DISTRICT
DATA QUALITY ASSURANCE

A training package for monthly use of DHIS2 data quality dashboards at district and health facility levels

World Health Organization

DISTRICT
DATA QUALITY ASSURANCE

A training package for monthly use of DHIS2 data quality dashboards at district and health facility levels

World Health Organization

District data quality assurance: a training package for monthly use of DHIS2 data quality dashboards at district and health facility levels

Contents

Acknowledgements

Abbreviations

ANC	antenatal care
app	application
DHS	Demographic and Health Surveys
DHIS	District Health Information Software
DPT	diphtheria–pertussis–tetanus vaccine
DQ	data quality
DQA	data quality assurance
HMIS	health management information system
MICS	Multiple Indicator Cluster Survey
OPD	outpatient department
Penta	Pentavalent vaccine
RHIS	routine health information system
SOP	standard operating procedure
WHO	World Health Organization

Introduction

This training package aims to improve the quality of routine health facility data through building capacity in the monthly use of District Health Information Software (DHIS)2-based data quality (DQ) tools at district level. The package includes tutorials with related exercises and uses a website (https://who-demos.dhis2.org/dq) to demonstrate the tools.

Rationale: data quality assurance should be decentralized and routine

Health facilities regularly report data on their activities through the routine health information system (RHIS). In many countries, the most common and most important routine data are reported by facilities in aggregate each month or quarter. High-quality routine data are essential for effective local management of health services. However, while routine data are necessary for decision-making and for assessing performance, experience shows that these processes are often compromised by problems with data quality.

For countries that manage their data using the DHIS2, new automated tools enable rapid screening of data to identify the most important missing data and the most suspicious values. These include:

- the **DHIS2 version of the WHO Data Quality Tool**[1] (also called the WHO DQ Tool or DQ app);
- specially configured **data quality dashboards** that are visible to users as soon as they log into the DHIS2 website.

Until recently, these tools have been deployed mainly at national level. National-level staff, however, often confront several significant challenges including:

- an unmanageably large number of suspicious or missing values may be identified, especially if desk review of data is conducted infrequently;
- suspicious values can often only be investigated and corrected at the local level where there is access to the primary paper records and the staff that complete and submit them;
- government policy often stipulates that health data belong to district authorities and thus, if a suspicious value is confirmed to be incorrect, it can only be edited at district level.

For these reasons, the most effective approach to data quality assurance (DQA) is to decentralize data desk review (e.g. to district/council level) and to make it routine, i.e. to do it each month following the completion of the data entry.

---

1   https://www.who.int/healthinfo/tools_data_analysis/dqr_desk_review/en/

Audience

The target audience for this training package includes:
- ▷ DHIS2 data managers at district level[2] that are responsible for entry and review of the quality of DHIS2 data;
- ▷ health officials at district and national levels that are responsible for supervision of district data managers and for the development and implementation of policies and procedures, including standard operating procedures (SOPs) for DQA.

Learning objectives

- ▷ District data managers will have the knowledge and skills to use the two data quality tools each month and to follow up on the most important suspicious and missing values identified.
- ▷ Health officials at district and national levels will have the knowledge and skills to:
  - • instruct additional district data managers in the use of the two data quality tools; and
  - • refine policies, procedures and capacity-building interventions to support widespread adoption of a district DQA approach.

Overview of the materials and how to use them

The training package consists of a series of three tutorials and related exercises that demonstrate and provide hands-on practice in using the DHIS2 data quality tools. Each tutorial uses the online DQ Trainingland website (https://who-demos.dhis2.org/dq) which contains realistic data for demonstration of the tools. The tutorials include step-by-step instructions and screenshots to enable first-time users to follow along easily.

▷ **Tutorial 1: An introduction to the WHO Data Quality Tool**

This tutorial is designed to introduce users at all levels to the WHO Data Quality Tool. It briefly demonstrates how each of four dashboard functions and the Analysis function can be used at national level to identify and investigate inconsistencies and important missing values.

▷ **Tutorial 2: An introduction to a data quality dashboard[3]**

This dashboard is visible as soon as a user logs into the DHIS2 website. The tutorial demonstrates how the data quality dashboard can be used at national level.

▷ **Tutorial 3: How to train district staff in the monthly use of DHIS2 data quality tools**

This tutorial demonstrates an approach for rapidly training staff at district level to use the data quality dashboard and the outlier dashboard function of the WHO Data Quality Tool.

---

2   Where data are entered in the DHIS2 at facility level, facility data managers are also an appropriate audience.
3   There are two distinct types of dashboards for review of data quality. Tutorial 1 demonstrates the dashboard function of the WHO Data Quality Tool. Tutorial 2 demonstrates a data quality dashboard that is visible as soon as a user logs into the DQ Trainingland website.

An exercise is provided for each tutorial and may be undertaken immediately after the tutorial. The exercises enable learners to practise use of the data quality tools demonstrated in the tutorial. Each exercise contains a set of questions. Answers to the questions are found at the end of the document.

The initial focus on the national level (Tutorials 1 and 2) is designed to acquaint staff at higher levels (including senior officials at district level) with data quality issues at national level. This approach can also help to make them receptive to a decentralized approach which, in the future, should help to prevent and address quality issues before they reach national level. It is these senior officials who must then refine the SOPs and develop an approach for scaling up district DQA activities. Such a rollout of capacity-building at district level could involve use of Tutorial 3, without the need for Tutorials 1 and 2.

The three tutorials can be used in several ways:

› In a **generic workshop** approach, the tutorials can serve as notes for presentations given by a facilitator. The accompanying exercises are intended for use immediately after each tutorial, enabling participants to practise what has been demonstrated. The exercises involve use of the DQ Trainingland instance of DHIS2 and include screenshots to guide first-time users through each step. A sample agenda for a 3-day workshop is provided at the end of this section.[4]

› A **country-adapted workshop** can be organized in the same way as a generic workshop. However, the data quality tools, tutorials and exercises should be adapted for the national DHIS2. This will increase the relevance of the training to the participants, which should include a selection of district data managers. Adaptation should involve a number of steps:
   • A list of up to 20 core indicators are identified for review.
   • The WHO Data Quality Tool and the data quality dashboard then need to be configured to assess the consistency and completeness of the data for these indicators.[5]
   • A special "District" group of indicators should be added to the dashboard of the WHO Data Quality Tool.
   • A further essential step in country adaptation is to update the data management SOPs to specify the ways that the data quality tools are to be used by staff at each level, as well as the ways in which they should further investigate the data quality problems revealed by the tools.

› In an **online distance learning** approach, users with internet access can log into the DQ Trainingland website at https://who-demos.dhis2.org/dq. The username and password are shown on the login page. Users can then use the tutorials for self-paced learning. The tutorials include screenshots that exactly match what users will see online.

---

4   For a workshop, multiple participants must be able to access the DQ Trainingland instance simultaneously and with minimal delays. It is recommended that qualified information technology experts set up, at the training venue, a local server with a copy of the DQ Trainingland instance. Facilitators and participants can then access this instance using local wifi.
5   For documentation on how to configure these data quality tools, refer to https://www.dhis2.org/who-dq

> In a **video instruction** approach, each tutorial can serve as a detailed script for a facilitator to pre-record a demonstration. The tutorial can also be translated into the local language. A video can then be used as a supplement to a workshop or for online distance learning.

> The **"30-minute training"** approach involves use of a country-adapted version of Tutorial 3 to quickly train district data management staff. Tutorial 3 can be used even if participants have not received Tutorial 1 or Tutorial 2 or any other training related to the two data quality tools. Tutorial 3 demonstrates each of the steps needed to introduce the tools and enable participants to gain confidence in using them. The proposed DQA process has been streamlined to shorten both the time required to teach it and the time required to implement it each month. As such, the 30-minute training approach could be taken to scale with minimal resources, for example, by adding it on to a regularly scheduled district meeting or a supervisory visit. The key first step is to adapt the approach to the local context, i.e. to focus the data quality assessment on the right set of indicators: those that are of greatest interest and those for which problems with inconsistency and incomplete reporting are most common.

This document shows screenshots taken from https://who-demos.dhis2.org/dq in June of 2020, When this website is viewed in a different month, data from a somewhat different range of months will be presented in the charts and tables.

Sample agenda for a 3-day workshop

Workshop to introduce DHIS tools for routine data quality assurance at district level

DAY 1		
9:00–9:30	• Registration	Facilitator
9:30–10:00	• Welcome and introductions • Workshop objectives	
10:00–10:30	Coffee break	
10:30–11:30	• A brief presentation on WHO's framework for data quality review • Discussion of current procedures for data quality assurance (DQA) and current problems with data quality	
11:30–12:30	• Demonstration of the dashboard of the WHO Data Quality Tool (Tutorial 1)	
12:30–14:00	Lunch break	
14:00–15:00	• Demonstration of the Analysis function of the WHO Data Quality Tool (Tutorial 1 continued) • Explanation of the exercises	
15:00–15:30	Coffee break	
15:30–16:30	• Exercises to practise use of the dashboard and Analysis function of the WHO Data Quality Tool • Questions	
DAY 2		
9:00–10:00	• Demonstration of skills acquired on Day 1 • Questions	Selected participants
10:00–10:30	• Demonstration of the standard data quality dashboard (Tutorial 2)	
10:30–11:00	Coffee break	
11:00–12:30	• Exercises to practise use of the standard data quality dashboard	
12:30–14:00	Lunch break	
14:00–15:00	• Demonstration of how to quickly train district data managers in routine use of the data quality tools (Tutorial 3)	
15:00–15:30	Coffee break	
15:30–16:30	• Exercises in use of the data quality tools at district level	
DAY 3		
9:00–10:30	• Demonstration of skills acquired on Day 2 • Questions	Selected participants
10:30–11:00	Coffee break	
11:00–12:30	• Presentation: How data quality tools are incorporated into data management procedures • Discussion	
12:30–13:00	Concluding remarks	

Tutorial 1: An introduction to the WHO Data Quality Tool

Overview

This tutorial demonstrates how to use selected components[1] of the DHIS2-based WHO Data Quality Tool (DQ Tool). The tool is an application that enables users to quickly assess the completeness and consistency of DHIS2 data.

The DQ Tool can greatly assist staff responsible for assuring the quality of data reported by health facilities, including staff at district or facility level with online access to the DHIS2, as well as staff at higher levels of the health system. At national and regional levels, this may include staff working for specific health programmes and also those with cross-cutting responsibilities for data management, e.g. staff of the health management information system (HMIS) unit or monitoring and evaluation staff.

This tutorial is designed for staff using the DHIS2 at national level. The screenshots show what national-level users will see if they log into https://who-demos.dhis2.org/dq. (Tutorial 3 provides guidance on use of the DQ Tool at district level.)

The demonstration begins with a focus on the dashboard of the DQ Tool. The dashboard is pre-configured and ready for use by staff who may not be very familiar with the DHIS2 data elements. It can be used to review the quality of a pre-defined set of priority DHIS2 data elements/indicators. The dashboard features four pages:
  ⊳ completeness
  ⊳ consistency - time (month-to-month consistency)
  ⊳ consistency - data (consistency between related indicators)
  ⊳ outliers (detailed examination of month-to-month consistency and missing data).

Next, the Analysis function of the DQ Tool is demonstrated. The Analysis function enables review of the quality of any DHIS2 data element/indicator. Thus, it is more flexible than the dashboard. However, to use the Analysis function, the user must first select correctly from among the very large number of data elements/indicators. The Analysis function is therefore better suited to staff of the HMIS unit or public health programmes familiar with the details of the DHIS2 data.

---

1   This brief overview does not discuss how to make use of the "Annual Report" function of the WHO DQ Tool, nor how to configure the dashboard of the application and the annual report. For further details on the WHO Data Quality Assurance framework and the full range of functions of the DHIS2-based WHO DQ Tool, refer to the following WHO and University of Oslo websites: https://www.who.int/data/data-collection-tools and https://www.dhis2.org/who-dq

If this tutorial is demonstrated by a facilitator, such as during a workshop, it is recommended that the participants practise each of the steps by following the exercises for the tutorial.

1. **Launch the Tool**
   Log into the DHIS2 instance: URL = https://who-demos.dhis2.org/dq
   Username: demo   Password: District1#

2. **Type** "WHO" in the "Search apps" line in the upper right of the DHIS2 home page. Click on **WHO Data Quality Tool.**

3. **Quickly review the main menu of the DQ Tool:** After the DQ Tool is launched, five tabs appear at the top of the page. These are used to select between several different functions. Note that the word "Dashboard" is highlighted, and that the Dashboard function is selected by default.
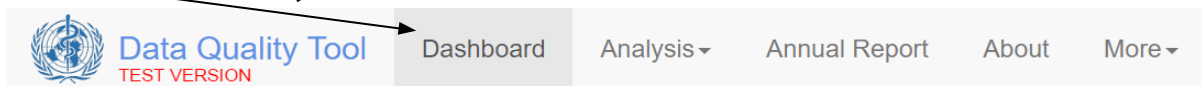
4. **Quickly review the Dashboard menu:** The Dashboard has its own set of four tabs. The Completeness tab is selected by default.
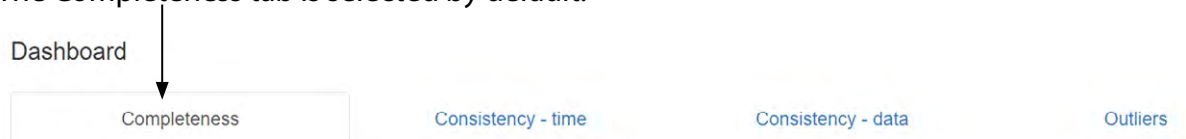
Use the Completeness dashboard

5. **Review the Completeness dashboard.** Scroll down and review the Completeness page of the data quality dashboard. The page shows completeness for each of five "core datasets".

a. Review the charts on the left side of the page: The charts on the left show, for each dataset, the completeness by month for each of the last 12 months.
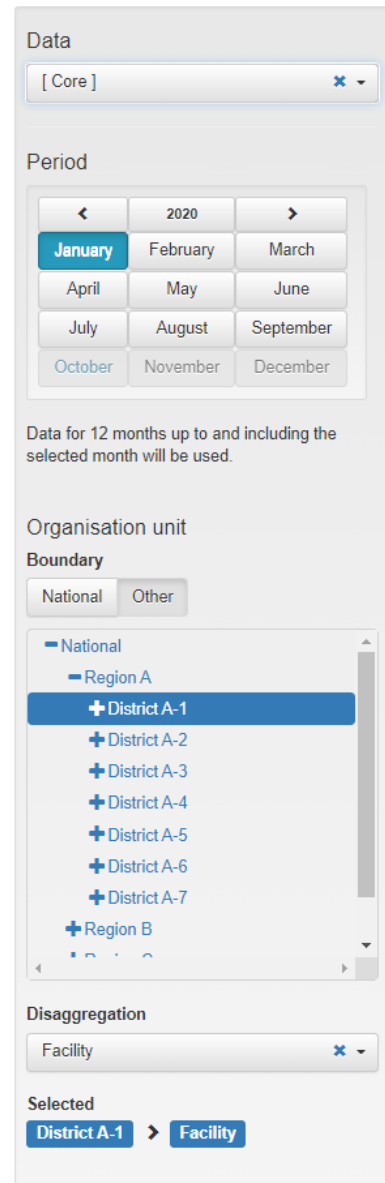
b. Review the charts on the right side of the page: The charts on the right show, for each dataset, the completeness by region *for the last month* of the period being analysed.

Click on the settings icon (▣) in the upper right of the screen. The reporting period that is charted can be changed by clicking on a different month. After experimenting with changing the month, reset it to January of the current year.[2]

c. Change the "Organisation unit" that is being analysed: By default, the DQ Tool analyses all data nationwide, disaggregated by region. The DQ Tool can also be used at district level with data disaggregated by individual health facility.

In the "Organisation unit" section of the settings window, click on "Other" and select "District A-1" of Region A. The charts on the left side of the completeness page now show the 12-month trend in reporting completeness of District A-1. By default, the charts on the right side of the completeness page show results disaggregated for one level below: Facility.

After experimenting with changing the "Organisation unit" and the month, reset it to "National" with **disaggregation by District** *and* **period = January of the current year.**

Data

[Core]   ✕ ▾

Period

‹	2020	›
January	February	March
April	May	June
July	August	September
October	November	December

Data for 12 months up to and including the selected month will be used.

