ANALYSIS AND USE OF HEALTH FACILITY DATA

General principles

WORKING DOCUMENT, FEBRUARY 2018

World Health Organization

ANALYSIS AND USE OF HEALTH FACILITY DATA

General principles

WORKING DOCUMENT – FEBRUARY 2018

World Health Organization

MODULE 1. **General principles**

LEARNING OBJECTIVES

This module provides an overview of the analysis curriculum, its objectives and structure and key principles that are relevant for all managers and analysts of health facility data. By the end of this module, participants will be able to:

- Understand the objectives and structure of the curriculum for the analysis of routine health facility data;
- Examine core indicators, based on international standards, recommended for collection through routine health information systems;
- Have a basic understanding of key data quality checks;
- Examine quality of population estimates/denominators and learn how to calculate alternate denominators;
- Have a basic understanding of key analytical concepts;
- Learn key principles for presentation and communication of data.

AUDIENCE

This module is relevant for different members of the health workforce including: (1) Policymakers and managers at different levels of the health system; (2) Monitoring and evaluation staff and analysts at different levels of the health system; (3) Trainers/facilitators who will support in-country training.

SUGGESTED REFERENCES

- World Health Organization (2018). Global Reference List of 100 Core Health Indicators (plus health-related SDGs). Geneva, Switzerland: WHO. Retrieved from http://www.who.int/healthinfo/indicators/2018/en/
- World Health Organization (2017). Data Quality Review (DQR) Toolkit. Geneva, Switzerland: WHO. Retrieved from http://www.who.int/healthinfo/tools_data_analysis/dqr_modules/en/
- WHO (2018). Reference list for HMIS Indicator List (draft). Geneva, Switzerland: WHO
- International Health Partnership + Related Initiatives (IPH+) and World Health Organization (WHO). (2011). Monitoring, evaluation and review of national health strategies: A country-led platform for information and accountability. Geneva, Switzerland: WHO. Retrieved from http://www.who.int/healthinfo/country_monitoring_evaluation/documentation/en/
- MEASURE Evaluation. (2011). Tools for data demand and use in the health sector: Framework for linking data with action. Chapel Hill, NC, USA: MEASURE Evaluation, University of North Carolina. Retrieved from http://www.cpc.unc.edu/measure/resources/publications/ms-11-46-b
- USAID, PEPFAR, World Health Organization (2018). Master Facility List Resource Package: Guidance for countries wanting to strengthen their MFL (pre-final). Washington, DC.

KEY AUTHORS

Kathryn O'Neill | Chelsea Taylor |Kavitha Viswanathan |Robert Pond |Ties Boerma

Contents

Acknowledgements

This guidance document has been developed by the World Health Organization, 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.

1. Introduction: objectives and structure of the curriculum

WHY CREATE A CURRICULUM FOR ANALYSIS AND USE OF ROUTINE FACILITY DATA?

In June 2015, at the Measurement for Accountability for Health Summit, 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."[1] This publication presents guidelines for managers and analysts to review, analyse, and present the findings from data that are routinely reported by health facilities[2]. The core purpose of this curriculum is to provide an overview of the most promising methods and tools that can be used to analyse and use facility data to assess the health care system as the level of facilities

WHAT IS MEANT BY ROUTINE HEALTH FACILITY DATA?

Routine health facility data are collected at clinics, hospitals and other health service points (public; private; community-based) at the time that services are provided. These data are processed at the health facility and summary reports are sent to the appropriate administrative authority. The system for collection, management and reporting on these routine data is sometimes referred to as the Health Management Information System (HMIS).

USES AND VIRTUES OF ROUTINE HEALTH FACILITY DATA

Routine health facility data are widely used for national and sub-national health sector reviews and planning. They form the basis of national annual reports of health statistics and periodic analytical reviews of health system performance, and they are used to assess health program at all levels of the health system. Analysts and program managers use routine health facility data to measure levels, study trends and assess geographic differences for a range of standard health indicators related to service delivery, coverage of interventions and the leading diagnoses among clients. (Analysts often construct summary indices by aggregating several of these indicators in order to compare health system performance over time or among regions or districts. Unlike periodic population and facility surveys which take place every few years using a limited sample, regular collection and analysis of routine health facility data provides frequent and current assessments of population health at sub-national (e.g. individual districts) level.

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

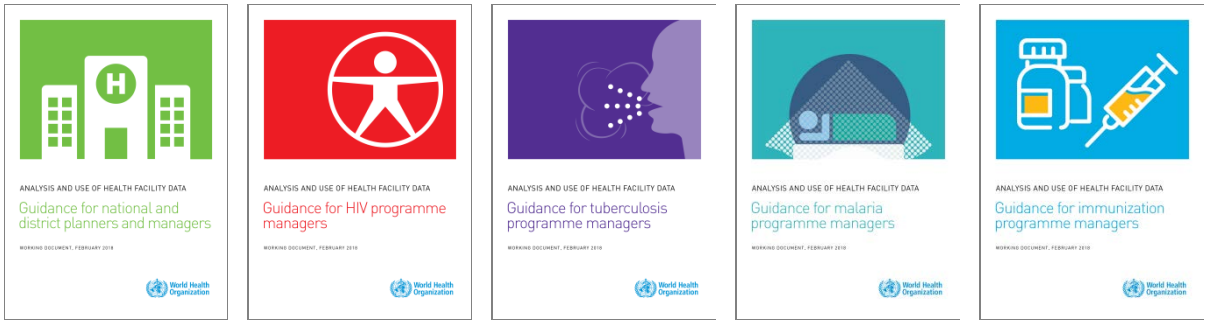
[2] In addition to data that are routinely collected and reported, information is collected through occasional facility assessments and surveys. Together these data constitute a Health Facility Information System which is one component of a broader health information system (HIS). The HIS brings together data from multiple sources, including from health facilities, household surveys, censuses, civil registration systems, surveillance systems, and other administrative data sources. For further discussion see the WHO document Facility Information Systems Resource Kit. For elaboration on the components of an HIS see the HMN Framework (http://www.who.int/healthmetrics/documents/hmn_framework200803.pdf)

HOW IS THE CURRICULUM STRUCTURED?

This curriculum is divided into seven key modules. The first two are cross-cutting. This introduction provides an overview of the basic steps and standards for analysing health facility data. These steps and standards relate to the selection of appropriate indicators, ensuring data quality, key analytic concepts, and effective presentation and communication of the results of analyses. Module 2 provides guidance relevant to all district and national planners and managers. This encompasses coverage and quality of care; inputs and outputs of the system; and morbidity and mortality.

The remaining modules are designed for specific programme areas and provide guidance for TB, HIV, malaria, immunization, and RMNCAH.

Each module introduces the core indicators that should be collected, provides proposals for assessing data quality, presents suggested analyses, including consideration for interpretation, and gives details on reference documentation and further reading.



