNT NUMBER OF NOTIFIED CASES OF TB IN SOURCE DOCUMENT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>YES, SOURCE DOCUMENT AVAILABLE WITH INFORMATIONRecorded FOR NOTIFIED CASES OF TB*</td>
<td>NO, SOURCE DOCUMENT NOT AVAILABLE OR INFORMATION ON NOTIFIED CASES OF TB NOTRecorded</td>
<td></td>
</tr>
<tr>
<td>01</td>
<td>Quarter (Month 1 to Month 3)</td>
<td>1 → 8</td>
<td>2</td>
</tr>
</tbody>
</table>

---

*Note: DV_203 and DV_403 are placeholders for specific data codes or variables.*
Ensuring quality

All members of the survey team are responsible for ensuring that the data collected at each facility are as accurate and comprehensive as possible. Each interviewer is responsible for:

- checking that filled questionnaires are complete at the end of each interview, ensuring that all answers are clear and reasonable, and that the handwriting is legible;
- returning to the original respondent(s) if questions are omitted or there appear to be errors, in order to complete the questionnaire (In this situation, the interviewer should apologize, explain that a mistake was made, and then ask the question again);
- notifying the field supervisor whenever there are problems in completing the daily assignment; and
- taking into account the field supervisor’s feedback and recommendations on the fieldwork/interview procedures based on the ongoing assessment work.

Correcting mistakes

If a mistake is made while recording an answer or the respondent changes his/her reply, draw two diagonal lines through the incorrect response (Figure 10). The interviewer should not try to erase an answer, use white-out or write over an answer. It is particularly difficult for data entry staff to understand which of two numbers is correct if the interviewer has tried to write over a response.

![Figure 10. Correcting a mistake](image)

Checklist

All questionnaires should be reviewed from beginning to end for the following:

- Verify that the interviewer has signed the verbal consent.
- Verify that all skip and filter instructions have been respected.
- Verify that the responses are legible.
Verify that only one response code is ticked for each question.
Verify that any corrections made by the interviewer are made legibly according to the instructions above.
Check that all questionnaires contain the correct number of pages.
Check that there are no missing responses.

**District-level**

The data collection teams will need to visit the district-level HMIS office in the districts with sampled health facilities. Data will be collected for accuracy, timeliness and completeness of reporting from the health facility to the district, and a system assessment specific to the district will be conducted.

The district data collection tools (both paper and electronic) function in a similar way to the health facility versions. The data verifications are specific to indicators and health programmes, and the system assessment is generic – unless reporting is programme-specific from health facilities (i.e. data from different health programmes at facilities are reported on separate, programme-specific forms).

This element will be verified on the data collection tool. Question DVD_010 asks:

- We need to know whether the reporting from facilities is “integrated” or **programme-specific**.
- If all data for the different health programmes are entered/aggregated by the same team at the district office (i.e. “Yes”), then the reporting is integrated and all questions can be asked of the same staff once. In this case, proceed to question DVD_011 on integrated system assessment.
- If the reporting is programme-specific (i.e. “No”), the different staff managing data for the different health programmes will all need to be interviewed. In this case, go to question DVD_100.

**System assessment**

The system assessment is expanded at the district level to include categories of questions not found on the health-facility system assessment (e.g. reporting of data, and data analysis and use).
Sections of the district-level system assessment include:

- Reporting of data. How does the district handle reports received from facilities, and when are the reports due?
- Availability of trained staff. Does the district office have adequate numbers of trained staff with specific responsibilities for data quality?
- Availability of guidelines. Does the district office have (observable) written guidance for data management, data quality control and the use of data for action?
- Availability of reporting forms. Does the district supply blank forms to the facilities, and do stock-outs of forms occur? Does the district provide feedback to reporting sites?
- Analysis and use of data:
  - Does the district prepare and maintain data visuals?
  - Does the district produce reports of analysed data (e.g., bulletins)?
  - Does the district carry out follow-up actions based on evidence?
  - Does the district use HMIS data for performance evaluations?
  - Does the district use HMIS data for planning?

Document review

The data verification tool is used to examine, at district level, the completeness, timeliness and accuracy of monthly or quarterly reports received from the health facilities.

- Note: Question DVD_X20: Does the district enter facility-level data into an electronic health information system (i.e., DHIS) and does the system automatically aggregate data to create a district report?

If data are entered into a database at district level, use the DV/SA district aggregation tool to recount the value of the indicator for the district for each programme area.

For countries that are implementing the district DV/SA and have an electronic health information system in which facility-level data are entered directly and the system automatically aggregates the data to create a district report, the district DV/SA tool does not implement recounting at the district level. Instead, it is recommended to compare the recounted values from the source documents at facility level to the HMIS recorded values, and compare the monthly report values at facility level to the HMIS recorded values. An Excel chartbook has been created to facilitate the analysis (See Chapter 4: Data analysis).

If the data are aggregated by hand at the district level and a district-level value is reported through the system, use the templates provided and enter the following information into the data collection tool (Figure 11):
Chapter 3. Implementation

For each programme area:

- Record how many reports there should have been for the three months.
- Count the actual number of reports for the same months and enter the value.
  - Check the dates on the reporting forms and ascertain how many reports were received on time (where “on time” means that a report was received by the due date).
  - Check how many reports were complete (where “complete” means that a report contains the reported count relevant to the indicator).
- The time period for verification is three months – e.g. January, February, March 2020.

Recounting

For each selected indicator, record on the data collection tool:

- whether the facility is part of the sample for the HFA;
- whether the facility is expected to report for the indicator;
- whether the report is available (observed) at the district level;
- whether the report was received at the district level by the due date (i.e. on time),
- the value on the facility monthly report at the district.

Aggregation and comparison

For each selected indicator, record on the data collection tool (Figure 12):

- the sum of the values from all facility monthly reports;
- the value reported by the district on the district monthly report (or in the database);
- reasons for discrepancies and missing reports.
Tool adaptation

Survey instruments should be adapted to the local health system. In particular, naming conventions for health facility types should be adapted, as should names and definitions of indicators, and source documents and reports. If the list of core indicators is modified, ensure that survey questions are appropriate for the indicator. For instance, while most service delivery output indicators are “cumulative”, some indicators are classified as “current”. A cumulative indicator is one for which monthly values are added to the previous month’s value in order to derive a running total (e.g. number counselled and tested for HIV). A current indicator is one where the current month’s value updates or replaces the previous month’s value (e.g. currently on ART, where “lost”, “stopped”, “transferred out” or “died” are all subtracted from the total, new patients are added, and those counted this month are most likely to have also been counted last month). Thus, a quarterly value for a cumulative indicator would be the aggregate of the three months constituting the quarter, while a quarterly value for a current indicator would be the value of the indicator for the last month in the quarter. Ensure that the data collection tools prompt for three values (one for each month of the quarter) in the case of cumulative indicators, and for one value in the case of a current indicator.

Typically, tool adaptation is informed via a workshop with programme managers and other health programme personnel (e.g. data managers, M&E officers, etc.). These personnel are knowledgeable regarding the intricacies of data collection and reporting for the different health programmes involved in the DQR, and will therefore be invaluable for providing specifics for the appropriate data collection methods.
local adaptation of the survey instrument (both paper and electronic versions). Ensure there is adequate representation of health programme personnel in the tool adaptation workshop.

Once the adaptations are identified and agreed on by all stakeholders, a subgroup should be tasked with updating and finalizing the survey instruments for implementation. Ensure that the revised tools are correctly labelled so there is no ambiguity as to which version of the tool is being used to collect data.

Editing the structure of the questionnaire

The DV/SA questionnaire is also available in electronic format along with automated tools for data processing and production of results. If these automated tools are to be used, editing of the structure of the questionnaire should be carried out as follows:

- **Adding a question**: Country-specific questions that are key to measuring tracer elements for service delivery can be added to the questionnaire. A practical and recommended way to number these specific country questions is to use the country ISO.2 code. For example: **SL_01**, where SL corresponds to the ISO.1 code for Sierra Leone plus numbering (sequential according to the number of questions added).

  - **Deleting a question**: It is possible that certain questions might not be relevant or applicable to a particular country. In this case, a question may be deleted. It should be removed from the questionnaire and the question number should also be deleted and should not be re-used. This procedure should remain occasional – the DQR aims to measure a minimum of tracer elements that are defined for data quality. Deleting too many questions will change the measurements’ parameters.

  - **Changing the text of a question**: Question text should not be replaced by other question text. If needed, clarification may be added in parentheses to help the respondent understand the question. It is very important to keep each question with its original numbering. Therefore, while you may add or delete questions, DO NOT change the content of existing questions.

  - **Skip patterns**: Any addition or deletion affecting a skip pattern in the questionnaire should be updated accordingly.

Important tips

- **Do not change numbering**: the original numbering structure of the standard questionnaire should be maintained. Changing the numbering will affect the links to the existing tools for automated data processing and results production.

- The goal of the DQR is to measure, on the basis of key tracer items, the data quality and system capabilities in health facilities. It is important not to stray from the DQR concept by adding a long list of additional items.

---

It is also important to remember that adding more to the tool will have an impact on the training, the data collection and the data analysis. Any additional question should also be considered in terms of the outputs of the analysis. Before a question is added, it should first be added to the analysis plan so that it is clear how it will be used in the analysis.

Training of data collectors and supervisors

The DQR requires high-quality training for the personnel who will take part in implementing the survey. Data collectors will recount indicator values at health facilities for up to five programme areas. Each programme area has a separate set of source documents (e.g., tally sheets, registers, etc.) and different methods for aggregating data to derive indicator values. The exercise is complicated and requires great attention to detail. The training of data collectors should include ample time for practising indicator compilation on example forms.

The training venue should be large enough to accommodate all data collectors and supervisors, with sufficient space to spread out examples of data collection tools on tables. The venue should be reserved well in advance to ensure the availability of adequate space for the training.

Experienced facilitators should be recruited to conduct the training. Facilitators should have sufficient experience in programme M&E and health-facility assessment. Enough facilitators should be engaged so that, as far as possible, they can work individually with participants. A good rule of thumb is that there should be at least one facilitator for every 10 participants.

A training plan should be developed and budgeted as part of the overall DQR planning process. All personnel should be identified, recruited and notified well before the start of the DQR. The training itself should be conducted just prior to the data collection teams going to the field so the training content is still fresh in participants’ minds.

Standardized training curriculum

WHO has developed a standard training curriculum for the DV/SA. The curriculum includes guides for facilitators and participants, presentations, and exercises for a training event for up to six days for different staff levels.

Goals and objectives of training

The training curriculum is designed to help participants to develop the knowledge and skills required to assess the accuracy, timeliness and completeness of data, and the extent to which the data management system is ready to produce quality data on health-service delivery. The training course focuses on data quality for routine health information systems (RHIS) – i.e. the HMIS of the Ministry of Health, and/or health or disease programme information systems.

The training includes up to five days of instruction and practice, though the length may be modified according to the specific needs of local implementers. The whole training course may
not be suitable for every participant and the content of the training should be modified within the country to suit the needs of the assessment. For instance, data collectors do not need to know about data compilation and cleaning, or data analysis, and data managers do not need to know how to collect the data. Relevant elements of the training should be extracted to meet the needs of the participants.

**Target audience**

The course is designed for leaders, programme managers, data managers and data collectors from the national HMIS, health and disease programmes and donor-funded projects. Table 8 shows the intended target audiences of each session in the module.

Table 9 shows suggested content, duration and emphases for DV/SA training for different target audiences. For national-level stakeholders responsible for planning and implementation of the DQR, an overview training is all that is needed, perhaps lasting 1–2 days. For programme managers and data managers, who will carry out and oversee essential aspects of

**Table 8. Intended audiences for each training session**

<table>
<thead>
<tr>
<th>Training session</th>
<th>Intended target audience</th>
</tr>
</thead>
<tbody>
<tr>
<td>Session 1: Overview of Data Quality Review (DQR) (content to be adapted to the type of trainee)</td>
<td>Programme managers, Data managers, Supervisors, Data collectors, DQR coordinating group</td>
</tr>
<tr>
<td>Session 2: Implementing the DQR (content to be adapted to the type of trainee)</td>
<td>DQR coordinating group, Programme managers, Supervisors, Data collectors</td>
</tr>
<tr>
<td>Session 3: DV/SA data collection tools</td>
<td>Supervisors, Data collectors</td>
</tr>
<tr>
<td>Sessions 4–8: Data verification</td>
<td>Supervisors, Data collectors</td>
</tr>
<tr>
<td>Session 9: DV/SA at the district level</td>
<td>Supervisors, Data collectors</td>
</tr>
<tr>
<td>Session 10: Tablet-based data entry with CSPro</td>
<td>Supervisors, Data collectors</td>
</tr>
<tr>
<td>Session 11: Roles and responsibilities, and quality assurance</td>
<td>Supervisors, Data collectors</td>
</tr>
<tr>
<td>Session 12: DV/SA – Field practice</td>
<td>Supervisors, Data collectors</td>
</tr>
<tr>
<td>Session 13: Data compilation, cleaning and quality</td>
<td>Programme managers, Data managers</td>
</tr>
<tr>
<td>Session 14: Data analysis</td>
<td>Programme managers, Data managers</td>
</tr>
<tr>
<td>Session 15: Interpretation and dissemination of results, and action planning to improve data quality</td>
<td>DQR coordinating group, Programme managers</td>
</tr>
</tbody>
</table>
Table 9. Suggested training lengths and emphases for different target audiences

<table>
<thead>
<tr>
<th>Target audience</th>
<th>Suggested content</th>
<th>Suggested length of training</th>
<th>Suggested emphasis</th>
</tr>
</thead>
<tbody>
<tr>
<td>DQR coordinating group</td>
<td>DQR overview and implementation, Quality assurance, Dissemination of results, Action planning to improve data quality</td>
<td>1–2 days</td>
<td>Planning and coordination, Monitoring and troubleshooting implementation, Using results for action</td>
</tr>
<tr>
<td>Programme managers</td>
<td>DQR overview and implementation, Quality assurance Data analysis, Interpretation and dissemination of results, Action planning to improve data quality</td>
<td>2–3 days (depending on data management skills and familiarity with tools)</td>
<td>Quality assurance, Accurate data compilation and cleaning, Analysis and presentation of data, Standardized tools</td>
</tr>
<tr>
<td>Data managers</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supervisors</td>
<td>DQR overview and implementation, Data verification and system assessment (facility and district levels), Practice in a health facility</td>
<td>5 days</td>
<td>Appropriate survey implementation, Accurate data verification, Quality assurance</td>
</tr>
<tr>
<td>Data collectors</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 10. Illustrative agenda for 5-day data collector training

<table>
<thead>
<tr>
<th>Time</th>
<th>Day 1</th>
<th>Day 2</th>
<th>Day 3</th>
<th>Day 4</th>
<th>Day 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>8:30 – 9:00</td>
<td>Welcome, workshop objectives and expected</td>
<td>Data validation at health-facility level –</td>
<td>Data validation – TB cases notified</td>
<td>Tablet-based data entry with CSPro</td>
<td>Field practice</td>
</tr>
<tr>
<td></td>
<td>outputs</td>
<td>ANC 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Review of source documents /reports</td>
<td>• Review of data collection tools/reports</td>
<td>• Demonstration</td>
<td>• Teams visit health facilities to practise</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Practical exercise</td>
<td>• Practical exercise</td>
<td></td>
<td>data collection for data verification and</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>system assessment</td>
</tr>
<tr>
<td>9:00 – 10:30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10:30 – 10:45</td>
<td>Break</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10:45 – 13:00</td>
<td>DV/SA Data collection tools</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• HF Data verification</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• HF System assessment (paper &amp; electronic)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13:00 – 14:00</td>
<td>Lunch</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14:00 – 16:00</td>
<td></td>
<td>Data validation – DTP3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Review of source documents /reports</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Practical exercise</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16:00 – 16:15</td>
<td>Break</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16:15 – 18:00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
implementation, more in-depth training is required, with the length depending on the previous experience of the participants with the established tools and methods. For data collectors, the emphasis is on accurate data collection, so much of their time should be spent practising recounting indicator values at facility level and filling the data collection forms (5 days).

An sample agenda for a 5-day training course for data collectors is shown in Table 10. The standardized training curriculum can be accessed here: URL

Roles and responsibilities

DQR coordinating group
The DQR coordinating group is responsible for determining the scope and content for the survey, and establishing the timeline and budget. The group will identify and/or recruit by contract a survey implementing agency and will engage consultants for technical assistance if required. The group will monitor planning and implementation of the survey and take corrective measures if necessary. Members of the group will organize meetings of stakeholders to help discuss and determine the content, and will preside over meetings for adaptation of tools, validation of data and dissemination of results. They will help ensure that the results of the DQR are used for planning purposes and that any weaknesses in data quality found during the assessment are addressed in action plans and corrected by appropriate interventions.

Implementing agency
During the planning phase the agency identified to implement the DQR will assist the DQR coordinating group to plan and budget for the survey, and recruit and train data collectors and supervisors. During implementation the agency will monitor the progress of data collection teams, resolve problems as they arise, and check the quality and completeness of the data as it comes in. After data collection is completed the agency will ensure data entry for paper-based surveys, and data compilation and cleaning for the electronic data. The agency will participate in data analysis, validation, reporting writing and dissemination of results.

Supervisors
Field supervisors have a crucial role to play in ensuring data quality and consistency. Field supervisors will oversee all aspects of data collection in the survey area(s) for which they are responsible.

These responsibilities include:

- organizing data collection visits to facilities (making initial contact, preparing a schedule of data collection visits etc.);
- ensuring the availability of paper forms and/or functionality of electronic data collection tools;
- supervising data collection activities, namely:
  - making sure that data collection protocols are followed,
  - ensuring regular communication with data collection teams,
• checking data collection forms at the end of each day for completeness and legibility,
• making sure that electronic data are transferred to the national level via secure
  electronic transmission as often as possible (following an established survey protocol);
• validating data collection by re-conducting the survey at a small percentage of facilities
  (e.g. 10%) and comparing results to those of the data collectors;
• collecting and storing data collection forms and sending them to the survey manager;
• transferring electronic data from electronic data collection devices to the survey area
  computer/laptop (if applicable).