Organisation unit

Boundary

| National | Other |

– National
   – Region A
     ✚ District A-1
     ✚ District A-2
     ✚ District A-3
     ✚ District A-4
     ✚ District A-5
     ✚ District A-6
     ✚ District A-7
   ✚ Region B

Disaggregation

Facility   ✕ ▾

Selected

District A-1 ❯ Facility

---

2   When the period is set this way, the screen should match the screenshots shown in this guide.

Use the Consistency - time dashboard

6. **Review the Consistency - time dashboard:** Click on the tab for "Consistency - time".
    Scroll down and review the page. Move the cursor around on each chart to view the pop-
    up details.
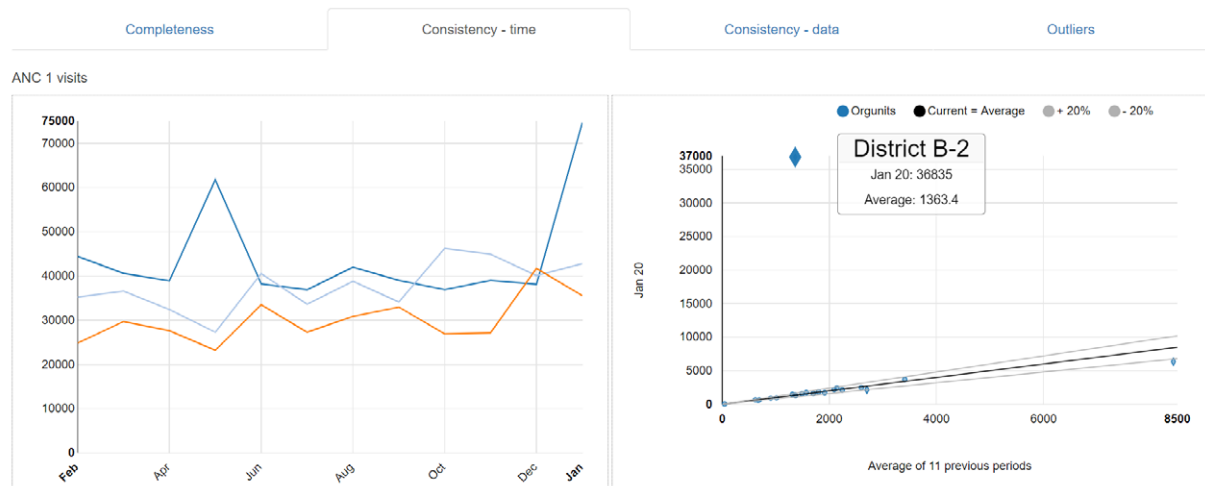
a. **Review the charts on the left side of the page:** The charts on the left side of the screen
    show, for each core numerator, the trend in the numerator value over the period 1 to
    12 months previously (dark blue in the chart above), the period 13 to 24 months previously
    (light blue in the chart above) and the period 25 to 36 months previously (orange in the
    chart above). Note the sharp rise in the value of antenatal care (ANC) 1 visits in May of last
    year and again in January of this year. Such a dramatic rise at national level is certainly
    suspicious.

b. **Review the charts on the right side of the page:** Each chart on the right is an example
    of a "scatter plot". When the Tool is set to disaggregate by district, each dot represents the
    value for a single district. In this chart, the position of the dot on the vertical axis represents
    the value of the numerator for the month selected. (Remember that the settings were
    reset to show results for Period = January of the current year.) The position of the dot on
    the horizontal axis is the average value in the same district over the 11 previous months
    (February of last year to December of last year). Dots that appear below the lower grey line
    or above the upper grey line are suspicious. (In the example above, the number of ANC first
    visits in District B-2 was 36 835 in January of this year, compared with an average of
    1 363.4 visits per month for the preceding 11 months.)

c. **Modify the period to investigate an outlier:** ANC 1 has an outlier value for May of last
    year. To investigate this outlier, the settings window can be used to set Period to May of
    last year. The charts will then change as shown below. Place the cursor over the diamond-
    shaped dot above the lines in the chart on the right. Which district does the dot represent?
    Before continuing, change the period back to January of this year.
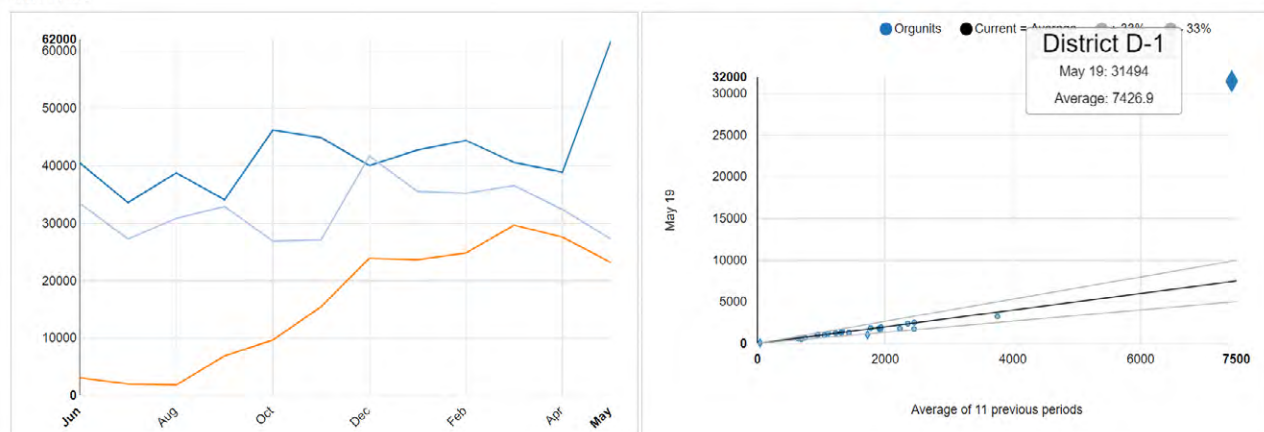
Use the Tool for a programme-specific review of data quality:

7. **Review the list of "Core" data element/indicators:** In the settings window of the Tool, Data is now set to "Core". As a result, the Tool presents findings for the data elements/indicators which were included in the "Core" grouping. Scroll down the "Completeness - time" page to review the list. Note that only one immunization indicator is featured.

8. **Change Data to "Immunization":** In the settings window (click on the settings icon if the window is not open), change Data from "Core" to "Immunization". Note what happens to the list of data elements/indicators featured. There are now 10 immunization indicators featured on the "Completeness - time" page. In this way, the Tool can be used for a programme-specific review of data quality – to review the full range of programme-specific indicators.

Before continuing to the next step, use the settings window to change Data back to "Core".

Use the Consistency - data dashboard

9. **Review the Consistency - data dashboard:** Click on the tab for "Consistency - data". This page shows whether there is a good match between the values of two indicators that are related. Review the chart at the bottom of the page showing the Penta 1 to Penta 3 dropout rate. The chart shows that two districts (A-2 and A-3) had a negative Penta 1 to Penta 3 dropout rate over the 12 months up to and including January 2020 (i.e. these two districts reported a higher value of Penta 3 than Penta 1). A negative Penta dropout rate for a district for an entire year is usually a sign of poor data quality and should be investigated.

Penta 1 to Penta 3 dropout - Feb 19 to Jan 20

District A-3
-4.54% dropout

Use the Outliers dashboard

10. **Use the "Outliers" dashboard to identify suspicious district-level values:** Click on the tab for "Outliers". If necessary, use the settings window to set Disaggregation to District. To free up space on the page, click on the settings icon to make the settings window disappear. The page shows a "Result" table with many rows. Each row shows 12 monthly values for one indicator for a single district. Each row has one or more values highlighted in red. How are the values highlighted in red different from the other values in the same row? The Tool finds the values highlighted in red to be suspicious because they are either significantly higher or significantly lower than the other values in the row. The rows are sorted in order of "weight". The "weight" can be considered the difference between the suspicious value and the average of other values in the row.

There are so many rows that the table will not fit on a single page. To view the subsequent pages, click on the controls at the bottom of the page.

Region	Unit	Data	Feb 19	Mar 19	Apr 19	May 19	Jun 19	Jul 19	Aug 19	Sep 19	Oct 19	Nov 19	Dec 19	Jan 20	Missing	Outlier	Total
Region B	District B-2	ANC 1 visits	1372.0	1311.0	1374.0	1289.0	1413.0	1330.0	1513.0	1335.0	1322.0	1434.0	1304.0	36835.0	0	35472	35472
Region D	District D-1	ANC 1 visits	7519.0	5927.0	6613.0	31494.0	4866.0	6016.0	6803.0	5280.0	5794.0	6026.0	6349.0	6359.0	0	25353	25353
Region C	District C-2	HIV tests done	9275.0	5977.0	5199.0	5560.0	3836.0	4337.0	26652.0	4743.0	4101.0	7089.0	6632.0	9120.0	0	20667	20667
Region C	District C-3	HIV tests done	4122.0	3893.0	3347.0	3428.0	2580.0	3512.0	4936.0	4124.0	2491.0	4779.0	7751.0	5048.0	0	3909	3909
Region D	District D-3	ANC 1 visits	3809.0	1794.0	1652.0	1592.0	1681.0	1489.0	2011.0	1818.0	1722.0	1791.0	1672.0	1723.0	0	2087	2087
Region D	District B-4	HIV tests done	6269.0	5657.0	3847.0	3946.0	4233.0	3862.0	4961.0	4531.0	4103.0	4316.0	4197.0	4771.0	0	1867	1867
Region A	District A-2	Penta 3 given < 1	1599.0	1483.0	1434.0	1139.0	1483.0	1701.0	1751.0	1417.0	1318.0	3002.0	1166.0	1249.0	0	1571	1571
Region B	District B-5	ANC 4 visits	250.0	267.0	239.0	255.0	258.0	224.0	1384.0	377.0	343.0	250.0	264.0	250.0	0	1352	1352
Region C	District C-6	HIV tests done	2291.0	2612.0	2241.0	1883.0	2165.0	2256.0	3050.0	2396.0	2334.0	2429.0	2557.0	3687.0	0	1304	1304
Region B	District B-5	Penta 3 given < 1	1543.0	1497.0	1154.0	1395.0	1162.0	1447.0	1588.0	1528.0	1612.0	1430.0	2573.0	1344.0	0	1146	1146
Region C	District C-1	ANC 1 visits	2677.0	2629.0	2640.0	2509.0	3329.0	2325.0	2973.0	2707.0	2761.0	2672.0	2503.0	2330.0	0	717	717
Region C	District C-4	ANC 1 visits	2030.0	2114.0	2084.0	1947.0	2214.0	2361.0	2226.0	2170.0	1678.0	2161.0	2123.0	2181.0	0	568	568
Region D	District D-2	Penta 3 given < 1	1787.0	1775.0	1829.0	1859.0	2061.0	1972.0	1780.0	1592.0	1698.0	2078.0	1796.0	1356.0	0	483	483
Region A	District A-3	HIV tests done	5020.0	4025.0	4495.0	3988.0	4483.0	2800.0	4132.0	4301.0	4095.0	5716.0	4344.0	5636.0	0	388	388
Region C	District C-4	Institutional deliveries	1463.0	1457.0	1430.0	1395.0	1265.0	1188.0	1354.0	1439.0	1033.0	1557.0	1562.0	1455.0	0	382	382

a. Use "Z-scores" to identify the most suspicious values: The WHO DQ Tool uses a "Z-score" to assess how suspicious a monthly value is. The higher the Z-score, the more different the value is from what has been reported by the same facility or district in other months. Use the following steps to filter the Result table so that it shows only rows with a high Z-score:

- Click on the Options button to open the Options window.
- To see the Z-score for each row, click on the box "Include Z-score". Review the Modified Z-score values in the Result table. Note that the values range from less than 2 to more than 400.
- Click on the following buttons:

1) Outliers; 2) Modified Z-score; and 3) Extreme.

Note the menu icon at the end of each row.

b. Inspect each row: How has the Result table changed? How have the Modified Z-scores changed? Note how the Result table reduces to only a few rows. Note that the Modified Z-score for each row is greater than 5.0.

Are *these* suspicious values due to errors? How could the data be investigated to determine whether a highlighted value is due to an error?

Click on the menu icon at the end of the first row (for values of ANC 1 visits reported for District B-2; see the above figure) and select "Visualize" from the drop-down menu.
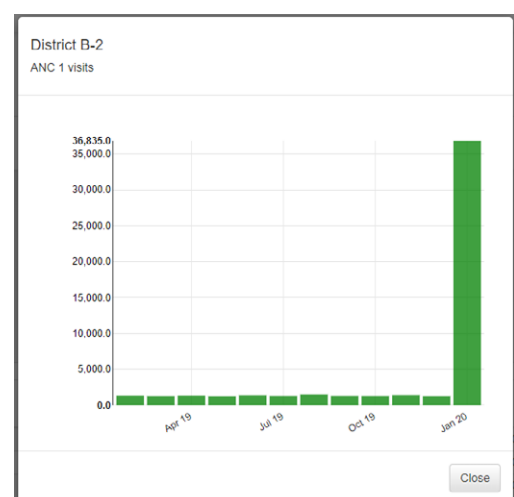
c. "Drill down" to investigate the source of the suspicious value: Click on the Close button to close the chart, click again on the menu icon at the end of the first row and select "Drill down" from the drop-down menu. If necessary, again use the Options feature to filter the table to Outliers, Standard Score and Extreme. How has the Result table changed? Each row of the table shows values for which level of the health system? Now each row shows 12 monthly values for a single indicator for a single health facility of District B-2. Are these suspicious values due to errors? Very large extreme outliers such as the one in the first row are very likely to be errors.

d. Consider sending a message to the person responsible: Click on the menu icon at the end of the first row and select "Contact" from the drop-down menu. A window should appear showing the Contact person for notifying of the extreme outlier. Click on "Write message". Use the drop-down menus to fill in the information below Recipient, write a message and click on "Send message". A message will then be sent via the DHIS2 internal messaging service.

As this tutorial has shown, the dashboard of the DQ Tool is easy to use. A limitation of the dashboard, however, is that it can only be used to review those data elements/indicators for which the dashboard has been configured (see "How to configure the WHO Data Quality Tool"). The next section will show how the Analysis function of the DQ Tool can be used to review the quality of any data element or indicator for which DHIS2 has data.

Use the Analysis - Consistency function

11. **Review the Analysis - Consistency function:** In addition to the dashboard, the Tool provides a convenient way to customize the review of data quality.

    Click on the "Analysis" tab at the top of the page, then select "Consistency". A window appears for configuring a customized assessment of consistency over time or consistency between related numerators.

    By default, the consistency analysis type "Between Indicators" is selected and compare organisation units to "Overall result" is selected. Leave "Consistency analysis type" set to "Between Indicators" but click on "Expected result" and select "Dropout" (see figure).

13

Click on Data to choose the indicators for comparison. Configure the window as shown.

Click on Period to customize the period – set to last year.[3]

Click on "Organisation unit: and set disaggregation to District. Click on the blue "Analyze" button in the lower right of the page and the results will appear. Note the following:

- The chart on the left side of the page is similar to the chart on the Consistency - data dashboard. Two districts are shown to have a negative Penta 1 to Penta 3 dropout rate in 2019. To view details, move the cursor over the red bars.
- The table on the right side of the page has a row for the entire country ("National"), followed by one row for each district. The

Data

Select data element/indicator

Data element	Indicator

Immunisation

Penta 1 given < 1

Select data element/indicator for comparison

Data element	Indicator

Immunisation

Penta 3 given < 1

Period

Period type		Year	
Years	▼	2019	▼

districts are sorted according to their dropout rates. The two districts with a negative dropout rate are at the top of this list. The rows continue on two more pages, but it is less important to view these other pages because these dropout rates are all positive.

Penta 1 given < 1 to Penta 3 given < 1 dropout. 2019.

	Download	Previous

Unit	Penta 1 given < 1	Penta 3 given < 1	Dropout rate (%) ❶	Weight ❶ ↓
National	479411	451771	5.8%	
District A-2	16532	18986	-14.8%	2454
District A-3	10714	11219	-4.7%	505
District D-4	30699	29000	5.5%	0
District D-3	23025	21676	5.9%	0
District D-2	24519	21930	10.6%	0
District D-1	89320	78435	12.2%	0
District C-6	14302	13648	4.6%	0
District C	44000	20027	5.0%	0

3   The screenshot was taken in 2020 when "last year" was 2019. In 2021 and future years, select whichever year is last year.