2. Data quality

All data have limitations that affect their reliability and interpretation; facility data are no exception. Before conducting analysis and interpretation, the analyst should review the facility data for completeness and quality to determine inconsistencies and errors and make adjustments if necessary. Data quality reviews can be completed as a 1) "desk review" or 2) involve a field investigation. A **desk review** of data quality often focuses on checking and analysing the statistics aggregated and reported by each district. A **field investigation** is a more extensive and revealing review that disaggregates data by facility and by month for the period of analysis and helps define strategies for improving data quality. A field investigation includes a survey of a sample of districts and health facilities to determine the extent to which reported data match the data collected in source documents (i.e. facility registers and tally sheets). If time and resources permit, the desk review should be complemented by field investigation because it will provide insights into the effectiveness of the data management system and inform data quality improvement strategies as well as determine data quality.

WHO, with partners, has developed a toolkit for conducting data quality checks. This toolkit, Data Quality Review (DQR) Toolkit, has recommended guidelines, data collection and analysis tools (see Key References section). The following section gives a brief overview of the key metrics included in the DQR guidelines document. Please refer to this document for a more in-depth discussion of each metric. Additionally, modules 2 to 7 include data quality checks pertinent to that specific programme area.

There are four dimensions to consider when assessing data quality: completeness, internal consistency, external comparisons and external consistency of population data. Each of these is described below.

BASIC DATA QUALITY ASSESSMENT

Assess reporting completeness

Completeness is the percentage of expected reports[3] which have been submitted to a higher level. The analysts should assess both the completeness of facility reports (submitted to district level) and the completeness of district reports (aggregated data from multiple facilities which have been submitted to the national level). To key steps in assessing completeness are:

Assess completeness of reporting of each form and each key data element

Different forms are often used to report different types of services. For example, outpatient morbidity is often reported separately from immunizations, antenatal care, etc. The number of facilities delivering each type of service and expected to report each form (which determines the denominator for calculating completeness) may vary from one service to another. The number of reports submitted (the numerator) may also vary by form.

Each cell of a reporting form is called a data element. Faced with the burden of completing numerous cells, a substantial percentage of health workers may consistently leave certain cells blank, using only a sub-set of cells to report all the data[4].

[3] For example, 12 monthly reports are expected per facility per year.

[4] In one country, data on administration of doses of DTP vaccine was split into 24 separate data elements to provide data disaggregated by dose sequence number, gender, <1 versus ≥1, and inside the catchment area versus outside the catchment area of the health facility. Review of the data for this country showed that, in the course of 2014, 59% of health facilities filled in only half or fewer of the vaccination cells each month and consistently left the others blank. For example, 37% of health facilities never reported giving DTP to a child who was 12 months or older. This has obvious implications for the validity of an analysis which depends upon these disaggregated data elements to estimate, for example, immunization coverage by 12 months of age.

Assess completeness of reporting from hospitals and from private sector facilities

Hospitals report the great majority of inpatient deaths and admissions and a significant percentage of outpatient services. Yet, in some health systems, the completeness of reporting is significantly lower from hospitals than from health centres and health posts.

While private, not-for-profit facilities may reliably report routine data, this is often not true of for-profit health facilities. Especially in cities, such for-profit facilities may account for a significant percentage of select services such as delivery care. Assessment of the completeness of reporting from such private facilities begins with a robust inventory of all facilities as part of efforts to prepare a Master Facility List for the country. WHO has developed guidance for developing and maintaining a Master Facility List[5].

Internal consistency of reported data

Internal consistency of the data relates to the coherence of the data being evaluated. Internal consistency metrics examine: 1) coherence between the same data items at different points in time (outliers and consistency over time), 2) coherence between related data items (consistency between indicators), and 3) comparison of data in source documents and in national databases.

Data entry errors can occur either when a paper form is first completed or when the data is transcribed such as when it is entered into an electronic database. Some data entry errors can be identified by screening for outliers – values that are more than two or three standard deviations from the mean. Fortunately, most outliers are so small that they contribute much less than 5% to the annual district total and an even lower percentage to the regional total. A spreadsheet can be used to rapidly identify the outliers that are large enough to have a major influence on the district value of the indicator. It is then practical to carefully investigate each large outlier to determine whether, as with the above example, there is strong evidence that the value is erroneous.

Time permitting, the analysts could investigate a larger number of outliers. Investigations might involve communications with local staff at district and facility level or comparison, for specific health facilities, of records of doses administered with records of commodities supplied. This would be even more practical if investigations were conducted by district staff themselves as part of a monthly exercise.

Investigation of outliers is important for “cleaning” of the dataset prior to analysis. Such investigation is also important to identify regions, districts (see the above example) and health facilities with significant data problems.

Data from multiple years not only permit assessment of trends but also assessment of the internal consistency of the data. When an indicator fluctuates by 10% or more from one year to another and/or when the trend is not consistent in one direction, analysts should consider the possibility that the changes observed reflect data quality problems rather than valid trends.

Some services such as delivery of a third dose of DTP vaccine are preceded by another service such as delivery of the first dose of DTP vaccine. Some clients will receive the preceding dose (DTP1) but then fail to receive the third dose (DTP1) – the clients will “dropout”. Hence, DTP1 should be greater than DTP3. When DTP1 – DTP3 is negative this is called “negative dropout”. Negative dropout is a sign of poor data quality. Such findings should be discussed when presenting the results from analysis of the respective indicator.

[5] Master Facility List Resource Package: Guidance for countries wanting to strengthen their MFL (pre-final). USAID, PEPFAR, World Health Organization (2018).

External consistency with other data sources

The level of agreement between two sources of data measuring the same health indicator is assessed. The two sources of data usually compared are data flowing through the HMIS or the programme-specific information system and data from a periodic population-based survey. The HMIS can also be compared to pharmacy records or other types of data to ensure that the two sources fall within a similar range.

External comparison of population data

This involves determining the adequacy of the population data used in evaluating the performance of health indicators. Population data serve as the denominator in the calculation of a rate or proportion and provide important information on coverage. This data quality measurement compares two different sources of population estimates (for which the values are calculated differently) in order to ascertain the level of congruence between the two. If the two population estimates are discrepant, the coverage estimates for a given indicator can be very different even though the programmatic result (i.e. the number of events) is the same. The higher the level of consistency between denominators from different sources, the more likely it is that the values represent the true population value. A more detailed section on population denominator follows.

3. Core facility indicators

INTERNATIONAL STANDARDS FOR RHIS FACILITY INDICATORS

Analysis and use of routine facility start with the indicators and the related data elements. This module presents a recommended list of core indicators for health management information systems (HMIS) that include programme-specific indicators. Countries can add or modify this recommended list based on their country priorities and epidemiological profile. A country can use this list to see if their HMIS include these standard indicators or if they have the relevant indicators if the metadata correspond to international standards.

