



TOOLKIT FOR ANALYSIS AND USE OF ROUTINE HEALTH FACILITY DATA

General principles

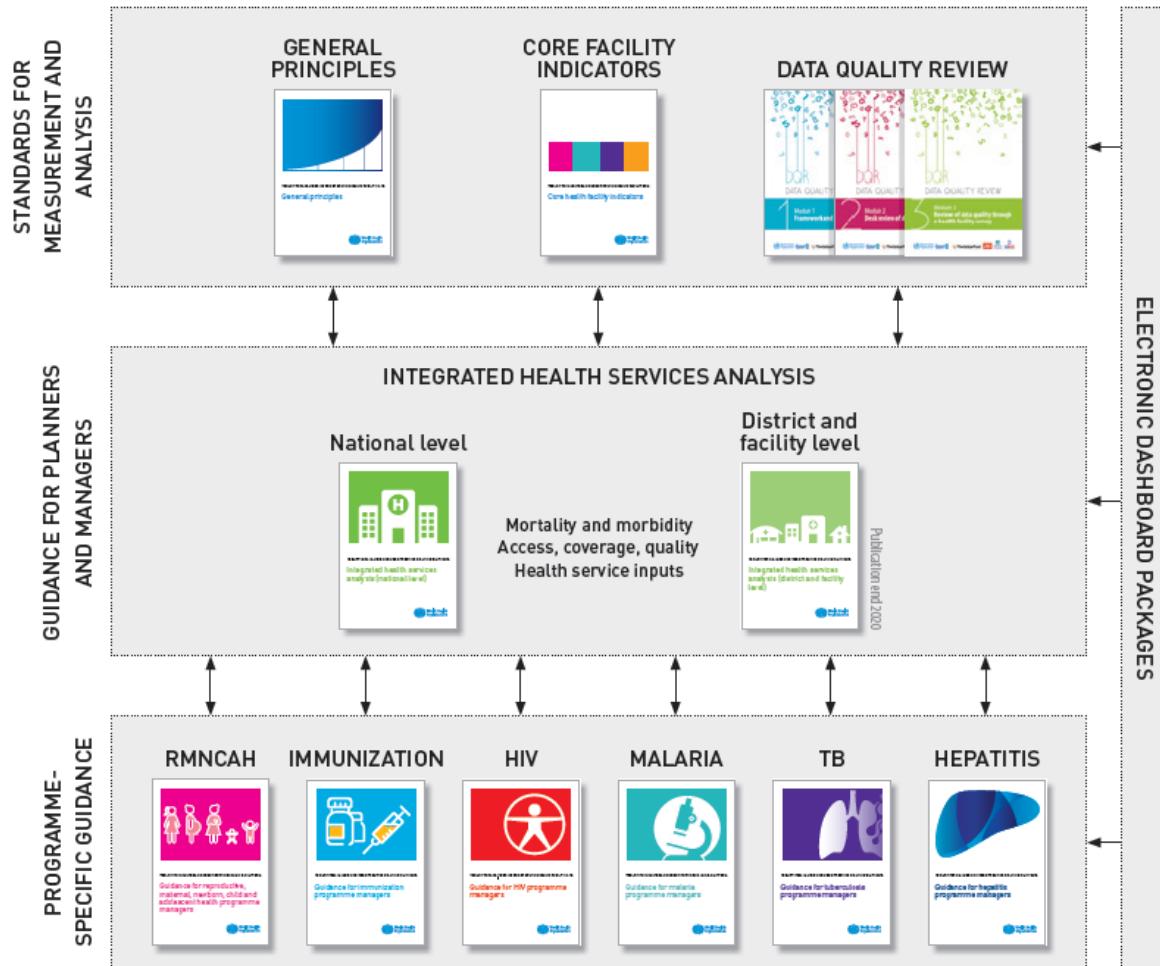
WORKING DOCUMENT June 2021



WHO TOOLKIT FOR ANALYSIS AND USE OF ROUTINE HEALTH FACILITY DATA

This document is part of the WHO *Toolkit for analysis and use of routine health facility data* – a set of capacity-building resources to optimize the analysis and use of data collected from health facilities through routine health information systems (RHIS).

The toolkit is a collaborative effort by multiple WHO technical programmes and partners. It promotes an integrated, standards-based approach to facility data analysis, using a limited set of standardized core indicators with recommended analyses, visualizations and dashboards.



The toolkit consists of a series of modules that can be used individually or together:

- *General principles* introduces key data analysis concepts that are applicable to all modules.
- *Core facility indicators* is a compendium of the indicators from the various modules.
- The *Data quality review (DQR) toolkit* includes guidance and tools for systematic review of the quality of routine facility data.
- *Integrated health services analysis* targets general health service managers, providing a comprehensive, integrated analysis of tracer indicators across multiple health service components and programmes.
- The *programme-specific guidance modules* are customized according to the needs of the programme. Each module contains a guidance document, training materials and an electronic configuration package for automated dashboard production.

The materials within the Toolkit will be periodically updated and expanded.

Further details: <https://www.who.int/data/data-collection-tools/health-service-data/toolkit-for-routine-health-information-system-data/modules>

TOOLKIT FOR ANALYSIS AND USE OF ROUTINE HEALTH FACILITY DATA

General principles

WORKING DOCUMENT June 2021



© World Health Organization 2021

All rights reserved. This is a working document and should not be quoted, reproduced, translated or adapted, in part or in whole, in any form or by any means.

Contents

Acknowledgements	5
Abbreviations	6
Guidance overview and references	7
1. Introduction	9
1.1 Toolkit for analysis and use of routine health facility data	9
1.2 Overview of routine health information systems	10
2. Core indicators	14
2.1 Standards for RHIS indicators.....	14
2.2 Core RHIS indicators.....	14
3. Facility representation in RHIS.....	15
3.1 Representativeness versus completeness.....	15
3.2 Master Facility List (MFL)	15
4. Data quality	16
4.1 Types of data quality assessment	16
4.2 Dimensions of data quality assessment	17
5. Population estimates/denominators	22
5.1 Census-based projections	22
5.2 Other methods.....	22
6. Key analytical concepts	24
6.1 Basic statistical terms	24
6.2 Aggregated versus individual data	24
6.3 Disaggregation.....	25
6.4 Equity analysis.....	25
6.5 Comparisons with other data sources.....	26
7. Presentation and communication	28
7.1 Tables	28
7.2 Charts	29
7.3 Maps.....	31
7.4 Presenting a story.....	32
8. Interpretation and use	35
8.1 Interpretation of RHIS data	35
8.2 Uses of RHIS data	35
Annex 1. Adjustment for incomplete reporting.....	37
Annex 2. Triangulation of data	38
Annex 3. Estimation of population targets.....	39

Acknowledgements

This *General principles* document from the *Toolkit for analysis and use of routine health facility data* has been developed by the World Health Organization (WHO), with the support of grants from Bloomberg Philanthropies Data for Health Initiative; Gavi, the Vaccine Alliance; The Global Fund to Fight AIDS, Tuberculosis and Malaria; and the Norwegian Agency for Development Cooperation.

Key contributors: Ties Boerma (University of Manitoba), Kathryn O'Neill (Universal Health Coverage/Life Course, WHO), Robert Pond (independent consultant), Chelsea Taylor (Universal Health Coverage/Life Course, WHO), Wendy Venter (Data, Analytics and Delivery for Impact, WHO), and Kavitha Viswanathan (independent consultant).

The June 2021 revision of this document was completed by Robert Pond and Wendy Venter, with inputs from Sandro Colombo (independent consultant), Teresa Diaz (Universal Health Coverage/Life Course, WHO), Hazim Timimi (Universal Health Coverage/Communicable and Noncommunicable Diseases, WHO), Elizabeth Katwan (Universal Health Coverage/Life Course, WHO), and Kavitha Viswanathan.

Abbreviations

ABER	annual blood examination rate
ACT	artemisinin-based combination therapy
ANC	antenatal care
ART	antiretroviral therapy
BCG	bacille Calmette–Guérin (vaccine)
CBR	crude birth rate
CRVS	civil registration and vital statistics
DHIS	District Health Information Software
DHS	Demographic and Health Surveys
DQR	data quality review
DTP	diphtheria–tetanus–pertussis (vaccine)
EPI	Expanded Programme on Immunization
FBO	faith-based organization
HIS	health information system
HIV	human immunodeficiency virus
HMIS	health management information system
IPTp	intermittent preventive treatment for malaria during pregnancy
LMIS	logistics management information system
MCV	measles-containing vaccine
MFL	Master Facility List
MICS	Multiple Indicator Cluster Surveys
MR	measles-rubella (vaccine)
NGO	nongovernmental organization
OPV	oral polio vaccine
Penta	pentavalent vaccine
RHIS	routine health information system
RMNCAH	reproductive, maternal, newborn, child and adolescent health
RDT	rapid diagnostic test (malaria)
SD	standard deviation
SOPs	standard operating procedures
TB	tuberculosis
UHC	universal health coverage
USAID	United States Agency for International Development
WHO	World Health Organization

Guidance overview and references

This document introduces the *Toolkit for analysis and use of routine health facility data* and provides an overview of key data analysis principles that are relevant to all the modules of the toolkit.

Learning objectives

The guidance aims to promote understanding of:

- the concept of a standardized core indicator list;
- issues concerning health facility representation in routine health information systems (RHIS);
- key dimensions of data quality assessment;
- challenges related to population estimates and denominators for calculating indicators;
- basic analytical concepts, including disaggregation, equity analysis and comparisons of data from various sources;
- principles for presentation and communication of data;
- basic concepts for data interpretation and use.

Audience

The guidance is relevant for workers in ministries of health and other organizations at various levels of the health system, including:

- decision-makers that use RHIS data for planning, management and review of health services;
- staff responsible for the analysis and presentation of health data, including analysts and monitoring and evaluation officers;
- health information systems (HIS) staff involved in data management and data quality improvement;
- research institutes and academic institutions involved in the analysis of RHIS data and/or efforts to improve data quality;
- trainers and workshop facilitators.

Suggested references

Core health facility indicators. Geneva: World Health Organization; January 2021, (<https://www.who.int/data/data-collection-tools/health-service-data/toolkit-for-routine-health-information-system-data/modules>), accessed 10 May 2021).

Data quality assurance (DQA) toolkit. Geneva: World Health Organization; 2020 (<https://www.who.int/data/data-collection-tools/health-service-data/data-quality-assurance-dqa>, accessed 10 May 2021).

Global reference list of 100 core health indicators. Geneva: World Health Organization; 2018 (<http://www.who.int/healthinfo/indicators/2018/en/>), accessed 10 May 2021).

Health facility and community data toolkit [website]. Geneva: World Health Organization; 2014 (https://www.who.int/healthinfo/facility_information_systems/en/, accessed 10 May 2021).

International Health Partnership + Related Initiatives (IPH+) and WHO. Monitoring, evaluation and review of national health strategies: a country-led platform for information and accountability. Geneva: World Health Organization; 2011 (http://www.who.int/healthinfo/country_monitoring_evaluation/documentation/en/, accessed 10 May 2021).

MEASURE Evaluation. Health information system strengthening: standards and best practices for data sources. Authors: F Greenwell and S Salentine. Chapel Hill (NC): MEASURE Evaluation, University of North Carolina; 2018 (<https://www.measureevaluation.org/resources/publications/tr-17-225>, accessed 10 May 2021).

MEASURE Evaluation. RHIS curriculum. Chapel Hill (NC): MEASURE Evaluation, University of North Carolina; 2017 (<https://www.measureevaluation.org/our-work/routine-health-information-systems/rhis-curriculum>, accessed 10 May 2021).

1. Introduction

1.1 Toolkit for analysis and use of routine health facility data

Rationale for the toolkit

Routine health facility data comprise data that are reported at regular intervals from facilities providing health services. The system of regular recording, reporting, analysis and presentation of health facility data is known as the **routine health information system (RHIS)**.¹

RHIS data provide a picture of the services delivered in health facilities and the health status of the people using the services. The data can be used to assess the performance of individual facilities and also to assess service utilization and coverage of interventions in defined populations.