14

- Click on any single row; three things will happen: the colour of the row changes to pale yellow; a one-row table appears beneath the table; and another chart appears at the bottom of the page. This chart shows the results by month. In the example below, Penta 3 was greater than Penta 1 for all 12 months of last year. This district appears to have a recurrent problem with over-reporting of Penta 3. In addition, the Penta 3 value for November 2019 is suspiciously high.
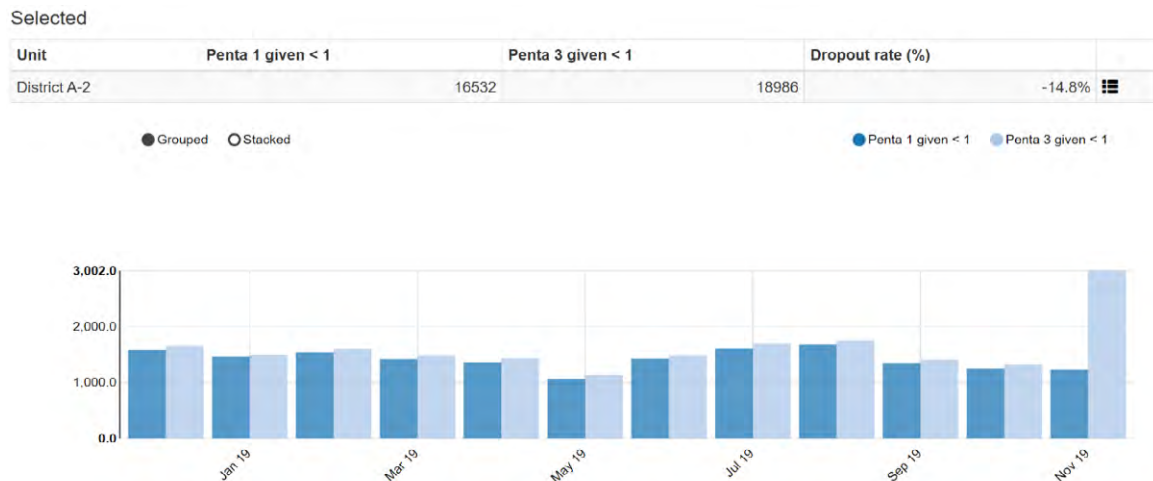
- Click on the menu icon at the end of this single row to "drill down" and see the results disaggregated to a level below. The chart shows that the negative dropout rate is not the result of data reported by only a single health facility: 27 out of 31 facilities in District A-2 reported more Penta 3 doses than Penta 1 doses in 2019. District A-2 therefore appears to have a problem with systematic over-reporting of Penta 3. The problem is seen for multiple facilities and multiple months. This may be due to a tendency of facilities in District A-2 to misclassify Penta 1 doses as Penta 3 doses.

The dropout rate for Facility 197 is so greatly negative that it warrants further investigation. Click on the row for Facility 197 so that it turns yellow. The chart appearing at the bottom of the page shows that, for November 2019, Facility 197 reported a value of Penta 3 that is so much greater than the value of Penta 1 that it seems likely to be an error. This error partially explains the negative dropout rate for District A-2. However, District A-2 has many health facilities and many months for which the number of Penta 3 doses exceeds the number of Penta 1 doses. This may suggest systematic over-reporting of Penta 3 doses by the facilities of this district. Such a highly suspicious finding warrants further investigation.

Penta 1 given < 1 to Penta 3 given < 1 dropout. 2019.

Unit	Penta 1 given < 1	Penta 3 given < 1	Dropout rate (%) ❶	Weight ❶ ↓↑
District A-2	16532	18986	-14.8%	
Facility 197	2320	4151	-78.9%	1831
Facility 403	747	836	-11.9%	89
Facility 469	953	1007	-5.7%	54
Facility 202	573	627	-9.4%	54
Facility 474	634	683	-7.7%	49
Facility 468	930	975	-4.8%	45
Facility 466	616	651	-5.7%	35
Facility 460	541	576	-6.5%	35
Facility 198	636	669	-5.2%	33
Facility 465	861	893	-3.7%	32

First Previous **1** 2 3 4 Next Last

Use the Analysis - Outliers and Missing Data function

12. **Review the Outliers and Missing Data function:** Click on the "Analysis" tab at the top of the page, then select "Outliers and missing data". The Analysis - Outliers function works in the same way as the Outliers dashboard but is much more flexible. It can be configured to assess any data element or indicator, i.e. it is not limited to the data elements/indicators for which the dashboard of the DQ Tool has been configured. It can be configured to simultaneously assess any combination of data elements and indicators desired. It can even be configured to simultaneously assess all data elements or indicators in a dataset, data element group and/or indicator group.

It can be configured to assess data over any period and for any grouping of health facilities.

Tutorial 2: An introduction to a data quality dashboard

Overview

This tutorial demonstrates how to use a data quality dashboard that is visible as soon as the user logs into https://who-demos.dhis2.org/dq. It is not necessary to launch any application to view this dashboard. The dashboard enables users to quickly assess the completeness and consistency of DHIS2 data.

The data quality dashboard can be of great use to staff responsible for assuring the quality of data reported by health facilities – staff at district or facility level with access to the DHIS2 as well as staff at higher levels of the health system. At national and regional levels, this includes staff working for specific health programmes as well as those with cross-cutting responsibilities for data management, e.g. staff of the HMIS unit, or monitoring and evaluation staff.

Tutorial 2 demonstrates how the data quality dashboard can be used at national level. The screenshots show what national-level users will see if they log into https://who-demos.dhis2. org/dq.[1] (Tutorial 3 provides guidance on how to use the data quality dashboard at district level.)

Note: as https://who-demos.dhis2.org/dq is a live DHIS2 instance, the months shown in some charts and tables update each month to include data from the current "last month". For this reason, there will be differences between the items on the instance and some of the screenshots included in this document.

The dashboard is structured according to the four dimensions of the WHO Data Quality Assurance framework: 1) completeness; 2) internal consistency of data (do DHIS2 data values agree with each other?); 3) external consistency of data (do DHIS2 data agree with statistics from other sources such as a survey?); and 4) consistency of DHIS2 denominator/target estimates.

Quick overview of the data quality dashboard

Log into the DHIS2 instance: URL = https://who-demos.dhis2.org/dq
Username: demo   Password: District1#

---

1   Several of the charts and tables of the data quality dashboard are configured to present data for the "last 12 months". The months included in "last 12 months" will change each month. As a result, the screenshots included in this tutorial (which were captured in July 2020), will differ somewhat from the dashboard seen on https://who.dhis2.net/dq. When preparing for a training session in a specific month, facilitators may choose to update Tutorial 2 and Tutorial 3 so that the "last 12 months" shown in the screenshots of charts and tables is current and identical to what will be seen on the DHIS2 website.

In the dashboard ribbon at the top of the screen, find and click on the button for "Data quality".
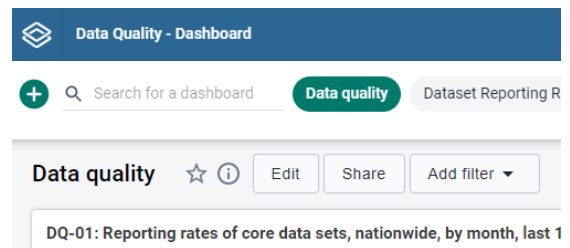
Quickly review the Data quality dashboard. Note the following:

Dimension 1: Completeness

The first three items (**DQ-01, DQ-02** and **DQ-03**) are used to assess completeness, also known as the reporting rate for a dataset.
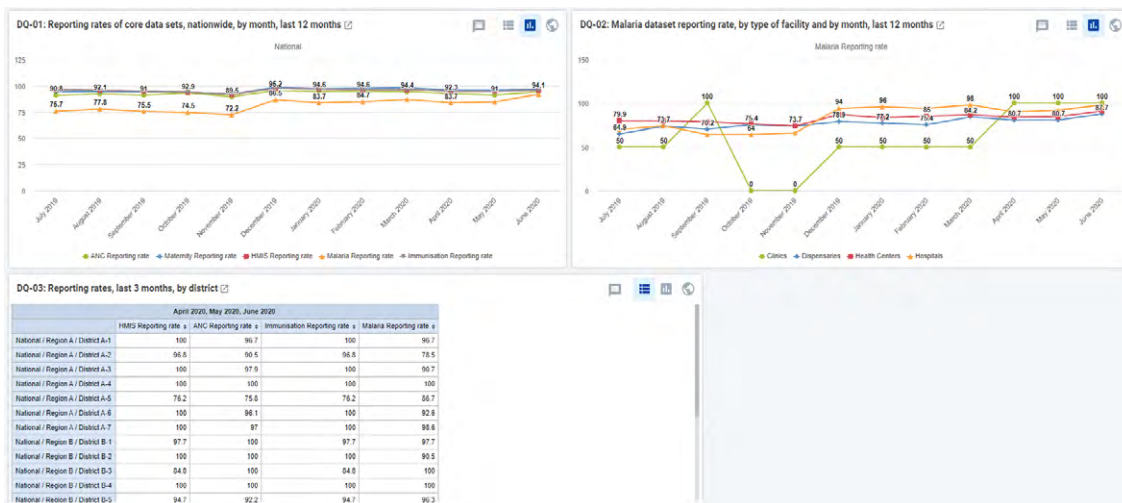
Dimension 2: Internal consistency

The next five items (**DQ-04, DQ-05, DQ-05b, DQ-06** and **DQ-06b**) are used to visualize month-to-month consistency or year-to-year consistency and to identify outliers. Outliers are values that are very different from the values reported by the same district or facility for other periods.
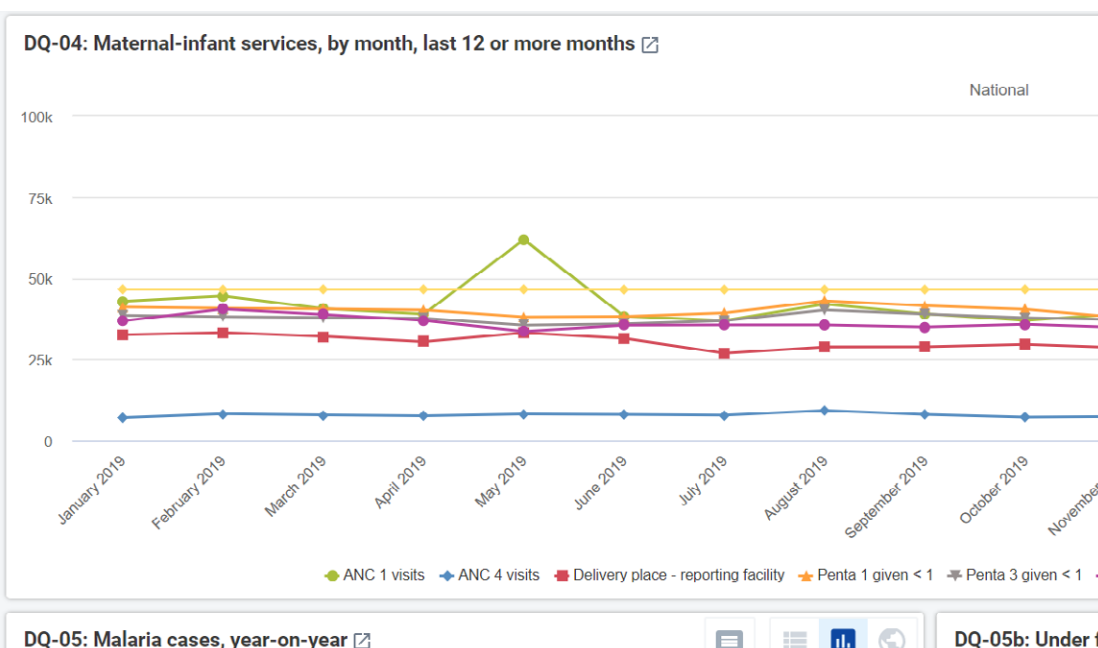
DQ-07 is an example of a chart for visualizing consistency between related indictors. In this case, the chart identifies districts that, over the last 12 months, had a negative Penta 1 to Penta 3 dropout rate.
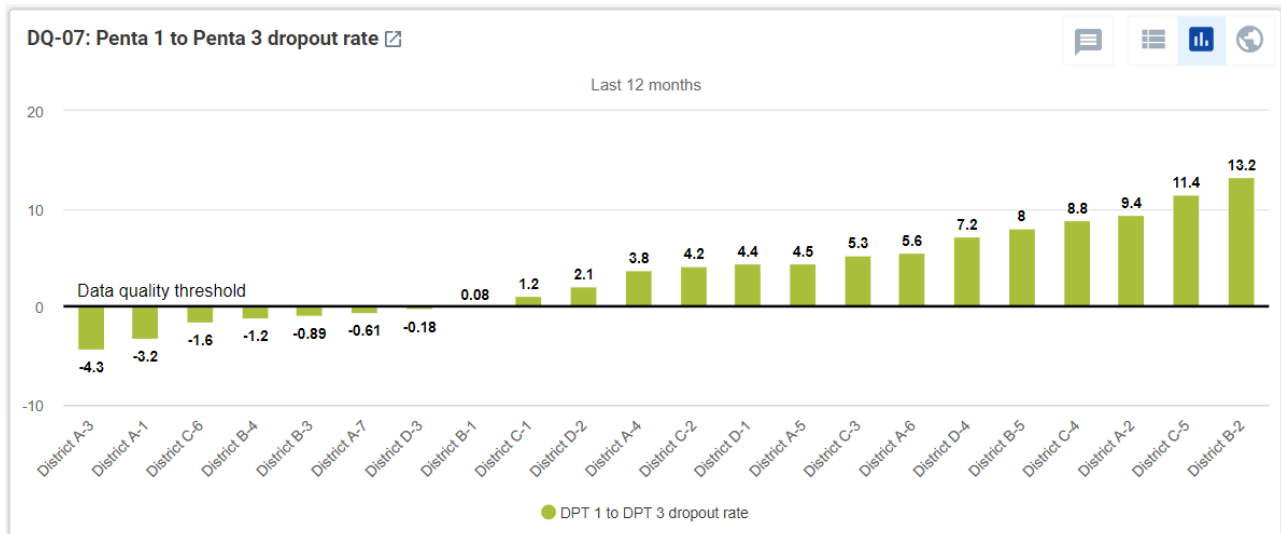
Dimension 3: External consistency

DQ-08 assesses the consistency of Demographic and Health Survey (DHS) estimates of coverage (green columns) with routine estimates of coverage, i.e. coverage that is calculated by dividing the data reported from facilities by the estimated target population.
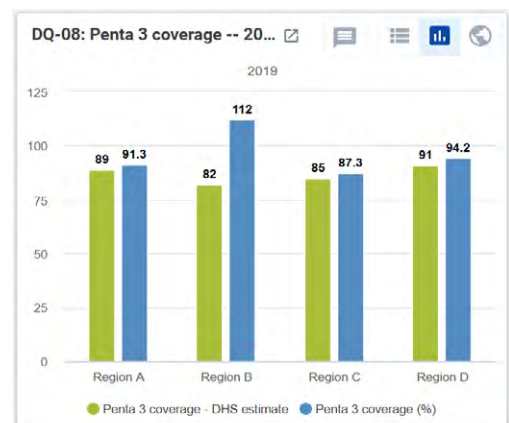
Dimension 4: External comparison of population data

DQ-09 is used to assess the consistency of estimates used by DHIS2 for various target populations for maternal and infant health services.
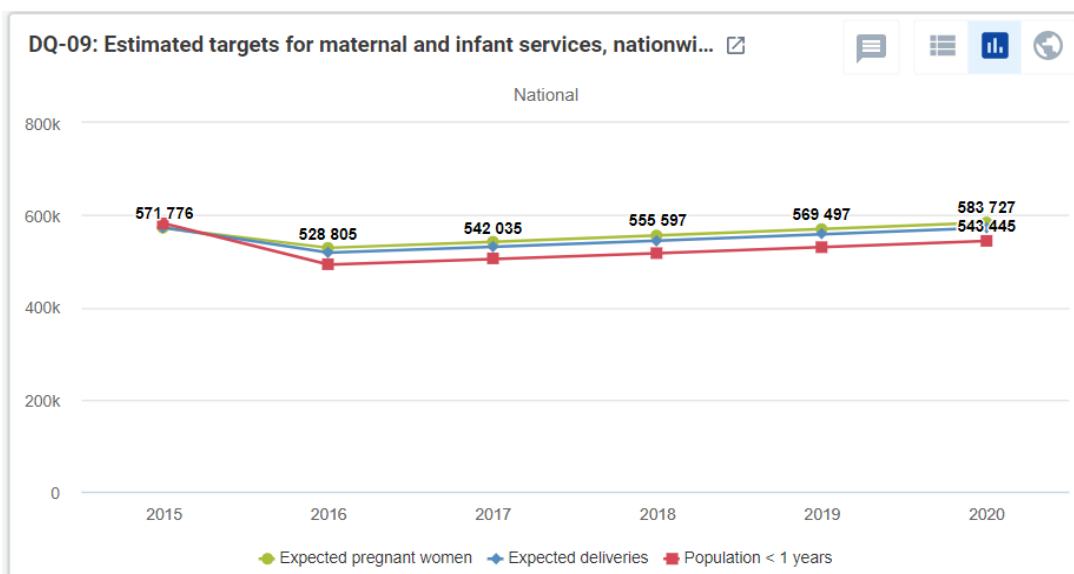
Dimension 1: Assess for incomplete reporting of key datasets