The indicators included in this list have been adapted from WHO's Global Reference List of 100 Core Health Indicators (see Key References section) that can be measured at the facility level as well as other key program indicators that are part of country commitments for both global and national monitoring (references for these key documents can be found in the individual program modules). This list includes indicators/data elements that are reported each month/quarter or data gathered through sentinel sites, disease surveillance systems or annual inventories of health infrastructure and human resources.

These indicators have also been mapped to the results chain framework (Figure 1). Such a mapping allows either a national or sub-national manager to see how they can select indicators in the different domains (Inputs, Outputs, Outcomes, Impact) for monitoring their national or programme-specific performance.

Figure 1: Results Chain Framework for HMIS Core List of Indicators

Inputs and processes	Output	Outcome	Impact
Health workforce Health worker density and distribution **Health infrastructure** Health facility density and distribution Hospital bed density Bed occupancy rate Functional Status of Cold Chain Equipment Temperature Alarms **Health information/governance** Birth registration Death registration Completeness of reporting by facilities Documented birth weight	**Service access and availability** Outpatient Department Service Utilization Inpatient Service Utilization Surgery Rate Caesarean section rate Annual blood examination rate TB notifications vs cohort Service-specific availability and readiness Full availability of vaccines and supplies Full availability of malaria control commodities Access to a core set of relevant essential medicines Density of medical devices and essential technologies Vaccine Wastage (Open Vial) Vaccine Wastage (Closed Vial) **Service quality and safety** Perioperative mortality rate Institutional maternal mortality ratio Institutional stillbirth rate Pre-discharge neonatal death rate Institutional neonatal mortality rate Institutional under-five mortality rate Maternal death reviews Neonatal death reviews PLHIV retained on ART for 12 months PLHIV retention rate over 12 months (%) Ratio of new on ART to newly diagnosed Registered new and relapse TB patients with documented HIV status Registered new and relapse TB patients with documented HIV-positive status TB case notification and rates Incident TB case notification and rates TB treatment success rate TB treatment success rate in new and relapse HIV positive patients TB treatment success rate in RR-/MDR-TB patients PLHIV tested that are virologically suppressed (%) Proportion of malaria cases with symptoms diagnosed within 24 hours Proportion of malaria cases notified within 24 hours of diagnosis Proportion of malaria cases investigated Proportion of malaria cases classified Proportion of malaria foci investigated Proportion of malaria foci classified Adverse events following immunization (AEFI)	**Coverage of interventions** Antenatal care coverage ANC syphilis screening coverage ANC clients tested for HIV PMTCT testing coverage rate Pregnant women receiving 2 doses of Tetanus Toxoid Intermittent preventive therapy for malaria during pregnancy (IPTp) Institutional delivery coverage Postpartum care coverage – women Postpartum care coverage – newborn PMTCT coverage rate during breastfeeding Received antibiotics for pneumonia Oral rehydration solution (ORS) and Zinc for diarrhoea Appropriate treatment among children treated for malaria Malaria cases given ACT Immunization coverage rate by vaccine for each vaccine in the national schedule Immunisation session completion rate Number of patients tested for malaria HIV tests performed DPT Dropout rate BCG Dropout rate Measles Dropout rate PLHIV new on ART PLHIV currently on ART HIV-positive new and relapse TB patients on ART during TB treatment HIV-positive new and relapse TB patients on CPT during TB treatment TB patients with results for drug susceptibility testing Confirmed RR-/MDR TB cases enrolled on MDR-TB treatment regimen Unconfirmed RR-/MDR TB cases enrolled on MDR-TB treatment regimen **Risk factors and behaviours** Early initiation of breastfeeding Incidence of low birth weight among newborns Facility distribution of mosquito nets	**Health status** Adolescent birth rate Leading discharge diagnoses (by discharge) Leading discharge diagnoses (by population) New cases of vaccine-preventable diseases Sexually transmitted infections incidence rates New cases of IHR-notifiable and other notifiable diseases New cases of neglected tropical diseases Cancer incidence, by type of cancer HIV test positivity HIV tests positive Malaria test positivity Malaria diagnostic testing ratio ANC clients who are HIV postive and are on ART before pregnancy TB drug susceptibility testing results Confirmed outpatient malaria diagnoses Presumed outpatient malaria diagnoses Incidence of outpatient malaria Inpatient malaria diagnoses Incidence of malaria admissions Distribution of cause of death in health facilities Inpatient malaria deaths Incidence of inpatient malaria mortality Cause-specific case fatality rates for major causes Deaths from neglected tropical diseases Proportion of malaria foci classified as active Proportion of malaria cases which are indigenous versus imported Proportion of malaria foci with zero local cases % of positive malaria tests with P. falciparum

4. Population estimates/denominators

Denominators, estimates of the affected or target population, are required to derive rates (e.g. disease incidence per 1,000 population per year) and coverage (e.g. % of infants vaccinated) from the routine data reported by health facilities. Population estimates are typically based upon projections from the most recent national population census. The reliability of these projections declines as years pass since the last census. Due to internal migration (especially in rapidly developing countries or those affected by some crisis), estimations of the populations of administrative divisions (i.e. regions and districts) become especially unreliable with the passage of time.

Whichever denominators are used, the methods and assumptions for estimating should be presented along with the rest of the analysis. A table of estimates of the key denominators (total population, children less than 5, infants, pregnancies, surviving infants, etc.) by geographic region should be included in the full report of the analysis (as shown in the following example in Table 1). The assumptions used to calculate the denominators should be explained. Where denominators are based upon projections of census figures, the annual growth rate should be stated explicitly.

Table 1: Extract from a table of denominators used to calculate core indicators

Sub-populations during 2012, by region of Tanzania. Based upon projections of the 2002 national census.

Region	Annual growth rate	Total population	Pregnancies = Births	Births	Surviving infants = Birth * 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 that the assumption that the number of pregnancies = the number of births will cause the ANC coverage to be somewhat under-estimated.