Conducting field activities

Preparing for data collection

• Schedule survey visits and identify replacement facilities.
  • The survey manager will provide a list of health facilities and replacement facilities for
    your survey area.
    – Contact (in person or by telephone) each health facility and replacement facility to
      introduce the survey and seek permission for data collection:
      ‣ Introduce the survey and its objectives.
      ‣ Use the letter of endorsement and introduction provided by the survey
        manager.
      ‣ Stress that individual facilities will not be identified in the results
    – Make an appointment for data collection at a date and time which is convenient for
      the facility, avoiding peak hours:
      ‣ Plan for approximately 3–4 hours for each data collection visit, plus travel time.
      ‣ More time should be allotted for large facilities/hospitals (1 day per hospital).
    – Note the name and telephone number of the contact person at each health facility.
    – Explain about the possibility of a second visit for “validation”, which may take place
      in 10% of the surveyed health facilities.
    – If a facility refuses to participate, alert the survey manager who will contact the
      health facility directly and, if necessary, will provide you with an alternative site.
      Luckily, this rarely happens.

• Prepare a schedule of data collection visits for each pair of data collectors.
  • Prepare a schedule for each pair of data collectors, including:
    ‣ the date and time of each visit;
    ‣ the name, number, sector and contact person of each health facility;
    ‣ the address and location of each health facility; and
    ‣ contact information for a replacement facility to be visited if necessary.
Example:

<table>
<thead>
<tr>
<th>Date and time of appointment</th>
<th>Name of facility</th>
<th>Contact person</th>
<th>Location</th>
<th>Managing authority</th>
<th>ID Number</th>
<th>Replacement facility name and contact details</th>
</tr>
</thead>
<tbody>
<tr>
<td>April 20 10:00</td>
<td>ABC Health centre</td>
<td>Mrs Nguyen</td>
<td>45 Main Street, Eastern City Tel: +22 414 00</td>
<td>Government</td>
<td>01234</td>
<td>XYZ Health Centre, 59 Main Street, Eastern City</td>
</tr>
</tbody>
</table>

- Call each facility and confirm an appointment on the day before the data collection.

Arrange for regular communications and transport. Once all of the survey sites are known, transportation should be arranged according to the number of sites to be visited, the number of teams going into the field and the number of persons per team.

**Preparing the necessary materials for data collection**

- Prepare a data collection form for each facility to be visited. An example is shown in Figure 13.
  - The survey manager will provide you with a separate data collection form for:
    - each sample health facility in your survey area;
    - each replacement facility; and
    - each validation visit.

Make sure that there are enough forms for all facilities listed in the survey area prior to starting the field data collection.

- Complete the front page of the DV/SA questionnaire data collection form with the identifying information of each sample facility.
  - Complete the following fields in the cover page of the form:
    - name of health facility
    - health facility unique ID
    - name of town/village
    - region and district
    - type of facility
    - managing authority.

**Do not** complete these fields, as these will be completed by data collectors during facility visits:
- date
- name of person(s) who provided information
- name of data collectors.

- For the material checklist, the supervisor should make sure that the following material is available each day so that the field data collectors to conduct the survey correctly.
  - Checklist of materials for data collectors:
    - contact details of the field supervisor, including a mobile telephone number to call in case of difficulty in the field:
    - data collectors’ guides and relevant handouts;
Figure 13. Data collection form

<table>
<thead>
<tr>
<th>Number</th>
<th>Question</th>
<th>Result</th>
<th>Skip</th>
</tr>
</thead>
<tbody>
<tr>
<td>DV_001</td>
<td>Facility number</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DV_002</td>
<td>Is this a supervisor validation check of a facility?</td>
<td>DATA COLLECTION FOR FACILITY ASSESSMENT .... 1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SUPERVISOR VALIDATION</td>
<td></td>
</tr>
<tr>
<td>Date</td>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interviewer Name</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DV_003</td>
<td>Name of facility</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DV_004</td>
<td>Location of facility</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(Town/City/Village)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DV_005</td>
<td>Region/Province (name and code)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DV_006</td>
<td>District (name and code)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DV_007</td>
<td>Type of facility</td>
<td>NATIONAL REFERRAL HOSPITAL ... 1</td>
<td>96</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DISTRICT/PROVINCIAL HOSPITAL ... 2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>HEALTH CENTRE/CLINIC ... 3</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>HEALTH POST ... 4</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>MATERNAL/CHILD HEALTH CLINIC ... 5</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>OTHER (SPECIFY)</td>
<td></td>
</tr>
<tr>
<td>DV_008</td>
<td>Managing authority</td>
<td>GOVERNMENT/PUBLIC ... 1</td>
<td>96</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NGO/NOT-FOR-PROFIT ... 2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>PRIVATE-FOR-PROFIT ... 3</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>MISSION/FAITH-BASED ... 4</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>OTHER (SPECIFY)</td>
<td></td>
</tr>
<tr>
<td>DV_009</td>
<td>Urban/rural</td>
<td>URBAN ... 1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RURAL ... 2</td>
<td></td>
</tr>
</tbody>
</table>
– a schedule of visits to survey sites and contact details of the sites;
– a list of data collection teams and contact information when in the field;
– copies of letter(s) of endorsement and letter of introduction;
– a data collection form for each facility that may need to be visited that day;
– extra copies of the DV/SA data collection form;
– electronic data collection devices (fully charged with CSPro application installed and loaded with the DV/SA questionnaire) batteries and power cables;
– memory cards or USB keys for data backup;
– GPS units (fully charged and accurately configured, if relevant);
– pens (pencils should not be used to record data), a clipboard and other supplies;
– a notebook to record any significant events or findings;
– field allowance for local expenses;
– identity document with a photograph;
– a mobile telephone with credit for each team.

• Checklist of materials needed by supervisors for daily meeting with the data collectors:
  – detailed planning of site visits for each data collection team;
  – electronic data collection software installed on the laptop (CSPro 6.0) (see instructions in Section 4);
  – a fully charged laptop computer with appropriate software for copying data from the electronic data collection device to the laptop computer (using memory cards or USB keys).

**Supervising data collection activities**

Supervisors’ responsibilities during data collection are to supervise data collectors and to make sure that data collection forms are complete and accurate. Supervisors should go into the field regularly with data collectors to make sure that the survey protocols are being followed. Problems relating to the data collection process should be identified and resolved promptly. Problems that cannot be resolved should be reported to the survey manager as soon as possible.

The supervisor is responsible for the accuracy of the data collected by data collectors.

▷ Make sure data collection protocols are followed:
  • Ensure that data collection teams are conducting interviews at the facility.
  • Keep track of facilities that have been covered from the sample.

▷ Arrange for regular communication with data collection teams:
  • Provide data collectors with a mobile telephone and a phone number where they can contact the supervisor during data collection.
  • Arrange to meet with data collectors at the end of each day of data collection.
  • Ensure that data are transferred to the computer at the end of each day.

▷ Tracking facilities
The field supervisors should keep a running tally of facilities that have been assessed from the list of facilities in the sample assigned to them. Use a table such as the one in Figure 14.

- Include information on which facilities were selected for supervisor validations.
- Any problems encountered with the data should also be documented in the tally list.
- The table should be submitted to the survey manager with the electronic data files at the end of the fieldwork.

**Daily meeting with data collectors**

- At the end of each day, supervisors have the responsibility to:
  - meet with data collectors to collect completed forms and resolve any problems encountered;
  - review each data collection form completed that day to make sure they are complete and legible, and verify missing or suspicious information;
  - check uncertain or illegible data with data collectors and re-contact the facility to clarify the issue if necessary;
  - sign the last page of each questionnaire to record that it has been checked – but only after you are sure that data are complete and legible, and there are no obvious mistakes;
  - use CSPro tools to check for completeness of the electronic data collection forms.

**Storing completed data collection forms**

- Completed paper forms should be stored in waterproof plastic bags at the field location.

---

**Figure 14. Maintaining a list of facilities that have been assessed**

<table>
<thead>
<tr>
<th>Facility Code</th>
<th>District</th>
<th>County</th>
<th>Sub-County</th>
<th>Parish</th>
<th>Health Unit</th>
<th>Owner</th>
<th>Authority</th>
<th>Level</th>
<th>Status</th>
<th>PDA</th>
<th>Validation Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>12345678</td>
<td>Central Region</td>
<td>Mzimba</td>
<td>Mzimba District</td>
<td>Mzimba</td>
<td>Mzimba</td>
<td>Mzimba</td>
<td>Mzimba</td>
<td>Mzimba</td>
<td>Mzimba</td>
<td>Mzimba</td>
<td>Mzimba</td>
</tr>
</tbody>
</table>

**GREEN** = facility has been assessed, data collectors have entered data into CSPro, data have been checked.

**RED** = facility could not be assessed.

**BLUE** = replacement facility.

**WHITE** = not yet covered.
until fieldwork is completed, at which time they should be sent to the survey manager.

- All original data collection forms, including those for validation visits (labelled clearly!), should be transferred to the survey manager upon completion of fieldwork.

- Field supervisors should retain copies for use in the event that the originals become lost or damaged.

Transferring electronic data collected (using USB flash drive)

- Electronic data should be backed up on a memory card/USB key and transferred to the computer of the field supervisor at the end of each day to prevent data loss.

- Transferring final data files from laptops/netbooks/etc. with a USB flash drive:
  
  A USB flash drive should be inserted into the laptop/netbook computer that was used for data collection. The file explorer can be used to navigate to where the DV/SA data files are stored and then to the folder FACILITY_DV_SA\Data (or the folder DISTRICT_DV_SA\Data as the case may be). This location was selected at the beginning of data collection.

  The Facility_DV_SA_data.dta file should be copied from the laptop/netbook computer to the USB flash drive. The USB flash drive can now be ejected from the laptop/notebook.

  The USB flash drive can be inserted into the desktop computer or laptop computer that will be used for data processing. The DV/SA data files can now be copied and pasted to a folder on the desktop/laptop.

  This should be repeated for all laptop/netbook computers used for data collection.

- Organizing data files from the field collection

  The back-up procedure (steps as described above) should be done every time supervisors meet their teams. It is important to save the data files in an organized manner to make sure that the latest files contain data for all surveyed facilities by each team. A saving procedure as below could be used to save in an organized manner data in a specific folder for each team:

  FACILITY_DV_SA_Team1_Date1
  FACILITY_DV_SA_Team1_Date2          FACILITY_DV_SA_team1
  FACILITY_DV_SA_Team1_Date3

  The latest file for each team should correspond to the final file. This will be validated by the supervisor. A copy of the final file should be created and renamed:

  FACILITY_DV_SA_Team1_FINAL

  The final data set will contain the final files from each team:

  FACILITY_DV_SA_Team1_FINAL
  FACILITY_DV_SA_Team2_FINAL          FACILITY_DV_SA_FINAL (REGION X)
  FACILITY_DV_SA_Team3_FINAL
This final data set should be sent to/shared with the specified data manager at central level in charge of the compilation of the data from field collection. A back-up of all data files (final and stamped with dates) should be carefully preserved as a back-up and should remain accessible during the cleaning and data processing phase.

### Validation of data collection

- The supervisor will make a validation visit to 10% of health facilities. This means that the supervisor will return to some of the sample facilities (10% randomly selected within the list) and collect data again in order to make sure that the data obtained by the data collectors is accurate and reliable. To do this, supervisors will:
  - select facilities for validation at random (randomly select one public facility and one private facility);
  - conduct the validation visits on the same day as the visits to these facilities by the data collectors (or as soon as possible afterwards);
  - compare the data obtained with that collected by the data collectors;
  - identify and resolve any issues/mistakes and discuss them with the data collectors;
  - enter the data collected for validation electronically in CSPro. The consistency of responses (exact matching) will be analysed for quality control.

### Data collectors

#### Objectives of the facility reporting data verification tool

The DV/SA tool guides a record review in the health facilities being surveyed. The goal of the data verification module is to provide information on the quality of monthly reported data from health facilities to a senior level by comparing discrepancies between data in the primary source and in the monthly report). This information is integrated into the DQR.

#### Data collection instruments

A standard core questionnaire has been developed for the DV/SA. This questionnaire is typically adopted in its entirety and then adapted in the country with changes to certain elements – such as names, types of facilities, indicators etc. The questionnaire is usually tested during an in-country pilot visit for final adjustments and validation.

This tool is often used in conjunction with other survey modules such as the DV/SA. The collection instruments are paper-based or used in conjunction with electronic collection devices (tablets, laptops etc.).

The principal responsibility of the data collector, or interviewer, is the appropriate use of the questionnaire to collect information that is as accurate as possible by asking questions of the appropriate respondents and accurately recording their responses.
The health facility assessment will be completed in teams. Typically, each team will include two persons responsible for data collection who work closely with a field supervisor.

**Role of the interviewer**

The data collectors’ main tasks include:

- to visit health facilities and collect **information**;
- to verify geographic coordinates (if relevant);
- to complete a DQR data collection paper form and/or an electronic form, as well as the facility **reporting**;
- to validate indicator values for the facility by re-aggregating the service delivery results for the selected period and making comparisons with the reported values (data verification);
- to back up electronic data on a memory card/USB **key**;
- to report back to the field supervisor at the end of each day.

For accuracy in data collection and validity in the assessment findings, it is imperative that data collectors **correctly** re-count service delivery results for the selected indicators and period(s). The level of experience and knowledge required of data collectors is substantial; they should have working knowledge of data collection tools for up to five programme areas and should know the protocols for monthly compilation of the five indicators. Attention should be paid to the quality of training for data collectors and the level of experience of staff selected to this job. Training should include ample practice with sample data collection tools to build capacity for this critical task.

**Survey regulations**

The following survey regulations have been established to ensure the success of the DV/SA tool and must be closely adhered to by all interviewers.

- The interviewer’s attendance is required during each day of the fieldwork. Any person who is late or absent during any part of the fieldwork (whether it is a whole day or part of a day) without prior approval may be dismissed from the survey.

- Throughout the fieldwork period, interviewers are representing the implementing agency. As an interviewer, your conduct must be **professional** and your behaviour must be congenial when dealing with the public. You must always be aware that we are able to do our work only with the goodwill and cooperation of the people we interview. Therefore, any team member who is consistently overly aggressive, abrupt or disrespectful to others may be dismissed from the survey team.

- For the survey to succeed, the members of each team must work closely together. Any team member who, in the judgment of the survey manager, is a disruptive influence on the team may be asked to transfer to another team or be dismissed.
The data gathered during the fieldwork must be both consistent and accurate. Field staff may be dismissed at any time during the fieldwork if the quality of their work is inadequate.

Vehicles and gasoline are provided for the survey for official use only. Any person using a vehicle for an unauthorized personal reason will be dismissed.

Data are confidential. Under no circumstances should confidential information be passed to third parties. Persons breaking these rules, and therefore the confidence placed in them by respondents, will be dismissed.

Planning the DV/SA fieldwork

The following describes the activities concerning data collectors in the planning the DV/SA fieldwork.

Fieldwork schedule

The field supervisor will assign to each team a list of facilities to be visited for data collection. The list will include the name and location of each facility as well as the facility identification information required in the DV/SA questionnaire.

If the information is available, the list may include the name of the person in-charge at the facility, telephone numbers or other information on how to contact the facility, and the hours during which the facility is open and/or various services offered. The field supervisor will also provide the team with a map showing the location (or approximate location) of all of the facilities on the team’s list.

The team generally will need to arrive at a facility on or before the official opening hours. Consequently, the lodgings that the team will use each night must be within a reasonable distance of the facility that is to be visited on the next day.

Advance contact with authorities/facilities

Generally, the survey manager or another senior member of the survey team will have notified the appropriate authorities of the nature and purpose of the health facility assessment in advance of the fieldwork. An official letter from the managing authority of the facilities being surveyed should have been sent to the regional or district offices for that organization. Each team should also have a copy of the letter to show at facilities if necessary.

Logistical arrangements

Prior to departure for fieldwork, the field supervisor must ensure that the team has the questionnaires and other materials necessary to complete the assignment.
Each morning before leaving for field visits, the team should check that they have all necessary materials.

Checklist of materials for data collectors:

- a list of data collection teams and contact information:
  - contact details of the field supervisor, including a mobile telephone number to call (with sufficient credit) in case of difficulty in the field;
- a schedule of visits to survey sites:
  - the contact details of the sites to be visited;
  - details of backup facilities to be visited if scheduled visits are not possible;
  - copies of the letter of introduction;
  - the DV/SA Implementation Guide (guide for data collectors);
  - a DV/SA data collection form for each health facility to be visited that day (with the cover page filled by the field supervisor);
  - a DV/SA data collection form for each backup site that may need to be visited that day;
  - an EDC/tablet/laptop (fully charged with CSPro installed and loaded with the DV/SA questionnaire) a charger and battery;
  - a memory card/USB stick for data backup;
  - a fully charged and accurately configured GPS unit (if relevant);
  - a fully charged cell phone with credit;
  - pens (pencils should not be used to record data), a clipboard and other supplies;
  - a notebook to record any significant events or findings;
  - a field allowance for local expenses;
  - an identity document with a photograph for each data collector.

Activities during a facility visit

There are many general procedures to be followed by a survey team during a visit to a survey facility. These procedures are outlined in the following sections.

Locating and verifying the survey facility

The field supervisor has provided you with a list of the facilities for which you are responsible. Every attempt should be made to conduct the survey at each facility on your list. If, after contacting local authorities, you cannot locate a health facility on your list, or are not sure whether a facility that you have found is actually the one identified on the facility list, contact your field supervisor. If a facility included in the assignment has closed, no interview will be necessary, just note that fact on the cover sheet of the assigned questionnaire. Finally, no facility that is not listed in the sample should be visited and interviewed unless specifically approved by the survey manager.

Validating the cover page of the questionnaire

Before starting the collection of data, the data collectors should check that the information on the first page of the data collection form (cover page) is complete and correct (pre-filled by the
field supervisor). If there are any mistakes, inform the field supervisor at the end of the day. If applicable, enter this information on the electronic form on the electronic data collection device/tablet or laptop.