DQ-01 shows the trend, over the last 12 months, in the nationwide reporting rates for multiple datasets.
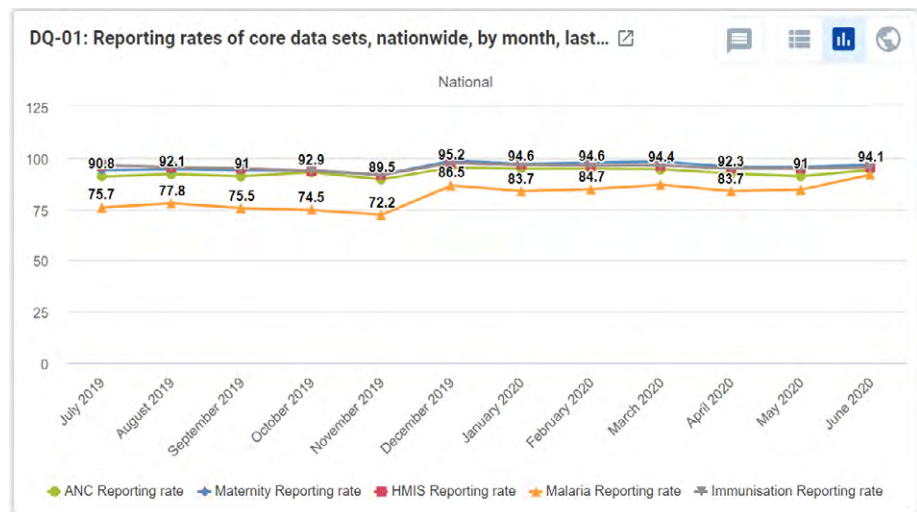
A chart such as DQ-01, with multiple lines, is useful but may sometimes be difficult to read. To simplify the chart, click on the names of some of the data elements/indicators in the legend at the bottom of the chart. This will hide the lines for those data elements and the names in the legend will become light grey.
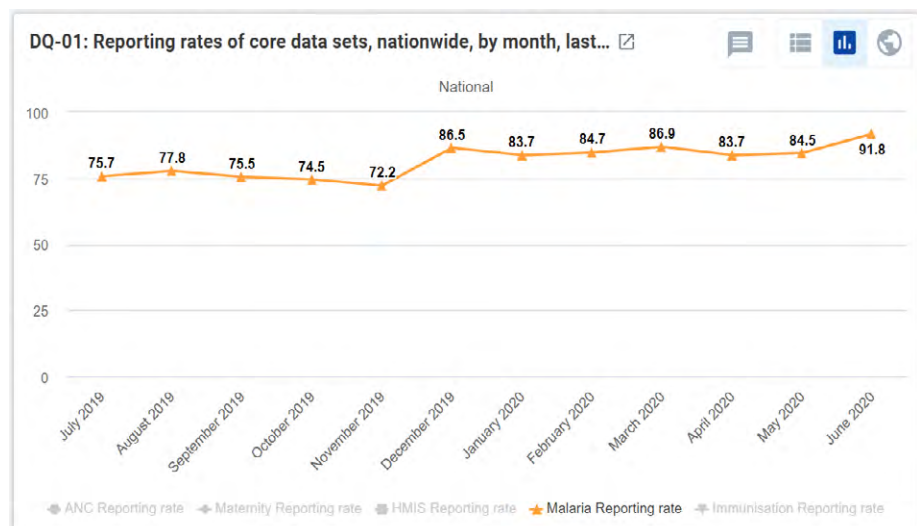
Alternately, if the cursor is placed over a single line, the other lines will become lighter.
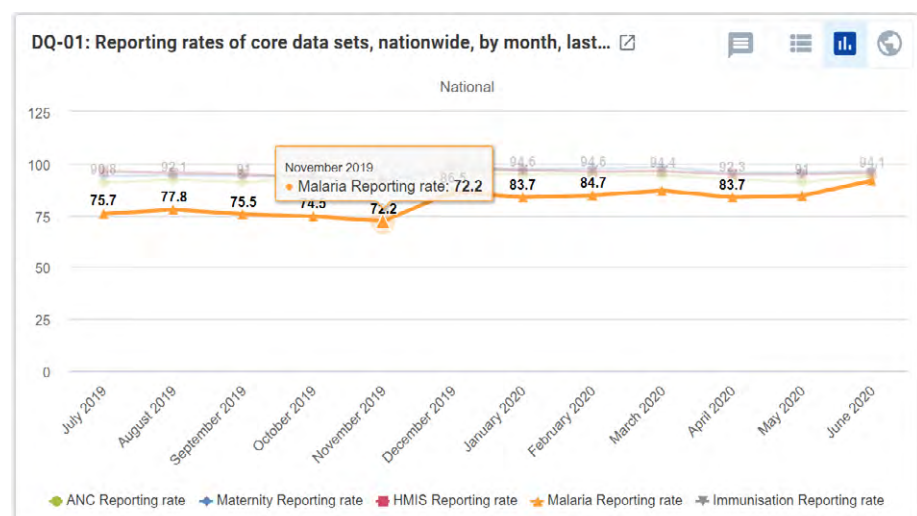
DQ-02 shows that the Data Visualizer app can be used to produce a chart that disaggregates a reporting rate *by type of facility*. In this example, reporting appears to be especially low for dispensaries and clinics.

DQ-02: Malaria dataset reporting rate, by type of facility and by mon...
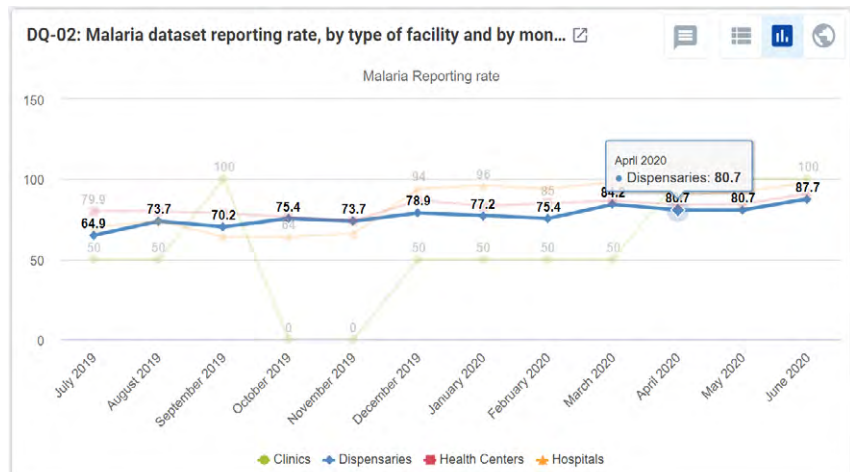
Q-03 presents reporting rates for multiple datasets for the last 3 months for each district. By clicking twice on a column header (e.g. "ANC Reporting rate"), the rows can be sorted from low to high for a specific dataset.

Question: When this table is viewed at the beginning of a month, the numbers will be lower than at the end of the month. Why?[2]

DQ-03: Reporting rates, last 3 months, by district

	April 2020, May 2020, June 2020			
	HMIS Reporting rate ⇕	ANC Reporting rate ⇕	Immunisation Reporting rate ⇕	Malaria Reporting rate ⇕
National / Region A / District A-5	76.2	75.8	76.2	86.7
National / Region D / District D-1	90.3	76.5	90.3	71.2
National / Region C / District C-5	98.4	84.7	98.4	88.9
National / Region D / District D-2	75	88.1	75	66.7
National / Region D / District D-4	78.4	89.3	78.4	85.4
National / Region A / District A-2	96.8	90.5	96.8	78.5
National / Region B / District B-5	94.7	92.2	94.7	96.3
National / Region C / District C-6	100	92.9	100	90.5
National / Region A / District A-6	100	96.1	100	92.6
National / Region A / District A-1	100	96.7	100	96.7
National / Region A / District A-7	100	97	100	98.6
National / Region D / District D-3	100	97.6	100	97.6

Dimension 2a: Identify month-to-month inconsistencies

DQ-04 presents the nationwide totals, over the last 12 to 24 months, for several related maternal health and immunization indicators.

DQ-04: Maternal-infant services, by month, last 12 months

---
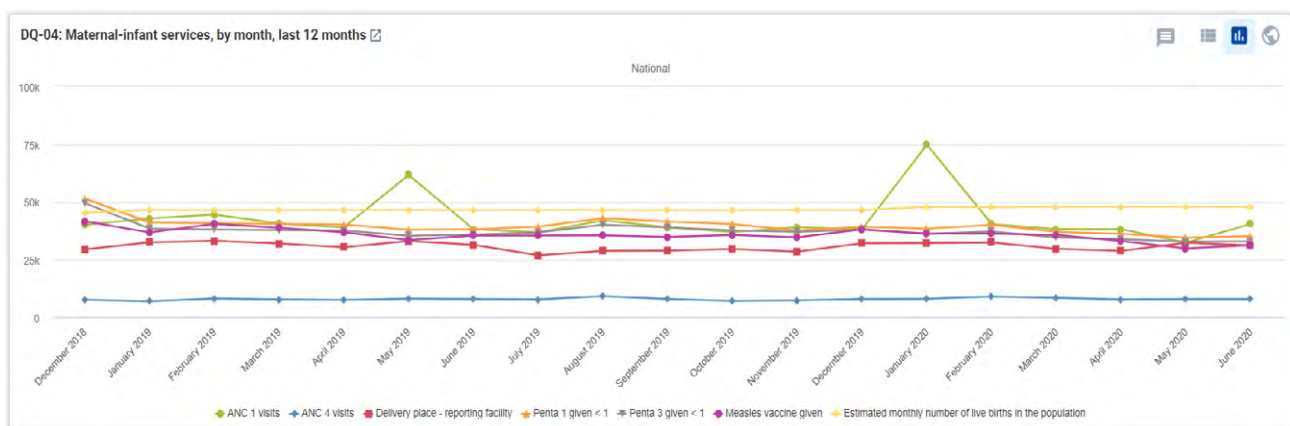
2   When DQ-03 is viewed at the beginning of a month, the data will not yet have been entered for the previous month. As a result, the highest reporting rate for any health facility for the 3-month period will be only 66.7%. For this reason, it is best to wait until the data for the previous month have been entered for most health facilities before reviewing the data quality dashboard.

Again, a chart with multiple lines may be difficult to read. Click on different lines to simplify the chart. Is there anything suspicious in the data shown in the chart below?[3]
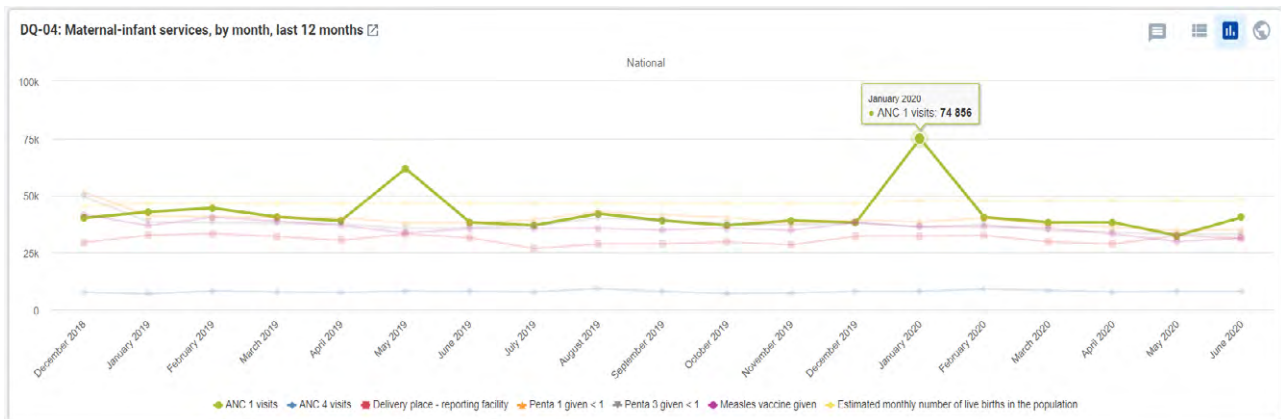
In a month-to-month chart like DQ-04, only the largest and most suspicious values can be identified. As discussed later, the "Outlier" function of the WHO Data Quality Tool is much more sensitive for detecting suspicious values than a month-to-month chart.

DQ-05 shows the month-to-month trend in malaria cases. Note that the values change a lot between one season (June to November) and another (December to May). This great variation in the values, based on the season, makes it very difficult to identify suspicious data. A high value may be due to a seasonal increase or it may be due to an error in the data. When DQ-05 is viewed at national level, no suspicious values are seen. However, when this chart is viewed at district level, it may reveal suspicious values – especially if a high value is reported during the non-malaria season (December to May).
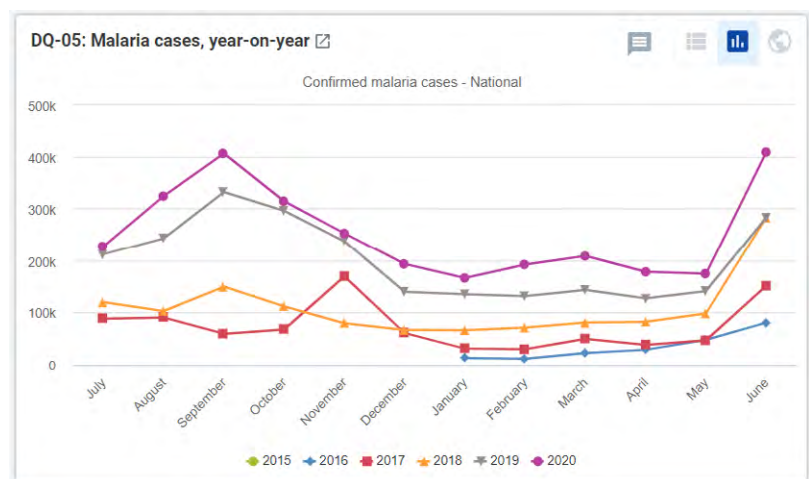
DQ-05b shows a way to identify suspicious malaria values when data are disaggregated by age group. It presents, for each district, the trend in the under-5 years outpatient department (OPD) malaria cases as a percentage of total OPD malaria cases. At first glance, the chart looks like a mess of lines. However, if the cursor is placed on the points which appear above or below all the other lines, what is seen?

---

3   The values for ANC 1 visits for May 2019 and January 2020 are suspiciously high. These are examples of extreme outliers.