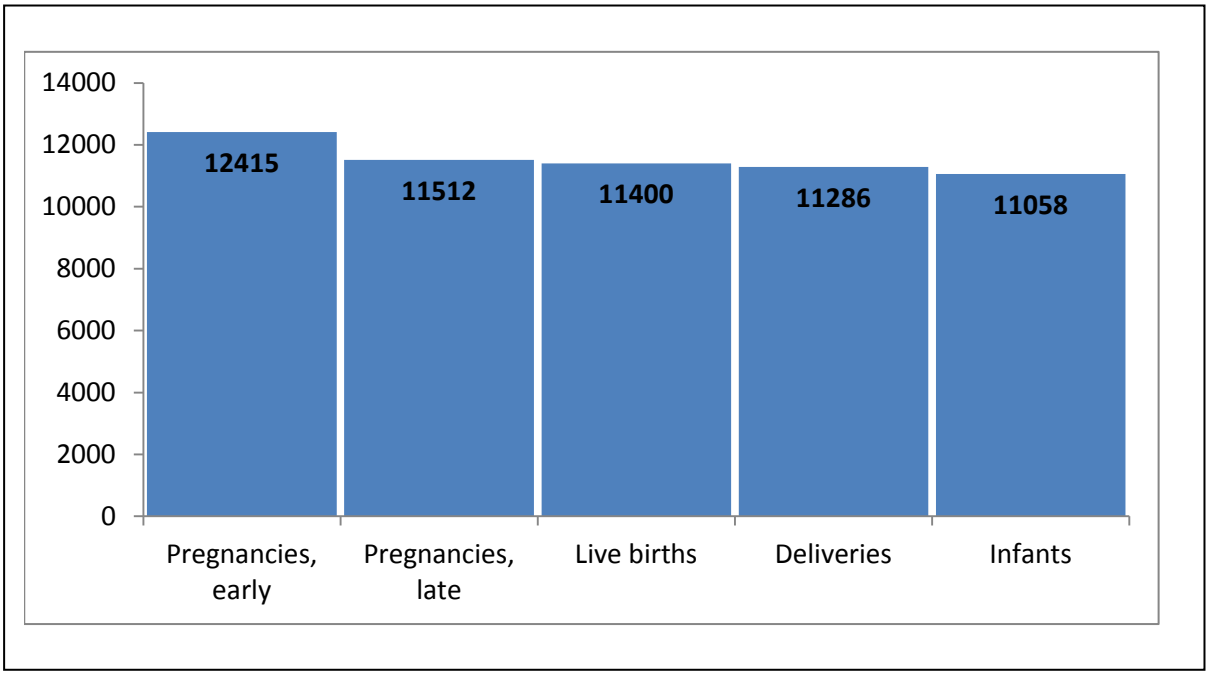
Source: Mid-term Analytic Review of the Health Sector Strategic Plan II

A good way to estimate the number of births is to multiply the total population of an area by the best estimate of the Crude Birth Rate (CBR)[6]. Whichever method of estimation is used, the denominators used to calculate coverage with antenatal care, institutional delivery, post-natal care and childhood immunization must be mutually consistent:

- As the denominator for calculating coverage with first ANC visits prior to 16 weeks of gestation, use the number of early pregnancies = number births + total pregnancy loss. Total pregnancy loss may be roughly estimated as 10% of the number of births;
- As the denominator for calculating coverage with later ANC visits[7], use the number of late pregnancies = number of births + number of still births. The number of stillbirths may be roughly estimated as 2% of the number of births;
- As the denominator for calculating coverage with the institutional deliveries and coverage of mothers with post-natal care, use the number of deliveries = number of births – number of twins. The number of twins may be roughly estimated as 1% of the number of births;
- As the denominator for calculating coverage of children with post-natal care, use the number of births;
- As the denominator for calculating coverage with childhood immunization, use the number of surviving infants = number of births – number of infant deaths. The number of infant deaths may be estimated as the Infant Mortality Rate x total population / 1,000.

The inter-relation of these various denominators is illustrated in Figure 2 below with an example from a district in Tanzania.

Figure 2: Number of pregnancies, deliveries, live births, infants. Tanzania district example, 2014



[6] Demographers estimate the CBR based upon specialized analysis and modeling of data derived from a national population census. An alternative way to estimated surviving infants and births (surviving infants + infants deaths) is to use a projection (based upon the annual population growth rate) of the number of infants counted during a national population census. This alternative approach can result in under-estimation of the number of surviving infants and births due to undercounting of infants during a national population census.

[7] In countries where the majority of first ANC visits occur after 16 weeks of gestation it would be acceptable to use the number of late pregnancies as the denominator for calculating coverage with first ANC visits.

Some health programs may use their own estimations that differ from those of the National Bureau of Statistics. If this is the case and if program denominators are used for the analysis then the analytic report should include a table of these program denominators along with an explanation of the methods used to calculate them.

It is often difficult to use census projections to estimate appropriate denominators for individual districts and health facilities. This is because people often seek care from health facilities that are outside of their area. The result can be that some districts and some health facilities have coverage that is significantly greater than 100% while other districts and health facilities have very low coverage when census projections are used to estimate denominators. As an alternative to use of census data*, when there is consistently very high coverage (> 95%) for a service such as ANC1 or DTP1 and when the data are felt to be of high quality* then these data can be used to estimate the number of pregnancies or the number of surviving infants. For example, the ANC 4 coverage for a district can be calculated by dividing the number of fourth ANC (ANC4) visits in the district by the number of first ANC visits (ANC1) in the district. As another example, the DTP3 coverage for a health facility can be calculated by dividing the number of third doses of DTP (DTP3) administered by the health facility by the number of first DTP doses (DTP1) administered by the health facility. The analyst 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. An example of this is presented as Annex 1.

5. Key analytical concepts

BASIC STATISTICAL TERMS

Once data are collected, they must be analysed to uncover the story that they tell and to create output that can be used to inform strategy and policy. Data analysis can help answer questions pertaining to many aspects of a health system. Some common questions answered include:

- The level of workforce and infrastructure available in the system
- The extent to which target populations have access to services and interventions
- The health status of the population
- The quality and safety of the services provided

Within these questions, it is important to also compare performance across time, between facilities, and across dimensions of disaggregations (e.g. sex, income, and geographic area).
Depending on the question being asked, various basic statistical terms can be useful. Some commons ones are outlined below.

A ***rate*** is the number of cases or events that occur in a population at risk over a given time period. A rate is often expressed per 1,000 discharges or admissions. Since the number of cases of a specified outcome depends upon the size of the population being considered, dividing by their population sizes makes two or more groups comparable.

A ***percentage*** is the number of events that occur in a population at risk, expressed per 100 people at risk. All people in the numerator must also be included in the denominator. A *proportion* is similar to percentage but expressed per person at risk. Percentages and proportions allow for comparison across facilities, and regions. They can also be used to track progress toward our targets, estimate coverage, measure outcomes, and understand our performance against quality-of-care indicators.

An ***average*** measures the central or typical value in a set of data and provides an easy point estimate for how the indicator scores across facilities, regions, or other groups. There are two common measures of average. The *mean* is calculated by dividing the sum of the indicator values by the number of observations. The *median* is calculated by ranking the scores from smallest to largest and taking the score in the middle of the list. In choosing between mean and median, it is important to note that mean is more sensitive to extreme values since, unlike median, it considers only the magnitude of each observation while median considers the ranking and relative magnitude.
The analysis described in this curriculum is descriptive analysis. That is, it can describe what is happening in the sample or target population, but cannot tell you why. Further, deep-dive, analysis, is required to understand the cause of any problematic findings. Common analyses include comparison of rates and percentages across facilities, regions, and time, ranking of disease burden or mortality, and investigations of differences by age, sex, income, or other levels of disaggregation.

EQUITY

Inequity in the delivery and access to health service persist even when overall service coverage is high. Measurement of equity is especially critical as it is fundamental to Universal Health Coverage (UHC) and Sustainable Development Goals (SDG).