The following information should be completed by the data collectors on the first page of the data collection form prior to starting the survey:

- date;
- names of the data collectors (interviewer);
- number/code of the interviewers' team (given by the field supervisor prior to the fieldwork)

**Global positioning system data collection**

On arrival at the health facility that is to be surveyed, the data collectors should fill out the geographical coordinates section of the questionnaire included in the cover page. The global positioning system (GPS), which is a satellite system, will be used to locate precisely the geographical position of the site. Step-by-step instructions on using a GPS device are available on page 64.

**Gaining permission to survey the health facility**

Data collection teams will be visiting health facilities operated by the government, by NGOs and perhaps also private facilities. All facilities must give permission for the survey to be conducted on their premises. On the day of the survey, data collectors may provide reassurance to facility staff that only the final results will be provided so that no individual respondent can be identified.

The first contact at the site should be made by asking to speak with the person in charge. If the official person in charge is not present on the day of the survey, the data collection team should ask to see the person temporarily in charge on that day. The data collectors will then:

- introduce themselves;
- explain the purpose of the visit and the activities that are part of the survey;
- give the person in charge the introductory letters from the relevant organization and the letters explaining the survey and giving the authorization to visit the facility;
- when consent is obtained, the person in charge (or substitute) sign the informed consent section on the cover page of the questionnaire to indicate that the consent of the person in charge has been obtained.

If you are refused an interview in the facility and nothing you say can make the person in charge reconsider, contact the field supervisor and provide the name of the facility, its managing authority, and location. The field supervisor will make every attempt to contact appropriate persons who can help to convince the health facility staff to allow the interview.
Meeting with the person in charge of the facility

An important objective of the survey is to obtain correct and consistent answers to the questions. As the questions relate to a facility and not to a specific person, the information can be obtained from a variety of respondents so long as they are knowledgeable about the topic. During the interview, interviewers may need to speak with various respondents in order to obtain complete and correct information.

The data collectors are responsible for working out a plan for completing all components of the questionnaire at each facility. They should discuss the plan with the person in charge. It may be helpful to meet with relevant supervisors (at large facilities) and other staff who may be requested to allow interviews and observations during the team’s visit. For a small facility this may be relatively easy but for larger facilities it may involve different departments.

Duration of the survey

The duration of the interview will depend on the size of the facility and the availability of suitable staff to provide the answers to the questions, but generally should be 3–5 hours for a small health post or health centre and up to a day for a hospital (depending on the number of indicators and volume of service delivery).

Interviewer skills

Skill and practice are required to collect data that accurately reflect the health services available at a facility, whether it is a small health post or an urban hospital. This section provides general instructions on the skills required for gathering this information. Survey findings are only as good as the data from which they are calculated. The quality of that data depends to a large extent on the interviewer.

Below are some basic instructions on the practices that should be used when interviewing respondents, together with tips on dealing with difficult situations that you might encounter during an interview.

General interviewing practices and techniques

In order to obtain accurate information from health providers at work in a health-service setting, it is very important that they become engaged in the data collection process. There are several basic ways to gain health providers’ cooperation while collecting accurate and specific information.

Show respect for the respondent

The interaction between the interviewer and all respondents is very important. All respondents should be treated respectfully and politely. The respondents should understand that their cooperation and the time they are taking to help make the survey successful is very much appreciated.
A respondent’s first impression of the interviewer will strongly influence his/her willingness to participate fully in the interview. It is important that the interviewer approaches each person to be interviewed and his/her colleagues at work in a friendly, respectful and professional manner.

One basic way to show respect for health workers at work is to be considerate of what they need to accomplish during their workday and to let them know that there will be no interference with their client-related tasks. Two ways to accomplish this are: 1) if the health worker to interview is busy with a client, the interviewer should wait until that visit is completed before approaching him/her; and 2) the interviewer should wait until there are no clients around or until there is a qualified person available to complete the questionnaire. The interviewer will discover other ways to fit smoothly into the health workers’ busy schedules while gaining more experience gathering data in a variety of health service settings during the survey.

If it appears that there will never be a convenient time for collecting the data, the interviewer should discuss with the health worker or the person in charge the best approach for collecting the required data with the least interference possible.

**Listen carefully to the respondent**

Listening carefully to what the respondents say is as important as asking the questions on the questionnaire, and demonstrates respect. Some questions in the questionnaire require the interviewer to listen to what the respondent says and record it by simply circling a category on the printed form. Sometimes, however, the interviewer must write down exactly the answer given by the respondent if it does not fit into any of the listed categories. In either case, the interviewer should listen well. He/she should not rush into circling the code category before he/she has really listened to the respondent. After all, writing on the form while the respondent is talking may be taken as a sign of disrespect or not paying attention. More importantly, people who rush into coding a response are often in danger of attributing their own biases, preferences and favorite response categories to their respondents.

**Request consent from the respondent prior to asking questions**

There is a consent form and some background information that should be read to the respondent prior to starting the interview. The consent form is located at the beginning of each questionnaire. The interviewer is required to read the information in its entirety to the respondent and then request his/her consent before starting to ask questions. Without the respondent’s consent the interview cannot proceed beyond the cover section.

Answer respondents’ questions without pressuring them

Some respondents may question the interviewer about the purpose of the survey before agreeing to participate. In this situation, the interviewer should answer the respondent’s questions as directly as possible. If respondents feel that the information is important and that the interviewer is sympathetic to their situation, they will be more straightforward in their responses and will be more likely to answer questions to the best of their ability. If they feel pressured to respond, or feel that the interview is a burden, they may not think carefully about their responses.
Offer no opinions or advice on specific health-facility practices or patient care issues
If a respondent has specific questions that require the interviewer’s medical opinion or advice, he/she should politely respond by saying that he/she is has come to collect information to provide an overview of the services, and that he/she is interested in the systems and practices at the facility. Explaining this and then simply stating, “I am not in a position to provide any advice or opinions” may be sufficient. It is important to remember that the job is not to educate respondents but only to collect information from them.

Read every question exactly as written and in sequence
The wording of each question has been carefully chosen and for that reason it is essential that the interviewer reads each question to the respondent exactly as it is written. It is very important for this survey that each question is asked to each respondent in exactly the same manner. Each section of the questionnaire also has an introductory paragraph that must be read to the respondent (when applicable) in its entirety.

The interviewer should speak slowly and clearly so that the people who are interviewed will have no difficulty in hearing or understanding the question. At times, the interviewer may need to repeat a question in order to be sure the respondent understands it. In these cases, the interviewer should not paraphrase the question but repeat it exactly as it is written. If, after the question has been repeated the respondent still does not understand it, the question may have to be restated but the interviewer should be very careful when changing the wording not to alter the meaning of the original question. During the practice sessions conducted at a facility not included in the survey sample, if interviewers find that they have to repeat the same question to several respondents, a note of this should be made and reported to the field supervisor so that, if necessary, the wording can be changed on the questionnaire.

Be straightforward
There are some questions in the survey that require the interviewer to ask about the availability of certain items, and then ask to see them. Providers will be more cooperative if they know beforehand what to expect. If the interviewer asks questions and then later asks to see items, people may think you are trying to trick them or are “checking up” on their answer. In order to have the greatest amount of cooperation, the interviewer should always tell the respondent what is coming. For example:

“I am interested in knowing if the following basic equipment and supplies used in the provision of client services are available in the general outpatient area of this facility. For each item, please tell me if it is available today and functioning. I will need to see the item so that I can completely fill in this questionnaire.”

Probe for a response
Occasionally, a respondent may answer a question incompletely or may seem to have misunderstood the question. The first thing to do is simply to repeat the question as written a second time. If this does not help, the interviewer will have to “probe” to obtain the response.
Probing is a way of asking for further information without influencing the response. For example, “Could you explain that a little more?” or “Could you be more specific about that?” The interviewer must never interpret a respondent’s answer and then ask the respondent if the interpretation is correct.

There is no uniform understanding, even between health-service providers within the same health facility, on some of the issues for which we are collecting data or on terms used to describe items or practices. If it appears that the respondent does not understand what the interviewer is asking, or the response does not seem consistent with other information the interviewer has collected, the interviewer may rephrase or describe in more detail the item or practice that is the subject of the question, using examples to ensure that the respondent completely understands the question to which he/she is responding.

In cases where it may be necessary to provide additional clarification, the interviewer should provide only the minimum information required for an appropriate response.

If, however, the respondent appears to understand the question and the response is still not consistent, the interviewer must record the response as given by the respondent.

Never suggest answers to the respondents
If the respondent’s answer is not relevant to a question, the interviewer should not prompt the respondent by saying something like “I suppose you mean that… Is that right?” In many cases, respondents will agree with the interviewer’s interpretation of their answer, even when that is not what they meant. Instead, in most cases, the interviewer should probe in such a manner that the respondents themselves come up with the relevant answer, e.g.

“Can you explain a little more?” “There is no hurry. Take a moment to think about it.”

Specific questions for which it may be necessary to provide additional clarification will be discussed in the detailed instructions for completing the DV/SA questionnaires. Even in these cases, the interviewer should provide only the minimum information required for an appropriate response. Except when specifically instructed, the interviewer should never read out the list of coded answers to the respondents, even if they have trouble in answering the question.

Remain neutral
The job of an interviewer is to obtain the facts. An interviewer should be friendly, but firm; neutral, but interested. The tone of voice, facial expressions and even bodily postures all combine to establish the rapport you create with your respondent. The interviewer should not express surprise, pleasure or disapproval at any response or comment made by the respondent.

Ask all applicable questions
In most cases, the interviewer will ask questions in the sequence in which they appear in the questionnaire. However, because the organization of health facilities often differs, the
interviewer may find that to complete one section he/she has to talk to more than one respondent or go to different areas of the facility. It is up to the interviewer to ensure that, when sections of the questionnaire are skipped because the information must be collected from a different respondent or location, those sections are completed before departure from the facility.

**Do not raise expectations of immediate changes in the situation of the staff or facility**

The interviewer should not raise expectations that he/she can immediately help to solve problems that the staff or clients raise. The interviewer is going to provide information to decision-makers, health planners and administrators. However, any changes as a result of the survey will most likely occur over an extended period of time and be implemented gradually. If clients or staff members complain about the poor state of repair of the facility, equipment or supplies, or mention other problems, the interviewer should provide a neutral or non-judgemental response (e.g., “I know these things are difficult”).

**Do not separate questionnaires**

The interviewer should never separate stapled or bound questionnaire forms to speed up the process of data collection. Experience has shown that this strategy may result in lost pages.

**Thank the respondent at the end of the interview**

At the end of every interview, the interviewer should thank the respondent for taking time out of his/her busy schedule, telling him/her it was very much appreciated. The respondent should be asked if he/she can direct the interviewer to the next appropriate clinic/unit and/or person.

**Tips on handling difficult interview situations**

**The respondent is reluctant to participate**

Occasionally, a potential respondent will refuse to participate in the survey. The interviewer should not take a respondent’s initial unwillingness to cooperate as a final refusal. The interviewer should try to put himself/herself in the position of the respondent and think of factors that might have brought about this reaction. The respondent may not be in the right mood at that particular time or he/she may have misunderstood the purpose of the visit. The interviewer should try to find out why the respondent is unwilling to participate, and respond accordingly. Some points can be used to persuade a respondent to participate, namely:

- The information he/she provides will help the Ministry of Health and the government to better understand the effectiveness of programmes and make improvements to the programme that will ultimately help the clients.
- If confidentiality is a concern, the interviewer should reassure the respondent that everyone working on the survey has pledged to maintain confidentiality and that the respondent’s name will not be shared with others, including his/her supervisors or colleagues.
- The respondent cannot be replaced by anyone else.
However, in some circumstances a respondent may continue to refuse. In this situation, the interviewer should respect the respondent’s right to refuse, and thank the respondent for his/her time. The interviewer should not take these refusals personally.

**The respondent seems bored**
There may be situations where the respondent simply says, “I don’t know”, gives an irrelevant answer, acts bored or detached, contradicts something they have already said or refuses to answer the question. This happens most when the respondent is concerned about their other clinic/unit responsibilities and wants to get back to them. In these cases the interviewer must try to re-interest the respondent in the conversation. For instance, if the interviewer senses that the respondent is growing restless, he/she should be reassured that there are not many more questions and that his/her responses are very valuable.

**The respondent is very talkative**
If informants give irrelevant or elaborate answers or complain about something, the interviewer should not stop them abruptly or rudely but listen to what they have to say. Then the interviewer should try to steer the person gently back to the original question. The interviewer can also write down what the respondent says and tell him/her that it is duly noted. A good atmosphere must be maintained throughout the interview. The best atmosphere for an interview is one in which the respondent see the interviewer as a friendly, sympathetic and responsive person who cares about him/her.

**Data managers**
Data managers at national level are responsible for receiving data from the field (paper and electronic) and reviewing it for completeness and quality. When gaps and other anomalies are found, the data managers should investigate the problem and notify supervisors about the affected health facilities so that they can follow up to resolve problems and fill gaps. Automated tools within CSPro can assist in the identification of gaps and inconsistencies in data collection. Data managers should be trained to use these tools to ensure that this critical task is performed. If capacity for data management is lacking in a country, external technical assistance can be sought (e.g. from WHO country and/or regional offices).

The data manager’s main tasks include:

- assisting in the establishment of a central data server to receive and store collected survey data;
- leading the process to enter into a computer database the data collected on paper forms;
- compiling data as it comes in from the field and reviewing it for completeness and quality;
reacting to data gaps and inconsistencies by informing relevant survey field personnel and following up to ensure that the situation is **rectified**;

- updating the survey sampling list frame to take into account facilities that have been dropped and those that have been added during implementation, and ensuring the appropriate use of unique IDs and relevant communications with field **personnel**;
- cleaning the data and ensuring a complete final data set for **analysis**;
- assisting with data analysis as **appropriate**;
- ensuring that the master data file for the survey is up-to-date and **complete**;
- calculating survey indicators from the raw survey data using the standard indicator batch file in CSPro and making country-specific adaptations to the batch file as **necessary**;
- exporting the DQR indicators file from CSPro to other software for analysis, including the standardized MS Excel-based DQR chartbooks for health facilities and districts.

**Data management**

**Using CSPro for data entry**

Electronic data collection facilitates the collection of more accurate and reliable data in a more efficient and timely manner. For the DV/SA survey, electronic data collection is carried out with the Census and Survey Processing System (CSPro) software. CSPro was developed jointly by the U.S. Census Bureau, Macro International and Serpro SA, with major funding from the U.S. Agency for International Development (USAID). CSPro is in the public domain. It is available at no cost and may be freely distributed.  

**Installing CSPro**

The following is based on a Windows 10 set-up. Steps may vary if using a different operating system.

- Download the CSPro application from https://www.census.gov/data/software/cspro. Download.html.

- Install CSPro 7.2 to your computer by double-clicking on **cspro7.2.0.exe** (the two last digits of the version number may change as new releases are published. This will start the installation wizard. (The version of CSPro that you download should be at least 7.2.0 or higher for correct functionality of the DV/SA application.)

- The rest of the default settings for the installation are OK, so just click “next” until finished.

---

2 For information about the Census and Survey Processing System (CSPro), including free download, visit: https://www.census.gov/data/software/cspro.html (accessed 7 October 2020).
DV/SA data entry application
Open the DV/SA data entry application in CSPro

- To open the DV/SA data entry application, click on the “Facility_DV_SA.pff” shortcut available on the desktop (or navigate to the folder where the DV/SA applications for facility and district levels have been installed and find the “Facility_DV_SA.pff” file there).
- A new form (Case) opens (Figure 15).

Filling in the DV/SA data entry forms

- You can now start filling the cover page information.
- The question text is located in the yellow top window. For each question, a pop-up window will appear. Select your answer and validate your choice by clicking on the green check mark (or press the ENTER button).
- Fill in all the responses for the cover page. When you complete all the questions on the cover page (ending with consent for the survey), the next form will appear with the next set of questions.

Note: saving the new case will be possible only when the cover page has been completed through question QINTERVIEWER.

Figure 15. Opening the DV/SA data entry application
Save a data entry form during data collection

➢ To save the file, go to: File/Save partial case.

➢ The following pop-up window appears. Click “Ok” to continue the data entry.

Remember to save regularly while filling in the case to avoid losing any data.

➢ At the end of the form the following pop-up window will appear:
Click “Yes” to validate the data entry for the form. You will then automatically return to the CSEntry main screen.

The completed form now appears in the tree on the left.

Stopping data entry in the middle of a form

If you are filling in a form and need to exit the form for some reason, click on the “X” at the top right of the form to close it. A pop-up window opens:

Select “Partial save” to save the data entered. Another pop-up window will open, indicating that “The current case has been saved”.

The incomplete form appears in the tree in the left column of CSEntry with a red plus (+) sign indicating that it is a partially saved case and may not be complete.
Note that all forms should be completed before leaving the facilities.

Finalizing an incomplete form

- To complete a form, double-click on it in the left-hand tree of the CSEntry window.
- The form will open and ask:

  ![Form Opened at Last Position]

- Click yes and the form will open at the point at which data were last entered. You can now continue to enter data and complete the questionnaire.

Starting a new case

- To start a new case, double-click on the DV/SA_2.0_Simple.pff icon to start CSEntry.
- In the menu bar, go to Mode -> Add case to begin a new questionnaire.

Using GPS for collection of geographical coordinates

This section introduces the required steps for collection geographical data using GPS devices. For the purpose of this description the Garmin eTrex device has been used. Complete functionality of the device is described in the user manual provided with the Garmin eTrex.

The functions described in this section focus only on the essential steps for collecting the geographical coordinates of health facilities to complete the national Master Facility List.
Overview of the GPS functions
All of the information needed to operate the GPS is found on four main “pages” (or display screens).

These pages are (Figure 16b):
1. Satellite
2. Map
3. Navigation
4. Menu.

Figure 16a. The Garmin eTrex GPS unit

Figure 16b. GPS instruction screens

Simply press the PAGE button to switch between pages.
Use the ENTER, UP and DOWN buttons to access the different functions of a page.