RHIS data serve multiple users and a wide range of purposes including patient/client management, facility management, disease surveillance, monitoring of service provision and resource use, and planning, resource allocation and policy-making.

Despite the importance of routine facility data and substantial investments in RHIS over the years, many contexts continue to face multiple challenges related to RHIS data, including:

- a multiplicity of indicators, data elements and disaggregations (in part, resulting from increasing global reporting needs and stakeholder demands);
- lack of standardization of indicators and data elements;
- gaps in specific data types (e.g. community-based, hospital and quality of care data);
- fragmentation and duplication of data systems (lack of interoperability);
- poor data quality with resulting lack of trust in RHIS data;
- capacity gaps in data analysis, presentation and interpretation; and
- failure to communicate data in formats that are appropriate to various users.

At the Measurement for Accountability for Health Summit in 2015, United States Agency for International Development (USAID), WHO and the World Bank called for action “to improve health facility and community information systems including disease and risk surveillance and financial and health workforce accounts, empowering decision makers at all levels with real-time access to information.”²

In response, WHO has led the development of the **Toolkit for analysis and use of routine health facility data**.³ This initiative represents the collaborative efforts of multiple WHO technical programmes and partners.

The toolkit emphasizes standardization, integration and a focused approach to data analysis. It aims to focus the analysis of RHIS data on a limited set of standardized indicators, using recommended standard analyses, visualizations and dashboards, and to provide guidance on data interpretation and use.

¹ The RHIS is also called the health management information system (HMIS). The term RHIS is used throughout this document.

² Health Measurement and Accountability Post 2015: Five-Point Call to Action, June 2015.

³ https://www.who.int/healthinfo/tools_data_analysis_routine_facility/en/

Structure of the toolkit

Refer to the inside front cover of this document for an illustration of the toolkit contents. The toolkit consists of three parts, each of which contains a number of modules. Each module contains a guidance manual(s) and may also contain downloadable electronic dashboard packages (“configuration packages”) and training materials.

Part 1: Standards for measurement and analysis

This part provides foundations for analysis and use of RHIS data that are applicable across all the toolkit modules. It includes three modules:

- general principles, which discusses basic concepts for analysing RHIS data;
- core indicators, which includes the indicators from all the modules; and
- data quality assurance tools, which countries can use to assess the quality of their RHIS data.

Part 2: Integrated health service analysis

This part contains two modules, targeting general health service planners and managers at national level and district and facility levels respectively. The modules provide an integrated (or cross-cutting) approach to analysis of health service performance, drawing on indicators from multiple toolkit modules.

Part 3: Programme-specific guidance

This part contains programme-specific modules including:

- reproductive, maternal, newborn, child and adolescent health (RMNCAH);
- Expanded Programme on Immunization (EPI);
- human immunodeficiency virus (HIV);
- tuberculosis (TB);
- malaria; and
- hepatitis.

Each of these modules contains a set of core indicators specific to the programme, addresses related data quality issues, presents recommended analyses and visualizations, and includes considerations for interpretation and use.

The indicators and analyses presented in the guidance manuals can be applied using various software packages. Electronic configuration packages for the District Health Information Software 2 (DHIS2) are included in the toolkit; each programme-specific module has an accompanying DHIS2 package.

1.2 Overview of routine health information systems

As health workers go about their daily work of delivering care, they record data in health facility records such as individual patient records (e.g. antenatal cards), registers, tally sheets and log books. Selected data are extracted from these records, aggregated in tally sheets or counted from registers, and then consolidated in monthly paper-based report forms. The reports are submitted at regular intervals through the RHIS to successive levels of the health system, with further aggregation, analysis and use at each level, e.g. district, province and national levels.

Data reported through the RHIS may be called “RHIS data”, “routine health facility data” or simply “routine data”. A defining feature of RHIS data is that they are reported **regularly**, usually at intervals of up to 1 year. In practice, this often refers to monthly or quarterly reports. However, less frequent reports are also sometimes considered part of the RHIS, e.g. annual facility reports on infrastructure, staffing, etc.

The RHIS may be paper-based or electronic or a hybrid of both system-types. In most health systems, aggregate data from the monthly reports are entered into an electronic database. This data entry may occur at various levels of the system, e.g. health centre, district office, etc. In some RHIS, aggregate data from all programmes are entered into the same electronic system; in other cases, specific programmes have separate systems.

Data collected at community level (e.g. by community health workers) and reported to health facilities may also be included in the RHIS, as well as data produced by other health service delivery sites such as prisons, schools and workplaces.⁴ Some of these data from other sites may be aggregated with facility data. However, the facility and non-facility data should also be analysed and presented separately.

This document refers specifically to the analysis of **aggregated RHIS data generated within health facilities**.

The RHIS within the national health information system (HIS)

The RHIS is an integral part of the overall national HIS. HIS data sources are usually either population-based or institution-based.

Population-based sources include population censuses, civil registration and vital statistics (CRVS) systems and population-based surveys. These sources relate to the entire population (i.e. not only to people using health facilities). **Institution-based sources** include the RHIS, health facility surveys/assessments and data systems for health service resources (e.g. infrastructure information systems, logistics management information systems (LMIS), health workforce information systems, financial management information systems).

Some types of data are obtained from both population-based and institution-based sources (“mixed sources”), e.g. surveillance data are obtained from community sources and other public health surveillance systems as well as from health facility reports.

All data sources have inherent strengths and limitations. An essential function of the overall HIS is to match each data element/indicator with the most appropriate and efficient means for generating it. Some data can be generated from multiple sources. The type and maturity of the various HIS components, as well as the frequency at which the data are needed, should determine the most appropriate data source for a specific context. Regarding the RHIS, these decisions have important implications for the workload of facility staff as well as for data quality.

Advantages of RHIS data

The RHIS is an established component of the health service delivery system, providing information on a wide range of services. It applies across the entire country, at all levels of the health system. RHIS data are collected by and can be analysed by the service providers themselves, providing them with insights into the outputs and outcomes of their work. As RHIS data are collected continuously and analysed regularly, they provide up-to-date information on health services and on health conditions occurring the populations using the services, thus facilitating early identification of problems and enabling managers to take timely actions.

The RHIS can be a data source for informing subnational planning and resource allocation. As the RHIS supplies data representing subnational (e.g. district) and facility levels, it contributes to monitoring of geographic inequities and progress toward targets at these operational levels.

⁴ If these services are not included in the analysis of RHIS data, populations such as school-age children and adolescents may be unaccounted for in the country HIS.

Limitations of RHIS data

Health facility data can provide information only on the people that use the facilities. However, substantial parts of the population may not have access to the facilities (e.g. nomadic or marginalized groups) or may choose to use alternate care options (e.g. self-medication, traditional healers). Furthermore, some facilities (notably private providers) may not be included in the national RHIS. Facility data are therefore not representative of the entire population, unless all facilities are included in the RHIS; reporting rates are high; and facility utilization rates are higher than 95%, which is rarely the case.

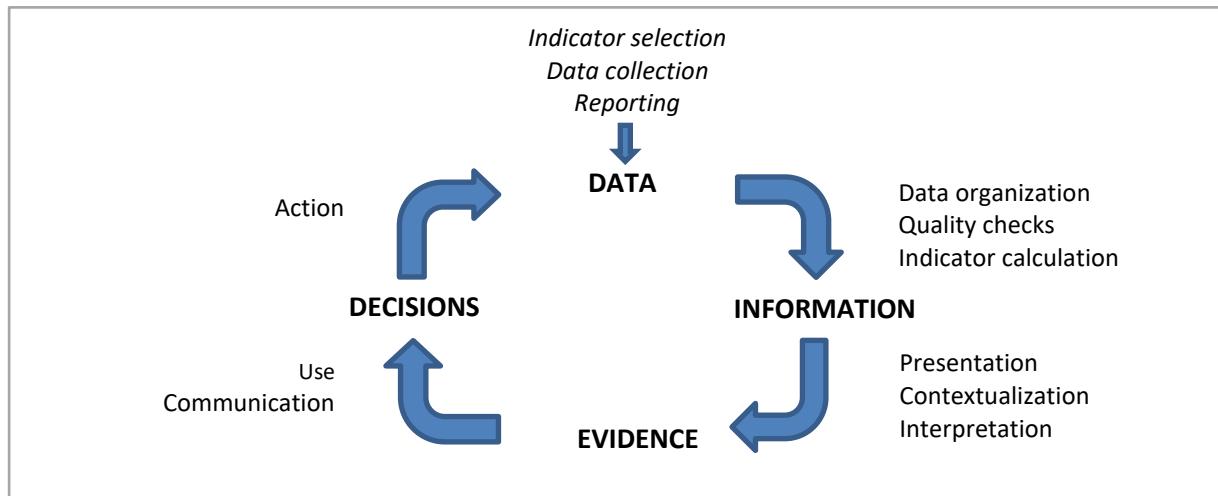
There are ongoing challenges in establishing reliable population estimates for use as denominators for rate and coverage indicators. Official population estimates are derived from census projections. However, significant demographic changes may have occurred since the last census and, over time, census projections may differ substantially from true population numbers. Furthermore, census estimates may not be available for small subnational areas such as districts. (Chapter 5 provides further discussion on population estimates.)

Depending on the maturity level of the system, in some contexts the RHIS may not be the most appropriate means of obtaining some types of data. For example, where the RHIS is mainly paper based, the collection of detailed quality-of-care or outcome data adds a substantial workload to health workers already overstretched by clinical and reporting duties. This overburdening also risks contributing to the deterioration of data quality. Such data could be collected instead through periodic facility assessments, supervision visits, sentinel facilities or special studies, and used in conjunction with RHIS data.

Analysis of RHIS data

Data do not speak for themselves. Data analysis is the process of transforming data into information and evidence that can be used for decision-making and action. This process involves a cycle of several steps, all of which are equally important. The steps are summarized in Fig. 1.1.

Fig. 1.1 Data analysis and use cycle



The terms “data” and “information” often are used interchangeably. However, “**data**” refers to raw, unprocessed numbers or text, while “**information**” refers to data that have been processed, organized and presented within a specific context to give the data meaning.

Selection of appropriate indicators and the establishment of functional systems for data collection and reporting are the foundations that make data available for analysis. The data analysis process includes compilation, organization and review of the raw data, assessment of the data quality, making corrections

and adjustments where necessary (data “cleaning”) and then, through statistical calculations, transforming the data into **indicators**.

The indicators can be presented in the form of visualizations (e.g. charts, tables and maps), along with information on the limitations of the data as well as relevant supplementary information and explanations. When the indicators are interpreted within the context of issues confronting the health system, the information (provided that it is of sufficient quality) becomes **evidence**. Evidence may be defined as “...any form of knowledge, including, but not confined to research, of sufficient quality to inform decisions...”⁵

Analysis of RHIS data includes review of trends over time, assessment of performance against targets and benchmarks, comparisons among facilities or geographic/administrative units, and assessment of differences by age, sex or other types of disaggregation. Analysis also involves comparisons between interrelated services (e.g. TB and HIV services) as well as review of RHIS data in relation to data from other sources.

Appropriate communication of information and evidence to decision-makers is critical. This includes formatting the information into dashboards, reports and presentations with user-friendly visuals, as well as providing explanations of the findings and interpretation within the specific context. Only then will the data have an opportunity to be used to inform decision-making that can result in actions.

The steps of the data analysis and use cycle are described in the various sections of this document.

⁵ Buse K, Mays N, Walt G (2012). Making health policy (second edition) United Kingdom: McGraw-Hill Education.

2. Core indicators

Analysis and use of routine health facility data can be strengthened by focusing on a limited, standardized set of core indicators.

2.1 Standards for RHIS indicators

Core indicator list

A core indicator list consists of a limited set of key indicators. The core RHIS indicator list focuses data analysis on the most important indicators needed on a regular basis to monitor health services and inform decisions at various levels of the health system. Establishment of the core indicator list should be a collaborative process involving multiple programmes and stakeholders. The core indicator list should be reassessed periodically to ensure that it reflects current global standards and country priorities.