Commonly equity has been measured by comparing coverage or access to services based on a household's wealth/socio-economic status and usually measured through household surveys (which are both infrequent and cost intensive). Measuring socio-economic status is difficult in facility

reporting systems. Other measures of equity are need for facility reporting systems. Area-based units of analysis are recommended for measuring inequity in facility reporting systems.[8]

Area-based unit of analysis include the use of sub-national geographical regions (such as provinces, districts, etc.) to measure inequality. As most routine health information systems (RHIS) are structured to examine data by sub-national administered, they are an important vehicle for measuring equity. Furthermore, as interventions to reduce inequity are implemented at the local administered level, the use of data collected through the RHIS can be used for resource allocation and planning.
Although routine health information systems are structured to measure area-based analysis of equity and present planners and policy makers with actionable data, some caution is needed when interpreting the data with an equity lens. There is a risk of committing ecological fallacy (in this case drawing erroneous conclusions about the health of individual using area-based data). For example, if richer districts have a higher prevalence of road traffic injuries it does not mean that road traffic injuries are more prevalent amongst richer individuals.

OTHER ANALYTICAL CONCEPTS: TRIANGULATION AND COMPARISON OF DATA WITH SYSTEMS REPORTING SAME HEALTH EVENTS

A manager/planner should aim to compare/reconcile results from the RHIS with other data sources (e.g. household surveys). While estimates derived from household surveys are frequently cited as the "gold standard" measurements of coverage, analysts should keep in mind that these estimates are subject to both sampling error (i.e. for sub-national estimates the confidence intervals of estimates can be wide due to a small sample size) and non-sampling errors (for example, there can be a recall bias when vaccination cards are reviewed for less than half of children surveyed). Demographic Health Surveys (DHS) and Multiple Indicator Cluster Survey (MICS) reports include annexes estimating the confidence intervals for key indicators at national and regional levels.

Sometimes the most valid estimate may not be provided by a single data source but rather by **"triangulation"** of findings from multiple data sources, each of which provides a partially valid picture of levels and trends. For example, WHO's estimates of trends in the incidence of malaria and tuberculosis are derived from such triangulation. This is illustrated by a case study presented in Annex 2.

While one of the objectives for the development of the standards for routine health facility information systems is to reduce duplication and redundancies in data collection, sometimes there are parallel systems that report on the same health events. For example, some countries have forms and data management systems for the Expanded Programme for Immunization (EPI) that collect data on immunizations on one form and also collect data on immunizations on a separate child health form. Such parallel reporting increases the reporting burden and can cause confusion when the data collected by the two systems do not agree. As another example of such parallel reporting, epidemic prone diseases may be reported on both a disease surveillance form and a monthly outpatient morbidity form. It is important to review and analyse the data from each of the parallel systems and present findings as part of the analysis. Possible reasons for any discrepancies should be discussed. With each related table or graph it is essential to specify which of the parallel systems was used for the analysis that is presented.

Data may also be available for sentinel sites such as hospitals and clinics that can assure higher quality diagnosis and reporting. Worthwhile data may also come from demographic surveillance sites (DSS) where regular tracking of household demographics and health status permits more reliable monitoring of the population and health events. The report of the analysis can compare findings from such sentinel sites and DSS's to those derived from routine health data.

[8] Hosseinpoor, A and N. Bergen (2016). Area-based units of analysis for strengthening health inequality monitoring. *Bull World Health Organ* 2016;94:856–858 | doi: http://dx.doi.org/10.2471/BLT.15.165266

6. Presentation and communication

At present, many health information systems are “data-rich” but “information-poor”. This is a consequence of the belief that data can be used directly for decision-making. Raw data alone are rarely useful. The point of a health information system is not just to generate high-quality data and hope that it will be used, but to convert it into credible and compelling evidence that informs local health system decision-making.

Only after **data** have been compiled, processed and analyzed do they produce **information** which can be integrated with other information and interpreted in terms of the issues confronting the health system. Information then becomes **evidence** that can be used by decision-makers.

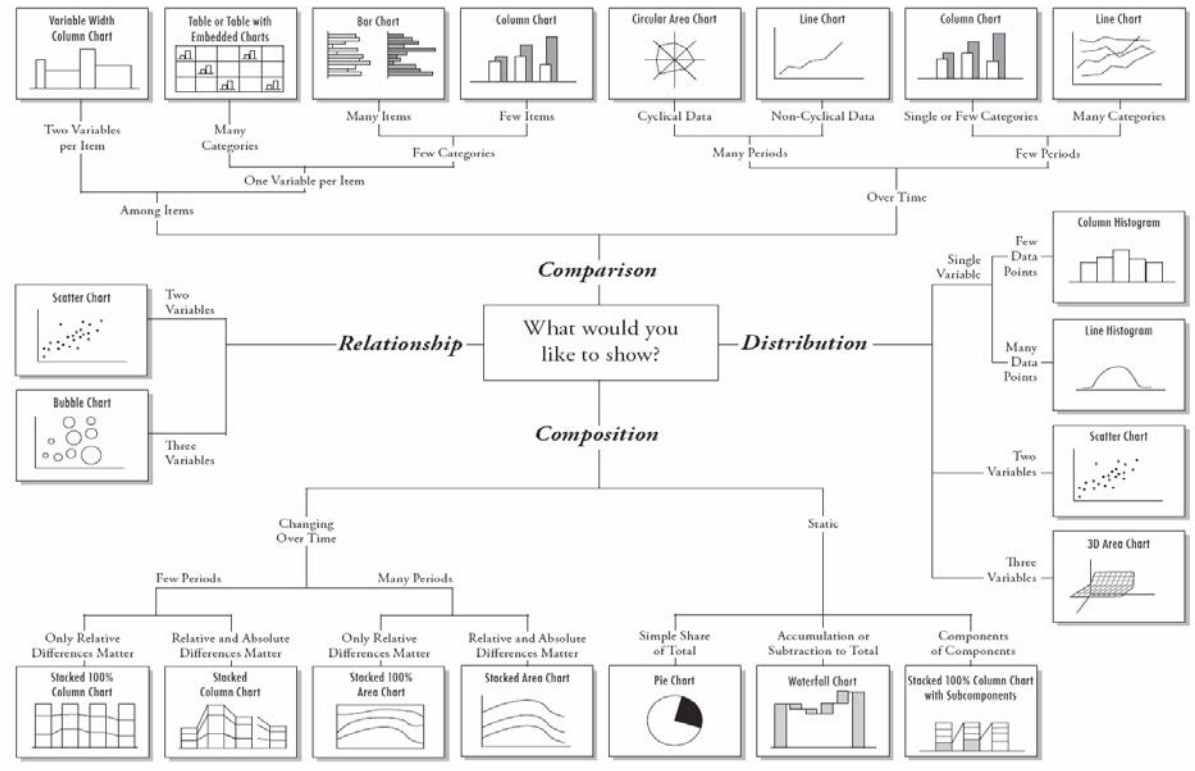
Evidence must be effectively communicated to decision-makers in order to shape their understanding of health issues and needs. It must be properly formatted into reports and presentations with user friendly graphs, tables and maps. This is the process of transforming evidence into **knowledge**, and once applied can result in **decisions** which will directly **impact** upon health and health equity. The impact on health can then be monitored by the health information system by measuring changes in health indicators.