For each month and for all districts, under-5 malaria cases represent a roughly constant percentage of the total malaria cases (25–50%). But District A-2 reported a much lower percentage than usual for February 2020. This is quite suspicious and suggests that the malaria data for District A-2 for February 2020 should be reviewed for a possible error.
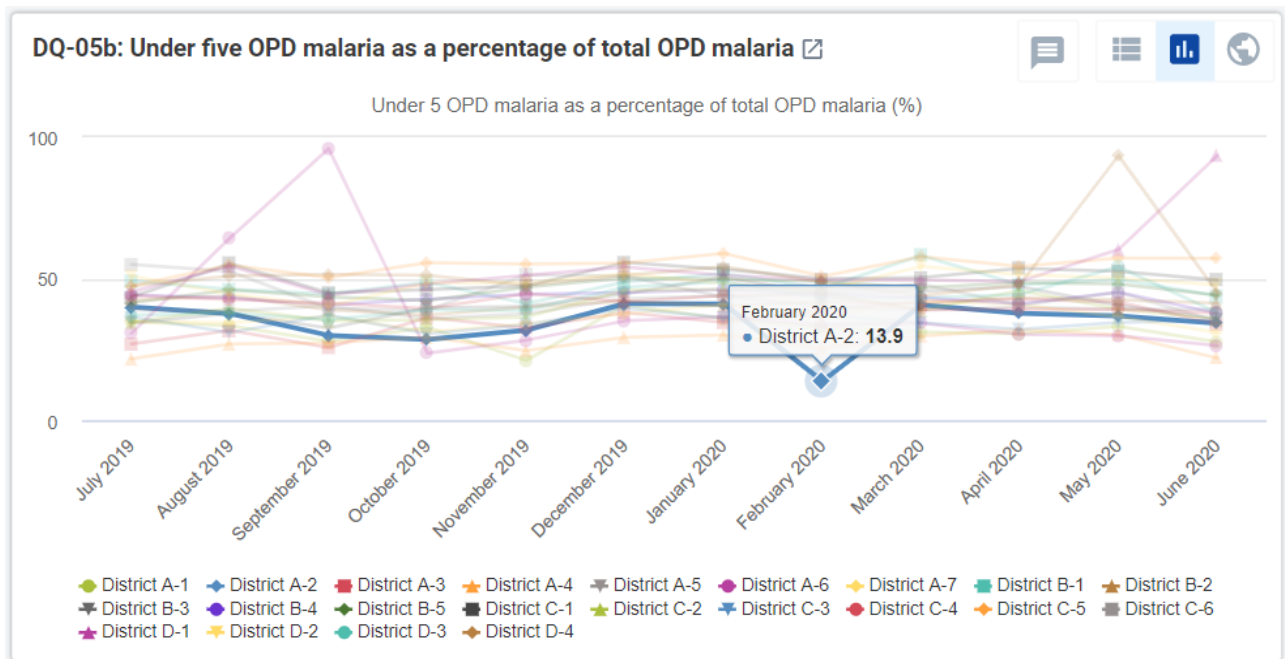
If **DQ-05b** is reviewed at the level of District A-2, a suspicious value is found for Facility 197.
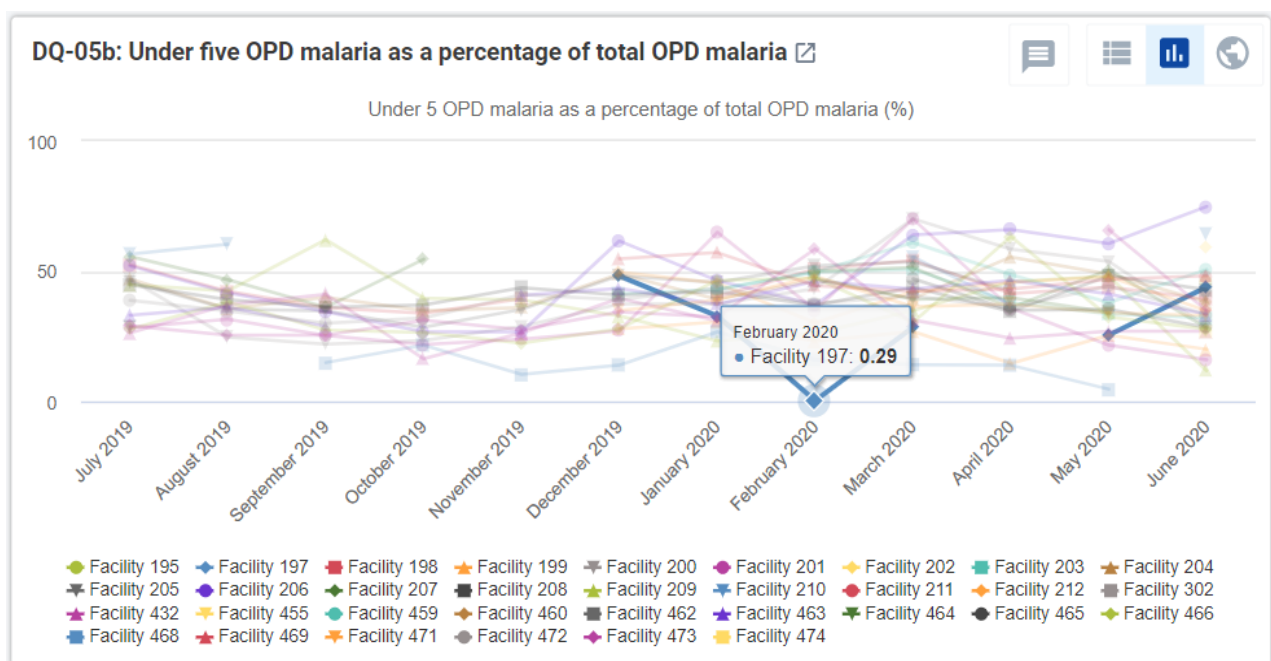
DQ-06 summarizes all the very large and extreme outliers reported for a selection of indicators. When this table is viewed at national level, it shows only the largest of extreme outliers found for any month for each of the indicators listed.[4] Each of these very large extreme outliers should be investigated.

Organisation unit	Data / Period	January 2019	February 2019	May 2019	June 2019	August 2019	September 2019	December 2019	January 2020	February 2020	August 2020	September 2020
National	ANC 1 outlier > 1,000	1 515	2 050	25 125						35 888		19 800
	ANC 4 outlier > 1,000					1 172						
	Penta 1 outlier > 1,000										1 155	1 323
	Penta 3 outlier > 1,000				1 230			1 117			1 337	
	OPD presumed malaria outlier > 100,000						653 536					
	Currently on ART outlier > 10,000									16 555		

Table title: DQ-06 Very large outliers, by month, last 12 to 23 months -- values reported for a SINGLE health facility for a SINGLE month

DQ-06b provides a way to begin such an investigation. The first column (on the left) lists individual health facilities. There is a very large and extreme outlier for each of the columns immediately following this first column (ANC 1 outliers > 1000; etc.) To identify the health facility that reported the outlier, click on the cell specifying the month beneath the name of the indicator. For example, click on the cell for May of last year beneath ANC 1 outlier > 1000. The rows of the table will then be sorted so that the very large outlier appears at the top. The table then shows that the very large outlier (25 125) was reported by Facility 599 of District D-1.

DQ-06b: Very large outliers, by health facility and by month, last 12+ months

	ANC 1 outlier > 1,000					ANC 4 outlier > 1,000	Penta 1 outlier > 1,000	
	January 2019	February 2019	May 2019	January 2020	September 2020	August 2019	February 2020	August 2020
National / Region A / District A-1 / A-1 District Hospital								
National / Region A / District A-1 / A-1 NGO Hospital								
National / Region A / District A-1 / Facility 313								
National / Region A / District A-1 / Facility 323								
National / Region A / District A-1 / Facility 327								
National / Region A / District A-1 / Facility 328								

DQ-06c is configured for use at district level, to identify the specific health facility that reported each extreme outlier, including smaller extreme outliers. This table may take a long time to appear. Do not spend time reviewing it now. It will be reviewed as part of Tutorial 3.

---

4   Ignore the final column for "June 2020". This appears to be due to a "bug".

Dimension 2b: Look for inconsistencies between related indicators
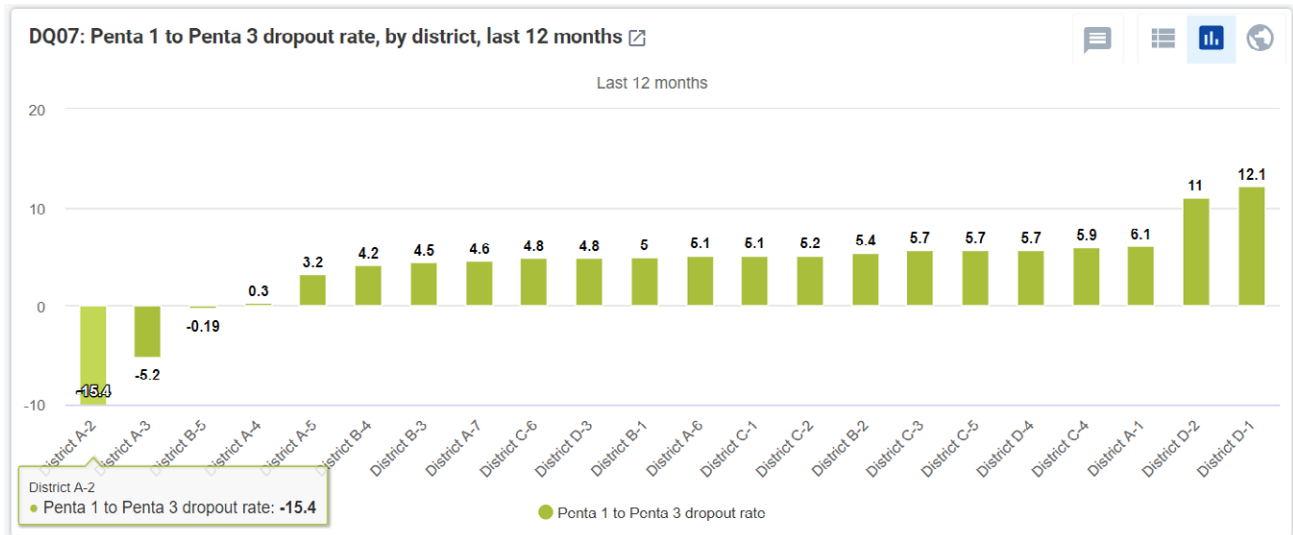
DQ-07 represents an example of how to compare the values of related indicators. In this case, the chart compares, for each district, the number of Penta 1 doses to the number of Penta 3 doses. This is done by calculating the "Penta 1 to Penta 3 dropout rate" which equals (Penta 1 doses minus Penta 3 doses) times 100 divided by (Penta 1 doses). This represents the percentage of children who receive their first dose of Penta vaccine but who do not come back for their third dose. If the value is negative for an entire district for a full 12 months, it means that the district reported more third doses of the vaccine than first doses. DQ-07 ranks districts by their dropout rate from low to high; therefore, any districts with a negative dropout rate will appear on the left of the chart.

Dimension 3: Assess the consistency between coverage estimates based on a survey and coverage calculated using routine data

Coverage surveys such as the DHS or the Multiple Indicator Cluster Survey (MICS) are often considered to provide "the gold standard" for measuring coverage in childhood immunization, ANC and deliveries in health facilities. Where possible, statistics from such surveys should be entered/imported into DHIS2 so that coverage estimates based on a survey can be compared with coverage calculated using routine data.

DQ-08 provides an example of this, comparing the two estimates of Penta 3 coverage for each region. What can be concluded from this chart? Are the two coverage measurements consistent with each other? For which region are the measurements not consistent? If it is assumed that the survey estimates are reliable, what are some possible reasons why the routine estimate for this region is not consistent with the survey estimate?[5] What are some possible reasons why a survey estimate may not be reliable?[6]

---
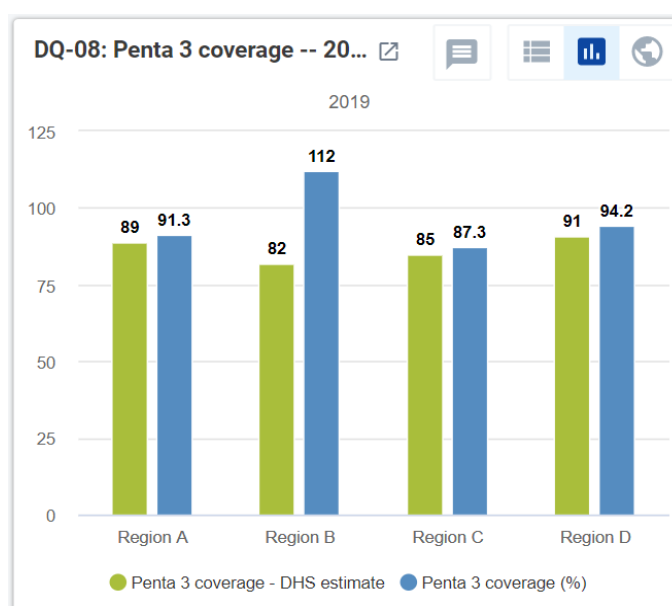
5   Some reasons why a coverage estimate based on routine data may be unreliable: 1) the number of Penta 3 doses may have been over-reported, either as a result of a transcription error when entering the data into DHIS2 or as a result of doses of Penta 1 and Penta 2 being misclassified and reported as Penta 3 doses; 2) the denominator (estimated population under 12 months of age) may have been underestimated.

6   To provide reliable estimates, household surveys must use methods that strictly adhere to standards. Survey measures can be less reliable for the following reasons: 1) sample size too small; 2) sample not selected randomly; 3) recall bias – reliance on the memory of caretakers when vaccination cards are not available for a large percentage of children.
     DHS and MICS surveys report the confidence intervals of survey estimates. These confidence intervals estimate the effect of sampling error, which decreases as the sample size increases. However, surveys are also subject to non-sampling errors such as those due to selection bias or recall bias. The direction and size of the non-sampling error of a survey can seldom be estimated even though it may significantly exceed the sampling error.

Dimension 4: Assess the consistency of denominator/target estimates

As previously discussed, if the estimate of the denominator/target is not reliable, the coverage estimates based on routine data will not be reliable. **DQ-09** shows the year-to-year trend in the estimated denominators used to calculate coverage in maternal and infant health services. A simple chart such as this one can be used to quickly assess two aspects of consistency:

▷ Have the estimated denominators grown consistently from year to year or has growth been significantly different in some years than in others?

▷ Is there a logical progression in the values of related denominators, i.e. "Expected pregnancies" are greater than "Expected deliveries" and "Expected live births" are greater than "Population < 1 year"?

In the example provided, there is a logical progression in the values of related denominators. The numbers of expected pregnancies are slightly greater than the numbers of expected deliveries. Expected deliveries are slightly greater than the estimated population under 1 year of age. This is expected, as some pregnancies are lost due to miscarriages and some births are lost due to stillbirths and infant deaths.

However, is there anything that is inconsistent about the data presented in DQ-09? What could explain why the targets dropped between 2015 and 2016?

DQ-09: Estimated targets for ma...

National

Year	Expected pregnant women	Expected deliveries	Population < 1 years
2015	571 776		
2016	528 805		
2017	542 035		
2018	555 597		
2019	569 497		
2020	583 727	543 445	
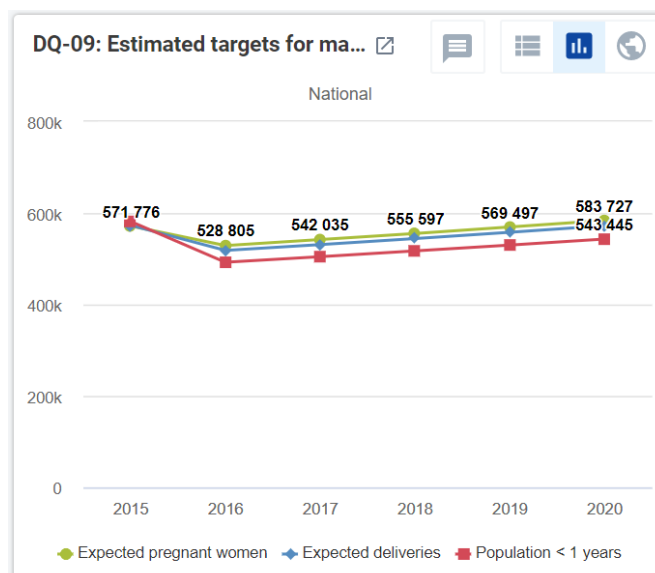
An abrupt drop or abrupt rise in targets between one year and the next can result when the national statistical office releases new results from a recent census. To avoid such year-to-year inconsistency, the correct practice is to go back and edit the target estimates for previous years so that they are consistent with the subsequent trend. In this case, the estimated targets for 2015 should be reduced.

Tutorial 3: How to train district staff in the monthly use of DHIS2 data quality tools

Overview

The best way to assure the quality of DHIS2 data is for data managers to check them each month soon after they have entered the data for most health facilities. DHIS2 provides two powerful tools for quickly assessing the completeness and consistency of data:

> ▷ a data quality dashboard which is visible as soon as the user logs into DHIS2; and
> ▷ the WHO Data Quality Tool – a DHIS2 application.