Consistency with international standards

The core facility indicator list should include a balanced set of both general and programme-specific indicators that are consistent with international health service and programmatic standards and reporting requirements.

Standardization of indicators and data elements

Well-defined, standardized indicators, data elements and metadata are essential. This avoids the creation of multiple similar but incompatible indicators and data elements and enables consistent analysis of data across programmes and partners and over time. A standard core indicator set also provides the basis for a set of standardized core analyses, visualizations and dashboards.

2.2 Core RHIS indicators

Each programme-specific guidance manual in the *Toolkit for analysis and use of routine health facility data* contains a list of recommended core indicators relevant to the programme. The *Integrated health services analysis* module includes general health services indicators (e.g. outpatient department attendance) as well as a selection of tracer indicators from the various programme-specific lists.

The *Core health facility indicators*⁶ document is a compendium of all the indicators in the various guidance manuals. The recommended core indicators in this document are consistent with global programmatic and health services standards. The list includes relevant indicators from WHO's *Global reference list of 100 core health indicators*⁷ as well as other key RHIS indicators required for planning and managing health services and programmes, and for reporting to national- and global-level stakeholders.

Countries can expand or modify this recommended core list based on local priorities and epidemiological profiles. A country can also use the list as a reference to assess whether their RHIS includes these recommended indicators and whether their existing indicators, terminologies and metadata correspond to international technical standards.

⁶ https://www.who.int/healthinfo/tools_data_analysis_routine_facility/en/

⁷ <https://www.who.int/healthinfo/indicators/2018/en/>

3. Facility representation in RHIS

3.1 Representativeness versus completeness

Appropriate interpretation of RHIS data requires an understanding of the extent to which all the facilities in the country are included in the RHIS. Assessment of facility representation in the RHIS is different from assessment of reporting completeness, which measures whether the facilities that are already included in the RHIS have submitted reports as required.

National RHIS data represent only those facilities that report into the RHIS. In some countries all facilities are included, while in others only the ministry of health facilities are part of the system. The private sector, including private-for-profit providers, nongovernmental organizations (NGOs) and faith-based organizations (FBOs), often delivers a substantial part of a country's health services but is often inadequately represented in the RHIS. Military facilities (serving military personnel and their families) may also be excluded. Particularly in cities, private-for-profit facilities may account for a significant percentage of select services such as delivery care. Failure to include private sector facilities in the RHIS may thus result in significant bias in routine facility data. Analysis of RHIS data should therefore always explicitly state the types of providers that are represented and should also provide an estimation of the number and proportion of facilities (with their ownership) that are not included.

Assessment of facility representation in the RHIS requires comparison of the list of facilities included in the RHIS with a truly comprehensive national Master Facility List (MFL).⁸

3.2 Master Facility List (MFL)

An MFL is a complete, up-to-date, authoritative listing of the health facilities in a country, including public, private-for-profit, NGO, FBO and military facilities, etc. It is the primary source from which other facility lists in the country are drawn, and must be validated, continuously updated and accessible. The MFL includes the data needed to accurately identify each facility, such as facility name, unique facility identifier, location and contact information, as well as administrative data to categorize the facility, such as facility type, ownership and operational status. The MFL may also include information about the service capacity of the facility, e.g. type of services offered and number of beds. Ideally, the MFL is stored in a facility registry service or software program that makes the list accessible to stakeholders such as ministries, donors and implementing organizations. WHO, with partners, has produced guidance for developing and maintaining an MFL.⁹ Where an updated MFL does not exist, a comprehensive inventory of all facilities in the country should be developed, as part of efforts to establish or strengthen the national MFL.

In the absence of a reliable MFL, it may be possible to obtain information from facility listings used for facility assessments. It may also be possible to use population-based survey data to obtain an estimate of the size of the private sector share in the provision of health services. The Demographic and Health Surveys (DHS) and Multiple Indicator Cluster Surveys (MICS) provide data on the private sector role in services such family planning, antenatal care (ANC), delivery and postnatal care, childhood treatment services, HIV and TB services, etc. This enables estimation of the relative contribution of the private sector to the provision of specific services. In some contexts, information from insurance providers could also be used.

⁸ The MFL is sometimes also referred to as the national health facility registry.

⁹ Master Facility List resource package: guidance for countries wanting to strengthen their MFL. USAID, PEPFAR, WHO; 2018. https://www.who.int/healthinfo/country_monitoring_evaluation/mfl/en/.

4. Data quality

All data have limitations that affect their reliability and interpretation. Therefore, before further analysis and interpretation, RHIS data should be assessed for quality. Errors can be corrected through communications with districts or health facilities. Adjustments to the data, e.g. for incomplete reporting, can also be made by analysts. Any such corrections and adjustments should always be documented. (Annex 1 provides further details on adjustment methods.)

All analytical reports should include information on the quality of the data presented and explanations of any adjustments. This enables the user to understand the limitations and to decide whether the data are of sufficient quality to be used as intended, i.e. whether the data are “fit for purpose”.

Good quality data are:

- **complete and timely:** all the required data are available within the required timeframe;
- **consistent:** there are no unexplained variations over time, and related data elements display expected relationships; and
- **correct:** the data are accurate, i.e. the values reflect actual events.

4.1 Types of data quality assessment

Assessment of the quality of routine facility data requires a multipronged approach that includes:

- routine, regular assessments (e.g. monthly);
- periodic assessments of a core set of tracer indicators (e.g. annual); and
- periodic in-depth programme-specific reviews.

Routine, regular assessment

Routine assessment of data quality can identify problems in close to real-time, enabling correction of errors as they occur. This involves regular quality checks at facility level and at each subsequent reporting level. Such checks should be part of the standard operating procedures (SOPs) of the RHIS and may include:

- checking and approval of monthly reports by facility supervisors before report submission;
- automated quality checks (e.g. pre-set minimum and maximum values; validation rules)¹⁰ that are built into electronic systems such as the DHIS2 to provide alerts at the time of data entry;
- simple visual scanning of data displayed in tables or trend charts to identify obvious problems such as missing values, unusual fluctuations and mathematical errors;
- automated data quality dashboards that display data quality metrics (indicators) along with the related routinely reported data; and
- routine and regular use of software such as the WHO Data Quality Tool (see below) to screen electronic data for inconsistencies.¹¹

In many contexts, however, such data quality checks are not applied in a consistent, systematic manner. In such circumstances, it is difficult to achieve significant improvements to data quality.

¹⁰ Validation rules are based on a logical relationship between two variables. For example, the number of suspected malaria cases tested should be greater than or equal to the number of confirmed malaria cases.

¹¹ WHO has developed guidance for district data quality assurance through monthly review of DHIS2-based data quality dashboards and the WHO Data Quality Tool.

Periodic assessments of tracer indicators

Periodic data quality assessments of a defined set of tracer indicators are often conducted annually but can also be carried out more frequently. Such assessments involve a comprehensive review of a limited set of tracer indicators representing key programmes. This provides an overview of the general quality of RHIS data and helps to identify problems that are common across multiple programmes. This information can be used to inform data quality improvement strategies. The assessment can be conducted as a desk review and/or a field investigation.

A **desk review** involves carrying out quality checks on the aggregated data reported by each facility or administrative unit (e.g. district); these data are usually available electronically. The quality checks involve the use of standardized data quality metrics. If time and resources permit, a desk review should be complemented by a field investigation.

A **field investigation** is a more extensive review based on assessment of a sample of districts and health facilities. It includes a “data verification exercise” to determine the extent to which reported data match the data found in the facility’s source documents (e.g. facility registers, tally sheets). The field investigation also includes an assessment of the facility’s data management system to determine its adequacy in producing quality data.

WHO, with partners, has developed a *Data quality assessment (DQA) toolkit*¹² which contains guidance as well as data collection and analysis tools to support desk reviews and field investigations. The toolkit includes an Excel tool that analyses the data quality dimensions of completeness, internal consistency and external consistency. The toolkit also includes the *WHO Data Quality Tool*,¹³ an application that can be installed on the national DHIS2 system to automatically generate and display data quality metrics.

Periodic in-depth programme-specific reviews

In addition to the quality metrics of the *DQA toolkit*, the programme-specific packages of the *Toolkit for analysis and use of routine health facility data* include discussions on data quality issues specifically relevant to the programme. Further information on in-depth programme-specific reviews can be found in relevant programme documents.

4.2 Dimensions of data quality assessment

Four dimensions are considered in a basic data quality assessment:

- reporting completeness and timeliness;
- internal consistency of reported data;
- external consistency with other data sources; and
- external comparison of population data.

The following section briefly describes the concepts underpinning each dimension. Refer to the *DQR toolkit* for an in-depth discussion of the dimensions, the metrics for measuring each dimension and visualizations that display the findings of the data quality assessment.

Dimension 1: Reporting completeness and timeliness

Completeness is expressed as the percentage of expected¹⁴ reports submitted to a higher level of the reporting system. Completeness is assessed at each reporting level, e.g. completeness of facility reports

¹² <https://www.who.int/data/data-collection-tools/health-service-data>

¹³ <https://apps.dhis2.org/>

¹⁴ For example, 12-monthly reports are expected from each health facility per year.

submitted to district level; completeness of district reports (aggregated data from multiple facilities)¹⁵ submitted to provincial or national level. Key steps for assessing completeness include:

- **Assess completeness of reporting for each form**

Different reporting forms are often used to report different services. For example, outpatient morbidity is often reported separately from immunizations, ANC, etc. Completeness is assessed separately for each service or report. The expected number of reports may vary according to the service. For calculation of reporting completeness, the numerator is the number of reports received and the denominator is the number of facilities expected to submit each form.

- **Assess completeness of reporting for key data elements**

This step assesses the completeness of specific data elements within a reporting form. Each cell of a reporting form represents a data element. (Faced with the burden of completing many cells, health workers may consistently leave certain cells blank, using only a subset of cells for reporting.)

- **Consider completeness of reporting from hospitals and private sector facilities**

Hospitals report the great majority of admissions/discharges and inpatient deaths as well as a significant proportion of outpatient services. Yet in some health systems, reporting completeness is significantly lower from hospitals than from primary care facilities. There may be incomplete reporting from all hospitals or from only some large referral hospitals. Some hospitals may not report at all. This incompleteness can introduce an important bias in nationwide statistics on the services offered by such facilities, such as inpatient services, surgical services, deaths, diagnoses that require specialized services (e.g. cancers, heart diseases, neglected tropical diseases, etc.). As noted in Chapter 3, some or all private facilities may not be represented in the RHIS. Among private facilities that are included, NGOs generally have good reporting rates, while reporting by private-for-profit facilities is often problematic.

- **Assess the impact of incomplete reporting on coverage**

Incomplete reporting and variations in completeness over time affect the interpretation of levels and trends in coverage and service utilization indicators. Analysis of RHIS data should include an assessment of the impact of incomplete reporting on coverage indicators. Based on this assessment, analysts may consider adjusting the number of events reported in the RHIS to obtain a more realistic picture of the levels and trends. Annex 1 provides further details on adjustment methods.