There are numerous ways in which data can be presented and communicated. The choice can have a significant impact on how the data are interpreted and received; therefore, the choice is not trivial.

TYPES OF GRAPHICAL PRESENTATION

Figure 3 depicts a broad spectrum of charts and shows how choice depends on the types of questions you want to address and the types of variables you have in hands.

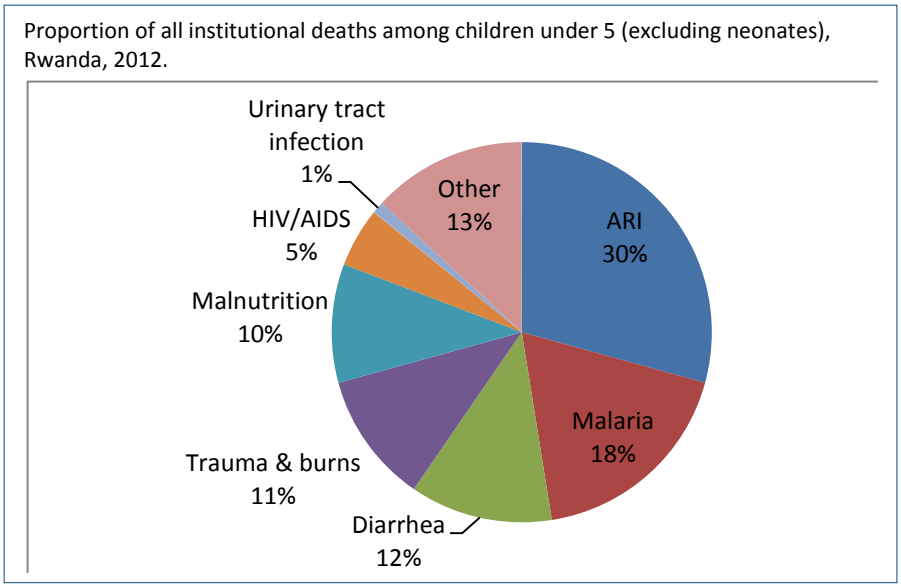
Figure 3: Examples of types of charts to select to explain specific messages



Some examples of charts and when they are useful are outlined below.

- **Pie charts** can be useful in depicting how individual parts make up the whole of something (such as overall mortality broken down by causes of death) (see Figure 4)
- **Tables** present precise numerical statistics in an orderly fashion. The interested reader can conduct some further analysis of these statistics if they so choose. However, it is often difficult for the reader to appreciate the most important conclusions to be derived from a large table of numbers (see Figure 5).
- "Thematic" **maps** give a different color to each geographic region for quick appreciation of regional disparities in a key indicator. Alternatively, or in addition, the size or color of points placed on a map can represent the value of a key indicator (see Figure 6).
- **Line graphs** can be useful for showing trend over time.

Figure 4: Pie-chart showing proportional mortality of institutional deaths



Source: Rwanda Annual Health Statistics Bulletin for 2012.

Figure 5: Tabular presentation of coverage statistics

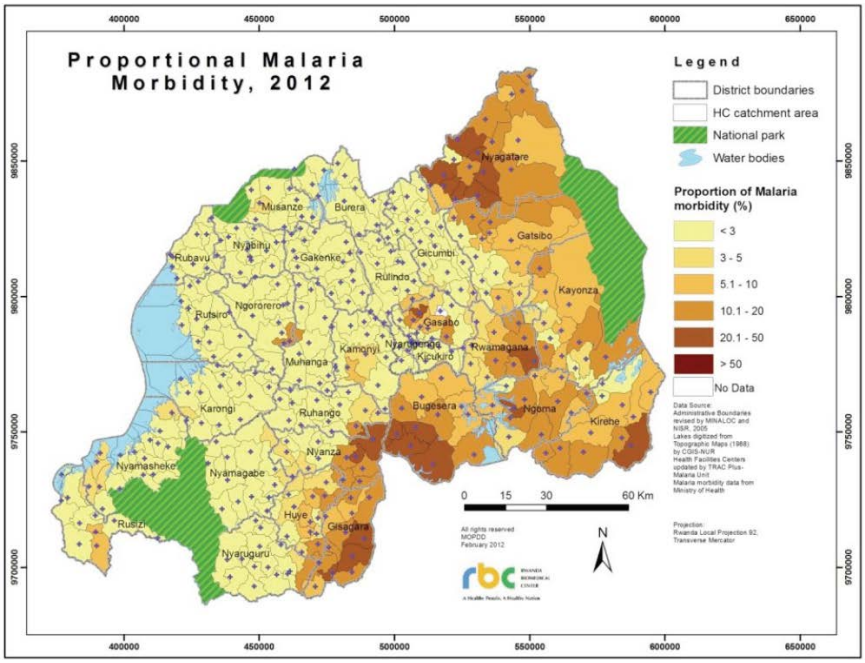
Childhood vaccination coverage by region/district and by antigen, Burkina Faso, 2014:

Tableau 4.02: couverture vaccinale (%) par antigène en 2013 (suite)

Régions/districts	VPO0	VPO1	VPO2	VPO3	DTC+HepB-Hib1	DTC+HepB-Hib2	DTC+Hep Hib3
Boucle du Mouhoun	**112,6**	**104,1**	**100,4**	**101,4**	**104,1**	**100,4**	**101,4**
Boromo	114,6	100,9	97,7	99,1	100,9	97,7	99,1
Dedougou	108,9	104,5	99,8	102,4	104,5	99,8	102,4
Nouna	124,3	110,2	106,7	107,1	110,2	106,7	107,1
Solenzo	114,2	103,2	99,3	99,8	103,2	99,3	99,8

Source: an extract from the Annual Statistical Report of the Ministry of Health, May 2014.

Figure 6: Example of a thematic map - Proportion of all outpatient visits attributed to malaria, Rwanda, 2012.