Satellite
The Satellite page shows the eTrex gathering all the necessary satellite information in order to work. The Satellite page has two display options: Normal Skyview and Advanced Skyview. Normal Skyview shows the satellites, satellite signal strength, and the eTrex’s estimated location accuracy. Advanced Skyview shows the numbered satellites the eTrex is using, their proximity to your current position, and their individual signal strengths. To navigate from one to the other, select the ENTER button and choose the preferred skyview.
Map
The Map page shows where you are (the animated figure) and provides a real picture of where you are going. As you travel, the animated figure “walks” and leaves a “trail” (track log). Waypoint names and symbols are also shown on the map.

Navigation
The Navigation page helps guide you to a destination. When you are moving with no particular destination in mind, the Navigation page shows you your moving direction and speed. When you are moving towards a specific destination, the Navigation page shows you the name of the location, the distance and time left before arrival, and displays a direction arrow in the compass ring. To navigate, simply follow the arrow.

Menu
The main Menu gives you access to the eTrex’s more advanced features. With the main Menu, you can create and view waypoints, create a route, save and track logs, or access the system set-up features.

Set-up for map units
From the MENU page, select SETUP, then select the category UNITS in order to specify units of measure.

<table>
<thead>
<tr>
<th>POSITION FRMT</th>
<th>hddd.ddddd</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAP DATUM</td>
<td>WGS 84</td>
</tr>
<tr>
<td>UNITS</td>
<td>metric</td>
</tr>
<tr>
<td>NORTH REF</td>
<td>magnetic</td>
</tr>
<tr>
<td>ANGLE</td>
<td>degree</td>
</tr>
</tbody>
</table>

Where to collect GPS coordinates
The rules for GPS data collection are as follows:

Single facility in a building

- The geographical coordinates should be recorded in front of the main sign attached to the building of the facility.
- If there is no sign attached to the building, the geographical coordinates should be recorded in front of the main door or reception area of the facility.

Multiple facilities in a single building

- The geographical coordinates should be recorded in front of the sign(s) that lists what facilities are located in that building (if the sign is outdoors and attached to the building).
If there is no sign listing what is in the building (or if the sign is indoors), the geographical coordinates should be recorded in front of the main entrance door or reception area of the building.

Single facility in multiple buildings

The geographical coordinates should be recorded in front of the door or main entrance to the reception area of the facility (preferably in front of the main sign of the facility). If there is no reception area, the coordinates should be recorded in front of the door to the administrative offices of the facility.

**Using the GPS for collecting geographical data**

Once you have configured the settings on the GPS receiver, you can use it to record the geographical coordinates of a facility. Make sure that the internal memory has been cleared if the receiver had been used previously to collect data.

- Move to the main entrance of the building and stand within 30 metres of the main door. It is necessary to be in an open area that has a clear view of the sky and to hold the GPS receiver parallel to the ground so that its antenna is able to receive signals.
- Collect coordinates only when the receiver indicates that it has acquired signals from enough satellites to produce an accurate reading. The message “Ready to navigate” should appear on the GPS receiver (Satellite page) and the accuracy should be lower than 20 metres. In case you do not get the message “Ready to navigate”, wait at the same place for 5 minutes.
- When the signal is sufficient and the accuracy is at the recommended level, the geographical coordinates can be recorded as follows:
  - Go to the MENU page and select MARK.
  - Highlight the WAYPOINT number and press ENTER. You can now enter the facility code (maximum 6 digits or letters).
  - Press ENTER and scroll down to OK.
  - Press ENTER to go back to the MARK page.
  - Highlight OK using the UP and DOWN button and press ENTER.
  - The WAYPOINT is now registered.

**Retrieving a waypoints list from the GPS receiver**

To retrieve the waypoints from the list of GPS coordinates collected:

- Go to the MENU page and select WAYPOINTS.
- Using the UP and DOWN button you can now look for a waypoint just or previously entered.

You can now report the geographical coordinates on the DV/SA paper questionnaire (Figure 17) as well as on the electronic form (if applicable).
Features of the Garmin eTrex GPS unit

The eTrex GPS device has five main buttons, as described in Figure 16.
Latitude and longitude

Latitude is the north/south value measured from the equator. Longitude is the east/west value measured from the prime meridian that runs through West Africa and Western Europe. Latitude and longitude together can identify an exact location on the earth’s surface. Thus, based on the location of the health facility, positive (+) (North/East) or negative (-) (South/West) coordinates should be properly reported.

In order to conserve battery life, switch off the GPS receiver once the geographical coordinates are recorded.

Checklist for GPS units

Each morning before you leave for visits, check that you have all the necessary materials with you. Remember the following when collecting GPS coordinates:
- Check the battery level and verify that the device is properly working.
- Check the settings of your GPS receiver.
- Have the questionnaire for data collection ready to be used.
- Make sure to be properly placed for reliable data collection.
- Report coordinates, taking into account the location to the Equator and Greenwich meridian.
- Make sure that all information on the collecting point has been reported (in the paper DV/SA form as well as electronic form if applicable).

Data processing

After data entry, data should be processed in order to compile all data in a single file and to check for inconsistencies and possible errors. Any inconsistencies or errors should be addressed and reconciled in order to create a final, clean data set that is ready for analysis.

Note: If the standardized DV/SA analysis tools are to be used, all data processing must be done in CSPro and the final, cleaned data file must be in CSPro format.

The following are the major steps that must be taken when processing and cleaning the DV/SA data set:

- Concatenation
- Data cleaning
- Data verification for completeness
- Calculating sample weights
- Calculating the DV/SA indicators
- Exporting data.

The following sections provide details on DV/SA data cleaning and processing. These steps are based on the use of CSPro for data collection and data processing but they also provide some generic guidance and principles.
Chapter 3. Implementation

**Concatenation**

**Gathering data files into a single folder on a desktop/laptop computer**

After the data have been captured electronically, the data files have to be moved to a single desktop or laptop computer for further processing. Copying data files from the data collectors’ computers to a back-up computer/laptop is usually done by the field supervisors. It is also their responsibility to transfer data collected in the field to the central level.

At the end of data collection, supervisors should have a folder for each team with the backup files, as follows:

FACILITY_DV_SA_Team1_Date1
FACILITY_DV_SA_Team1_Date2
FACILITY_DV_SA_Team1_Date3

The latest file for each team should correspond to the final file. After validation by the supervisor, a copy of the final file should be created and should be renamed:

FACILITY_DV_SA_Team1_FINAL

The final data set should gather the final files from each team, as follows:

FACILITY_DV_SA_Team1_FINAL
FACILITY_DV_SA_Team2_FINAL
FACILITY_DV_SA_Team3_FINAL
FACILITY_DV_SA_data_collection_FINAL (REGION X)

This final data set should be transferred to the central-level data manager/focal point in charge of the compilation of the data from field collection. A back-up of all data files (final and stamped with dates) should be carefully saved and should remain accessible during the cleaning and data processing phase.

A USB flash drive can be used to transfer data from the supervisors’ computers to the data manager/focal point at the central level.

When all data have been transferred, there should be only one final data folder containing all data files from all the field teams, as follows:

FACILITY_DV_SA_Team1_FINAL
FACILITY_DV_SA_Team2_FINAL
FACILITY_DV_SA_Team25_FINAL
FACILITY_DV_SA_data_collection_FINAL
Data concatenation

If the CSPro data entry application has been used, the data will need to be concatenated (combined) into a single data file before any further processing can be done.

Before consolidating the data files, any duplicate facility IDs should be identified. If there are two or more data files containing the same facility IDs, CSPro will be unable to consolidate the files since no two cases are allowed to have identical ID items. If there are duplicates, one copy should be deleted prior to consolidating the data files.

Merge the data files

There are two potential processes for merging data files, depending on how the survey was deployed. If the data were collected on android tablets, the Data Viewer tool will be used to download the complete data set. If the data were collected on Windows computers, the Concatenate tool will be used to merge the data files into a single master file.

Data Viewer tool

CSPro stores data on the server in its own specific format. CSPro uses this format to allow synchronization at the case level. To access the data, the Data Viewer tool must be used to concatenate all the data files from the tablets and to convert the resulting file to a CSPro database file.
To do this, open CSPro, navigate to the Tools menu and select Data Viewer. From the file menu of the Data Viewer, select “Download”. Select the server you used and enter your login credentials. Then select the data file you wish to download and select the location where you would like to save the data set.

If the data have been entered on non-android tablets in the field, they will need to be merged into one data file.

Download the CSPro data files from the server. There should be one file per data collection team. Put all files in the same folder on your computer. Open CSPro and navigate to the Tools menu, and then select Concatenate Data (Figure 18).

1. Specify the output file name in “Output file:”.
2. Specify the CSPro data dictionary FACILITY_DV_SA.dcf.
3. Click “add” and navigate to the folder containing the data files.
4. Select all the data files and click “run”.

![Concatenate Data screenshot](image)
Data cleaning

Tracking facilities
After all the data have been concatenated, the first step in the review process is to take stock of the data and determine what has been collected. It is important to check that all facilities in the sample have been covered and, if not, to keep track of those that are missing. An Excel sheet such as shown in Figure 19 should be kept by the supervisors during the field data collection to help in this process:

In this example, green highlights the facilities that were in the original data set and have been assessed. It also indicates that data collectors have entered data in CSPro and that the data have been validated (in accordance with the steps described in the following section). The facilities highlighted in blue are replacement facilities. It is important to indicate which of the facilities from the original sample have been replaced. Finally, the facilities in red are those which could not be assessed; information on why the facility could not be assessed should be included in the tracking table.

This information is extremely useful in understanding what happened during the field data collection compared with the original plan. The information will also be helpful later when sample weights are being calculated.

Figure 19. Keeping track of facilities sampled
Reviewing data files

Each data file should be opened and the following key items checked:

- The facility name and ID number correspond to each other.
- The facility ID information is correct (facility type, managing authority, as on the master facility list).
- The data collector ID is correct.
- All forms have been fully completed.
- Empty cases have been deleted.
- Duplicate cases have been identified and reconciled.
- “Other” responses have been recorded as applicable.
- Geographical coordinates have been checked (if applicable).
- Supervisor validations have been compared with originals and have been reconciled.

If errors are found in the facility name, facility ID number, facility type, or managing authority or any other element, these should be corrected. Key items should be reviewed as follows (instructions based on electronic data collection using CSPro):

Reviewing facility identification

Reviewing facility identification is the first step in data cleaning. It consists of validating the name of the facility and its corresponding identification code, type, managing authority and location, as defined in the master facility list. If any updates have been made to the facility identification during field data collection, all these changes should be captured in the log of the facilities surveyed (as described in the previous section). This will assist the supervisor and others in charge of the data handling to understand the facility identification and why it differs from the initial master facility list.

It is also important to make sure that the interviewer ID is properly entered. In some cases, data processors find discrepancies that need to be reconciled and will need to contact the team responsible for the original data collection. Having the interviewer ID properly recorded greatly facilitates this process.

Reviewing completeness of a case

It is important to verify that all forms relating to a case (CSPro electronic version of the paper questionnaire) have been filled properly. The completeness of a case should be verified before leaving the facility and should be double-checked by the supervisor when backing up the data so that any gaps can be identified and filled if necessary.
Recode “other” responses as applicable
All responses that have “other, please specify” as an option for the answer should be reviewed to make sure that the response written in the “other” category does not fall into one of the precoded categories. If the “other” response does fall into a precoded category, it should be recoded as such.

Delete empty cases
Occasionally a case that contains no data will be stored by accident. These cases should be removed from the data set.

Final check for any duplicate cases
Duplicate cases are cases with the same facility code. If two cases appear to be duplicates according to the facility name but do not contain the same data, a list of criteria must be used to determine if they are true duplicates. The following data elements could be used as the criteria for determining duplicates:
- district;
- facility code/name;
- GPS coordinates (if collected);
- facility type;
- managing authority;
- interviewer’s code.

If these are all the same in both cases, it is safe to consider the cases as duplicates. At this point, the most complete case should stay in the data set. If both cases are complete, the case with latest time stamp should be kept.

Check the validity of GPS coordinates, if applicable
GPS coordinates should be checked to ensure that they fall within the boundaries for the country. Sometimes latitude and longitude coordinates can be entered incorrectly. All GPS coordinates should be double-checked to ensure that they are valid for the area being surveyed.

For instance, all facilities in Kenya should fall within the following ranges:
- Latitude: 5°N and 5°S (-5.0000 to 5.0000 in decimal degrees);
- Longitude: 34° and 42°E (34.0000 to 42.0000 in decimal degrees).

One common mistake is not to record correctly the positive (+) and negative (-) values for coordinates. North and East coordinates should be positive (+). South and West coordinates should be negative (-). Another common mistake is to reverse the recording of longitude and latitude coordinates. Review and edit the GPS coordinates using the same method as used above for reviewing and editing the key items for each facility.
Identify supervisor validation records and reconcile with the original record

If supervisor validations have been conducted, it is important to identify them in the data set and make sure that they have been labelled correctly (DV_002). A comparison between the supervisor validation record and the original record should be carried out and any differences should be reconciled so that there is ONE record per facility in the final data set. See the guidance below on using the CSPro Compare Data tool to compare two data files and identify the differences.

Dependent verification if survey was conducted on paper and entered into CSPro at a later time (if applicable)

Dependent verification is used to check that the electronic data are consistent with the responses in the paper version of the questionnaire. When you verify a case, you key the case a second time as if you were in “Add” mode. Even if there is already data in the data file, CSEntry does not show this to you. All fields on the current form start out blank. Each time you key a field, the system compares the value you keyed with the value in the data file. If these two values match, you move to the next field. If the values do not match, you see a message telling you so. When this happens, simply rekey the field. One of the following situations will occur:

- The second value you key matches the value in the data file. The system assumes your first value is in error and moves to the next field. There will be no change to the data file for this field.
- The second value you key matches the first value you keyed. The system assumes the value in the data file is in error and moves to the next field. The new value, which you keyed twice, will replace the original value in the data file.
- The second value you key matches neither the value in the data file nor the first value you keyed. The system will throw away the first value you keyed, show you the mismatch message and wait for you to rekey the field again.

Using CSPro for data checking and validation

Reviewing data in CSPro

Supervisors should review data transferred to the laptop for completeness and consistency. The steps for reviewing the data are as follows:

- Data files in CSPro should be checked for each facility.
- At a minimum, the following items should be verified:
  - the facility code and the facility name match;
  - the level and type of facility are correct (based on the facility inventory);
  - the data collector’s ID/team name is correct;
  - the location of the facility is correct;
  - the data file has no missing values.
Open FACILITY_DV_SA CSPro application (Figure 20):

- Click on the file named FACILITY_DV_SA.pff (there is a separate application for district level DV/SA).

**Figure 20. Opening the facility application**

Select the data file to open:

- Select the case containing the data for the facility you would like to review. Double-click to open it.

The case opens on the cover page form.
Review the information in accordance with the beginning of this section, i.e.

- The facility code and the facility name match.
- The level and type of facility are correct (based on the facility inventory).
- The data collector ID/team name is correct.
- The location of the facility is correct.

After reviewing the cover page (Figure 21), use the tree on the left window to review other sections if necessary.

When review and editing is complete, click on stop and be sure to save your changes.

**Check for completeness of data**

A batch edit application has been created to track inconsistencies in data. This allows more in-depth data cleaning and validation. The application identifies unanswered questions that should have been answered, as well as questions that should not have been answered but were (based on specific skip patterns). To run the batch application use the following steps:

- Browse to Facility_DV_SA\Batch\Batch_1 and click on the "Facility DV SA data cleaning.bch" file.
- Click on “Run” from the menu bar. Browse to Facility_DV_SA\Data and select the Facility_DV_SA_data.dta file as the input file. Then click on OK. No output file is required.
- When the application finishes running, a text file will open with the results of the batch such as the one in Figure 22.
Figure 22. Facility batch results in CSPro
Each case will be identified by the facility number and name and a list of error messages will be shown. These messages correspond to inconsistencies in the data. Each inconsistency should be reviewed with the survey team and the records edited appropriately to finalize the data set.

The completeness batch file should be used by supervisors in the field to monitor the completeness of the data as they come in (i.e. each night after data collection at facilities) so that field teams can make corrections, as necessary, before leaving the area for the next facility.

Data managers at national level should also use the completeness batch file on the compiled data to monitor data completeness at intervals during implementation of the survey. When gaps are uncovered, field teams should be contacted and alerted to the gaps in the survey data.

**Programme managers and M&E officers**

Programme managers and M&E officers have unique insights into the dynamics of service delivery for their specific health programmes. Their knowledge is invaluable for interpreting and determining the plausibility of results. Therefore they should be involved in the review and interpretation of findings (e.g. by participating in the results validation workshop).

During implementation of the survey, these personnel can play a valuable role as higher-level monitors and supervisors for designated areas.

**Quality assurance**

There needs to be quality assurance in all aspects of survey planning and implementation. Special attention should be paid to critical aspects such as data collection. Working in pairs, data collectors can assure the quality of the work of each other. Supervisors should review data collection forms for completeness and quality, as well conducting a repeat assessment of a small sample of facilities. In addition, an independent group can be engaged to repeat the survey at a small percentage of facilities. The results of these parallel assessments are compared and discrepancies quantified. Any programme areas or indicators with large discrepancies should be investigated further and the survey repeated in cases where the discrepancies are severe.

The CSPro DV/SA applications includes a tool which allows data managers to compare, as a measure of quality control, different records completed for the same facility. The “supervisor split” batch file compares each field in the data set for multiple records for the same facility. Values that do not match are flagged and an output file, sorted by facility, displays the fields with non-matching data.

To compare the records between data collectors and supervisors (or an external quality assurance team) the records must be divided into two data sets.

Records in the database can be identified as either ”interviewer” or ”supervisor” by the value in the field DV_002; 1 for interviewer and 2 for supervisor.
To split the database:

- Double-click on the “Supervisorsplit.bch” batch file in the Batch 2 folder and then click the stoplight.
- This will run a batch application to create two datasets: 2_FACILITY_DV_SA_FINALDATA.csdb (which includes all the original data collector records) and 3_FACILITY_DV_SA_SUPERVISORDATA.csdb (which contains only the supervisor validations).
- When the application finishes running, a report of the process will open. You can close this window when it is complete. Browse to FACILITY_DV_SA\Data and check to make sure that the two files are present.