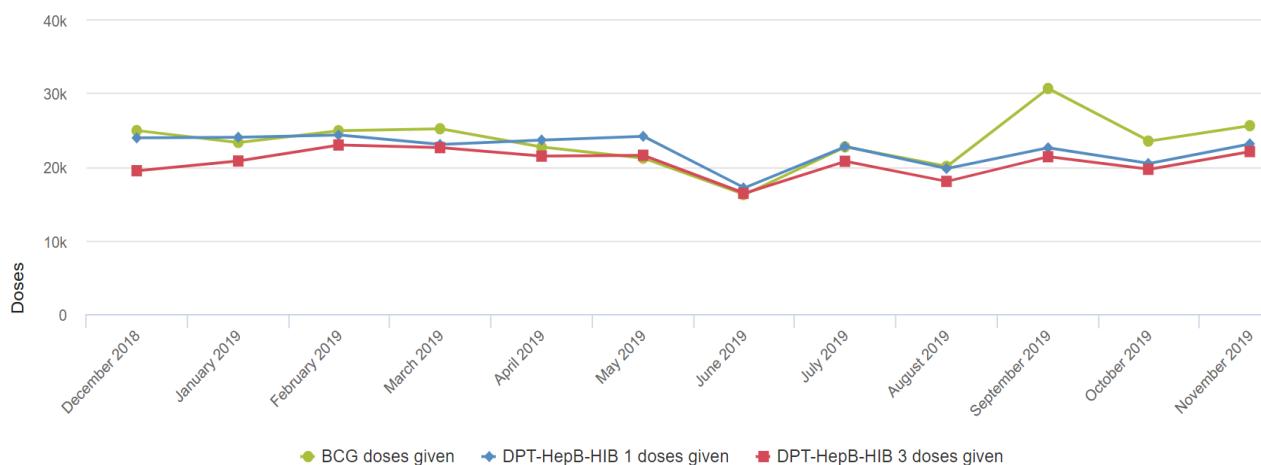
Dimension 2: Internal consistency of reported data

Internal consistency relates to the coherence between different RHIS data elements that have an expected relationship with each other. Assessment of internal consistency examines whether data values follow expected patterns over time and in relation to each other. Data entry error is an important cause of inconsistency. Errors may occur, for example, when data are added up or transcribed from a tally sheet or register to a monthly report or transcribed or entered from a monthly paper report into an electronic database. Key steps for assessing internal consistency include:

- **Assess coherence between the same data items at different points in time**

Outliers are values that are unusually high or low in comparison with historical trends. Major data entry errors can be identified by screening for outliers; tables or charts showing trends over time can be used to quickly identify outlier values. Fig. 4.1 illustrates how the expected month-to-month stability in the number of maternal health and immunization services delivered can be used to identify very large outliers (for example, the September 2019 value of BCG doses) based upon nationwide total values. To identify smaller outliers, the same chart can be viewed at district level. Outliers often reflect poor data quality, but they can also be the result of true changes in events.

¹⁵ Where facility reports are entered into an online electronic data management system such as DHIS2, the aggregation is done automatically and no distinction can be made between facility-level reporting completeness and higher-level (district or province) reporting completeness.

Fig. 4.1 Doses given for select vaccines, nationwide, last 12 months

In general, values of more than three standard deviations higher or lower than expected may be considered data errors, unless there is evidence that the value is correct. For highly consistent data, narrower limits than three standard deviations (SD) can be used to identify outliers.

Seasonal data, such as the number of malaria cases, do not show the same month-to-month consistency as data on maternal health and immunization services. However, to identify outliers the seasonal trend for the last 12 months can be compared with the seasonal trend for 12 to 23 months previously, 24 to 35 months previously, etc. Year-to-year consistency can also be assessed. For example, the annual fluctuation in TB data is likely to be less than 10%; if it is greater than 10%, the data may reflect quality problems rather than true trends.

Ideally, extreme outliers are identified and investigated at district level. Investigations may involve communications with facility staff or, for example, comparisons of doses administered with commodities supplied in specific facilities. At the national level, such investigations are often too time-consuming to be done at the time that an annual or multi-year report is prepared. Therefore, outliers identified at this level need to be corrected to obtain a “clean” dataset for further analysis. Analysts should consider replacement of the outlier by the expected value based on previous time periods. Any adjustments should be documented.

▪ **Assess coherence between related data items**

Data elements that have a predictable relationship are examined to check whether the expected relationship does in fact exist between them. For example, the number of children receiving the first dose of diphtheria–tetanus–pertussis vaccine (DTP1) is expected to be roughly equal to the number of children receiving the first dose of oral polio vaccine (OPV1), as these vaccines are given at the same time. The number of DTP1 doses given is expected to be greater than (or equal to) the number of DTP3 doses given. If DTP1 minus DTP3 is negative, there is a “negative dropout”. Negative dropout for a full year at the level of a district or higher is usually a sign of poor data quality. Similarly, the number of DTP1 doses should usually not be greater than the number of ANC1 visits, unless there is a special explanation, e.g. ANC1 may be under-reported because women attend non-reporting private clinics for ANC.

▪ **Assess consistency between reported data and source documents**

This is the only dimension of the data quality assessment that requires additional collection of primary data. It involves a field investigation to assess reporting accuracy through comparison of reported data with the same data in the source documents in health facilities. Source documents may include registers, tally sheets or individual patient records. This assessment (“verification exercise”) may provide evidence of over-reporting or under-reporting or it may reveal problems related to the aggregation of data.

- **Assess other potential consistency issues**

An unusual degree of uniformity in the data or the existence of certain patterns may also point to quality issues. Performance that seems “too perfect” may be worth investigating. When achievements are linked to financial incentives, the possibility of “over-reporting” should be considered. Datasets should be reviewed for double entry of data for time periods (e.g. duplication of entries for 2 consecutive months) or reporting units (e.g. duplication of entries for different facilities). Data with very many reported values that are multiples of 5 or 10 may point to guesses rather than reporting of true values. The data should also be checked for unlikely or impossible entries, e.g. males on oral contraception, females with prostate cancer, pregnancies in females under 10 years of age.

Dimension 3: External consistency with other data sources

This dimension examines the level of agreement between two data sources that measure the same health indicator. Indicators derived from RHIS data may be compared with indicators obtained through:

- estimates from population-based surveys;
- parallel data systems (e.g. vertical, programme-specific systems);
- sentinel site data; and
- statistics that have been officially reported to WHO.

The most important data sources for comparison with facility data are population-based surveys. Section 6.5 provides further details on such comparisons.

Dimension 4: External comparison of population data

This dimension examines the adequacy of the population data used to obtain denominators for calculating RHIS indicators on rate and coverage. Different sources of population estimates (for which the values are calculated differently) are compared to assess the level of agreement between them. The greater the consistency between the estimates from different sources, the greater the likelihood that the values represent the true population value. If the population estimates from the different sources are very different, the coverage estimates based on the respective denominators can be very different for a given indicator, even though the programmatic output (numerator) is the same.

Examples of population comparisons:

- population projections of live births from the national bureau of statistics are compared with United Nations projections of live births for the country;
- population projections from the national bureau of statistics are compared with population values used by programmes;
- population projections of live births or surviving infants from the national bureau of statistics are compared with estimates derived from facility data on DTP1 vaccinations received.

Population denominators are discussed in further detail in Chapter 5.

Summary of data quality measures

Table 4.1 Data quality dimensions, measures and review frequencies

Dimension	Data quality measure	Suggested frequency
Completeness and timeliness	Completeness and timeliness of reporting form submission	Monthly, annually
	Completeness of data elements	Monthly, annually
Internal consistency	Presence of outliers	Monthly, annually
	Consistency month-to-month and year-to-year	Monthly, annually
	Coherence between related data items	Annually or as needed
	Consistency between reported data and original records	Annually
External consistency with other data sources	Consistency between RHIS data and sources such as population-based surveys	Annually
External comparisons of population data	Consistency between population data used as denominator for calculating facility indicators and other sources of population estimates	Annually

5. Population estimates/denominators

Denominators, representing estimates of the target population, are required to calculate rates (e.g. disease incidence per 1000 population per year), coverage (e.g. percentage of infants vaccinated) and service utilization indicators (e.g. outpatient department attendance rates) from RHIS data. However, reliable estimates of target populations remain an ongoing challenge in many settings. Denominators are usually derived from census-based projections, but alternate methods may also be considered.

5.1 Census-based projections

Official population estimates are usually based on projections from the most recent national population census. However, censuses may contain inaccuracies, e.g. undercounting of some areas. Furthermore, the reliability of the projections declines as years since the census pass. The projections of the total population are based on population growth rates that were observed in the previous intercensal period (usually a decade). However, major changes may have taken place since the last census, including changes in fertility and mortality, migration into or out of the country, and changes in population distribution among subnational areas. Consequently, the census projections may be substantially different from the true population numbers, particularly for subnational areas.

Furthermore, it is often difficult to use census projections to estimate appropriate denominators for individual districts and health facilities. In addition to the challenges previously described, census data may not include small area estimates, such as for districts or health facility catchment areas.

Estimates of pregnancies, live births and surviving infants (the denominators for calculating coverage of maternal health and immunization services) are typically calculated by multiplying an estimate of the total population by an estimate of the “crude birth rate” (CBR: the estimated number of live births per year per 1000 persons in the total population). This may represent another source of inaccuracy, as the CBR may have declined significantly since the last census and there may be large variations in the CBR between regions and districts.

When census projections are used to estimate denominators, coverage may be significantly greater than 100% in some districts, but very low in others. In some cases, such findings may be true, for example, people living in one district may use services in a neighbouring district, as is often the case in urban areas. However, a common cause of coverage greater than 100% is an underestimation of the target population that resulted from incorrect or outdated census projections. The opposite situation (coverage that is much lower than in reality) is probably also common but is more difficult to detect.

Unexpected year-to-year changes in coverage and rates may occur when target population estimates are adjusted after new census data become available. The influence of such denominator adjustments can be illustrated by charting the numerators and the denominators by year, along with the coverage indicators. Countries should ensure that population estimates are retrospectively adjusted following a census in order to produce corrected time trends in coverage and rates.

5.2 Other methods

Alternate methods are sometimes used to improve denominator estimates and coverage calculations, for example:

Estimates derived from RHIS data: RHIS data on numbers of ANC1 or DTP1 are sometimes used to estimate the numbers of pregnant women or surviving infants in the population. However, this method

can be used only when coverage is consistently very high (> 90%) for ANC1 or DTP1 (as confirmed by population-based surveys) and when the facility data are believed to be of high quality. This methodology is discussed further in Annex 3.

Note: Some RHIS indicators assess the percentage of individuals receiving a specific intervention among those that accessed the service. The denominator in such indicators is based on facility data rather than on population data, e.g. “Antenatal syphilis screening (%)", where the denominator is the number of first ANC visits and the numerator is the number of ANC clients screened for syphilis.

Health programme estimates: Some programmes may use target population estimates that differ from those of the national bureau of statistics. If programme denominators are used for the analysis of RHIS data, the analysis reports should include a table of these denominators along with an explanation of the methods used to calculate them.

Local enumeration: Local headcounts are sometimes used to estimate the target population size. This approach can provide good operational targets for health workers at local levels. However, local enumeration is not recommended for coverage monitoring at national level as it lacks independence. Furthermore, local counts may also be inaccurate as some groups (e.g. nomadic or remote populations) may not be included.

Where reliable population estimates are not available, as is often the case for health facility catchment areas, presentation of numerator trends can provide useful information. Refer to *Integrated health service analysis: district and facility levels* for further discussion.

Presenting denominator sources

The methods and assumptions for projecting and estimating the target populations should always be presented along with the results of the analysis. A table of estimates of the key target populations (total population, pregnancies, surviving infants, children under 5 years of age, etc.) by administrative unit should be included in the analysis report. Where denominators are based on census projections, the annual growth rate should be stated. Table 5.1 provides an illustration.

Table 5.1 Extract from a table of denominators used to calculate indicators

Subpopulations by region, 2012, United Republic of Tanzania (based on projections of 2002 national census)							
Region	Annual growth rate	Total population	Pregnancies = births^a	Births	Surviving infants = birth x 0.95	Children < 5 years	Women 15–49 years
Arusha	2.74%	1 694 310	71 161	71 161	67 603	254 147	406 634
Dar es Salaam	5.76%	4 364 541	130 936	130 936	124 389	654 681	1 047 490
Dodoma	2.12%	2 083 055	87 488	87 488	83 114	312 458	499 933
Iringa	1.11%	1 643 335	69 020	69 020	65 569	246 500	394 400
Kagera	3.25%	2 773 054	122 014	122 014	115 914	415 958	665 533
Kigoma	2.43%	2 127 930	89 373	89 373	84 904	319 190	510 703
Kilimanjaro	1.82%	1 640 087	49 203	49 203	46 742	246 013	393 621
Lindi	0.90%	864 652	31 127	31 127	29 571	129 698	207 516

Note:^a The assumption that the number of pregnancies equals the number of births will result in ANC coverage being somewhat underestimated.