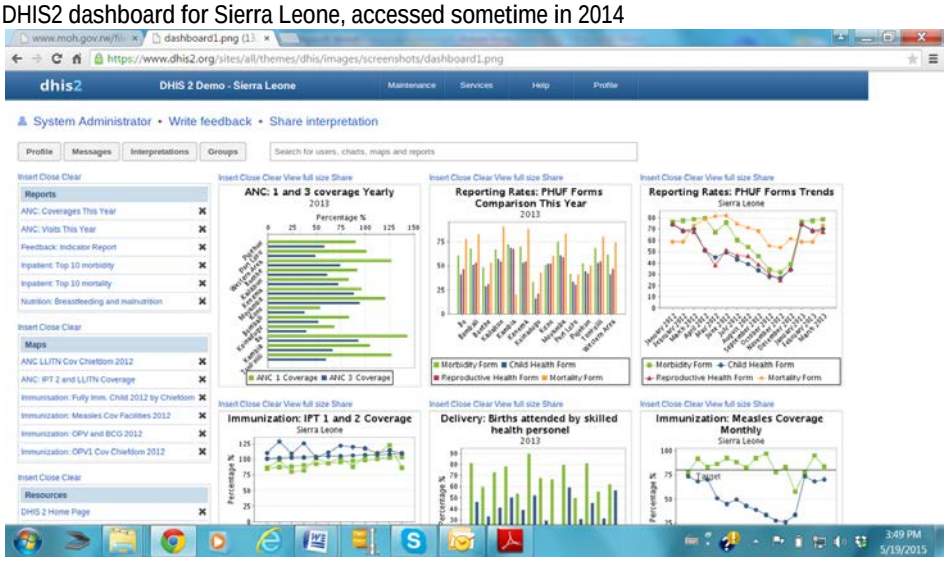
Source: Rwanda Annual Health Statistics Bulletin for 2012.

PRESENTING A STORY – WEBSITES AND DASHBOARDS

Computer code can be written with computer or web-based data management systems to display multiple table, graphs and maps on the same page. The idea is that the "driver" can glance at these items to get a quick impression of performance with key indicators each time that they log onto the data management system (see Figure 7).

Figure 7: Example of dashboard in DHIS

DHIS2 dashboard for Sierra Leone, accessed sometime in 2014



PRESENTING A STORY – SUMMARY MEASURES

A small set of standard indicators representing a range of health service functions can be selected to provide a comprehensive assessment of the overall performance of a national, regional or district health system. The scores from this fixed set of indicators can be combined mathematically (e.g. by averaging the coverage achieved with various health services) to calculate an index with which to judge trends from year to year or compare one district or regional to another. In this way, a "league table" can be generated (see Figure 8).

Figure 8: Example of a league table comparing performance of regions

Ghana's regional league table.

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
Ahsanti	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 POW 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, MoH Ghana, June 2013

KEY RULES FOR REPORTS AND GRAPHICAL PRESENTATIONS

1. On the cover page of a report, always specify the month and year that the document was finalized.
2. Don't forget to discuss the methods used for estimating denominators. Include in the report a table of key denominators.
3. In any report, and for each table or figure, always analyze data quality and present and discuss notable findings about date quality. Where relevant, discuss in particular the completeness of hospital data, the completeness of private sector data and striking inconsistencies over time. Tables and graphs of data from multiple years not only permit assessment of trends but also consistency. When an indicator fluctuates by 10% or more from year to year the text should acknowledge the possibility that the change reflects a data quality problem.
4. 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.
5. With graphs presenting estimates of immunization, ANC or delivery coverage also show recent survey estimates of the same indicator and, where possible, show the confidence interval for the survey estimate.
6. Every table and every figure needs a caption.
7. Specify the period and the geographic area for which the statistics apply.
8. Specify the data source – not just the publication or the organization that provided the data but the data source itself.
9. For each table, graph or map, include narrative in the report that interprets the most important findings and discusses how indicators are defined and any special limitations.
10. If findings are to be projected on a screen (e.g. with a PowerPoint presentation) do not include any text or number with a font size smaller than 16.

Annex 1: Use of service statistics as denominator

To calculate coverage, the size of the target population (i.e. the denominator) must be estimated. Population census projections are the preferred source. However, for small areas (for districts and certainly for individual health facilities) it can be difficult to reliably estimate the denominator. This is to some extent due to uncertainty about the size of the population (especially when a reliable census has not been recently completed or where internal migration has been high). However, the principal challenge to reliable estimation of the size of the "catchment population" is that people frequently seek care outside of defined administrative boundaries. This is shown by 2014 immunization data from Tanzania's HMIS. Figure 35 shows districts ranked according to their coverage with DPT3. DPT3 coverage is estimated using 3 different methods and for each method the districts are ranked separately on the horizontal axis from lowest to highest. Rankings for the conventional method of calculation, using census projections for each district to estimate the number of surviving infants, are shown by the light green line. For 52 (31%) of the 162 districts the DPT3 coverage by census was greater than 100%. Those designing the HMIS for Tanzania have attempted to address this anomaly by asking those delivering vaccinations to distinguish immunizations delivered to children living within a defined catchment area from children living outside of the catchment population. The result, using census projections for each district as denominators but including in the numerator only immunizations reported to have been given to children living within the catchment population, are shown by the red line. With this method, a somewhat smaller but still significant percentage of districts (18%) are found to have had coverage greater than 100%.

Recent household surveys have found that Tanzania's nationwide DPT1 coverage is 99%. Under these circumstances the number of children receiving DPT1 becomes a reasonable estimate of the true catchment population for childhood immunization services. DPT3 coverage can thus be calculated as DPT3/DPT1. The results are shown by the blue line. For 4 districts (2.5%; shown by the dashed blue line) during 2014 the reported number of doses of DPT3 exceeded the number of doses of DPT1 and thus the coverage calculated by this method was greater than 100%.

For both of the lines in Figure 36 districts are ranked according to their DTP3 coverage as calculated conventionally. In this way, the graph illustrates the effect on estimated coverage of using DTP1 as a denominator. For many of the districts that ranked low based upon the conventional estimate of DTP3 (i.e. those on the left of the graph), use of DTP1 as a denominator improves their estimated coverage. In fact, when DPT1 is used as the denominator, the DTP3 coverage for some of these apparently low performing districts is above the nationwide average (shown by the horizontal orange line). In contrast, for many of the districts that ranked high based on the conventional estimate of DTP3, use of DTP1 as a denominator reduces their estimated coverage. For some of the districts ranked highly based upon conventional DTP3 coverage, the coverage greater than 100% might reflect their popularity as service providers – their ability to attract clients from outside of their catchment area. Hence, use of DPT1 (or any other measure of the volume of services delivered) to estimate the target population arguably penalizes districts that attract clients from outside their boundaries (perhaps for reasons such as road access that have nothing to do with the quality of services) and rewards the districts where the population seeks care elsewhere. Nonetheless, when reliable population data are lacking (especially for individual health facilities or when census data old or suspect) data on services for which coverage is consistently very high can provide for more accurate denominators.