This tutorial details the steps to follow for using these two DHIS2 tools to rapidly identify "outliers" (suspicious data values) and important missing data values. By following the steps, it should be possible to train district-level staff (for example, data clerks) in less than 30 minutes to use these data quality tools.

Adapt this tutorial for the national instance of DHIS2

This document provides instructions and screenshots adapted for the "DQ Trainingland" instance of DHIS2 (https://who-demos.dhis2.org/dq).[1] To practise these steps, log in using the username and password which appear on the login page.

When using the tutorial for training in a specific country, several of the screenshots and some of the terms should be adapted to match the national DHIS2 instance.

Nationwide implementation of this approach to district DQA will require that it is incorporated into an update of the official SOPs for data management.

---

1 Several of the charts and tables of the data quality dashboard are configured to present data for the "last 12 months". The months included in "last 12 months" will change each month. As a result, the screenshots included in this tutorial (which were captured in July 2020, will differ somewhat from the dashboard presented on https://who.dhis2.net/dq. When preparing for a training in a specific month, facilitators may choose to update Tutorial 2 and Tutorial 3 so that the "last 12 months" shown in the screenshots of charts and tables is current and identical to what will be seen on the DHIS2 website.

Review the data quality dashboard

Log in and review the data quality dashboard

Log into the DHIS2 instance: URL: https://who-demos.dhis2.org/dq
Username: demo   Password: District1#

A data quality dashboard is visible as soon as the user logs into DHIS2; it is not necessary to launch any application to view the dashboard.
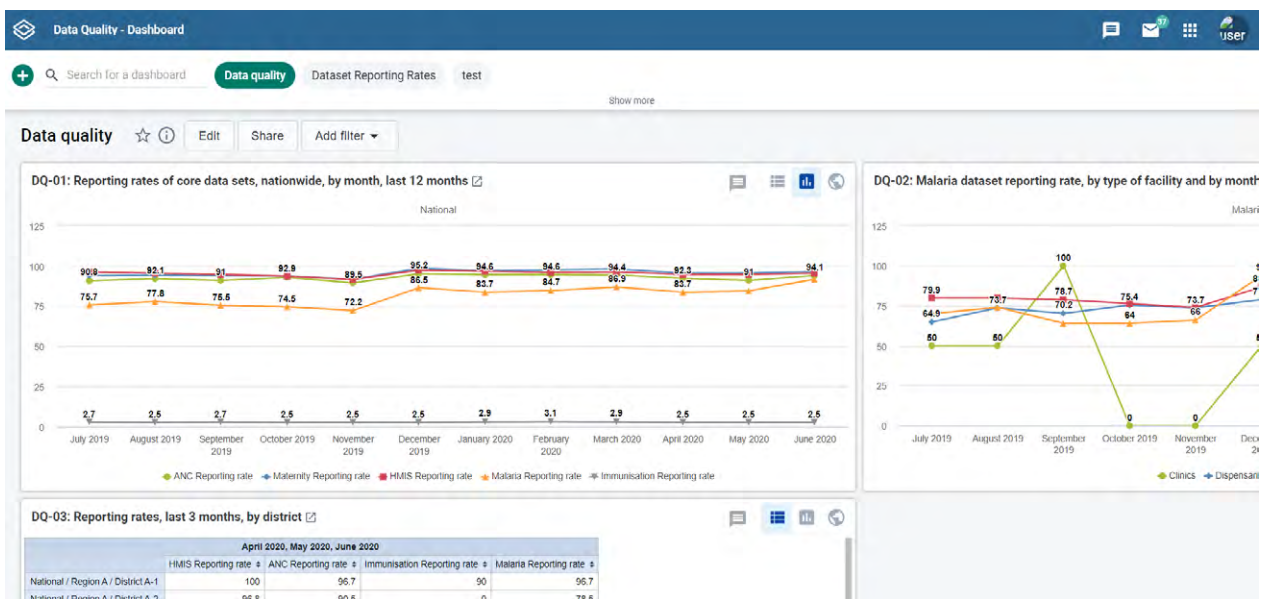
When a user logs in at the national level, they see charts and tables for assessing the data of the entire country.
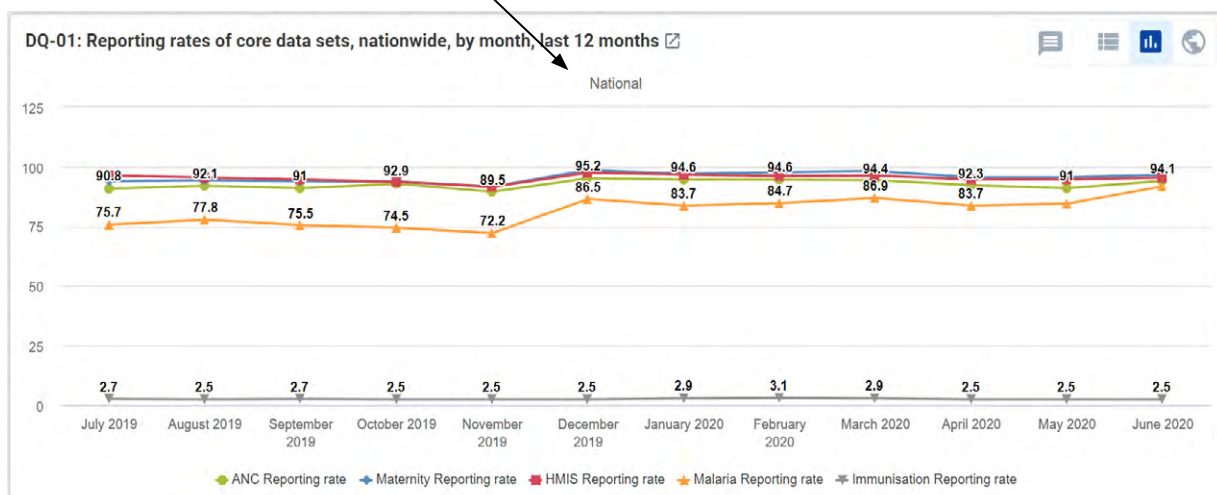
However, when a district data manager logs on, they usually see data only for their specific district.
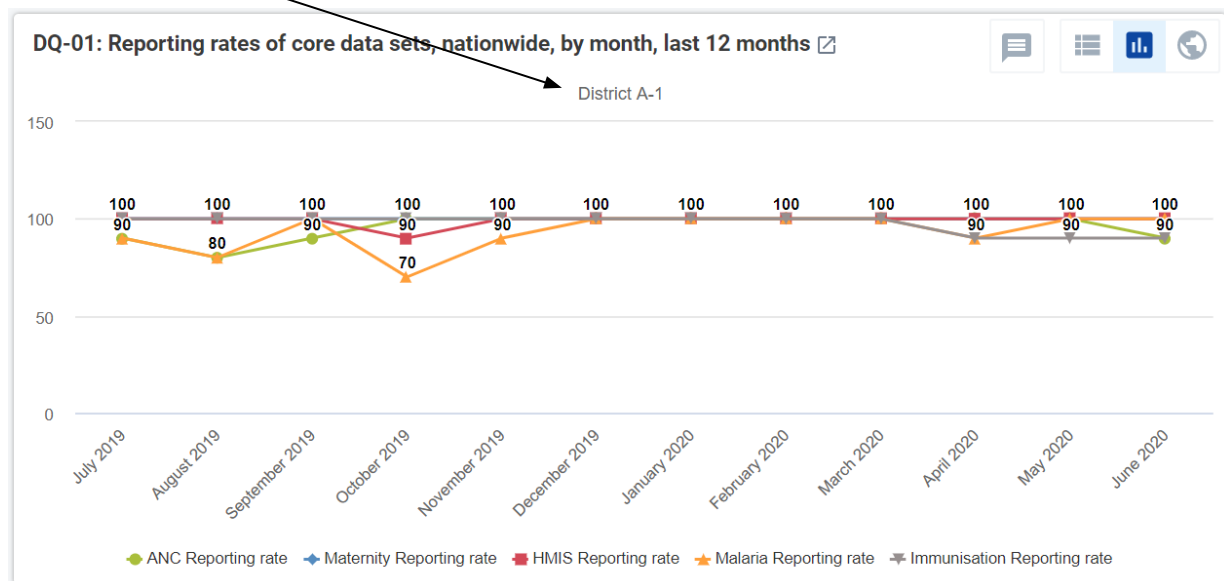
Filter the dashboard to see results for a single district

For supervisors and other staff monitoring data at higher levels, it is useful to "filter" the nationwide data quality dashboard, to view results for only one specific district.

To do this, click on the "Add filter" button. A drop-down menu appears. Click on "Organisation Unit".
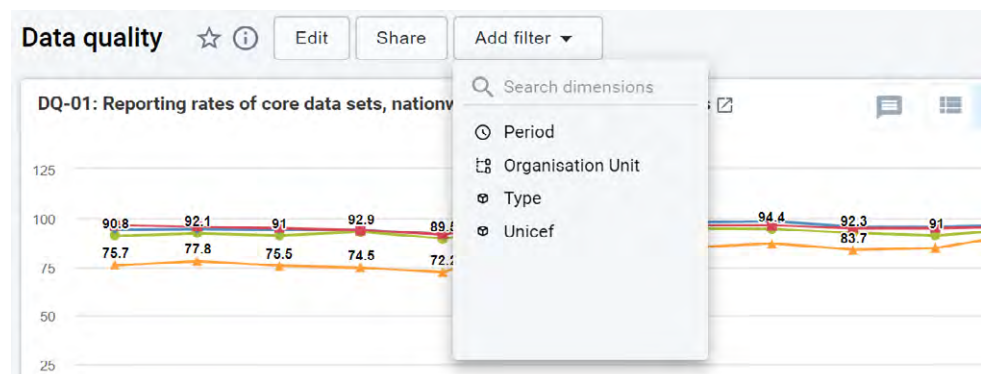
After clicking on "Organisation Unit" a window appears. To view a list of regions, click on the triangle to the left of the checkbox for "National".

To view a list of districts in a region, click on the triangle to the left of the checkbox for the region. Place a check in the box for any district.

To view results disaggregated by the facilities in the district, click on the triangle to the right of "Select a level" and choose "Facility". Click twice on "Confirm": the dashboard converts to show only the results for the selected district, disaggregated by facility.
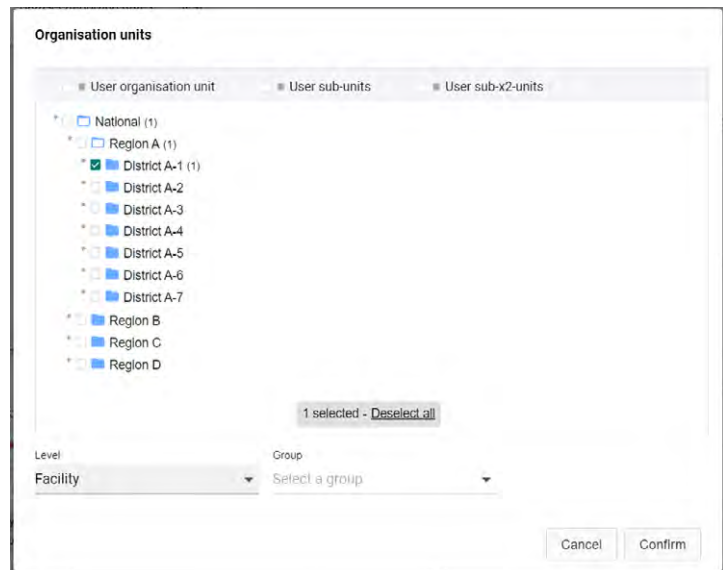
Assess the completeness of reporting

The first three items on the data quality dashboard can be used to assess the "reporting rates" (also known as "completeness") for key datasets. Reporting rate is defined as the number of monthly (or quarterly) reports submitted, divided by the number of reports expected – expressed as a percentage.

DQ-01 shows the trend, over the last 12 months, in the district's reporting rates for multiple datasets. When this chart is viewed at the beginning of a month, the reporting rates for last month (the month on the right) will be very low or zero. Why?[2]
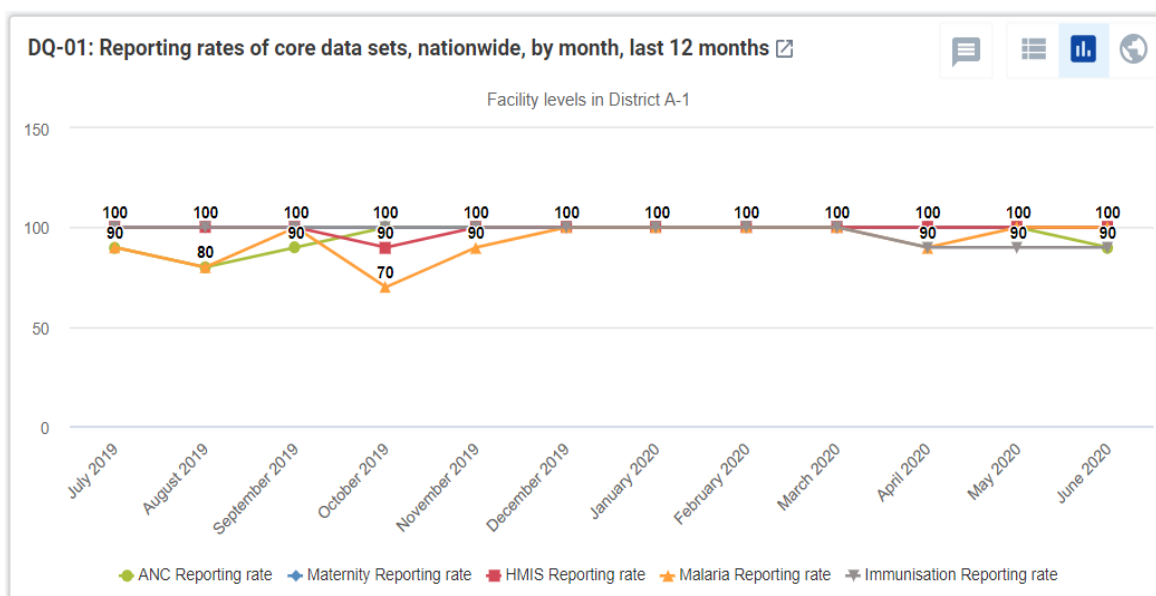
A chart like the one above, with multiple lines, is useful but may be difficult to read. To simplify the chart, click on the names of some of the data elements or indicators in the legend at the bottom of the chart. This will hide the lines for those data elements and the names become light grey.
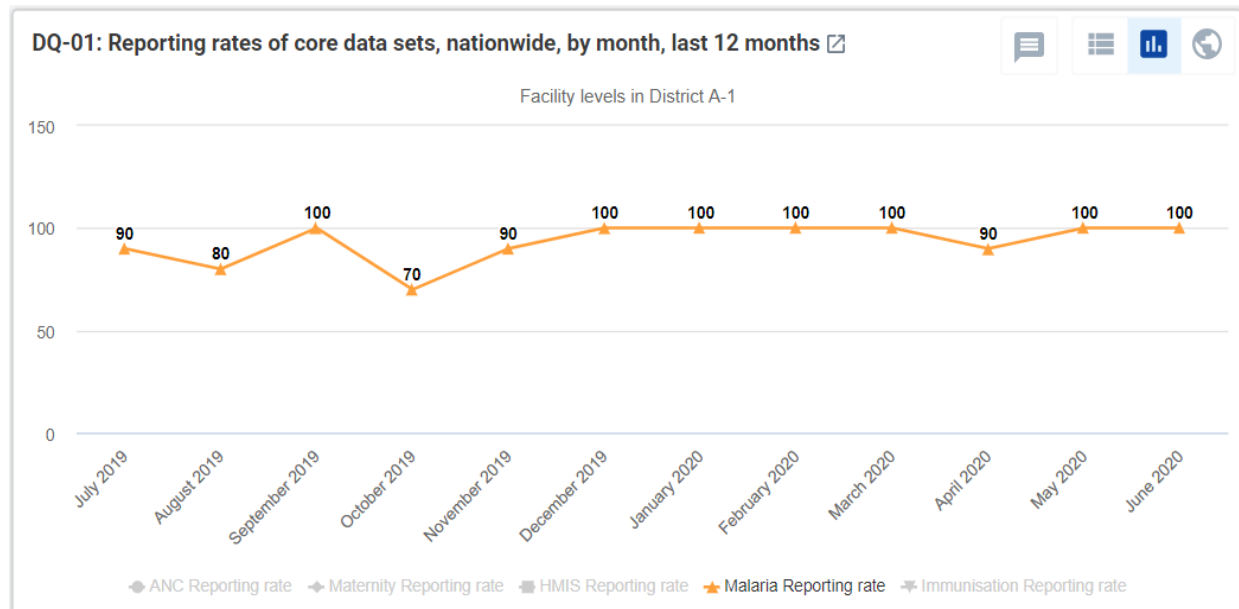
DQ-02 shows that DHIS2 can be used to configure a chart that *disaggregates* a reporting rate *by type of facility*. For District A-1, the reporting rate has been 100% for each of the last 12 months for hospitals but not for health centres.