Source: Midterm analytical review of the Health Sector Strategic Plan III 2009–2015. Ministry of Health and Social Welfare, United Republic of Tanzania; June 2013.

6. Key analytical concepts

6.1 Basic statistical terms

Table 6.1 Basic statistical terms

Term	Definition
Rate	A rate is the frequency with which an event or case occurs in a defined population over a specified period of time. A rate is often expressed as events per 1000 population per year. Some rates are expressed per 10 000 or per 100 000 population. All events or persons in the numerator are also included in the denominator.
Ratio	A ratio is a comparison of values, showing their size in relation to each other. It is calculated by dividing the first value by the second. The numerator is not contained in the denominator. Ratios are written in various ways, e.g. 3:4 or 3 to 4 or 0.75 or 75%.
Proportion	A proportion is the number of events or cases that occur in a defined population, expressed as a fraction, a decimal or a percentage. All events or persons in the numerator are also included in the denominator.
Coverage	Coverage is a measure of the extent to which the services provided cover the potential need for these services in a population. It is expressed as a percentage in which the numerator is the number of service units provided, multiplied by 100, and the denominator is the target population in need of the service.
Mean	The mean is the sum of all the values in a set divided by the number of values in the set. A single very large value in the set can result in a very high and unrepresentative mean.
Median	When values in a set are ranked from smallest to largest, the median is the value in the middle of the list, i.e. half of the values are greater than or equal to the median and the other half are less than or equal to it. When there is an odd number of values, the median is the middle value. When there is an even number of values, the median is the average of the two middle values. A single very large value in the set will not affect the median.

6.2 Aggregated versus individual data

A large proportion of RHIS data is based on reporting of total counts ("aggregates") of single events or characteristics, e.g. number of outpatient department visits, number of confirmed malaria diagnoses. These data can be collected using tally sheets and reporting does not require follow up of individual patients over time.

Some programmes (e.g. immunization, TB, HIV,) use tracking systems to record information on individual patients over time. Sometimes these tracking systems are electronic (e.g. electronic registers) and may be integrated with the RHIS. Often, however, they are separate systems and only selected aggregate data are extracted and submitted to the RHIS.

For some programmes, it is very difficult (or impossible) to obtain coverage or outcome indicators from aggregated RHIS data, e.g. family planning, HIV and noncommunicable diseases care. These programmes involve long-term care that needs repeated visits over time. In such cases, coverage is not based on receiving a single intervention (e.g. third dose of DTP vaccine), but on remaining in care over time. A

system for routine monitoring of individual longitudinal patient records is needed to know how many patients are active in the programme at any specific point in time. In programmes that do not implement such a system, simplified indicators for aggregate reporting can be used, e.g. the number of new contraceptive users or the number of newly diagnosed cases of HIV, hypertension, diabetes, etc.

6.3 Disaggregation

Disaggregation of RHIS data by, for example, age group, sex and geographic location is important for identifying differences in disease patterns and service utilization. In the RHIS, however, multiple disaggregations often substantially increase the reporting workload, particularly in paper-based systems, and may have a negative impact on quality data. Therefore, the purposes for which specific disaggregated data will be used and the frequencies at which such analyses are needed, warrant careful consideration.

If short-term variations are unlikely, routine facility reports may not be the most suitable means of obtaining detailed disaggregated data. Other data collection methods should be considered, for example, sentinel sites, periodic studies or population-based surveys. The following extract from *Immunization: guidance for programme managers* provides an illustration:

“...Sometimes the administered doses are further disaggregated, for example, by sex of the child, or the strategy that was used to vaccinate it (i.e. fixed versus outreach), or whether or not the child lives within the catchment area of the health facility. These additional disaggregations are not recommended as there is scant evidence that reliable data disaggregated in these ways can be collected or meaningfully used. Moreover, the recording and reporting workload doubles every time a new level of disaggregation is introduced. Therefore, the decision to further disaggregate immunization data needs to be weighed carefully against the benefit of the use that will be given to the collected data. Those designing routine reporting forms should aim to limit the number of cells and rely upon findings from household surveys to more reliably answer many questions...”¹⁶

6.4 Equity analysis

Equity analysis is fundamental to the monitoring of universal health coverage (UHC). Equity refers to fairness or impartiality: treatment of people in different ways, according to their needs, to achieve equality in outcomes.

RHIS data can identify differences in disease patterns, service utilization, coverage and health outcomes according to geographic/administrative units, sex, age groups and other disaggregations. These differences may point to underlying inequities related to, for example, risk factors, health service access and quality of care. Analysis of health service resource data in relation to, for example, service outputs and the population served, can highlight inequities in resource allocation.

Subnational analysis (or “area-based analysis”) is the most important equity dimension that can be assessed using RHIS data. As interventions to reduce inequity are implemented at subnational levels, RHIS data can play an important role in informing resource allocation and planning. However, some caution is also needed when interpreting subnational data with an equity lens. For example, population density, land area and physical terrain should be considered when comparing service utilization among geographic areas. Furthermore, the population in one subnational area may choose to use the services in a neighbouring area.

¹⁶ Immunization: guidance for programme managers. Section 4. Geneva: World Health Organization; 2020.

Analysis of health issues by socioeconomic characteristics is an essential aspect of equity assessment. However, socioeconomic data are not usually collected in the RHIS, as recording and reporting are time consuming and the analysis would still need a population-based component to assess health service coverage among, for example, the poor. Population-based household surveys remain one of the best sources of data for analysis by socioeconomic status.

6.5 Comparisons with other data sources

The synthesis and analysis of data from multiple sources and reconciliation of indicator values are essential to maximizing the value of all sources of data. Comparisons with data from other sources can provide insights into the quality of RHIS data. Such comparisons can also provide insights into the extent to which the RHIS data are representative of the population and are suitable for use in overall programme and health sector performance assessment. It is good practice to use tables and charts to present RHIS data alongside data from other sources.

▪ Population-based surveys

Population-based surveys, such as the DHS or the MICS, are the most important sources for comparison with RHIS data. Estimates from population-based surveys are often cited as the “gold standard” for measurement of coverage as they provide representative information on the entire population, including the people that do not use health services. However, surveys also have limitations, including:

- **time lag:** there may be a period of several years between the time that an event occurred and the availability of the survey results;
- **limited geographic disaggregation:** survey sample sizes are rarely large enough to provide valid estimates for lower subnational levels such as districts;
- **quality problems:** surveys vary in their adherence to sampling and interview protocols and other quality standards; all surveys contain sampling errors, which increase at subnational levels due to small sample sizes;¹⁷ non-sampling errors may also occur, e.g. selection bias or recall bias.¹⁸

▪ Health facility assessments

Health facility assessments/surveys are used to collect data not usually reported through the RHIS, such as data on quality-of-care and availability of equipment, medicines and human resources. Facility assessments can also be used to verify RHIS data. (Refer to Section 4.1 for further discussion on data verification.)

▪ Parallel data systems

A key objective in the development of RHIS standards is the reduction of duplication and redundancy in data collection. Unfortunately, however, the existence of parallel data systems that report on the same health events remains common. For example, in some countries the Expanded Programme on Immunization (EPI) collects immunization data both on an EPI form and on a separate child health form.

Such parallel systems increase the reporting workload and can result in confusion if there are discrepancies between the data reported by the two systems. It is important to review and analyse the data from each of the parallel systems and to present the findings as part of the analysis. Possible reasons for discrepancies should be discussed. Each table or chart must specify which of the parallel systems was used for the analysis presented.

¹⁷ DHS and MICS reports include annexes that provide confidence intervals for key indicators at national and regional levels. Confidence intervals estimate the effect of sampling error.

¹⁸ For example, recall bias may occur if vaccination cards were reviewed for fewer than half of the children surveyed and the survey relied on caregiver memory of vaccinations received.

- **Data from sentinel sites and demographic surveillance sites**

Data for comparison may also be available from sentinel sites at hospitals and clinics. Such sites often have the capacity to assure a higher quality of diagnosis and reporting than other facilities. Demographic surveillance sites that regularly track household demographics and health status in defined areas can also provide reliable data on the population and health events.

- **Data triangulation**

“Triangulation” of findings from multiple data sources and methods, including the RHIS, can be used to obtain the most valid estimate for an indicator. Annex 2 provides further details.

7. Presentation and communication

Appropriate packaging of information is a key requirement for influencing decisionmakers. However, the presentation of data is often aimed at technical experts, with little effort being made to ensure that the information is understandable to policy-makers, frontline health workers, non-health specialists or the public.¹⁹

Presentation involves the display of data in a format that facilitates its communication and interpretation. There are many ways of presenting data. The choice of format is important, as it can have a significant impact on how the data are received and interpreted. The format depends on the types of data to be displayed, the types of questions to be addressed, the communication method (e.g. printed report, power point presentation, on-screen dashboard) and the intended audience.

RHIS data can be visualized using a mix of formats such as tables, charts and maps. If used appropriately, these visualizations can deliver messages more effectively than text alone. Visualizations should be simple and uncluttered, keeping in mind Tuft's Principle: "Maximum amount of information for minimum amount of ink".²⁰

Visualizations should be accompanied by narratives that provide additional information and/or explanations. Examples of basic visualization types and their uses are outlined below.

7.1 Tables

Tables are used to present detailed data. They are also useful for displaying large amounts of data when the intention is to provide an overview of a situation, e.g. several indicators across several geographic units or time periods, as shown in Table 7.1. However, it is often difficult to draw conclusions from large tables of numbers. A user's understanding of a table can be facilitated by clear labelling, organization of columns or rows into specific sequences, or colour coding of values according to thresholds or categories. Tables are often too detailed to be used in presentations but are useful in printed documents and electronic dashboards.

Table 7.1 Selected RHIS indicators of a health district, 2015–2019

DA. 4.20 - Indicators (facility denominators)	2015	2016	2017	2018	2019
C-section rate: facilities (%)	3.5%	3.2%	3.3%	3.6%	4.2%
Bed occupancy rate (%)	79%	73%	70%	60%	66%
Average length of stay	2.9	2.9	3.2	3.0	3.4
ANC clients 1st visit < 12 weeks (%)	12%	18%	22%	25%	32%
ANC clients tested for HIV/known HIV+ (%)	95%	94%	91%	98%	93%
ANC clients screened for syphilis (%)	82%	82%	80%	77%	80%
ANC clients given IPTp3 (%)	0%	0%	7%	13%	22%
ANC clients with 4th ANC visit (%)	71%	66%	61%	59%	60%
DTP1-DTP3 drop out rate (%)	-1.5%	6%	7%	10%	7%
BCG-MCV1 drop out rate (%)	7%	6%	10%	11%	10%
MCV1-MCV2 drop out rate (%)	75%	65%	50%	40%	26%
TB cases with drug susceptibility test (%)	77%	77%	70%	80%	72%
TB cases with documented HIV status (%)	80%	87%	87%	88%	81%
TB treatment success rate (%)	70%	67%	73%	59%	
Suspected malaria cases tested (%)	82%	81%	81%	89%	95%
Malaria confirmed - given ACT (%)	109%	107%	109%	107%	107%

¹⁹ Framework and standards for country health information systems. Health Metrics Network and World Health Organization. Geneva: World Health Organization; 2012

(https://www.who.int/healthinfo/country_monitoring_evaluation/documentation/en/).

²⁰ ER Tuft. The visual display of quantitative information. Cheshire, CT: Graphics Press; 1983.

7.2 Charts

Charts can show a large quantity of data in a way that is quick and easy to understand. The shape of the chart helps the user to see trends over time, find patterns, see the relationships between variables and identify potential data quality problems.²¹ Different types of charts are used for different purposes.