Compare the data files:

- The Compare Data tool is a CSPro tool that allows you to compare two data files and identify the differences. The data files must have the same structure (i.e. they must be described by the same CSPro data dictionary). In addition, for comparisons to be made, the original case and the supervisor validation must have the same ID information and be in different data sets.
- In order to compare the supervisor validation record with the original facility record the two must have exactly the same ID information. In order to accomplish this, take the steps that follow.
  - Press the Start button and navigate to “Programs → CSPro7.0 → Data tools → Export data”.
  - The first screen in the CSPro Data Export application will ask for the Data Dictionary File. Navigate to the FACILITY_DV_SA folder, select the FACILITY_DV_SA.dcf file. Then click Open.
  - The panel on the left should now display the data dictionary’s records and items in a selectable dictionary tree. From this dictionary tree, select the data you would like to export. Click on the top box next to the dictionary icon to select all the data.
  - Next, open the tree called ID items. Deselect everything EXCEPT the facility number. The tree should look like the image to the right.
  - This screen will also display various export options – such as the export format, how many files you would like the application to create, and whether you want to include XML metadata. Keep the default options for everything except the export format. Select CSPro (.dat, .dcf) from the export format options.
  - To export the data, click on the stoplight on the toolbar, or select Run from the File menu.
  - The next screen will ask you to select the data file you would like to export. Browse to your concatenated, de-duplicated data file and click Open.
  - CSPro will then ask you to specify the name of the output data file and dictionary file. Browse to the FACILITY_DV_SA\Data\Supervisor validation folder and name the output
data file 4_FACILITY_DV_SA_FINALDATA_EXTRACT.csdb and save the dictionary in FACILITY_DV_SA\Data\Supervisor validation as FACILITY_DV_SA_data_compare.dcf (Figure 23). Then click OK to run.

Browse to the FACILITY_DV_SA\Data\Supervisor validation folder and name the output data file 4_FACILITY_DV_SA_FINALDATA_EXTRACT.csdb and save the dictionary in FACILITY_DV_SA\Data\Supervisor validation as FACILITY_DV_SA_data_compare.dcf (Figure 23). Then click OK to run.

Repeat these steps with the 3_FACILITY_DV_SA_SUPERVISORDATA.csdb data set, saving the file in the FACILITY_DV_SA\Data\Supervisor validation folder as 5_FACILITY_DV_SA_SUPERVISORDATA_EXTRACT.csdb.

The data files are now ready for the compare tool to be used.

Navigate to “Programs → CSPro7.0 → Data tools → Compare data”.

The first screen in the CSPro Compare Data application will ask you for the Data Dictionary file. Navigate to FACILITY_DV_SA\Data\Supervisor validation, select the FACILITY_DV_SA_data_compare.dcf file, and click Open.

The panel on the left should now display the data dictionary’s records and items in a selectable dictionary tree.

Click on the top box next to the dictionary icon to select all the data. The screen should look like the image to the right.

Figure 23. Specify the output data file and dictionary file
To run the Compare function, click “Run” on the toolbar; or press Ctrl+R; or, from the File menu select “Run”.

For the input file, select the 4_FACILITY_DV_SA_FINALDATA_EXTRACT.csdb file from the FACILITY_DV_SA\Data\Supervisor validation folder. For the reference file, select 5_FACILITY_DV_SA_SUPERVISORDATA_EXTRACT.csdb from the FACILITY_DV_SA\Data\Supervisor validation folder.

For the comparison method, make sure the “Compare Input to Reference and Reference to Input” box is selected (Figure 24).

For the comparison method, make sure the “Compare in indexed order” box is selected. The screen should look like the image below.

Click OK to run the Compare tool. An output summarizing the results of the file comparison will be shown.

Examine the output. The output should look like the image in Figure 25.

Figure 24. Run the compare tool
The input file and reference file are listed at the top. In both files each case appears listed on the left, identified by the facility code.

For each case, any difference between the input file and the reference file will be listed, with values for the input file under the column “Input File” and for the reference file under the column “Reference File” (at the right).

If the case exists in one file but not in the other, CSDiff will output “Case missing” in the relevant column.

In the screenshot on the previous slide, the input and the reference files contain data for the facility with facility code 01141234. The results will show only the differences between the two cases with the same ID.

If the differences are only in spelling of the facility name, facility location, "other" or text only fields, no edits need to be made.

If differences arise in other questions, make a list by facility of those questions that have a mismatch. This should be sent to the supervisor for resolution of discrepancies. When the discrepancies have been resolved, return to the original data set and edit the record to reflect the changes.
Chapter 4. Data analysis

Automated chartbooks in MS Excel

WHO has created an automated analysis template in MS Excel to facilitate the analysis of DQR survey data. Once the indicators for the analysis have been calculated using the Indicators batch file in CSPro, and the data file has been exported from CSPro to Excel format, the data can be pasted into the “chartbook” for ready analysis.

The chartbooks produce tables and graphs with sample estimates stratified by facility type, managing authority and milieu (urban/rural), as well as a user-specified subnational level of the health system (e.g., region or district). First, however, the raw data need to be compiled into indicators for use in the chartbook. Fortunately, there is a CSPro batch file to facilitate this task.

There are chartbooks for the health facility and for district-level analyses, and the processes for compiling the data and populating the tools are the same for both. At the district level there is an additional tool – the DV Chartbook for Countries with Electronic HMIS Systems – to facilitate the aggregation of facility indicator values for comparison to data aggregated to the district level in the HMIS.

In order to use this chartbook, data are required from two sources, namely:

- The FACILITY DV/SA indicators data generated in and exported from CSPro from the facility DV/SA survey.
- Facility-level HMIS data for the same time period, facilities and indicators, as included in the facility DV/SA survey.

After data from these two sources have been properly inputted into the chartbook, tables will automatically update and preliminary results will be available. See below for the steps needed to populate the chartbook.

The Health Facility Level DV/SA Chartbook can be accessed here: URL
The District Level DV/SA Chartbook can be accessed here: URL
The DV Chartbook for Countries with Electronic HMIS Systems can be accessed here: URL
Calculate DQR indicators

The DQR CSPro database application comes with software files designed to perform tasks related to data management, such as an evaluation of completeness and quality, and calculating indicators for data analysis. There are separate software files for both facility-level and district-level DQR analysis. The software files are run from the CSPro data management application and result in output files of calculated indicators ready to be exported from CSPro into Excel for pasting into the DQR chartbooks for both facility and district levels.

Run the indicator batch file

- Once the DV/SA data have been compiled, de-duplicated and cleaned, the indicators used in the analysis should be calculated.
- The facility and district DV/SA CSPro applications come with a batch edit software file designed to automatically calculate the indicators.
- Navigate to FACILITY_DV_SA\Batch\Batch_3\Facility DV SA Indicators.bch and double-click on the file (Figure 26).
- The left pane of the batch editor shows the data elements in the FACILITY_DV_DA data dictionary.
- The pane on the right shows the editor with the indicators software program.
- Notice the “STRATUM” variables – these will require local adaptation corresponding to the stratifiers in the sampling methodology.
- STRATUM 1 is for recodes of the subnational units (e.g. region or district).
- STRATUM 2 is for recodes of facility type.
- STRATUM 3 is for recodes of management authority, and
- STRATUM 4 is for recodes of urban/rural
- Scroll down in the editor pane. Notice the variables for Weight.

Programme-specific weights

- There is a generic weighting variable called “WEIGHT”. This applies weights to survey estimates that apply to system-wide indicators (such as the system assessment).
- WEIGHT1, WEIGHT2, WEIGHT3, etc. are programme-specific weights.
- Each programme-specific weighting variable is derived from: 1) the distribution of health facilities in the sample that provide that particular service, and 2) the service volume for the indicator.
- The weights are programme-specific since the different health programmes have different service volumes and not all health facilities offer all the services.
- See the next section for a detailed explanation of weighting of the survey estimates.
Run the indicator batch file

- Once the STRATUM variables are updated with local adaptations and the sample weights are added, run the indicator batch file by clicking on the stoplight, or select “Run” from the file pull-down menu.
- If any edits have been made for a country implementation of the Facility DV, those edits must also be made in the batch edit application. For example, if a question is deleted from the Facility DV questionnaire, all associated logic in the batch must also be deleted.
- Click “Yes” when asked to create the file.
A file called facility DV/SA Indicators.csdb is created and saved to the FACILITY_DV_SA\Data folder.

Check the facility DV SA_indicators.lst (Figure 28) to ensure that everything went as it was supposed to.

Figure 28. Checking the list of indicators
Export the indicator file to .txt

- Open CSPro, go to the Tools menu and select Export.
- Specify the FACILITY_DV_SA data dictionary when prompted.
- Click on ID Items to select all data elements. De-select “an underscore”.
- Click on Cover Page to select all items.
- Click on Section 12: Recodes Data Verification to select all data elements.
- Leave all other data elements unchecked.
- Use the default settings on the right and ensure the file is exported as a text file (tab delimited). Select “Run”.

Open the exported data file

- Open a blank workbook in Microsoft Excel.
- Go to File, Open, and browse to the FACILITY_DV_SA\Data folder. Make sure ALL FILES is selected in the browse window. Select the facility DV/SA Indicators.txt file that has just been exported from CSPro.
- In the text import wizard:
  - Make sure “delimited” is selected.
  - Under “delimiter”, select Tab.
  - Under column data format, select “general”.
- The txt file is now open in Microsoft Excel. Click on File -> Save as and change the “save as” type to “Excel Workbook” (Figure 29). The file is now saved as an XLS file.
- If the data set contains a decimal comma instead of a decimal point, replace the commas with points in order for the chartbook to work properly. To do this, select all the data using Ctrl+A, then Ctrl+F to open the Find/Replace window. In the “Find” box type , and in the “Replace” box type, click on “Replace all”.
Check that the Stratum and Weights columns are complete

- Select the header row and then in the menu ribbon click on Data -> Filter.
- On the column with the header STRATUM_1, right-click on the small icon at the right side of the cell. Scroll down and there will be a list of response values. Check that this list does not include “Blanks”.
- Repeat for columns STRATUM_2, STRATUM_3, STRATUM_4 and WEIGHTS.
- If there are blanks, return to the data-processing steps in CSPro to ensure that all questions are complete. If there are no blanks, move on to the next step.
Populate the Excel chartbook with data from the indicator file

Check that column headers match in the chartbook and in the exported data

- Select the header row of the exported indicator data and select “Copy”.
- Open a new, blank workbook. In cell A1, right-click and select “Paste/transpose”.
- Open the FACILITY DV/SA chartbook. Go to the blue tab called Indicators. Select the header row and select “Copy”.
- Go back to the blank workbook. Click on the cell B1, right-click and select “Paste/transpose”.
- In cell C1 type the following: =exact(A1,B1).
- Hover the mouse over the lower right-hand corner of cell C1 until a plus (+) sign appears, then double-click so that the formula is dragged to the bottom of the sheet.
- Check that all cells in column C say “True” (or that the column headers match). If FALSE appears anywhere in column C, edit the exported data file to add or subtract columns to match the chartbook columns.
- Some versions of the chartbook have three data elements each for latitude and longitude of the health facility. The FACILITY_DV_SA data dictionary has only two each. You may need to add an additional column for both latitude and longitude to the facility DV/SA indicators.xls in order to reconcile the difference in columns between the two files.
- Once the column headers match, go to the FACILITY DV/SA chartbook and click on the red tab called INSTRUCTIONS.
- Scroll down to row 41 and replace the text in <> quotes so that the name of the country and year of the assessment automatically show on the output printouts.

Copy/paste exported data into chartbook

- Go to the FACILITY DV/SA chartbook and click on the blue tab called Indicators. Highlight all data except for the header row and delete the existing data.
- Go to the exported data file and select all data except for the header row, and then select “Copy”.
- Go back to the FACILITY DV/SA chartbook Indicators tab, click on cell A2 and then click “Paste”. This should paste the data from the exported data file to the chartbook.
- The next tab – Weighted_Data – will automatically update on the basis of the information entered on the Indicators tab. The Weighted_Data tab is where the indicators are multiplied by the weights. The analytical outputs are based on the weighted data in this worksheet.
Enter information on the labels tab

- Go to the FACILITY DV/SA chartbook (Figure 30) and click on the purple tab called **Labels**.
- In column D – “Label of region” – enter the labels for each region in your survey. These should correspond to the order created for STRATUM_1 in the data processing step. Please note, the column is labelled “Region”, but this may correspond to any administrative unit such as zone, district, etc.
- In column H, Health facility types, enter the labels for each facility type in survey. These should correspond to the order created for STRATUM_2 in the data processing step.
- In column L, Managing authority, enter the labels for each managing authority in the survey. These should correspond to the order created for STRATUM_3 in the data processing step.
- In column P, Urban/Rural, enter the labels for each urban/rural category in the survey. These should correspond to the order created for STRATUM_4 in the data processing step.

Enter information on survey type

- Go to the FACILITY DV/SA chartbook, and click on the orange tab called **Survey type**.
- First click on the button 1- Check weighted data. This automatically extends or retracts the formulas in the **Weighted Data** tab to match the number of rows of data the in the **Indicators** tab.
- Then click on the button select “Hide empty rows”. This will hide the extra rows that will not be used (Figure 31).
- It may take up to 5 minutes for Excel to run this command. Please wait for Excel to finish the task.

**Figure 30. Inserting labels into the Excel chartbook**
If you do not change the Labels tab again, there is no need to run this step even if the data are updated.

However, if you change the Labels you will need to click “Show all rows” then “Hide empty rows” to update the chartbook.

Next, specify whether the survey is to have national-level results or subnational results. In the display level, please select either National or Subnational according to the sampling methodology chosen for the survey.

The National option activates sheets that have national estimates as well as stratification by facility type, managing authority and urban/rural.

The Subnational option activates sheets that include the same information as the national option plus stratification by region.

Note that, if you are interested in subnational information by region, you will need a chartbook for each region.

The chartbook should now be ready for review. Click on each tab in the chartbook to review all tables and charts (Tables 11 and 12).

Figure 31. Entering survey information into the Excel chartbook
Table 11. Example table from DQR facility-level Data Analysis Excel chartbook – verification factors

<table>
<thead>
<tr>
<th>Data verification factor</th>
<th>S1_06</th>
<th>S2_06</th>
<th>S3_06</th>
<th>S4_06</th>
<th>S5_06</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ANC (N=99)</td>
<td>DPT3/PENTA (N=109)</td>
<td>HCT (N=60)</td>
<td>Notified cases of TB (N=37)</td>
<td>Malaria cases (N=109)</td>
</tr>
<tr>
<td>Facility type</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hospital</td>
<td>0.99</td>
<td>0.92</td>
<td>0.97</td>
<td>1.00</td>
<td>1.70</td>
</tr>
<tr>
<td>CHC</td>
<td>0.99</td>
<td>1.02</td>
<td>1.00</td>
<td>0.77</td>
<td>0.96</td>
</tr>
<tr>
<td>CHP</td>
<td>0.99</td>
<td>0.95</td>
<td>1.12</td>
<td>1.00</td>
<td>0.95</td>
</tr>
<tr>
<td>MCHP</td>
<td>0.99</td>
<td>1.05</td>
<td>1.06</td>
<td>?????????</td>
<td>0.97</td>
</tr>
<tr>
<td>Managing authority</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Government/Public</td>
<td>0.96</td>
<td>1.01</td>
<td>1.05</td>
<td>0.82</td>
<td>0.98</td>
</tr>
<tr>
<td>Private</td>
<td>2.10</td>
<td>0.96</td>
<td>1.00</td>
<td>0.96</td>
<td>1.00</td>
</tr>
<tr>
<td>Urban/Rural</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urban</td>
<td>1.08</td>
<td>0.94</td>
<td>0.96</td>
<td>0.96</td>
<td>0.95</td>
</tr>
<tr>
<td>Rural</td>
<td>0.97</td>
<td>1.03</td>
<td>1.08</td>
<td>0.78</td>
<td>0.98</td>
</tr>
<tr>
<td>Total</td>
<td>0.98</td>
<td>1.01</td>
<td>1.05</td>
<td>0.82</td>
<td>0.98</td>
</tr>
</tbody>
</table>

Table 12. Example table from DQR facility-level Data Analysis Excel chartbook – System assessment

<table>
<thead>
<tr>
<th>Data quality and supervision</th>
<th>Routine process for checking quality of reports</th>
<th>Accuracy check of summarised data routinely conducted</th>
<th>Checks of timely entry and completeness routinely conducted</th>
<th>Written documentation of the results of data quality controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Facility type</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hospital</td>
<td>73%</td>
<td>73%</td>
<td>64%</td>
<td>82%</td>
</tr>
<tr>
<td>CHC</td>
<td>45%</td>
<td>40%</td>
<td>38%</td>
<td>62%</td>
</tr>
<tr>
<td>CHP</td>
<td>43%</td>
<td>29%</td>
<td>29%</td>
<td>49%</td>
</tr>
<tr>
<td>MCHP</td>
<td>26%</td>
<td>38%</td>
<td>38%</td>
<td>49%</td>
</tr>
<tr>
<td>Region</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>17%</td>
<td>27%</td>
<td>24%</td>
<td>36%</td>
</tr>
<tr>
<td></td>
<td>31%</td>
<td>44%</td>
<td>44%</td>
<td>58%</td>
</tr>
<tr>
<td></td>
<td>32%</td>
<td>25%</td>
<td>25%</td>
<td>39%</td>
</tr>
<tr>
<td></td>
<td>49%</td>
<td>44%</td>
<td>44%</td>
<td>65%</td>
</tr>
<tr>
<td>Managing authority</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Government/Public</td>
<td>37%</td>
<td>37%</td>
<td>37%</td>
<td>52%</td>
</tr>
<tr>
<td>Private</td>
<td>11%</td>
<td>11%</td>
<td>6%</td>
<td>39%</td>
</tr>
<tr>
<td>Urban/Rural</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urban</td>
<td>21%</td>
<td>21%</td>
<td>31%</td>
<td>46%</td>
</tr>
<tr>
<td>Rural</td>
<td>39%</td>
<td>38%</td>
<td>36%</td>
<td>53%</td>
</tr>
<tr>
<td>Total</td>
<td>36%</td>
<td>36%</td>
<td>35%</td>
<td>52%</td>
</tr>
</tbody>
</table>
District-level data analysis

The processes for compiling indicators and populating the chartbook for DV/SA at the district level is the same as the one followed for the facility level.