DQ-03 presents reporting rates for multiple datasets for the last 3 months for each of the health facilities in the district. By clicking twice on a column header (e.g. "ANC Reporting rate"), the rows can be sorted from low to high for a specific dataset.

DQ-03: Reporting rates, last 3 months, by district ⤢

	HMIS Reporting rate	ANC Reporting rate	Immunisation Reporting rate	Malaria Reporting rate
National / Region A / District A-1 / Facility 313	100	66.7	0	66.7
National / Region A / District A-1 / A-1 District Hospital	100	100	100	100
National / Region A / District A-1 / A-1 NGO Hospital	100	100	100	100
National / Region A / District A-1 / Facility 323	100	100	100	100
National / Region A / District A-1 / Facility 325	100	100	100	100
National / Region A / District A-1 / Facility 326	100	100	100	100
National / Region A / District A-1 / Facility 327	100	100	100	100
National / Region A / District A-1 / Facility 328	100	100	100	100
National / Region A / District A-1 / Facility 340	100	100	100	100
National / Region A / District A-1 / Facility 610	100	100	100	100

April 2020, May 2020, June 2020

This feature is especially useful if there are many health facilities in the district, as more than one can be viewed at the same time on the dashboard. **Question:** When this table is viewed at the beginning of a month, the numbers will be lower. Why?[3]

Assess the month-to-month consistency of the data

The next four items on the dashboard can be used to assess whether data are consistent from month to month.

DQ-04 presents the district totals, over the last 12 to 24 months, for several related maternal health and immunization indicators. The values for these indicators often do not change much from month to month. As a result, the month-to-month trend lines for these indicators are often almost straight, horizontal lines.

---

3   When DQ-03 is viewed at the beginning of a month, the data for the preceding month will not have been entered yet. As a result, the highest reporting rate for any health facility for the 3-month period will be only 66.7%. For this reason, it is best to wait until the data for last month have been entered for most health facilities before reviewing the data quality dashboard.
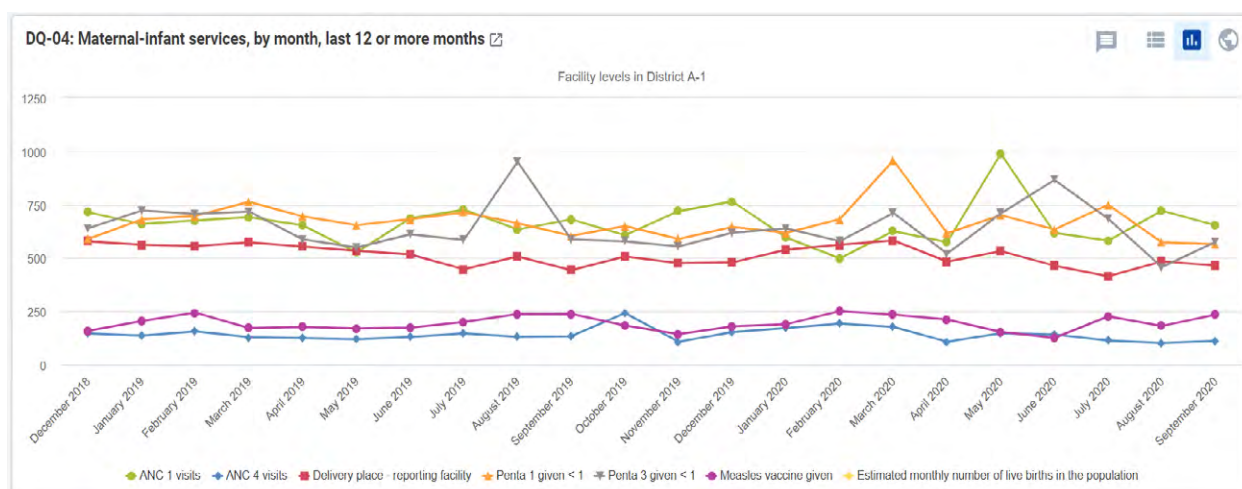
A chart with multiple lines may be difficult to read. To simplify the chart, click on the names of some of the data elements/indicators in the legend to hide the lines for those data elements. Is there anything suspicious with the data shown in the chart below?
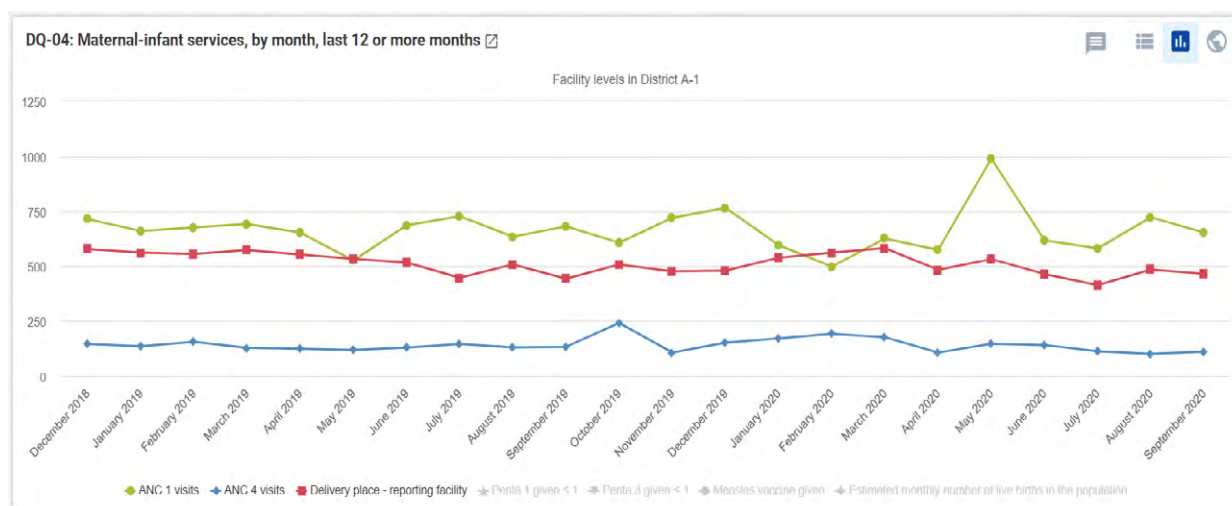
In a month-to-month chart like DQ-04, only the largest and most suspicious values can be identified, especially when viewed at national level. However, when this chart is viewed at the level of District A-1 the value of ANC 1 visits for May of the current year[4] is seen to be suspiciously high. There is also a suspicious rise in the value of ANC 4 visits for October of last year. These are examples of "outliers". As demonstrated in the next section, the "Outlier" function of the WHO Data Quality Tool is even more sensitive than district-level dashboards for detecting suspicious values.

---

4   If the dashboard is viewed during February to May it will not show results for May of the current year.

[Optional][5] **DQ-05** is called a "year-on-year" chart. It shows the trend in numbers of confirmed cases of malaria for multiple years. Note that each year the reported number of confirmed cases is much higher during the season of June to November than in other months.
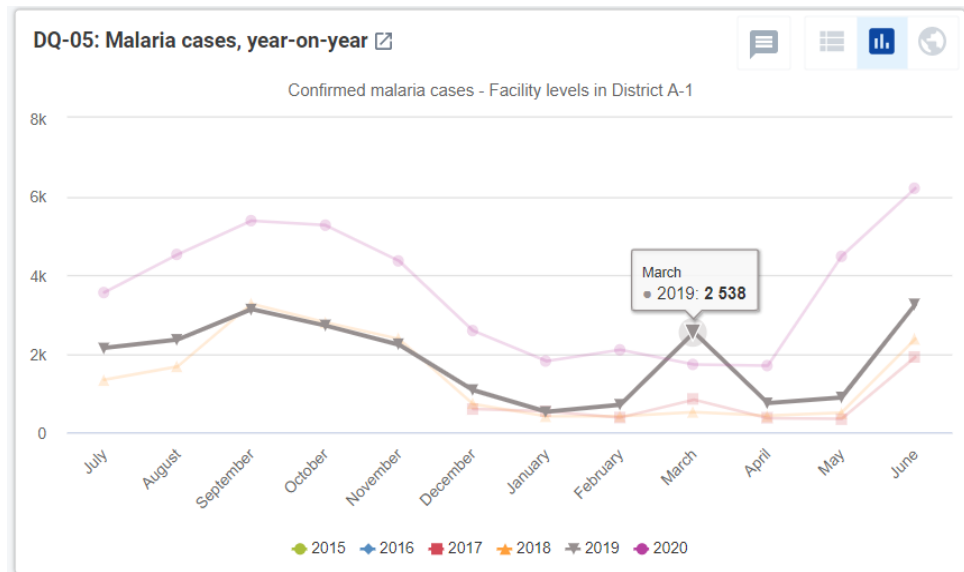
During this season, a substantial increase in the reported number of cases may be due to the real increase in the incidence of malaria. This makes it difficult to identify suspicious values of this indicator. However, a sudden increase in reported malaria cases during the non-malaria season (e.g. the value for March of last year) is suspicious and should be investigated.

DQ-05b shows a way to identify suspicious malaria values if data are disaggregated by age group. It presents, for each facility, under-5 years OPD malaria cases as a percentage of total OPD malaria cases.

The chart is difficult to interpret due to the number of lines. However, for almost all health facilities, the percentage of OPD malaria that was reported for children under 5 years does not change much from month to month, both during the malaria season and during the non-malaria season.

---
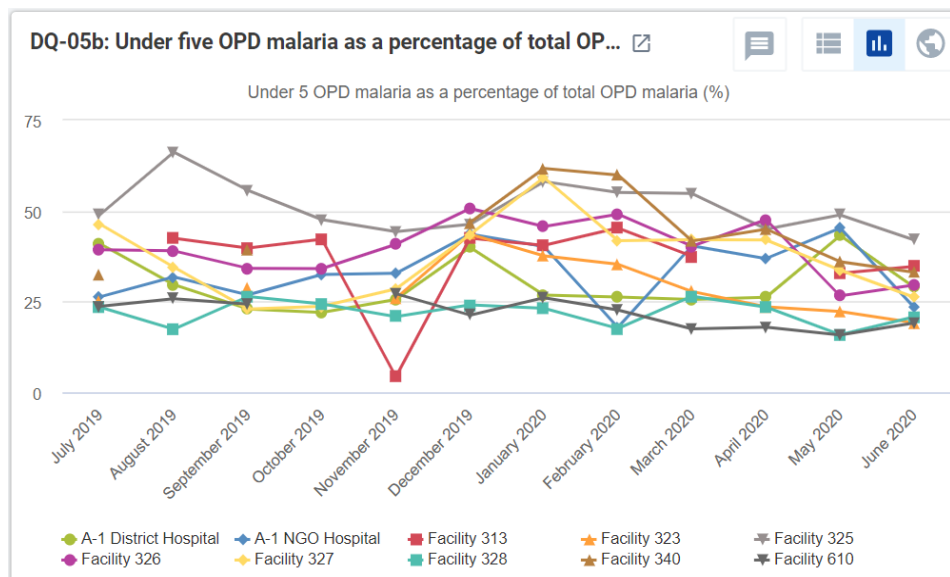
5   Year-on-year charts cannot be configured with versions of DHIS2 older than 2.31. However, similar charts can be viewed on the Consistency - time dashboard of the WHO Data Quality Tool.

However, one facility (Facility 313) shows an abrupt change in the percentage for November last year. This is highly suspicious. Data on confirmed malaria for Facility 313 for November of last year should be investigated.
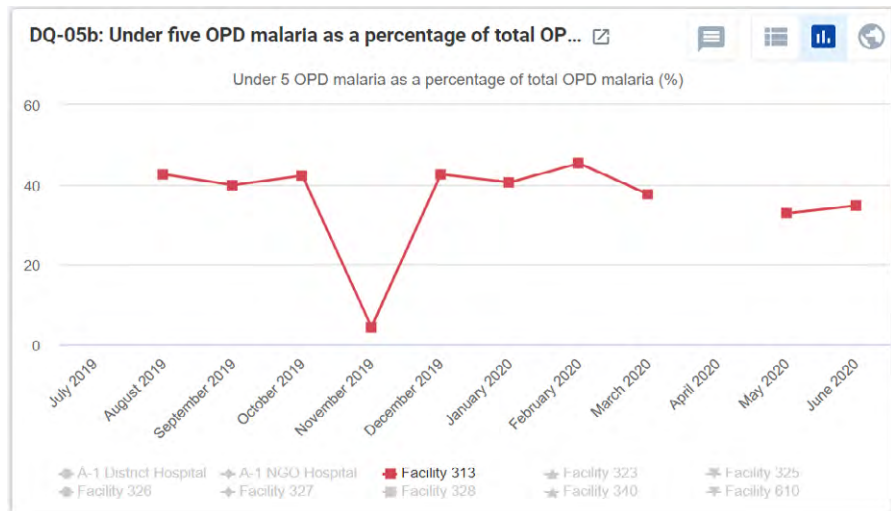
DQ-06 and DQ-06b are designed to detect very large and extreme outliers. This is useful at national level where data managers may want to focus on the most extreme outliers that have a value greater than 1000 (for antenatal visits or immunizations) or greater than 100 000 (for outpatient visits). None of the facilities of District A-1 reported any such very large and extreme outliers in the last 12 months.
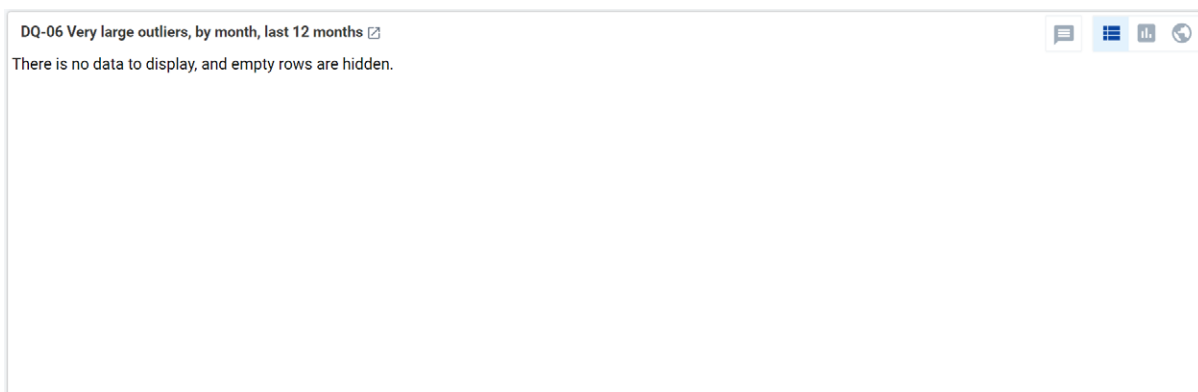
At district level, data managers have the advantage that they review data for a much smaller number of facilities than staff at national or regional levels. Therefore, they should investigate all extreme outliers, not only those that are very large.