- **Line charts**

Line charts are used for showing time series of data and for identifying trends. A trend is a pattern of gradual change or a general direction, e.g. upward or downward, shown by a series of data values over time. A minimum of three time periods should be presented when analysing trends. Fig. 7.1 illustrates seasonal trends in malaria cases based on rapid diagnostic testing (RDT). Note the peak from May to July of each year. Trend analysis can enable prediction of the pattern of the data in the future and can also quickly identify outliers or unusual fluctuations. Line charts can also be used to show the relationships between different data elements over time. For example, Fig. 7.2 shows that the numbers of DTP1 doses administered have consistently been higher than the numbers of DTP3 doses administered.

Fig. 7.1 RDT positive malaria cases, region X, 2017–2019

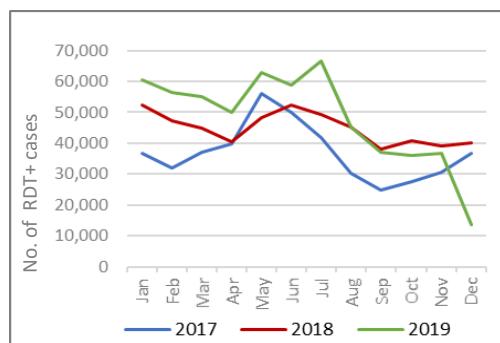
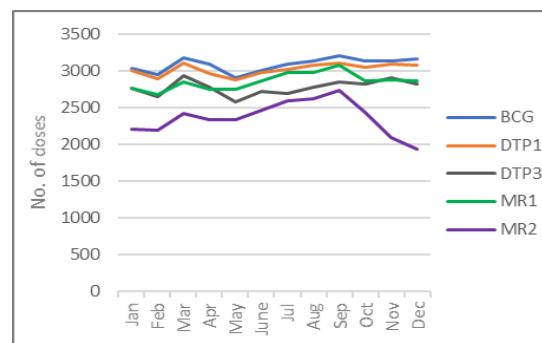


Fig. 7.2 Vaccine doses given, district Y, 2019



- **Column and bar charts**

Column (vertical) or bar (horizontal) charts are used to show comparisons, e.g. among different geographic areas or among different data elements or indicators. The columns or bars should be ordered by size (ranked), unless these are reasons for different arrangements. Such ranked charts are also called “league tables”. Bar charts can be used to avoid clutter when data labels are long or when there many items to compare.

Fig. 7.3 HIV treatment cascade, region Z, 2018

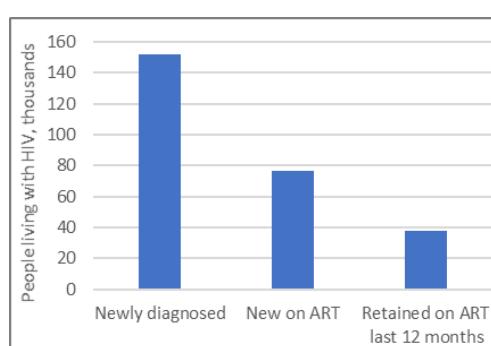
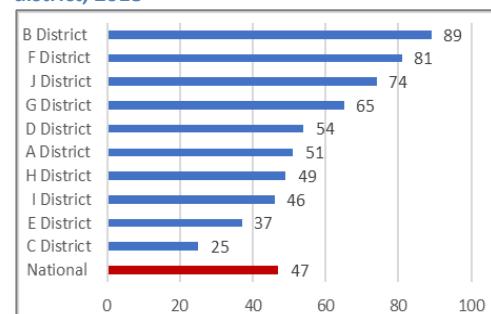


Fig. 7.4 Antiretroviral therapy (ART) coverage, by district, 2018



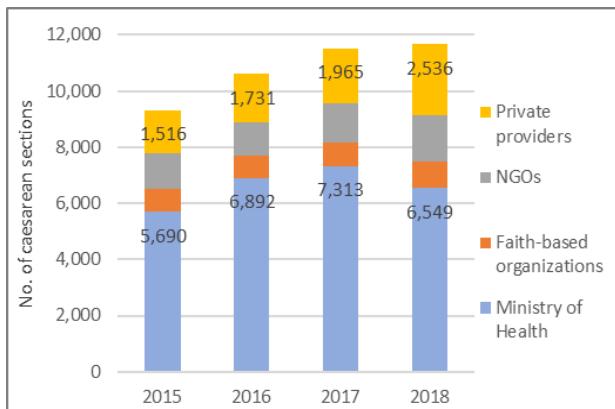
PLHIV: persons living with HIV

²¹ Evidence Based Health Management Manual. John Snow, Inc. and USAID. Enhancing strategic information project. March 2012 version.

- Stacked column and bar charts**

Stacked column or bar charts are useful for showing subcategories within the columns/bars.

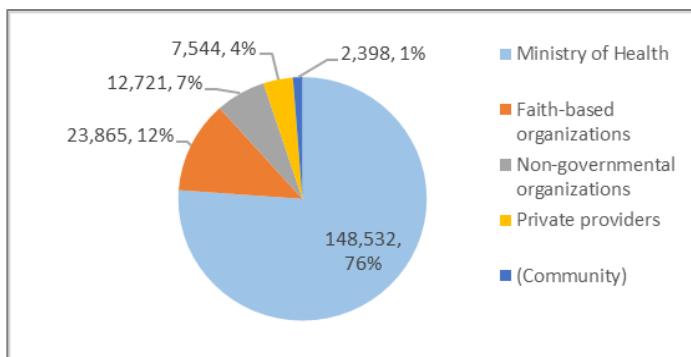
Fig. 7.5 Caesarean sections by provider, region X



- Pie charts**

Pie charts are used to show parts or percentages of a whole, with the segments totalling 100%.

Fig. 7.6 Deliveries in facilities by provider, region A, 2016



However, when pie charts contain more than about five segments, they may be difficult to understand quickly. In general, bar charts are preferred. For example, pie charts are sometimes used to display proportionate morbidity or mortality data (e.g. "top 10" causes of death or death), but a bar or column chart is a better choice, as illustrated in Figs 7.7 and 7.8.

Fig. 7.7 Leading causes of outpatient morbidity, clinic A, 2013

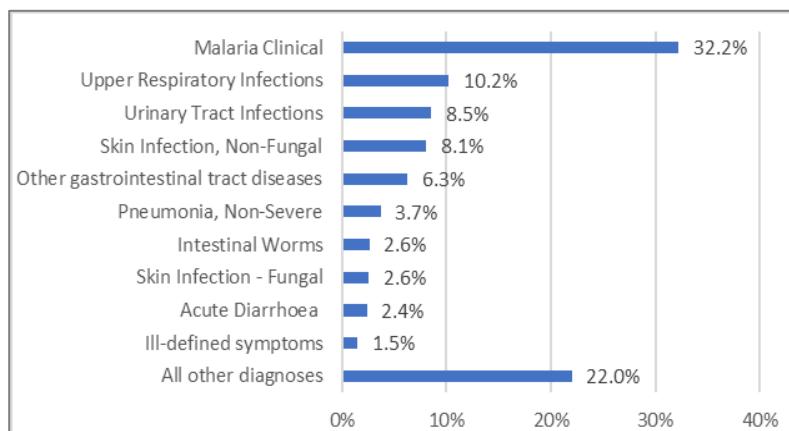
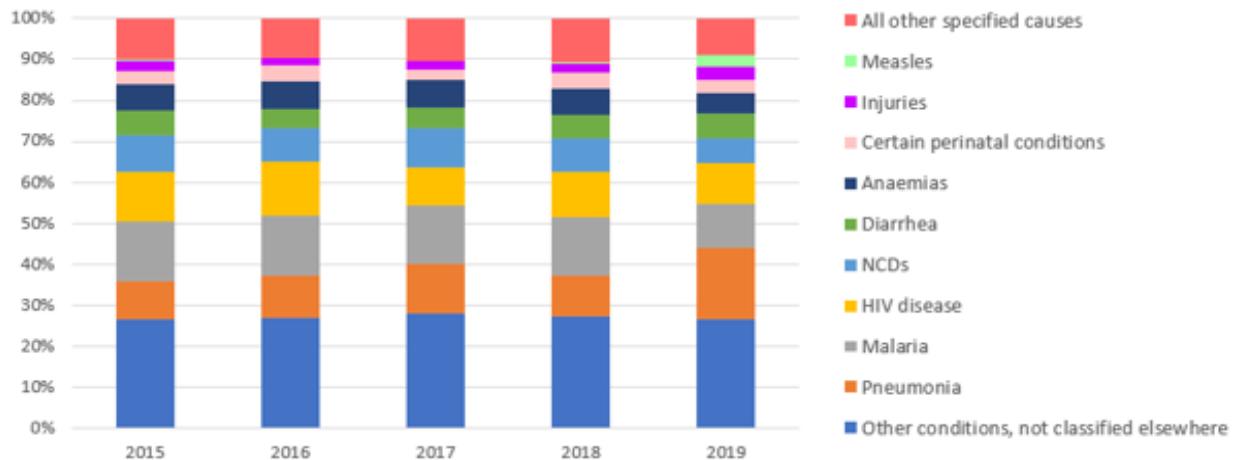


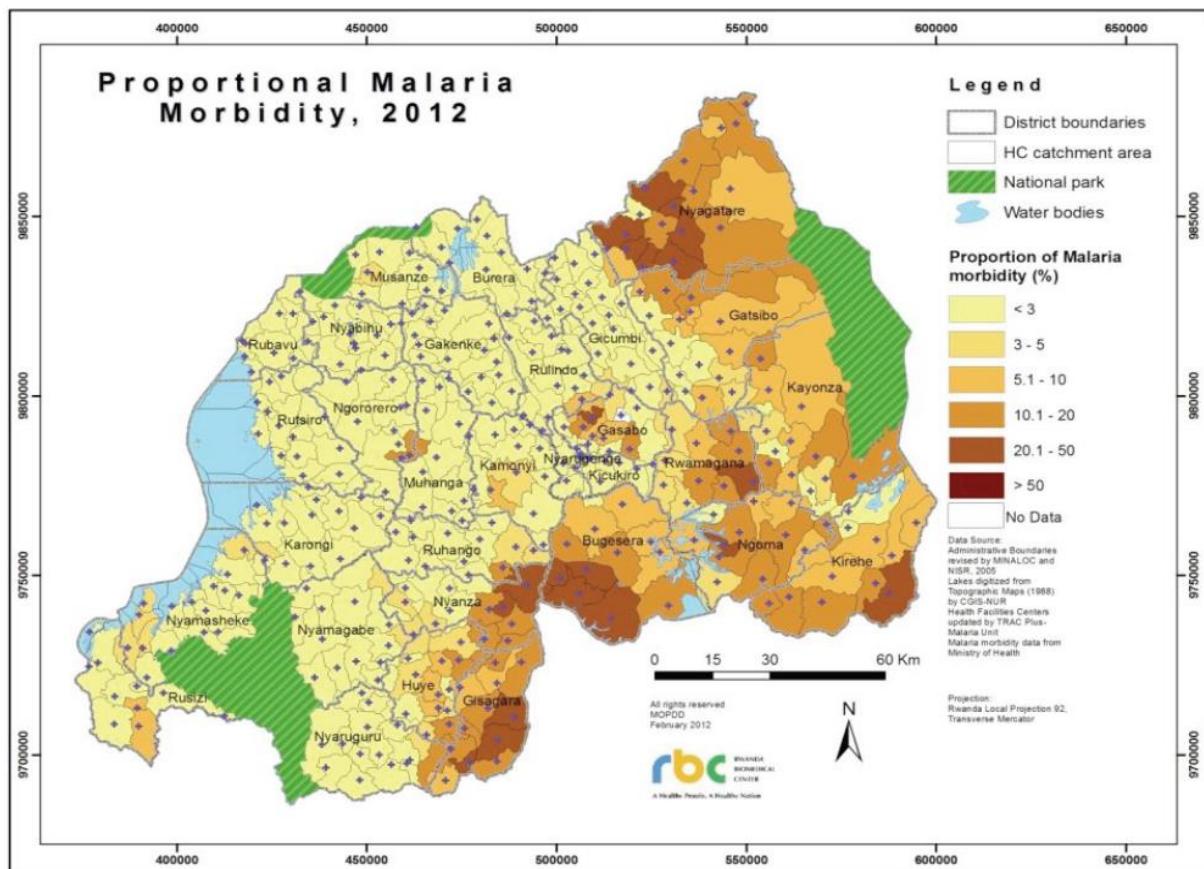
Fig. 7.8 Leading causes of outpatient morbidity, district X, 2015–2019



7.3 Maps

Maps can be used to show the locations of health facilities and outbreaks or other events, and to illustrate differences in indicator performance among geographic areas ("thematic" maps). In addition, the size or colour of points placed on a map can represent the value of a key indicator. However, maps can also mask the underlying data and should be displayed in conjunction with the numbers they represent as well as any potential limitations. Fig. 7.9 is an example of a thematic map.