For the DV chartbook for countries with electronic HMIS systems, follow the steps below to populate the tool and analyse the data.

Data from two sources are required in order to use this chartbook: 1) the FACILITY DV/SA indicators data generated in and exported from CSPro from the facility DV/SA survey; and 2) facility-level HMIS data for the same time period, facilities and indicators, as included in the facility DV/SA survey.

Steps to populate the DV chartbook for countries with electronic HMIS systems

Step 1. Enter the country name and year on the INSTRUCTIONS tab.
Step 2. Format the facility DV/SA data and copy/paste to the Facility Survey data tab.
Step 3. Check data tabs.
Step 4. Format HMIS data and copy/paste to the HMIS Data tab.
Step 5. Enter regions, facility, types, managing authorities and urban/rural on the Labels tab.
Step 6. Check data tabs, hide empty rows, and select national or subnational on the Survey type tab.

**STEP 1. Enter the country name and year in the chartbook**

- Open the DV_Electronic_HMIS chartbook and click on the red tab called INSTRUCTIONS.

Scroll down to row 41 and replace the text in <> quotes so that the name of the country and year of the assessment automatically shows on the output printouts.

**STEP 2. Format the facility DV/SA data and copy/paste to the Facility Survey data tab**

**Step 2a. Export the data from CSPro to use in the DV_electronic_HMIS chartbook**

- On your desktop or laptop computer, press the Start button and navigate to the CSPro Export Data application. This will most likely be located in “Programs → CSPro 7.2 → Export data”.

The first screen in the CSPro Data Export application will ask for the Data Dictionary file. Navigate to the CSPro Facility DV_SA folder, select the FACILITY_DV_SA.dcf file, and click Open.

The panel on the left should now display the data dictionary’s records and items in a selectable dictionary tree. From this dictionary tree, select the data you wish to export. You will see the following items (make sure the boxes next to these three sections are checked).

- ID items: Within the ID items section, the item called underscore needs to be unselected. To do this, click on the + next to ID items to expand the section. Next go to “an underscore” and uncheck the box.
Chapter 4. Data analysis

- Cover page: all items

- The following questions and indicators: DV_103_01B, DV_103_02B, DV_103_03B, DV_104_01B, DV_104_02B, DV_104_03B, DV_203_01B, DV_203_02B, DV_203_03B, DV_204_01B, DV_204_02B, DV_204_03B, DV_303_01B, DV_303_02B, DV_303_03B, DV_304_01B, DV_304_02B, DV_304_03B, DV_405, DV_407_01B, DV_503_01B, DV_503_02B, DV_503_03B, DV_504_01B, DV_504_02B, DV_504_03B, STRATUM_1, STRATUM_2, STRATUM_3, STRATUM_4, WEIGHT, WEIGHT1, WEIGHT2, WEIGHT3, WEIGHT4, WEIGHT5, S1, S1_01, S1_02, S2, S2_01, S2_02, S3, S3_01, S3_02, S4, S4_01, S4_02, S5, S5_01, S5_02

This screen will also display various export options, such as the export format, how many files you want the application to create, and whether you want to include XML metadata. Keep the default options for everything except the Export Format. Please select Tab delimited (txt) from the export format options.

Once you are ready to export the data, select Run from the File menu.

The next screen will ask you to select the data file you would like to export. Browse to the CSPro Facility DV_SA/Data folder, select the 6_FACILITY DV_INDICATORS_ file, and click Open. This file contains the facility DV indicators that were generated previously.

CSPro will then ask you to specify the name of the exported file. This is the TXT file that CSPro will create. Enter 8_FACILITY HMIS_INDICATORS_EXPORT.txt and ensure that the file is saved in the CSPro Facility DV_SA/Data folder. Press Save.

**Step 2b. Open the exported data file**

- Go to File, Open, and browse to the SARA and DV_SA\CSPro Facility DV_SA\Data folder. Make sure that ALL FILES is selected in the browse window. Select the 8_FACILITY HMIS_INDICATORS_EXPORT.txt, file that has just been exported from CSPro.

In the text import wizard:

- Make sure “delimited” is selected.
- Under “delimiter”, select Tab.
- Under “column data format”, select “general”.

The txt file is now open in Microsoft Excel. Click on “File -> Save as” and change the file type to Excel Workbook. The file is now saved as an XLS file.

If your data set contains a decimal comma instead of a decimal point, you will need to replace the commas with points in order for the chartbook to work properly. To do this, select all the data
using Ctrl+A, then select Ctrl+F to open the Find/Replace window. In the “Find” box type, and in the “Replace” box type, then click on “Replace all”.

**Step 2c. Check that the Stratum and Weights columns are complete**

- Select the header row and then in the menu ribbon click on Data -> Filter

On column AD, header STRATUM_1, right-click on the small icon on the right-hand side of the cell. Scroll down and there will be a list of response values. Check that this list does not include “Blanks”.

Repeat for columns STRATUM_2, STRATUM_3, STRATUM_4 and WEIGHTS (AE, AF, AG and AH-AM).

If there are blanks, return to the data processing steps in CSPro to ensure that all questions are complete. If there are no blanks, move on to the next step.

**Step 2d. Check that the column headers match between the chartbook and the exported data**

- Select the header row and select “Copy”.

Open a new, blank workbook and the in cell A1, right-click and select “Paste/transpose”.

Open the DV_Electronic_HMIS chartbook. Go to the blue tab called **Facility Survey Data**. Select the header row and select “Copy”.

Go back to the blank workbook, click on the cell B1, right-click, and select “Paste/transpose”. In cell C1 type the following: =exact(A1,B1).

Hover the mouse over the lower right-hand corner of cell C1 until a plus sign appears, then double-click so that the formula is dragged to the bottom of the sheet.

Check that all cells in column C say “True” (or that the column headers match). If FALSE appears anywhere in column C, edit the exported data file to add or subtract columns to match the chartbook columns.
Step 2e. Copy/paste exported data into the chartbook

- Go to the DV_Electronic_HMIS chartbook and click on the blue tab called Facility Survey Data. Highlight all data except for the header row and delete the existing data.

Go to the exported data file and select all data except for the header row, then select “Copy”.

Go back to the DV_Electronic_HMIS chartbook Facility Survey Data tab, click on cell A2, and then click “Paste”. This should paste the data from the exported data file to the chartbook.

STEP 3. Check data tabs

- Go to the DV_Electronic_HMIS chartbook and click on the orange tab called Survey type.

Click on the button 1- Check data tabs. This automatically extends or retracts the formulas in the HMIS data, Indicators, and Weighted Data tabs to match the number of rows of data the in the Facility Survey Data tab.

STEP 4. Format HMIS data and copy/paste to the HMIS data tab

The HMIS data tab should now be pre-populated with the following information for all facilities for which HMIS data is required: Facility number, Name of facility, Region/Province code, Region/Province name, District code, District name, Subdistrict code, Subdistrict name.

- From the country HMIS system, extract information corresponding to the remaining empty columns in the chartbook for the required facilities. This includes the monthly reported counts for three months, as determined by the survey data collection parameters for the assessed indicators. As an example, the monthly reported counts for three months could be Jan, Feb, Mar 2020. However, see the final facility DV/SA questionnaire to determine which months were assessed and if there were any specific indicator definitions.

Copy/paste the values into the DV_Electronic_HMIS chartbook HMIS Data tab in columns I – U.

STEP 5. Enter information on the Labels tab

- Go to the DV_Electronic_HMIS chartbook and click on the purple tab called Labels.

In column D, Label of region, enter the labels for each region in your survey. These should correspond to the order created for STRATUM_1 in the data processing step (see the CSPro manual for more information on this step). Note that the column is labelled “Region”, but this can correspond to any administrative unit such as zone, district, etc. However, please do not change the word “Region” as it is linked to macros within the chartbook that enable other functionalities. When the chartbook analysis is complete and finalized, you can find/replace to change “Region” to a label of your choosing. Once this is done, the chartbook will no longer have full functionality, so ensure that you take this step only after the results are final.
In column H, Health facility types, enter the labels for each facility type in your survey. These should correspond to the order created for STRATUM_2 in the data processing step (see the CSPro manual for more information on this step).

In column L, Managing authority, enter the labels for each managing authority in your survey. These should correspond to the order created for STRATUM_3 in the data processing step (see the CSPro manual for more information on this step).

In column P, Urban/rural, enter the labels for each urban/rural category in your survey. These should correspond to the order created for STRATUM_4 in the data processing step (see the CSPro manual for more information on this step).

**STEP 6. Enter information on the survey type tab**

Go to the DV_Electronic_HMIS chartbook and click on the orange tab called **Survey type**.

First click on the button 1- Check data tabs. This automatically extends or retracts the formulas in the **HMIS data**, **Indicators**, and **Weighted Data** tabs to match the number of rows of data the in the **Facility Survey Data** tab. The **Indicators** and **Weighted Data** tabs will automatically update based on the information entered on the Facility **Survey Data** and **HMIS data** tabs. The **Indicators** tab is where the raw data is used to calculate the data verification ratio indicators. The **Weighted Data** tab is where the indicators are multiplied by the weights. The analytical outputs are based on the weighted data in this worksheet.

Then click on the button to select “Hide empty rows”. This will hide the extra rows that will not be used. It may take up to 5 minutes for Excel to run this command. Please wait for Excel to finish the task. If you do not change the Labels tab again, you should not need to run this step even if the data are updated. However, if you change the labels you will need to click “Show all rows” and then “Hide empty rows” to update the chartbook.

Next, specify if the survey is to have national-level results or subnational results. In the display level, please select either National or Subnational according to the sampling methodology chosen for the survey. The national option activates sheets that have national estimates as well as stratification by facility type, managing authority and urban/rural. The subnational option activates sheets that include the same information as the national option as well as stratification by region. Note that if you are interested in subnational information by region, you will need a chartbook for each region.

The chartbook should now be ready for review. Click on each tab in the chartbook to review all tables and charts.
Deriving the sample weights

Data verification estimates based on the sample of health facilities must be weighted to adjust for discrepancies between the sample and the sample frame in the distribution of the number of health interventions of interest (e.g. births attended by skilled health personnel). If the sample is stratified, the stratum-specific estimates of data accuracy should be weighted. In general, the weights for each stratum for a given indicator are computed as the number of events in the stratum in the population divided by the number of events in the stratum in the sample. Since the number of events measured for the sample and the number measured in the population (i.e. in the HMIS) will be different for each indicator reviewed, the weighting of the estimates needs to be conducted separately for each indicator.

This is a form of post-stratification weighting. For instance, consider the setting where not all facilities in the sample provided immunization services and, among those who provided the service, not all are currently reporting or have provided a monthly report to the HMIS. In this situation, two corrections are necessary – 1) for non-coverage and 2) for non-response – which affect the overall national estimate of each indicator of interest.

Table 13 provides a hypothetical example of Country A, where the total number (N) of facilities is N = 900 distributed across four strata (facility types), where in each stratum a sample of about 35% was drawn for national representation. Column C displays a varying count of facilities providing the vaccination services across strata; among those, Column D gives the count of facilities for which both source documents and report are available in “Month X”, respectively. Column F summarizes the sampling weight for each facility by stratum type; and Column G and Column H are the necessary correction factors for non-coverage and non-response, respectively, by stratum. For example, for the stratum “General Hospitals” the correction factor adjustment for non-coverage = 1.12 (i.e. 65/48), and for non-response = 1.208 (i.e. 58/48), respectively. It is important to note that, in cases of both non-coverage and non-response, the information missing or unmeasurable is assumed to be randomly missing and non-informative missing.

In some settings, it might be more representative to adjust national estimates by service outputs (i.e. where outputs are typically higher in some stratum types than in others (e.g. hospitals versus health centres). This is a form of analytical weighting.

NOTE:
During the data verification exercise, the DQR coordinating group may encounter a situation whereby, for certain metrics or indicators, the service in question is available only in a subset of facilities within the sample – e.g. tuberculosis services. In this situation, the expected service coverage falls below 80% (i.e. the Column F adjustment factor in Table 13 will be greater than 1.20). Another situation might be that fewer than expected of the facilities providing a certain service have responded to the HMIS reporting in Month “X”, resulting in the response rate from facilities falling below 80% (i.e. the Column G adjustment factor in Table 13 will be greater than 1.20). If either or both of these situations occur, the DQR team is advised:
to use the crude verification factor as calculated by the actual numbers recounted and reported (i.e., do not use the weights); and, if required,

to further adjust the crude verification by the analytical weighting using the nationally reported service outputs to the HMIS.

Depending on the type of sampling used to select facilities for the survey component of the DQR, district values may or may not have sampling weights. Currently, the most common method for conducting the facility survey component of the DQR is to do so with another health-facility assessment – such as the SARA. The SARA most commonly uses a stratified sampling method for selecting health facilities where the primary sampling unit, and not the district, is the facility. Consequently, the district estimates presented are unweighted.

If a two-stage cluster sampling method is employed to selected health facilities, the cluster-specific (usually districts) verification factor is weighted on the volume of service in the cluster. An adjustment factor is applied to each cluster – i.e., the ratio of the district value found in the district office and the value for the district found at national level. A weighted average of the adjusted cluster-specific verification factors is then calculated to obtain the national-level estimate of accuracy on the basis of the sample.

Supplementary Word and Excel documents are available to facilitate calculation of facility DV/SA survey weights using the example in Table 13 below under the data analysis section at:


Regarding the sampling approach used for the district DV/SA (census or sample of district offices), the district DV/SA results are unweighted.

Table 13. Tabular summary of a representative sample survey of facilities (n = 310)

<table>
<thead>
<tr>
<th>Stratum</th>
<th>Facilities in the country (A)</th>
<th>Facilities in the survey sample (B)</th>
<th>Facilities in the sample providing immunization services (C)</th>
<th>Facilities in the sample providing immunization services &amp; responding to the HMIS (both source and monthly report are available in “Month X”) (D)</th>
<th>Probability of sampling each facility by facility type (E = n/N or B/A)</th>
<th>Sampling weight of each facility by facility type (F = 1/E)</th>
<th>Non-coverage weight of each facility by facility type (G = C/B)</th>
<th>Non-response weight of each facility by facility type (H = C/D)</th>
<th>Total weigh I = (G<em>H</em>I)</th>
</tr>
</thead>
<tbody>
<tr>
<td>General hospitals</td>
<td>185</td>
<td>65</td>
<td>58</td>
<td>48</td>
<td>0.351</td>
<td>2.846</td>
<td>1.121</td>
<td>1.208</td>
<td>3.854</td>
</tr>
<tr>
<td>Reference health centres</td>
<td>175</td>
<td>65</td>
<td>56</td>
<td>52</td>
<td>0.371</td>
<td>2.692</td>
<td>1.161</td>
<td>1.077</td>
<td>3.365</td>
</tr>
<tr>
<td>Health centres</td>
<td>400</td>
<td>130</td>
<td>120</td>
<td>100</td>
<td>0.325</td>
<td>3.077</td>
<td>1.083</td>
<td>1.200</td>
<td>4.000</td>
</tr>
<tr>
<td>Health posts</td>
<td>140</td>
<td>50</td>
<td>50</td>
<td>45</td>
<td>0.357</td>
<td>2.800</td>
<td>1.000</td>
<td>1.111</td>
<td>3.111</td>
</tr>
<tr>
<td>Total</td>
<td>900</td>
<td>310</td>
<td>284</td>
<td>245</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

N = 900; n = 310
Chapter 5. Dissemination and use of DQR results

Data validation workshop

After the data are cleaned and analysed using the Excel chartbooks, a data validation workshop should be conducted with health programme and data managers to review the results and interpret the findings. This workshop is critical to determining whether the results are plausible and within the range of expectations. Health programme managers have detailed knowledge of service delivery patterns for the specific health programmes and are the best placed to determine plausibility. They can also determine the most noteworthy assessment results to be highlighted in written reports. Data managers can help uncover data quality problems if necessary.

Results should be projected at the workshop for review and discussion of the findings. Open and honest discussion of the results among health-sector stakeholders will improve the quality and acceptability of the results. From the workshop participants, a smaller group can be identified to draft the final report. See Table 14 for a sample agenda of a data validation workshop.