Figure 1A: 2014 DTP3 coverage of Tanzania's 163 districts calculated by 3 different methods described in the text; districts ranked separately for each method; Tanzania DHIS2 data

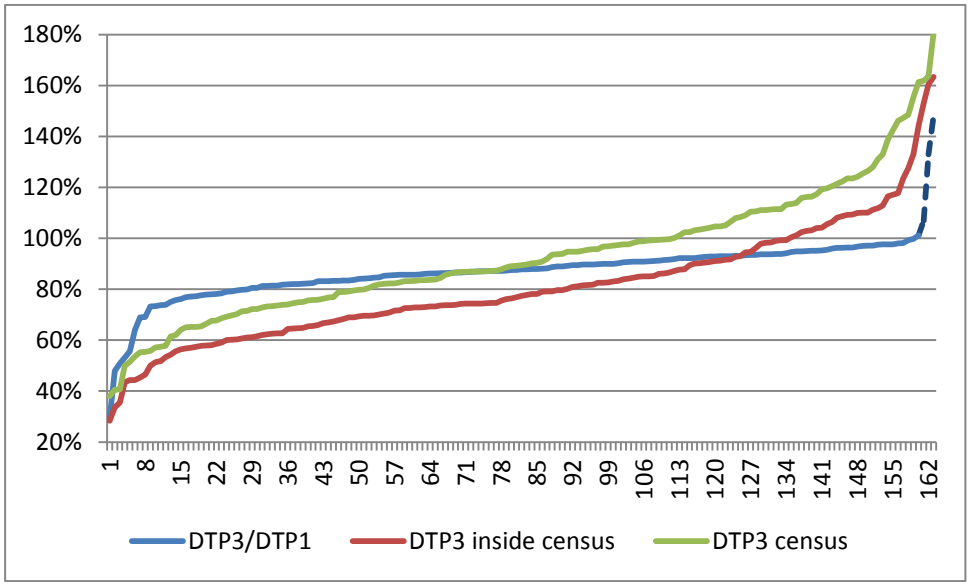
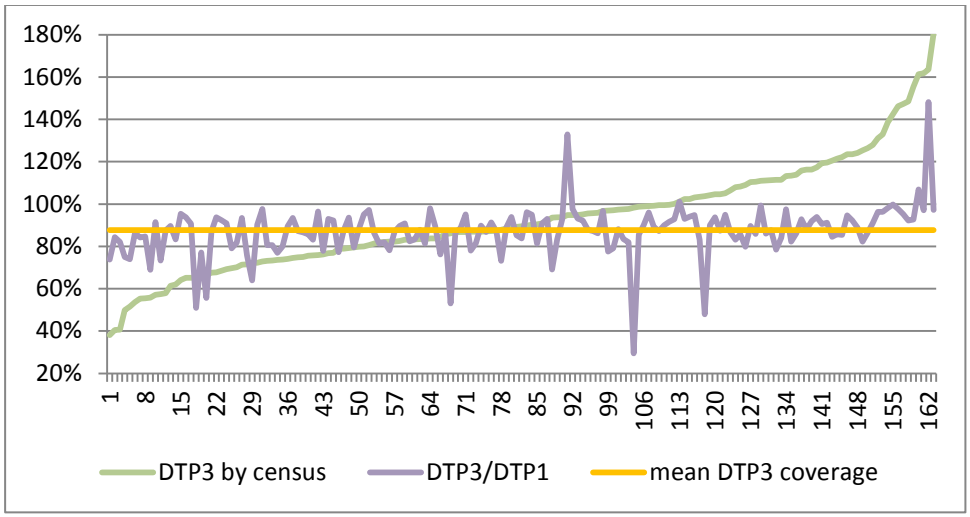


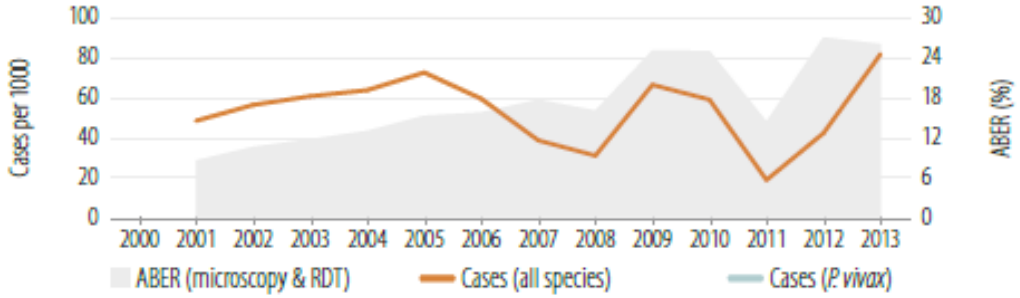
Figure 1B: 2014 DTP3 coverage as estimated by 2 different methods; districts ranked by DTP3 coverage as conventionally calculated; Tanzania DHIS2 data



Annex 2: Triangulation of data sources to assess trends in disease incidence

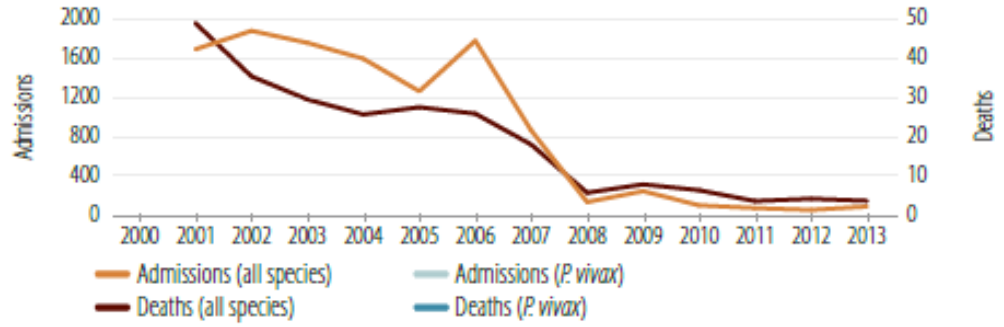
The country profile for Rwanda in the 2014 World Malaria Report includes the following two graphs. Figure 39 presents the trend, between 2000 and 2013, in confirmed cases of malaria reported per 1000 population per year. Shown on the same graph is the trend in the Annual Blood Examination Rate (ABER- the number of lab tests performed to confirm malaria per 100 population per year). Notice how confirmed cases reported has risen and fallen in parallel with the ABER.

Figure 2A: Confirmed cases of malaria reported /1000 population and ABER , Rwanda, 2000 to 2013. Source: 2014 World Malaria Report.



Rwanda's 2014 country profile also includes Figure 40 which presents trends in the incidence of hospital admissions and inpatient deaths due to malaria per 100,000. Both admissions and deaths from malaria have declined markedly since 2006. Based upon the trends in admissions and deaths, the World Malaria Report found that there is sufficient evidence to conclude that the true incidence of malaria in Rwanda has declined by more than 75% since 2000 even though confirmed cases reported has risen markedly over the last 2 years. The conclusion that there has been a marked reduction in malaria incidence is supported by survey findings that the prevalence of parasitemia among children 6 to 59 months fell from 2.6% in 2007-2008 to 1.4% in 2010. This example illustrates the value of triangulation of findings from multiple data sources.

Figure 2B: Trends in hospital admissions and inpatient deaths from malaria per 100,000, Rwanda, 2000 to 2013. Source: 2014 World Malaria Report.



World Health Organization
20, Avenue Appia
1211 Geneva 27
Switzerland