Fig. 7.9 Percentage of outpatient visits attributed to malaria, Rwanda, 2012



Source: Rwanda annual health statistics booklet. Ministry of Health; 2012.

7.4 Presenting a story

The presentation of data as a “story” helps the reader to engage with various aspects of the situation illustrated by the data. This may involve, for example, the presentation of several charts showing trends of related indicators such as TB notifications, the percentage of TB cases tested for HIV, and the TB treatment success rate. It may also involve displaying together indicators from different programmes to provide an overall picture of the health situation. Various tools and approach may be used to create data stories.

■ Websites and dashboards

Computer or web-based data management systems can be used to display multiple tables, charts and maps on the same page. This is called a dashboard. The intention is that the “driver” (the user/reader) can glance at this dashboard to gain a quick impression of the performance of key indicators. Dashboards can be customized to the needs of the user.

For RHIS data, the visualizations in the dashboard are linked to the raw data. The dashboard is continuously updated as new data are entered, providing a “close to real-time” snapshot of performance. Fig. 7.10 provides an example of a dashboard in DHIS2.

Fig. 7.10 DHIS2 dashboard showing fictitious data for Sierra Leone



■ Summary measures

Analysts sometimes combine several indicators to create a summary measure or index. A small set of standard indicators representing a range of health service functions can be selected to provide an assessment of the overall performance of a national or subnational health system. The scores from this fixed set of indicators are combined mathematically (e.g. by averaging the coverage achieved with various services or by assigning scores to percentage categories). This generates an index that can be used for assessing year-to-year trends or for creating a “league table” that ranks subnational units.

It is important to note that while an index may serve as a useful measure for comparison, it is strongly influenced by the selection and weighting of the indicators within the calculation, and by the limitations of each indicator. The index should therefore always be assessed in conjunction with its component indicators.

In the example presented in Fig. 7.11, each indicator value is assigned a score of 1, 0 or -1 according to defined levels of performance. The different colours indicate whether an indicator is on track, in process or not on track, in each region. This enables quick identification of underperforming geographic areas as well as underperforming interventions. The “Score” column provides the overall performance index. The regions are ranked according to the values in this column.

Fig. 7.11 Regional league table, Ghana

Region	Score	Penta 3	ANC 4+	Skilled delivery	FP acceptors	OPD/capita	iMMR	TB treatment
Central	5	0	1	1	1	1	1	0
Upper East	5	0	1	1	1	1	1	0
Eastern	5	0	1	1	0	1	1	1
Western	4	1	0	1	0	1	1	0
Greater Accra	4	1	1	0	0	1	1	0
Upper West	3	-1	-1	1	1	1	1	1
Volta	3	0	0	1	0	1	1	0
Ahanti	3	0	0	0	0	1	1	1
Brong Ahafo	2	1	0	1	-1	1	-1	1
Northern	1	1	0	1	-1	1	-1	0

Table 7: Holistic assessment of regional performance in 2012

“In the regional analysis of ... 2011, three regions came out with a score of zero or below. In the current review all regions have a positive score, which indicates a relative improvement over 2011 for these selected service delivery indicators.”

Source: Holistic assessment of the health sector programme of work 2012. Ministry of Health, Ghana; June 2013.

iMMR: institutional maternal mortality ratio; OPD: outpatient department (visits).

League tables can be presented as tables or charts and can be used to display a single indicator or an index. While a target or benchmark may be included where relevant, league tables are often used to display data for which there are no established benchmarks. In such cases, the highlighting of higher and lower performing areas can provide the basis for further investigation into the reasons for the differences and the identification of best practices as well as problems. However, such comparisons may also be viewed as unfair because of underlying differences between the units that may influence the ranking, e.g. there may be substantial socioeconomic differences.

Stratified (grouped) rankings may be useful, especially if large differences exist within the country. For example, districts can be grouped according to urban, peri-urban or rural character of the district, or epidemiological characteristics (e.g. with and without malaria; HIV prevalence). Districts can also be classified into socioeconomic quintiles, according to an official classification from the national bureau of statistics.

▪ **Ten tips for presenting RHIS data**

1. On the cover page of a report, always specify the month and year in which the document was finalized.
2. Every table, chart and map must have a title and data labels.
3. Specify the time period and the geographic/administrative area for which the statistics apply.
4. Specify the data source – not only the publication or the organization that provided the data but also the data source itself.
5. Present and discuss notable findings about the data quality for every table, chart and map.
6. Explain the methods used for estimating denominators and include a table of key denominators.

7. When presenting charts showing coverage, also present recent survey estimates of the same indicator and, where possible, show the confidence interval for the survey estimates.
8. For each table, chart or map in the report, include a narrative that interprets the most important findings and discusses indicator definitions and any special limitations.
9. If a table extends over more than one page of a report, always print the headers at the top of each column on each page of the report.
10. If findings are to be projected on a screen (e.g. PowerPoint presentation) do not include any text or numbers of a font size smaller than 16 point.

8. Interpretation and use

8.1 Interpretation of RHIS data

Interpretation involves drawing conclusions from or “making sense of” the data. It entails understanding what the data mean within a specific context and in relation to other information, including the limitations of the data. Interpretation requires insights into the technical and management aspects of the health issues/services under review, as well as knowledge of the context in which events have taken place.

In a RHIS, there are usually processes (e.g. SOPs) in place for data collection and reporting, and sometimes also for data analysis. However, a systematic approach to data interpretation is often lacking. There is no standard approach to interpretation, but the process of understanding the data can be guided by applying six questions: what, where, when, who, how and why? These questions can be applied both to the health/disease context as well as to health services performance. Table 8.1 provides examples of uses of the epidemiological questions.

Table 8.1 Epidemiological questions and related data analysis concepts

Questions	Related data analysis concepts
<ul style="list-style-type: none"> • What are the most common diseases/conditions for which people use health services? • Where are health events occurring/services delivered? • When are the events occurring/services delivered? • Who is affected by the diseases/using the services? • How significant are the findings? • How are the services performing in relation to: <ul style="list-style-type: none"> ○ Defined requirements ○ Other locations ○ Other services ○ Quality standards ○ Equity (resource allocation/health outcomes) ○ Efficiency (resources used per output/population served) 	<ul style="list-style-type: none"> • Rates (mortality, morbidity) • Disaggregation by geographic location • Trends, seasonality • Disaggregation, e.g. by age, sex • Comparisons with baselines, thresholds, trends, other populations • Targets/benchmarks • Geographic comparisons • Programme comparisons • Quality indicators • Disaggregation: geographic location, age, sex • Comparisons: geographic; different services

Addressing the “**why?**”, RHIS data are descriptive. They provide a picture of what is happening but do not explain why it is happening. Additional information is needed to understand the reasons for the findings. The various programme-specific guidance manuals in the *Toolkit for analysis and use of routine health facility data* provide insights into possible reasons underlying the performance of specific indicators and may assist in guiding further investigation.

8.2 Uses of RHIS data

“...most efforts to strengthen health facility and community health information systems are focused on digitization, improving data quality and data analysis, and identifying problems. But the ultimate goal of RHISs is that information is used to **solve problems** and to improve access to and delivery of quality health services...”²²

²² Lippeveld T. Routine health facility and community information systems: creating an information use culture. Glob Health Sci Pract. 2017;5(3):338-340 (<https://doi.org/10.9745/GHSP-D-17-00319>).

Decision-makers at different levels of the health system have different information needs and require different levels of detail, based on the functions of each level. At the national level, RHIS data contribute to national annual reports of health statistics and are used in periodic analytical reviews of health system performance. Evidence from RHIS data can be used to inform priority-setting, planning, resource allocation and policy-making.

At subnational (e.g. district) level, RHIS data are ideally presented and discussed at monthly or quarterly district management meetings. District managers can use RHIS data to follow overall district trends, identify problems needing immediate response (e.g. disease outbreaks) and compare performance among facilities. RHIS data can also be used to inform district service improvement, planning, resource allocation and management. At facility level, uses of RHIS data include disease surveillance, facility performance monitoring, work planning and quality improvement. Table 8.2 provides examples of information uses and needs at various health system levels.

Table 8.2 Examples of information uses and needs by health system level

Level	Information uses	Information needs
National	<ul style="list-style-type: none"> • Policy and strategy decisions • Health programme planning and management • Prioritization • Resource allocation • Ensuring equity and efficiency in distribution of services and allocation of resources (inputs) across the country • Disease surveillance • Monitoring and evaluation • Capacity building • Research and innovation 	<ul style="list-style-type: none"> • Events for immediate response (e.g. surveillance data) • Mortality and morbidity: national/subnational indicators • Access: national/subnational indicators • Service coverage: national/subnational indicators • Quality of care: national/subnational indicators • Health service resource management data: finance, workforce, infrastructure, medicines and commodities • Inputs versus outputs/outcomes • Data for global and regional reporting
District	<ul style="list-style-type: none"> • District health service planning and management • Resource management • Ensuring equity and efficiency in distribution of services and allocation of resources across the district • Achieving district targets • Monitoring mortality and morbidity; disease surveillance • Monitoring district and facility performance (access, coverage, quality, safety, efficiency) • Supervision • Capacity building 	<ul style="list-style-type: none"> • Events for immediate response • Mortality and morbidity: district incidence and absolute numbers; facility absolute^a numbers • Access: district indicators; facility absolute numbers • Service coverage: district indicators; facility absolute numbers • Quality of care: district and facility indicators • Health service resource management data • Inputs vs outputs/outcomes <p>^aAssumption: accurate denominator data are not available for facility level.</p>
Facility	<ul style="list-style-type: none"> • Facility work planning and management • Resource management • Ensuring service availability and quality • Achieving facility targets • Monitoring facility trends (mortality, morbidity, utilization, outputs, quality) • Alert and response to potential public health threats • Response to community needs • Client/patient care 	<ul style="list-style-type: none"> • Events for immediate response • Mortality and morbidity: absolute numbers • Access: utilization numbers • Coverage: utilization/output absolute numbers • Quality of care: facility indicators • Health service resource management data • Individual patient/client data (characteristics, health care needs, services delivered, outcomes, follow up)

Annex 1. Adjustment for incomplete reporting

Reporting rates are calculated for all relevant RHIS reporting forms and indicators. The expected number of reports may vary among forms, depending on the number of health facilities offering the services. Knowledge of reporting rates is particularly important in the calculation of intervention coverage and service utilization rates.

Regardless of how reporting completeness is used in coverage calculations, there is always an underlying assumption about the extent to which the non-reporting facilities are providing services. If no adjustments are made, it is assumed that no services were provided by these facilities. However, poor reporting does not necessarily mean that services were not provided. The quantity of services provided in non-reporting facilities may be less than those provided by reporting facilities or, in some cases, more. For example, large non-reporting hospitals may provide substantial quantities of services.