Objectives of the Data Validation Workshop

1. Review DQR findings with programme area experts and determine the plausibility of results.
2. Consider interpretation of results and identification of key findings.
3. Discuss and identify root causes of data quality problems.
4. Formulate recommendations to address data quality problems.
5. Begin action planning for system strengthening – begin work on a data quality improvement plan.
6. Finalize data analysis and presentation of data.
## Table 14. Template: Agenda for data analysis and verification workshop

<table>
<thead>
<tr>
<th>Time</th>
<th>Day 1</th>
<th>Day 2</th>
<th>Day 3</th>
<th>Day 4</th>
<th>Day 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>8:30 – 9:00</td>
<td>Welcome, Workshop objectives and expected outputs</td>
<td>Expected outputs (overview and summary report)</td>
<td>Report writing by indicator (Group work)</td>
<td>Report back from Group work</td>
<td>Presentation of Data Quality Improvement Plan</td>
</tr>
<tr>
<td>9:00 – 10:30</td>
<td>Field survey and data collection&lt;br&gt;- Data entry&lt;br&gt;- Response rate&lt;br&gt;- Lessons learned from the field experience: strengths and weaknesses</td>
<td>Data quality metrics – results and analysis</td>
<td>Report writing by indicator (Group work)&lt;br&gt;- Production of tables and graphics&lt;br&gt;- Draft of narrative</td>
<td>Session on cross-cutting data quality challenges&lt;br&gt;- Discussion on addressing data quality issues that affect all programme areas – interventions to address cross-cutting issues</td>
<td>Synthesis and next steps</td>
</tr>
<tr>
<td>10:30 – 10:45</td>
<td>Break</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10:45 – 13:00</td>
<td>Overview of DQR data processing and analysis&lt;br&gt;- Steps in data processing&lt;br&gt;- Data cleaning&lt;br&gt;- Validation by field supervisors</td>
<td>Data validation by indicator (Group work)&lt;br&gt;- Review of results&lt;br&gt;- Discussion of plausibility&lt;br&gt;- Stakeholder buy-in and intervention planning</td>
<td>Report writing by indicator (Group work – continued)&lt;br&gt;- Production of tables and graphics&lt;br&gt;- Draft of narrative</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13:00 – 14:00</td>
<td>Lunch</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14:00 – 16:00</td>
<td>DQR indicator calculation&lt;br&gt;- Adaptation of the « batch edit » in CSPro (DQR specificities)&lt;br&gt;- DQR indicators calculation&lt;br&gt;- Demonstration and practice</td>
<td>Data validation by indicator (Group work – continued)&lt;br&gt;- Review of results / findings&lt;br&gt;- Discussion of plausibility&lt;br&gt;- Stakeholder buy-in and intervention planning</td>
<td>Report writing by indicator (Group work – continued)&lt;br&gt;- Production of tables and graphics&lt;br&gt;- Draft of narrative</td>
<td>Drafting the Data Quality Improvement Plan&lt;br&gt;- Issues and interventions&lt;br&gt;- Budget&lt;br&gt;- Stakeholders&lt;br&gt;- Mechanism of intervention&lt;br&gt;- Timeline</td>
<td></td>
</tr>
<tr>
<td>16:00 – 16:15</td>
<td>Break</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16:15 – 18:00</td>
<td>Use of the Excel tool for automated production of « standard » DQR tables and graphs&lt;br&gt;- Demonstration and practice</td>
<td>Use of the Excel tool for automated production of « standard » DQR tables and graphs&lt;br&gt;- Demonstration and practice&lt;br&gt;- Report back from Group Work&lt;br&gt;- Presentation of findings from data quality assessment and proposed interventions</td>
<td>Report writing by indicator (Group work – continued)&lt;br&gt;- Production of tables and graphics&lt;br&gt;- Draft of narrative</td>
<td>Drafting the Data Quality Plan (continued)</td>
<td></td>
</tr>
</tbody>
</table>
Final report

The validated results of the DQR should be written up in narrative form as a report, with graphics depicting results to support the narrative. Graphics can be cut and pasted from the DQR Excel chartbooks. Key survey findings should be included, as should recommendations for interventions to address shortcomings in data quality. The report should be disseminated to all staff expected to participate in health-sector planning initiatives (e.g. the health-sector review) several weeks prior to the planning event. Other stakeholders – such as donors, technical assistance organizations, relevant national and international NGOs, private-sector bodies (e.g. universities, civil society organizations), and concerned ministries – should receive copies of the report.

The report should contain the following sections:

- Overview – to place the assessment and findings in the proper context for the reader.
- Methods – to describe how the assessment was designed, especially departures taken from the standardized methodology.
- Results – what was found on the DQR health facility survey? This should include:
  - completeness and timeliness of reporting;
  - verification factors for tracer indicators;
  - distribution of discrepancies among health facilities;
  - reasons for discrepancies;
  - reasons for missing source documents and reports;
  - system assessment findings.
- Discussion – to let the reader know why the highlighted results are important.
- Recommendations – to let the reader know what possible remedies can be applied to rectify data quality problems and to facilitate the drafting of the Data Quality Improvement Plan.

Outline for Data Quality Review final report

1. Introduction – what are the goals and objectives of the assessment?
2. Background – to place the assessment and findings in the proper context for the reader and to relate what has happened before.
3. Methods – describes how the assessment was conducted, and especially any departures from the standard methodology:
   3.1 Indicator selection
   3.2 Master facility List
   3.3 Sampling
      3.3.1. Weighting of indicators
   3.4 Data collection
3.5 Data validation and analysis
3.6 Quality assurance.

4. Results – what was found by the DQR health facility survey? This should include:
   4.1 Completeness and timeliness of reporting
   4.2 Verification factors for tracer indicators
   4.3 Distribution of discrepancies among health facilities
   4.4 Reasons for discrepancies
   4.5 Reasons for missing source documents and reports
   4.6 System assessment findings.

5. Discussion – to let the reader know why highlighted results are important:
   5.1 Principal findings and what they mean
   5.2 Unexpected results
   5.3 Challenges encountered
   5.4 Limitations to the survey results (if any)

6. Recommendations – to let the reader know what possible remedies can be applied to rectify data quality problems.

7. Annex of data tables
   7.1 Survey estimates by indicator
   7.2 Other results.

Develop a Data Quality Improvement Plan¹

Purpose

The purpose of the Data Quality Improvement Plan is to outline the steps and inputs required to address the causes of data quality problems found in the DQR. Needs should be identified and prioritized, and interventions developed and costed to address those needs. A mechanism for monitoring and coordination should be identified or created to ensure that interventions are implemented in good time and within the allotted budget. The goal of the plan is to improve both the quality of the data and the performance of the routine health information system. To meet this goal the plan should provide specific and practical actions that, when implemented, will improve the quality of the data.

Best practices for developing a Data Quality Improvement Plan include:

- The development and implementation of the Data Quality Improvement Plan should be led by the Ministry of Health (or other government ministry responsible for the management and upkeep of the health information system).

---

The improvement plan should be developed in collaboration with important stakeholders, such as donors, partners and NGOs, to ensure consensus and stakeholder buy-in.

The activities and interventions in the improvement plan should be relevant to the country context and should address the priority needs of the country or organization, including its subunits.

The interventions should build on what already exists, should be feasible, and should be appropriate for the context of the health information system and health system workforce.

The activities and interventions should promote and facilitate the sustainability of the information system, so that the system can satisfy the information needs of the present and evolve as those needs change.

Data Quality Improvement Plan process

Engage stakeholders

To ensure optimum development and implementation of the Data Quality Improvement Plan, important stakeholders should be invited to participate. Stakeholders who are part of the development process will invest in the success of the plan and can help ensure continued support and buy-in. The interests, requirements and priorities of stakeholders should be understood, as should their capacity to commit resources to ensure success. There are likely to be many stakeholders, though not all will need to be involved. Sometimes having too many stakeholders can inhibit the development of a responsive plan. Know your stakeholders and choose them strategically – i.e. those that give the plan the best chances for success. Stakeholders can help advocate for changes that are necessary and can mobilize resources to assist with implementation.

A stakeholder engagement matrix can help identify organizations and individuals who have a stake in the improvement of the information system.

A stakeholder engagement matrix (Table 15) can help you identify the organizations, people and groups who are the stakeholders in a data quality improvement process, whether as contributors, influencers or beneficiaries. The matrix is a structured way of defining the roles that stakeholders play in the activity and of assessing the resources they could bring to bear. The matrix also provides a framework for assessing the stakeholders’ interests, knowledge, positions, alliances, resources, power and importance. Who will resist the initiative? Who will support it? What are their reasons? The matrix helps you to assess which stakeholders to include in the process by determining their relative importance – i.e. which stakeholders have the highest priority for the success of the plan?

The identification and engagement of relevant stakeholders contributes to the development of an improvement plan that meets everybody’s expectations and needs.
### Table 15. Stakeholder engagement matrix

<table>
<thead>
<tr>
<th>Name of stakeholder organization, group or individual</th>
<th>Stakeholder description</th>
<th>Potential role in the issue or activity</th>
<th>Level of knowledge of the issue</th>
<th>Level of commitment</th>
<th>Available resources</th>
<th>Constraints</th>
<th>Engagement strategy</th>
<th>Follow-up strategy</th>
</tr>
</thead>
<tbody>
<tr>
<td>National, regional or local?</td>
<td>Primary purpose, affiliation, funding</td>
<td>Vested interest in the activity</td>
<td>Specific areas of expertise</td>
<td>Support or oppose the activity, to what extent, and why?</td>
<td>Staff, volunteers, money, technology, information, influence</td>
<td>Limitations: need funds to participate, lack of personnel, political or other barriers</td>
<td>How will you engage this stakeholder in the activity?</td>
<td>Plans for feedback or continued involvement</td>
</tr>
<tr>
<td>Government sector</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Political sector</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Commercial sector</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nongovernmental sector</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other civil society</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Donors and partners</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Review of DQR results

A formal review of the DQR’s results is a good way to understand and prioritize the data quality problems that the assessment has identified, discuss potential solutions, prioritize recommendations and prepare a realistic action plan. To encourage and promote ownership of the assessment results, it is recommended to begin by conducting an internal review with the Ministry of Health, followed by a review in a workshop setting with a broader group of participants. In this phase of reviewing and analysing the assessment’s results, it is important to ensure that the participants from the Ministry of Health and other stakeholder organizations have the ability to analyse the DQR findings, are knowledgeable about the country context and the country’s HMIS and, therefore, have the ability to recommend appropriate actions to improve data quality. Give the assessment report, charts, graphs and other reading materials to the workshop participants in advance so that they can prepare. Also identify facilitators who have the skills to keep participants focused and on track to achieve the workshop’s expected outcomes.

This review workshop can be combined with the action planning phase. If you decide to make action planning part of the workshop, be sure that the participants have the authority to make decisions. Alternatively, the first part of the workshop could be for health information system or health programme experts to review and validate the quality and relevance of the assessment results and prepare summaries and presentations for the decision-makers. In the second part of the workshop, relevant decision-makers can join the health information system experts, can be briefed on the assessment results and the recommendations, and can then contribute to identifying actions and interventions to address the findings, the definition of the timelines, the responsible persons and organizations, and the required resources.

For the effective review and formulation of recommendations, it is best to conduct discussions in small groups. The groups should have equal representation from the following categories of participant:

- decision-makers and other users;
- health programme managers;
- data managers and M&E specialists; and
- health-care providers.

The composition of and tasks assigned to each group may be as follows:

- by health programme area (e.g. maternal health, child health, HIV/AIDS, TB, malaria, etc);
- by level of the health system (e.g. national, regional, district, health facility/community).
Results to review

- Output from the DV/SA:
  - accuracy by indicator;
  - timeliness and completeness;
  - system assessment.

- Desk review of data quality:
  - completeness;
  - internal consistency
    - outliers
    - trends over time
    - between related indicators;
  - external consistency
    - comparisons with population-based survey data
    - comparisons with alternate data sets (e.g., programme-specific databases);
  - review of population data
    - comparisons between official government statistics and alternative sources (e.g., United Nations, country health programmes).

A plenary session should follow the group discussions to enable all participants to provide feedback and input on all groups’ ideas and proposals, and to learn from one another.

If the review of the DQR assessment results and the formulation of recommendations are conducted separately from the action planning session, the results and recommendations should be disseminated to the relevant decision-makers to guide them in identifying the appropriate strategies and actions for strengthening the RHIS.

**Action planning**

The planning process for data quality improvement follows the review and discussion of the DQR results and recommendations, and the identification and prioritization of strategies to achieve an improved quality of data for the HMIS.

The planning process also requires good facilitation to develop an action plan that describes specific, measurable, achievable, relevant and time-bound (SMART) objectives and activities, and with responsibility for the implementation of each activity assigned to a specific person or organization.

**Prioritizing data quality improvement interventions**

When formulating recommendations and developing the action plan, it is important to prioritize activities that will lead to the greatest improvement in data quality with available resources – or for which resources can be mobilized. The sustainability of the interventions should also be considered when identifying priority activities for improving data quality.
Participants in the action planning session can use the prioritization matrix (Table 16) to score the proposed activities, based on their expected impact on data quality and on the ability of the organization and stakeholders to implement the activities. The scores help prioritize the interventions that are the most feasible and likely to yield the greatest results.

The prioritization exercise is conducted through a consensus process. Participants in the action planning session agree on the level of impact that each recommended intervention will have and the ability of the stakeholders to implement it, taking into account the available resources (human, financial, technical, communication etc.). Participants can work in small groups to discuss and complete the matrix and can then come together in a plenary session to produce one completed and mutually agreed matrix.

The prioritization matrix is arranged with a scale for impact on the vertical axis and a scale for ability to implement with the required level of investment (human and financial resources, effort and time) on the horizontal axis. Each axis is divided into four scores: 1 represents the lowest score for the attribute and 4 represents the highest. The interventions with the most impact that are the easiest to implement and that require minimal investment are put in the top right cells of the matrix, and the interventions with the least impact and that are least feasible (i.e. those that require a high level of human or financial resources or efforts) are put in the lowest cells of the matrix on the left.

**Table 16. Intervention prioritization matrix**

<table>
<thead>
<tr>
<th>High impact</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low impact</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Difficult to implement (low ability)**

**Easy (high ability)**
Depending on the context, the use of this matrix helps distinguish relevant interventions that are easy or relatively easy to implement and that produce moderate-to-high impact from those that are less feasible or yield low impact.

Once an intervention has been determined, it should be broken down into well-defined activities so that the person or organization responsible for implementation and funding can be assigned. Table 17 provides an example of how to break down the main intervention into sub activities that result in data quality improvement.

**Scheduling and budgeting activities**

The purpose of scheduling and budgeting is to elaborate the overall Data Quality Improvement Plan, thereby providing a roadmap for the activities under each recommended intervention. Understanding the required work efforts for the implementation of each recommended intervention allows participants in action planning to break activities down and estimate the resources and time required for implementation accurately. Aligning the activities with the resources they require makes it possible to estimate the costs of data quality improvement efforts and to determine the requirements in time and the timetable for implementation.

<table>
<thead>
<tr>
<th>Priority actions</th>
<th>Objective</th>
<th>Activities</th>
<th>Short-term</th>
<th>Medium-term</th>
<th>Long-term</th>
<th>Responsible entity</th>
<th>Supporting partner</th>
<th>Budget</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Intervention 1</strong></td>
<td><strong>Action 1</strong></td>
<td>Obj 1</td>
<td>Activity 1</td>
<td>✘</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Activity 2</td>
<td>✘</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Activity 3</td>
<td>✘</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Action 2</strong></td>
<td>Obj 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>✘</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✘</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| Table 17. Example table for scheduling and budgeting activities
Monitor and follow up

M&E helps measure performance and assess the impact of different strategies, interventions and inputs on the data quality improvement efforts. The results of M&E contribute to the learning experience and help decision-makers to improve the interventions.

The action plan itself serves as a monitoring tool to follow up the implementation of the interventions and activities it comprises. Moreover, the use of DQR routine data quality supervisory checklists can help track progress in improving data quality. For evaluation purposes, the next implementation of the DQR health facility assessment (DV/SA) can be used to measure the success of interventions. Regular review of action plan implementation and monitoring of findings helps stakeholders to identify any mid-course corrections that may be necessary.

Below are some common data quality problems and potential solutions.

Example data quality problems and potential solutions

In order to address data quality problems accurately, the nature of the problems should be understood as far as is possible. Knowing the causes of data quality problems means that solutions can be more accurately targeted, making data quality improvement easier and more effective.

Data quality problems should be defined as clearly as possible. The solution may not require a large intervention or costly inputs. Sometimes the solution requires only a visit to a health facility by a supervisor who will work with those managing the data to refresh their skills. The first step is to define the scope of the problem – i.e. how pervasive is it? Is it occurring in all, or most, facilities? Or is it limited to just a few facilities? The issue is: is the problem limited or is it systematic? If it is limited, the interventions can be targeted at just those facilities experiencing the problem. If it is systematic, however, a broader approach is required.

The nature of the data quality problem must be identified. Is the problem a result of standard practice at the facility – i.e. does the error result from the facility’s usual data management practices and therefore occurs repeatedly? Or does the error result from a departure from standard practice and only happened once, or occasionally? Again, the solution will differ depending on the nature of the problem found and will depend on an accurate description of the causes of the problem.

Solutions to data quality problems will depend on the resources available and the political will to intervene. If the causes of data quality problems are well defined, it will be easier to make the case for resources to address the problems adequately. Problems should be fully described in writing, scoped and budgeted, and then presented to stakeholders (e.g. the HMIS technical...
working group, or inter-agency coordinating committee) at the appropriate time (i.e. when budgetary priorities are being discussed). The Data Quality Improvement Plan is a mechanism by which data quality problems, and their potential solutions, can be presented to policy-makers.

Accuracy of reporting

The accuracy of reporting is the extent to which the results reported by the facility represent the true level of service delivery for a selected period and indicator. The true level of service delivery is represented by the data in source documents (e.g. register). While there may be errors in recording service delivery in the source document, this is the best record we have of that service delivery. For most data quality assessments we must make the assumption that the data in the source documents are sufficiently accurate for our purposes.

If there are suspicions of problems with the data in source documents, methods exist to determine how accurately they reflect actual service delivery. For instance, we can observe how facility staff record the service delivery in the source document. Is it recorded contemporaneously (i.e. at or around the same time)? Is it done by a trained staff member? Are the staff using standard data collection tools?

The value of the indicator for a selected reporting period – generally a month, or several months if the reporting is monthly – as reported by the facility to the next level (in the monthly HMIS or programme report) is compared to a validated value. The validated value is the value of the indicator for the selected period recalculated by the assessment team from the source documents. We calculate a verification factor (VF) to represent the reporting accuracy for the facility, which is the ratio of the recounted value to the reported value. The VF can range from zero to infinity but a perfect congruence between the source documents and reports yields a value of 1. Values greater than 1 represent under-reporting (i.e. more service delivery was found in the source documents than was reported by the facility) while values less than 1 indicate over-reporting (i.e. less service delivery was found in the source documents than was reported by the facility).

The most common causes of poor accuracy of reporting (typically a VF more than 10% different from 1 – i.e. VF ≤ 0.9 or VF ≥ 1.1 – although the standard may depend on the agency commissioning the DQR) are missing source documents and/or reports, incorrect compilation of data, inadequate tools for data collection, and calculation errors. There are more precise causes within these general categories so it is important to fully understand data quality problems in order to target solutions effectively.

Missing source documents or reports

Missing documentation is one of the primary concerns when evaluating data quality problems. The effects of missing documents usually depend on which document is missing. A missing source document tends to result in a VF < 1 (over-reporting) since less service delivery could be verified than is reflected in the monthly reports. Conversely, missing monthly reports tend
to result in a $\text{VF} > 1$ (under-reporting) since more service delivery is recorded in the source documents than is evident in the available monthly reports.