If the completeness of reporting varies over time (e.g. 75% in 2015 and 85% in 2016), the coverage trend may be affected, even if there is no true change. Therefore, it is recommended to assess the impact of adjustment on levels and trends of coverage indicators. The adjustment is based on the assumption that non-reporting facilities are in fact providing services. The analyst can make decisions about the adjustment based on comparison with survey data or based on knowledge of the non-reporting facilities compared with the reporting facilities. Adjustment involves the selection of an adjustment factor and may be expressed as follows:

$$p_{\text{adjusted}} = p_{\text{reported}} + p_{\text{reported}}(1/(c)-1) \times k$$

where: p = number of service outputs; c = reporting completeness; k = adjustment factor.

If missing reports are considered an indication that zero services were provided during the reporting period, the value of $k = 0$ and no adjustment is made. In some cases, services may have been provided, although not at the same level as in prior reporting periods, in which case the apparent incomplete reporting is an indication of a lower level of service provision; in this case, k is between 0 and 1. In other cases, it may be assumed that services were provided at the same level in non-reporting facilities as in reporting facilities; in this case k equals 1. Box 1 illustrates potential adjustment factors.

Important considerations in the selection of a value for k are the extent to which large health facilities and private health facilities are reporting and providing the specific services. This is likely to be different for different services, resulting in different adjustment factors. The selection of the most likely value of k can be done through a comparison with survey results, by selecting a value of k that brings the adjusted health facility statistic close to the survey statistic for a particular year.

Box 1. Adjustment for non-reporting facilities

$p_{\text{adjusted}} = p + p(1/(c)-1) \times k$ where: p = number of service outputs; c = reporting completeness; k = adjustment factor.

- $k = 0$: no services in non-reporting facilities
- $k = 0.25$: some services, but much lower than reporting facilities
- $k = 0.5$: half the rate compared with reporting facilities
- $k = 0.75$: nearly as much as reporting facilities
- $k = 1.0$: same rate of services as reporting facilities.

Annex 2. Triangulation of data

Sometimes the most valid estimate may not be available from a single data source but rather through the “triangulation” of findings from multiple data sources and methods, each of which provides a partially valid picture. For example, WHO’s estimates of trends in the incidence of malaria and TB are derived from such triangulation. The combination of survey and health facility data estimates of key indicators can provide a good picture of levels and trends in, for example, coverage. Triangulation draws upon the strengths of each data source. Population-based survey statistics are less frequently available but have greater accuracy than RHIS data, whereas RHIS data are available continuously and for all levels of the health system.

The following example illustrates the value of triangulation. The country profile for Rwanda in the *World malaria report 2014* includes the following two charts. Fig. A2.1 presents the trend, between 2000 and 2013, of confirmed cases of malaria reported per 1000 population per year. Also shown is the trend of the annual blood examination rate (ABER) – the number of laboratory tests performed to confirm malaria per 100 population per year. Notice how the reported confirmed cases increased and decreased in parallel with the ABER.

Fig. A2.1 Confirmed cases of malaria reported per 1000 population

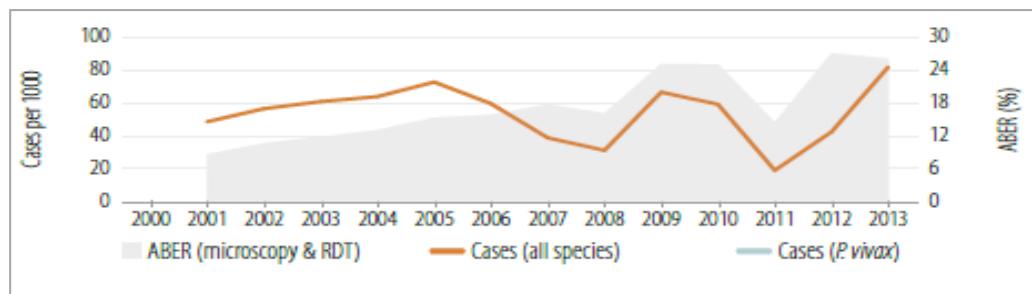
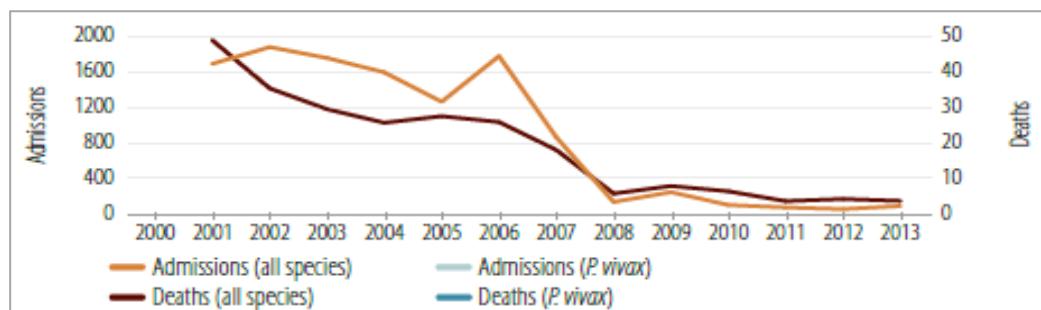


Fig. A2.2 presents trends in the incidence of hospital admissions and inpatient deaths due to malaria per 100 000. Both admissions and deaths from malaria declined markedly since 2006. Based on the trends in admissions and deaths, the *World malaria report 2014* found that there was sufficient evidence to conclude that the true incidence of malaria in Rwanda had declined by more than 75% since 2000, even though confirmed cases reported increased markedly from 2011 to 2013. The conclusion that there was a marked reduction in malaria incidence is supported by survey findings that the prevalence of parasitaemia among children age 6 to 59 months fell from 2.6% in 2007–2008 to 1.4% in 2010.

Fig. A2.2 Trends in hospital admissions and inpatient deaths from malaria per 100 000, Rwanda, 2000–2013



Annex 3. Estimation of population targets

Population targets for maternal health and immunization services (estimated pregnancies, estimated live births, estimated surviving infants) are typically provided by the national statistical office and are based on analysis of data from a national population census, supplemented with data from a nationwide household survey such as a DHS or MICS. However, as described in Chapter 5, various challenges may be associated with population projections and health service target populations. Therefore, it is useful to compare official population projections with estimates derived from RHIS data. Such comparisons must be done only after adjustment for incomplete reporting of RHIS data, as described in Annex 1.

First antenatal visits (ANC1): Where national ANC1 coverage is 95% or higher (according to population-based surveys), coverage is often high in almost every district in a country. This implies that if the number of ANC1 visits is accurately reported in the RHIS, the reported numbers of ANC1 visits can be used as an independent estimate of the target population – the actual number of pregnant women. A small proportion should be added to take into consideration non-attendance of ANC. For example, if ANC1 coverage is 95% in the survey, the target population for ANC1 is:

$$N_{\text{pregnancies}} = N_{\text{ANC1,reported}} / \text{survey ANC1 coverage}^{23} = N_{\text{ANC1,reported}} / 0.95$$

First DTP/pentavalent vaccine doses given (DTP1): DTP1 coverage is also almost universal in many countries and has been high for many years. This implies that, after adding a small proportion for non-immunized children, the number of children that received DTP1 should be a good indicator of the number of surviving infants, i.e. those live births that have survived the neonatal period.

$$N_{\text{infants}} = N_{\text{DTP1, reported}} / \text{DTP1 coverage}$$

BCG vaccinations: BCG vaccination should be given at birth. In some countries the reported number of BCG vaccinations can be a good indicator of the number of births. However, the number of BCG vaccinations in the RHIS may be higher than expected, possibly due to older children being recorded as receiving BCG vaccination.

Comparison of official census population projections with target populations derived from reported ANC1 and DTP1 numbers can provide information about the denominators needed for indicator calculations. Consistency between the two sources is generally an indication of good quality of both the census population projections and the facility reports for ANC1 and DTP1. If the two sources are not consistent, there are two possible explanations:

- If the health facility reports are considered complete and accurate, the population projections are inaccurate. In this case it is best to use the health facility reports as the basis for the estimation of target populations.
- If the population projections are considered more accurate than the RHIS reports, the size of the difference between the numbers is an indication of the quality of the health facility reports.

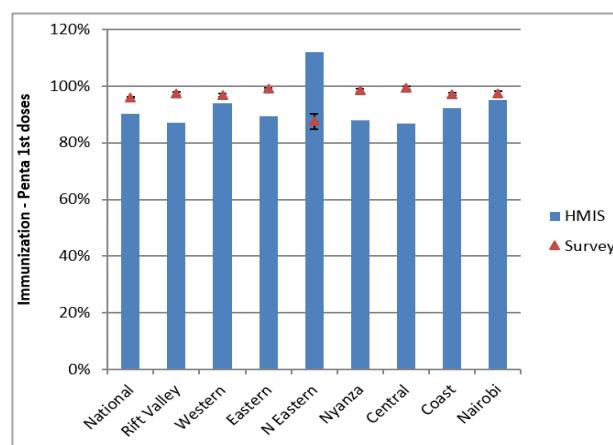
The ultimate choice of the denominator is often a qualitative judgment. Objectively, it is possible to calculate the population denominator with the smallest variation compared with the coverage statistics derived from the surveys. If the differences are small, it is best to use the population projections.

²³ As an alternative, the service data (ANC1 visits or DTP1 doses) can be adjusted by dividing by the survey estimate of coverage for the specific geographic area rather than the survey estimate of the nationwide coverage.

However, consistency at national level does not necessarily mean that these denominators work well for subnational units such as districts. Ultimately, the choice needs to be based on district analysis, which may lead to the identification of groups of districts where the population projections do not work well as target populations. For example, a large number of districts with DTP1 and ANC1 coverage of over 100%, or a substantial number that have coverage that is much lower than expected, are reasons to consider using facility data derived denominators.

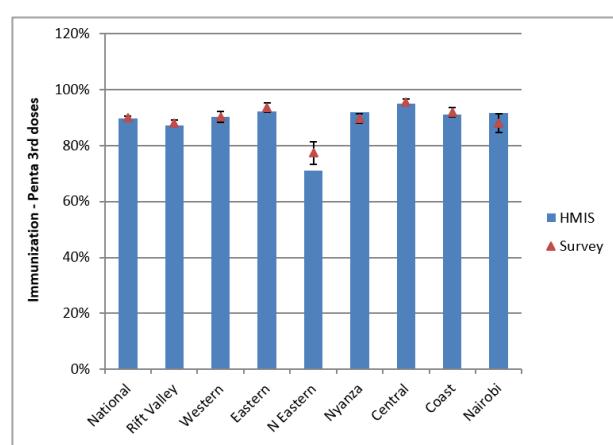
Analysts should be aware that such use of service data to estimate the size of the target population can modify conclusions reached about which districts are strong performers and which districts are weak performers. Figs A3.1 and A3.2 illustrate the use of alternative denominators based on DTP1/pentavalent vaccine first dose (Penta 1).

Fig. A3.1 Penta 1 coverage: DHS results versus HMIS estimates using official surviving infant denominators



Kenya's 2014 DHS found Penta 1 coverage to be greater than 95% in all regions except the North Eastern Region where the coverage was 88%. In contrast, as shown by Fig. A3.1, when the official estimate of the number of surviving infants was used as the denominator, routine data from 2013 suggested a Penta 1 coverage of 112% for this region. Penta 1 coverage in most other regions was substantially lower than that found through the DHS. There were similarly large discrepancies between the DHS and HMIS estimates of Penta 3 coverage.

Fig. A3.2 Penta 3 coverage: DHIS results versus HMIS estimates using alternative denominators



For the mid-term analytical review of Kenya's Health Sector Strategic Plan (2009–2015), a team of analysts chose to use an alternative estimate of surviving infants that was based on the number of Penta 1 doses administered in each region. As shown by Fig. A3.2, when this alternative denominator was used, Penta 3 coverage estimates based upon routine (HMIS) data closely matched the Penta 3 coverage estimates measured by the DHS.

World Health Organization
20, Avenue Appia
1211 Geneva 27
Switzerland

Contact: healthinfo@who.int