**Potential causes:**

- **Non-adherence to data storage protocols** of the national programme or HMIS. Either some staff members do not know the protocol or do not understand the importance of maintaining a data archive. The national programme or HMIS probably has a protocol which states how long filed source documents should be kept by the facility (e.g. 5 years).

- **Inadequate storage space.** Does the facility have a space for a data archive? Ideally, the facility should have a room, or a corner of a room, which is secure -- i.e. only those who need access have access – as well as being clean and dry.

**Potential solutions:**

- **Non-adherence:**
  - If the non-adherence is limited in scope or is non-standard practice, this may be addressed during supervisory visits.
  - If non-adherence is systematic or is standard practice, a memo could be sent to all facilities to remind them of their data archiving responsibilities. If the failing persists, refresher training could be required for data management staff. Or there could be more severe penalties

- **Inadequate space:**
  - Institute a programme for filing and archiving, including designating a room (clean, dry and secure) in the facility and purchasing filing cabinets and/or shelving units, file folders, etc.

**Misunderstanding of indicator definitions**

Problems with reporting accuracy can arise when data managers (or those compiling the monthly report at the facility) do not understand, or misunderstand, what to count. For example, how is the indicator supposed to be aggregated, or disaggregated? What constitutes a client served, or a service delivered?

**Potential causes:**

- **Inadequate training.** The training may have been *substandard* or it took place too long ago. Was the current data compilation protocol for the relevant indicator included in the training?

- **Inadequate documentation of reporting protocols at sites.** Is there a document that describes the indicator compilation process for each indicator (e.g. a job aide, or HMIS manual) and is there a printed copy accessible at the site?

- **Staff turnover.** The trained staff member may have left the job for another one and has been replaced by someone who has not had the benefit of training.
Workload. There may be no staff member designated to do the data compilation and the task is left to whoever is available, regardless of whether or not they have the appropriate training.

Potential solutions:

- If the problem is limited in scope it can be addressed by supervisory visits from the district to bolster the skills of the staff involved.
- If the problem is more systematic it may require refresher training on protocols for aggregating monthly results by indicator at facilities.
- If there are job aides describing how to compile the indicators, these can be distributed to facility staff. If there are no such job aides, they should be developed and distributed to facility staff.
- All staff members who compile monthly data should have had the appropriate training. A training database should be maintained to so that the training needs of all personnel can be tracked. A training database can show which staff have had which training and when, and which staff have not or are overdue for training.
- A staff member should be designated as the data manager. A stand-in should be trained to take over in the event that the designated staff person is not available.
- A poster can be put on the wall at the facility as a reminder.

Inadequate data collection tools

Up-to-date, well-designed, and always-available data collection tools are critical for good quality reporting.

Potential causes:

- Indicators have changed since the last time the tools were printed and distributed and the data collection tools no longer meet the needs of reporting.
- There are insufficient copies of blank data collection tools available at the facility so staff use improvised forms.
- If a coding scheme is used, the codes are not clear and concise, or not used consistently.
- There is insufficient space on the tools to enter all the required information.

Potential solutions:

- Ensure an adequate supply of blank data collection tools. If reproduction and distribution of the tools is more expensive than the HMIS or programme can bear, donor partners should be approached to obtain commitments for supporting reproduction and distribution of essential forms, reports and registers.
Conduct a review of the data collection tools to ensure that they still meet the needs of reporting for the current list of priority indicators.

Redesign, reprint and distribute new data collection tools.

Avoid buying too many blank copies of the tools. Although it may be cheaper to print several years’ worth of tools at the same time, this makes the system less adaptable to changes in indicators.

Supervisors should verify (i.e. take inventory) the availability of blank forms/tools at the sites they visit.

Ensure the process for ordering new forms/tools is well known and transparent. This should be written up and distributed to all health facilities.

Investigate the possibility of electronic data collection tools and the transfer of aggregate results.

**Calculation errors/recording errors**

Errors are inevitable but should be kept to a minimum. If the calculation errors are random, and they do not systematically inflate or reduce indicator values, they may go unnoticed. Even large errors are often masked in the aggregate data at district level and higher. It is better to prevent the errors from getting into the system than to try and find them in the aggregate data afterwards.

**Potential causes:**

- Key punch errors (for computerized systems at facility level) or transcription errors for paper-based systems. Key punch errors (typographical errors) and/or writing the wrong number, or writing an illegible number that is misinterpreted by a data entry clerk somewhere else, are bound to happen. Humans are error-prone.

- Arithmetical errors may occur – especially for indicators with a large volume of service (e.g. immunization in large facilities) which require aggregation of data across daily or weekly summary forms or tally sheets.

**Potential solutions:**

- For key punch errors, controls can be introduced into the data entry software to restrict entries that are implausible or impossible. For instance, a computer could prevent you from entering service delivery for antenatal care if the client is male, or not of child-bearing age. Some systems can compare the value entered to the mean of values for the previous year and flag those that are more than two or three standard deviations from the mean of values.

- For transcription errors, the best way to prevent these is for a designated person (ideally a supervisor) to conduct data verification on the form prior to submission to the next level. Each value should be checked for plausibility (e.g. by assessing whether the value seems
likely given the size of the facility catchment area and the values reported for that indicator by that facility in the past).

- If such as system is not in place, a data quality checking protocol can be introduced and facility staff trained to implement it. The DQR includes health facility data quality checklists for this purpose.

- At district level and higher, analyses such as those proposed for the Desk Review of data quality are a good way to identify extreme values in the data set. Analyses to identify outliers, anomalous trends and implausible relationships between related indicators are all ways to identify data that have been entered in error.

- The best way to identify calculation errors is by conducting data quality checks on the forms before submission to the next level. Also, a calculator can be purchased at minimal cost for the data manager.

**Missing or incomplete data**

“Completeness of reporting” measures the extent to which all health facilities that are expected to report actually do report on a monthly basis. “Completeness of indicator data” measures the extent to which indicator values are included on the reporting form from facilities that are expected to report on a particular indicator. Missing reports and data cause gaps in the understanding of the true levels of service delivery and hinder the ability to make informed decisions based on evidence. The less complete the data, the less useful it is for planning, monitoring and evaluation.

**Potential causes**

- *Non-reporting by health facilities.* Some facilities do not report when they are supposed to, for a variety of reasons such as staff absences, the lack of means to transmit the report (no Internet connection, no fuel for the car, etc.), or withholding of data to extract concessions during employment disputes.

- *Late reporting (lack of timeliness in reporting).* Timeliness is a form of completeness. If the report is late it is not available when needed for decision-making.

- *Values for certain indicators not included.* The causes of missing values are many and are similar to the reasons for missing reports. It may be the data were not compiled in time from source documents, or the source documents were missing. It is important, however, to be able to distinguish missing values from a valid report of zero service delivery. For some service areas, it is possible to have no services provided during a given reporting period. For instance, there could be a stock-out of vaccines making immunization impossible before the facility is resupplied. Many data managers are taught to include zero values when there is no service delivery so that no one at higher levels will misinterpret a missing value for a zero value. Increasingly, however, computerized information systems tend not to store zero values since they use a lot of space in the database.
**Potential solutions**

- The best way to address incomplete reporting is to try to avoid it. Legislation should be in place to compel public and private facilities alike to report routinely and on time. Facilities that do not report should be contacted immediately by the established mechanism (telephone, email, or other communication platform) to determine why the report is missing and to encourage the facility to send the report.

- Most often, a lack of data does not mean there was no service delivery. The report was just not compiled and sent, which means that subsequent decision-making and policy formulation at national level will not have that evidence available at the required time. One way to improve incomplete data is to adjust the completeness of the data based on the extent of missing data. For instance:
  - How much service is provided by the non-reporting facilities? None? Some? About half as much as facilities that do **report**? The same as the reporting facilities? More than the reporting facilities?
  - Adjustment depends on assumptions about the number of service outputs (pregnancy care, vaccinations, etc.) provided at non-reporting facilities compared to those that reported.
  - The adjustment can be expressed as follows: $N_{\text{adjusted}} = N_{\text{reported}} + N_{\text{reported}} \times (1/(c)-1) \times k$, where $N=$number of service outputs, $c=$reporting completeness, and $k=$adjustment factor.
  - An example would be 1000 DPT1 vaccinations reported, but 80% completeness. It is assumed that the non-reporting facilities provide services at half the rate of reporting facilities ($k=0.5$): $N_{\text{adjusted}} = 1000 + 1000 \times (1/0.80 - 1) \times 0.5 = 1000 + 1000 \times (1.25-1) \times 0.5 = 1125$.
  - Selecting the best adjustment factor for $k$:
    - 1. $k=0$: no services in non-reporting facilities;
    - 2. $k=0.25$: some services, but much lower than in reporting facilities;
    - 3. $k=0.5$: half the rate compared to reporting facilities;
    - 4. $k=0.75$: almost as much as in reporting facilities;
    - 5. $k=1.0$: the same rate of services as in reporting facilities.
  - Considerations:
    - 1. Stock-outs (e.g. vaccines) are an example of an actual absence of service delivery.
    - 2. It is important to consider the proportion of all services that are delivered by private facilities that may or may not be compelled to report?
    - 3. Large facilities represent a large proportion of all service delivery, and data missing from these facilities will have a large effect on overall completeness. Is it sufficient to adjust values from these facilities alone?
  - While adjustment is likely to provide a more accurate picture of the level and trends of service delivery for priority indicators, it should not be done hastily or without due regard to the integrity of the data set. For instance, do the adjusted values become official values for the health facilities or districts for the periods for which reports
were missing and data were adjusted? What if a report comes in late, and the values are different than those predicted? National and subnational HMIS and programme data managers should agree on when and how to adjust data, and how to deal with adjusted values for the official record.

Example outline of a Data Quality Improvement Plan

1. Introduction
   1.1 Background
   1.2 DQR methodology

2. Results of DQR
   2.1 Accuracy by indicator
   2.2 Timeliness & completeness
   2.3 System Assessment
   2.4 Results of Desk Review of Data Quality
   2.5 Results of routine data quality checks during supervision
   2.6 Review of systematic data quality problems identified through routine supervision

3. Cross-cutting interventions to address cross-cutting data quality problems
   3.1 Activities
   3.2 Responsible agencies and partners
   3.3 Budget
   3.4 Timeline
   3.5 Agency or unit responsible for monitoring & follow-up of implementation

4. Maternal & Child Health interventions to address MCH data quality problems
   4.1 Activities
   4.2 Responsible agencies and partners
   4.3 Budget
   4.4 Timeline
   4.5 Agency or unit responsible for monitoring & follow-up of implementation

5. Immunization programme interventions to address immunization programme data quality problems
   5.1 Activities
   5.2 Responsible agencies and partners
   5.3 Budget
   5.4 Timeline
   5.5 Agency or unit responsible for monitoring & follow-up of implementation
6. HIV/AIDS Programme interventions to address HIV/AIDS programme data quality problems
   6.1 Activities
   6.2 Responsible agencies and partners
   6.3 Budget
   6.4 Timeline
   6.5 Agency or unit responsible for monitoring & follow-up of implementation

7. TB Programme interventions to address TB programme data quality problems
   7.1 Activities
   7.2 Responsible agencies and partners
   7.3 Budget
   7.4 Timeline
   7.5 Agency or unit responsible for monitoring & follow-up of implementation

8. Malaria Programme interventions to address malaria programme data quality problems
   8.1 Activities
   8.2 Responsible agencies and partners
   8.3 Budget
   8.4 Timeline
   8.5 Agency or unit responsible for monitoring & follow-up of implementation

9. Monitoring and oversight of implementation
   9.1 Agency responsible for implementation
   9.2 Plan for monitoring and evaluation

10. Overall programme budget

11. Overall programme timeline

12. Conclusion
# Annex 1: Recommended indicators

## Core indicators

### Recommended DQA indicators

<table>
<thead>
<tr>
<th>Programme area</th>
<th>Indicator name</th>
<th>Full indicator</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maternal health</td>
<td>Antenatal care 1st visit (ANC1) coverage</td>
<td>Number (%) of pregnant women who received antenatal care at least once during their pregnancy</td>
</tr>
<tr>
<td>Immunization</td>
<td>DTP3/Penta3 coverage</td>
<td>Number (%) of children &lt; 1 year receiving three doses of DTP/Penta vaccine</td>
</tr>
<tr>
<td>HIV</td>
<td>Currently on ART</td>
<td>Number and % of people living with HIV who are currently receiving ART</td>
</tr>
<tr>
<td>TB</td>
<td>TB notification rate</td>
<td>Number of new and relapse cases of TB that are notified per 100 000 population</td>
</tr>
<tr>
<td>Malaria</td>
<td>Total confirmed malaria cases¹</td>
<td>Confirmed malaria cases (microscopy or RDT) per 1000 persons per year</td>
</tr>
</tbody>
</table>

Note: ANC = antenatal care; ART = antiretroviral therapy; DTP3 = diphtheria-tetanus-pertussis three-dose vaccine; Penta = pentavalent vaccine; RDT = rapid diagnostic test.

## Additional indicators

### Recommended DQR indicators

<table>
<thead>
<tr>
<th>Programme area</th>
<th>Indicator name</th>
<th>Full indicator</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIV</td>
<td>People living with HIV who have been diagnosed</td>
<td>Number (%) of people living with HIV who have been diagnosed</td>
</tr>
<tr>
<td></td>
<td>HIV care coverage</td>
<td>Number (%) of people living with HIV who are receiving HIV care (including ART)</td>
</tr>
<tr>
<td></td>
<td>PMTCT ART coverage</td>
<td>Number (%) of HIV-positive pregnant women who received ART during pregnancy</td>
</tr>
<tr>
<td></td>
<td>ART retention</td>
<td>Number (%) of people living with HIV and on ART who are retained on ART 12 months after initiation (and after 24, 36, 48, and 60 months)</td>
</tr>
<tr>
<td></td>
<td>Viral suppression</td>
<td>Number (%) of people on ART who have suppressed viral load</td>
</tr>
<tr>
<td>TB</td>
<td>Notified cases of all forms of TB</td>
<td>Number of new and relapse cases of TB that are notified per 100 000 population – Assess if quarterly case notification report blocks 1 and 2 are correct as per standards and benchmarks (B1.4) for paper-based systems²</td>
</tr>
<tr>
<td></td>
<td>TB treatment success rate</td>
<td>Number (%) of TB cases successfully treated (cured plus treatment completed) among TB cases notified to the national health authorities during a specified period – Assess if quarterly treatment outcome report block 1 is correct as per standards and benchmarks (B1.4) for paper-based systems²</td>
</tr>
<tr>
<td></td>
<td>Second-line TB treatment success rate</td>
<td>Number (%) of TB cases successfully treated (cured plus treatment completed) among all confirmed RR-TB/MDR-TB cases started on second-line treatment during the period of assessment</td>
</tr>
</tbody>
</table>


# Annex 1. Recommended indicators

**Additional indicators, continued**

<table>
<thead>
<tr>
<th>Recommended DQR indicators</th>
<th>Programme area</th>
<th>Indicator name</th>
<th>Full indicator</th>
</tr>
</thead>
<tbody>
<tr>
<td>TB-HIV</td>
<td>Proportion of registered new and relapse TB patients with documented HIV status</td>
<td>Number of new and relapse TB patients who had an HIV test result recorded in the TB register, expressed as a percentage of the number registered during the reporting period</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Proportion of HIV-positive new and relapse TB patients on ART during TB treatment</td>
<td>Number of HIV-positive new and relapse TB patients who received ART during TB treatment expressed as a percentage of those registered during the reporting period</td>
<td></td>
</tr>
<tr>
<td>Malaria</td>
<td>Proportion of malaria suspects tested</td>
<td>Number (%) of all suspected malaria cases that received a parasitological test (microscopy or RDT) [=\ Number tested / Number of suspected cases of malaria]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Confirmed malaria cases treated with 1st line treatment course (including ACTs)</td>
<td>Number (%) of confirmed malaria cases treated that received first-line antimalarial treatment according to national policy at public-sector facilities</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Malaria cases (suspected and confirmed) treated with 1st line treatment course (including ACTs)</td>
<td>Number (%) of malaria cases (presumed and confirmed) that received first-line antimalarial treatment</td>
<td></td>
</tr>
<tr>
<td></td>
<td>IPTp3</td>
<td>Number (%) of pregnant women attending antenatal clinics who received three or more doses of intermittent preventive treatment for malaria according to national policy at public-sector facilities</td>
<td></td>
</tr>
<tr>
<td>General</td>
<td>Service utilization</td>
<td>Number of outpatient department visits per person per year</td>
<td></td>
</tr>
<tr>
<td>Maternal health</td>
<td>Antenatal care 4th visit (ANC4)</td>
<td>Number (%) of women aged 15–49 years with a live birth in a given time period who received antenatal care, four times or more</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Institutional delivery coverage</td>
<td>Number (%) of deliveries which took place in a health facility</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Postpartum care coverage</td>
<td>Number (%) of mothers and babies who received postpartum care within two days of childbirth (regardless of place of delivery)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tetanus toxoid 1st dose coverage</td>
<td>Number (%) of pregnant women who received the 1st dose of tetanus toxoid vaccine</td>
<td></td>
</tr>
<tr>
<td>Immunization</td>
<td>DTP1–3/Penta1–3 coverage</td>
<td>Number (%) of children &lt; 1 year receiving 1st dose, 2nd dose, 3rd dose of DTP/Penta vaccines</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MCV1 coverage</td>
<td>Number (%) of infants who have received at least one dose of measles-containing vaccine (MCV) by age 1 year</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PCV 1–3(^\text{rd}) coverage</td>
<td>Number (%) of children &lt; 1 year receiving 1st dose, 2nd dose, 3rd dose of pneumococcal vaccines</td>
<td></td>
</tr>
</tbody>
</table>

Note: ANC = antenatal care; ART = antiretroviral therapy; DTP = diphtheria-tetanus-pertussis; MCV = measles-containing vaccine; MDR-TB = multidrug-resistant tuberculosis; PCV = pneumococcal conjugate vaccine; PMTCT = Prevention of mother-to-child transmission; RR = rifampicin-resistant.