DQ-06c, designed for use at local level, identifies all suspicious values from the last 12 months for a range of data elements. Values greater than 100 are highlighted in red. Each of these extreme outliers needs investigation. Note that some of the extreme outliers shown in DQ-06c explain some of the suspicious findings shown in DQ-04. For example, consider the extreme outlier value of 130 ANC 4 visits reported by A-1 NGO Hospital in October of last year. Is there evidence of this extreme outlier in DQ-04 (see above). As another example, consider the extreme outlier value of 341 ANC 1 visits reported for Facility 323 in May of the current year.[6]

---

6   When the dashboard is viewed from February to May, the values for May of the current year will not appear.

If the district has many health facilities and DQ-06c has many rows, some of the rows can only be viewed by scrolling downwards through the table. As an alternative, to see the largest extreme outlier for each month shown (months without any extreme outlier are hidden from the table), click once on the cell for the name of the month. The rows will then be re-sorted from high to low and the row for the health facility which reported the largest extreme outlier will be at the top. **Question:** How might these values be investigated?[7]

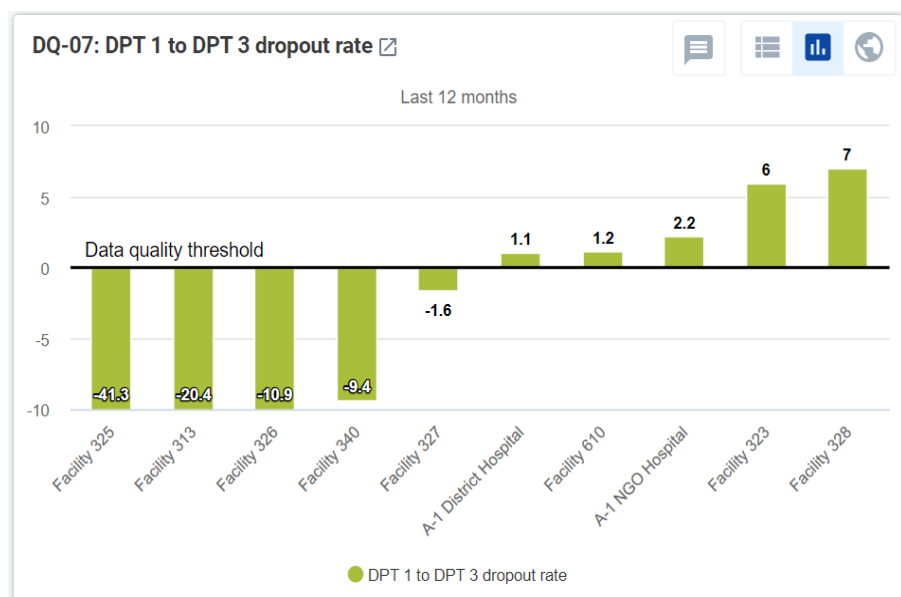| DQ-06c: Extreme outliers, by health facility and by month, since January of last year | | | | | | | | | | | | | | | | | | | | |
| Data | ANC 1 outlier | | | ANC 4 outlier | | | | | Penta 1 outlier | | | Penta 3 outlier | | | | | | | | |
Organisation unit / Period	September 2019	February 2020	May 2020	January 2019	August 2019	October 2019	February 2020	September 2020	January 2019	March 2020	July 2020	August 2019	January 2020	March 2020	May 2020	June 2020	July 2020	May 2019	August 2019	O
A-1 District Hospital					54					325										
A-1 NGO Hospital						130	41			177										
Facility 313		14																		
Facility 323			341																	▮
Facility 325	57							18				335		106					1 229	
Facility 326											86				101					
Facility 327																	109			
Facility 328				47					47					55						
Facility 340					36								92							
Facility 610																390		2 291		

Additional data elements can easily be added to this table.

Assess whether there is a good match between the values of related indicators

The final item on the data quality dashboard checks for consistency between related indicators. **DQ-07** identifies facilities which, over the last 12 months, had a negative Penta 1 to Penta 3 dropout rate. The dropout rate is the percentage of children who receive their first dose of vaccine but do not return for their third dose of vaccine. Facilities with a negative dropout rate (shown on the left side of the chart) reported a higher number of third doses of Penta vaccine than first doses of Penta vaccine. Sometimes a child may receive their first dose of vaccine at one facility and then receive the third dose at a different facility. However, a dropout rate that is negative over a full 12 months is suspicious and should be investigated.

DQ-07: DPT 1 to DPT 3 dropout rate — Last 12 months

Data quality threshold

Facility 325: -41.3; Facility 313: -20.4; Facility 326: -10.9; Facility 340: -9.4; Facility 327: -1.6; A-1 District Hospital: 1.1; Facility 610: 1.2; A-1 NGO Hospital: 2.2; Facility 323: 6; Facility 328: 7

● DPT 1 to DPT 3 dropout rate

---

7   The investigation can begin with further DHIS2-based desk review using the Pivot Table app and other applications.
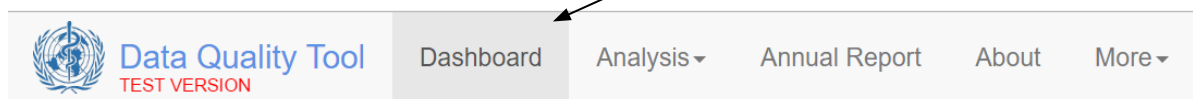
Review the WHO Data Quality Tool

Launch the DQ Tool and explore

1. **Log into the DHIS2 instance:** URL = who-dev.dhis2.org/dq
   Username: demo   Password: District1#

   Type "WHO" in the "Search apps" line in the upper right of the DHIS2 Home page. Click on **WHO Data Quality Tool.**
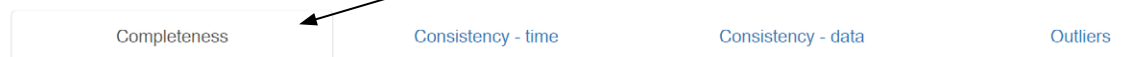
Search apps	⠿	aa

2. **Quickly review the menus of the DQ Tool:** After the DQ Tool is launched, five tabs appear at the top of the page. These are used to select between several different functions. Note that the word "Dashboard" is highlighted, and the Dashboard function is selected by default.

   Data Quality Tool TEST VERSION   Dashboard   Analysis ▾   Annual Report   About   More ▾

   The Dashboard has its own set of four tabs. The Completeness tab is selected by default.

   Dashboard

   Completeness   Consistency - time   Consistency - data   Outliers

   Do not spend long reviewing the Completeness dashboard. The completeness tab of the DQ Tool shows the reporting rate for select datasets. The same information is presented more comprehensively in the charts and tables on the data quality dashboard.

3. **Explore the settings window of the DQ Tool dashboard:** Click on the settings icon (▤) in the upper right of the screen to show a settings window on the right of the screen. If the settings window appears briefly but then disappears, click again until the window stays on the screen. Under "Data" there is a white space where [Core] appears. Click on the down arrow to the far right of this space. A drop-down menu will appear showing the various groups of indicators that can be reviewed on the DQ Tool dashboard: [Core], General Service Statistics, Maternal Health, HIV/Aids, Malaria, District and Immunisation.

   Data

District	✕ ▾

   Period

‹	2020	›
January	February	March
April	May	June
July	August	September
October	November	December

   Data for 12 months up to and including the selected month will be used.
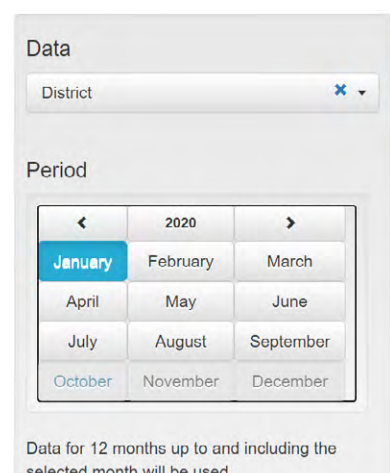
4. **Change "[Core]" to "District":** The reporting period that is shown on the dashboard can be changed by clicking on a different month. After experimenting with changing the **Period**, reset it to **January of the current year.**

5. **Change the "Organisation unit" that is being analysed:** By default, the DQ Tool analyses all data for the user's assigned organization unit. For example, if the user is assigned to district level, the DQ Tool will show all data for the user's district. If the user is assigned to national level, the DQ Tool will show results for all data nationwide. In the "Organisation unit" section of the settings window, under Boundary, "National" will then be highlighted.

    If the user is assigned to a higher level than district (e.g. regional or national level) but wants to view a specific district, they should use the settings window of the dashboard of the DQ Tool to change the "Organisation unit" Boundary from "National" to one specific district. To do this, under Boundary, click on "Other", then click on the "+" sign to the left of "National". The list of regions will then appear. Click on the "+" sign to the left of "Region A" and the list of districts in Region A will appear. Click on and select "District A-1".
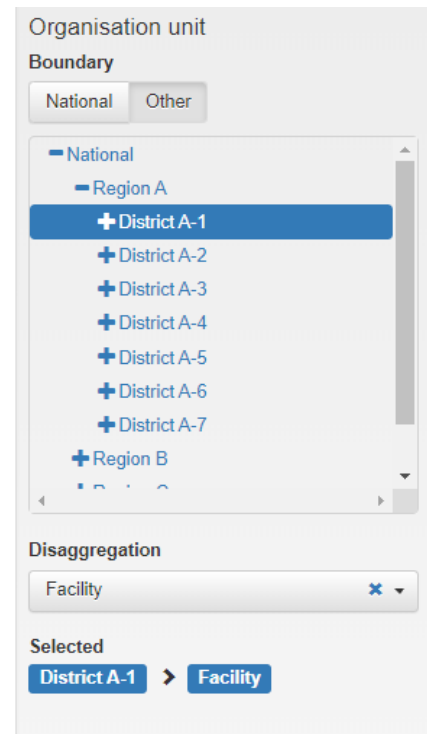
    When staff assigned to District A-1 first launch the DQ Tool, they do not need to reset "Organisation unit". It will automatically show the name of their district. By default, Disaggregaton will be set to "Facility".[8]

Use the Outliers page of the dashboard of the DQ Tool

6. **Use the "Outliers" dashboard to identify suspicious values:** Click on the tab for "Outliers". At first a bar with moving diagonal stripes may appear in the upper right of the screen, to the left of the settings icon. This indicates that data are being loaded into the Outlier detection tool. Wait for this bar to disappear.

___
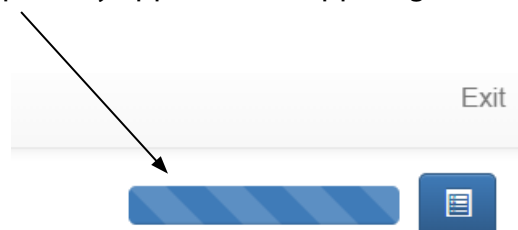
8   If there is a level in the organizational hierarchy between district (e.g. subdistrict) and facility then, by default, Disaggregation will be set to this intermediate level. However, the drop-down menu can be used to set Disaggregation to Facility or whichever level is two levels below District.

To free up space on the page, click on the settings icon to make the settings window disappear. The page shows a "Result" table with many rows. There are too many rows to all fit on one page. There are buttons at the bottom of the page to see other pages.

Each row is for one health facility. The name of the health facility is shown in the column headed "Unit". Each row contains one core indicator reported by that health facility. The name of the indicator is shown in the column headed "Data". Each row has the 12 values for that indicator reported by that health facility during a 12-month period.[9]

Each row has one or more values highlighted in red and/or one or more cells shaded yellow. Values highlighted in red are suspicious – they are significantly higher or significantly lower than the other values in the row. Data are missing for the cells shaded yellow.

The rows are sorted in order of "weight". The row with the highest weight appears at the top. The "Outlier weight" is the difference between the suspicious value and the average of other values in the row. The "missing weight" is equal to the average value reported by that facility multiplied by the number of months with missing (yellow) values.

---

7. **Use the Options window to filter the Result table and see the most important outliers and missing values:** A useful way to find the most suspicious values is to use their "Z-score".[10] Click on the Options button to open the Options window. To see the Z-score for each row, click on the box "Include Z-score". Review the "Modified Z-Score" values in the Result table. They range from zero (not suspicious) to more than 5 (extremely suspicious). Next, click on the following buttons: 1) Outliers; 2) Modified Z-Score; and 3) Extreme. Note how the Result table reduces to only a few rows. Note that now all the rows have a modified Z-score greater than 5. These are the most important values to investigate. Start by investigating the rows at the top of this filtered table.

8. **Inspect each row:** Click on the menu icon at the end of the first row and select "Visualize" from the drop-down menu. Is the red value suspicious? How could the value highlighted in red be investigated? After reviewing the bar chart, click on Close.

    Note the following outliers in the table:
    - Facility 325, Penta 3 < 1, Aug 2019 = 335;
    - A-1 NGO Hospital District, ANC 4 visits, Oct 2019 = 130;
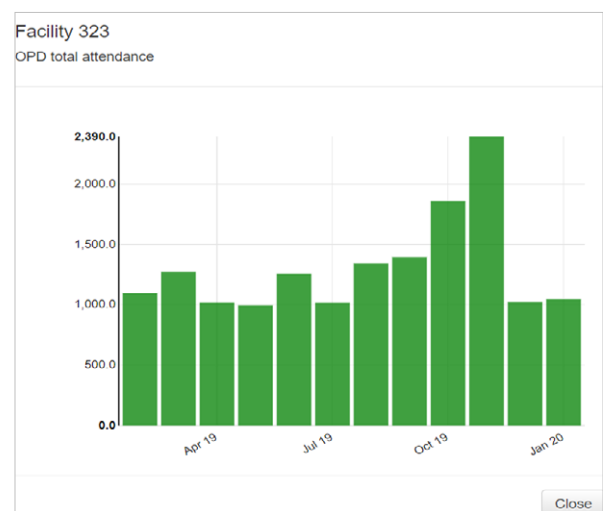
    Do DQ-04 and DQ-06c on the data quality dashboard also show these outliers?

---

10 The Z-score measures how different an outlier is, compared with the "standard deviation" (variation) seen in the other monthly values. The higher the Z-score, the more suspicious the value. Outliers can be filtered based on either the "Standard Z-Score" or the "Modified Z-Score". For the purposes of this guide, the "Modified Z-Score" is preferable because it is more sensitive and will detect a larger number of suspicious values than the "Standard Z-Score".

9. **Use the "Outliers" dashboard to identify important missing values:** To review the missing values, return to the Options window and click on "Missing Data" and "Off" (to turn off the Outlier filter).

Note the many rows in the Result table; there are so many rows that they do not all fit on a single page. **Focus on the rows with the highest Weight Missing.** Note that the rows on the first page, and especially the rows near the top of the first page, have the highest Weight Missing. This means that if values could be obtained for the yellow cells of the top rows, the aggregate total at district level would increase the most.

Download the data to a spreadsheet and use the spreadsheet to view the rows

Click on "Download" and a CSV file (like an Excel spreadsheet) will be downloaded with all the rows: rows for outliers as well as rows for missing values. **This file can be printed out and used during supervision visits or in meetings where access to the online DHIS2 is not easily available.**

When the Download window appears, leave the default settings as they are and simply click on "Download".

The file will appear in the bottom left corner of the screen. Click on the file name once to open the file.

Region A	District A-1	Facility 325	HI
Region A	District A-1	A-1 NGO Hospital	AN
Region A	District A-1	Facility 340	Un
Region A	District A-1	Facility 326	HI

Outlier and missin....csv

Visually scan the table. To identify health facilities with missing data, sort the table on "Gap weight", from largest to smallest.[11] To identify the most suspicious outliers, sort the table on "Max modified Z-Score" from largest to smallest. Outliers and missing values will not be highlighted on this CSV file, so the numbers must be visually scanned to determine which monthly values are most suspicious. Highlight in yellow the values which appear most suspicious.

---

11 To sort the rows of an Excel file, click on "Data" then "Sort". Make sure there is a check in the box for "My data has headers". Click on the down symbol (v) to the right of "Sort by", then select "Max modified Z-Score". Change "Order" to "Largest to smallest".

Exercises

Tutorial 1 exercise: Use the dashboard and the Analysis function of the WHO Data Quality Tool

Introduction

The exercise for Tutorial 1 provides practice in how to use the District Health Information Software (DHIS)-based WHO Data Quality (DQ) Tool at national level. The screenshots show what national-level users will see if they log into https://who-demos.dhis2.org/dq. (The exercise for Tutorial 3 provides practice in how to use of DQ Tool at district level.)

The exercise begins with a focus on the dashboard of the DQ Tool. The dashboard features 4 pages: Completeness; Consistency - time (month-to-month consistency); Consistency - data (consistency between related indicators); and Outliers (a detailed examination of month-to-month consistency as well as missing data).

Next, the exercise provides practice in the use of the Analysis function of the DQ Tool. The Analysis function permits review of the quality of any DHIS2 data element or indicator. Thus, it is more flexible than the dashboard. However, to use the Analysis function, the user must first select correctly from among the very large number of data elements/indicators. Thus, the Analysis function is best suited to staff of the health management information system (HMIS) unit or public health programmes, who are familiar with the details of DHIS2 data.

Launch the Tool

1. **Log into the DHIS2 instance:**
    URL: https://who-demos.dhis2.org/dq
    Username: demo   Password: District1#

2. Type "WHO" on the Search apps line in the upper right of the DHIS2 Home page. Click on **WHO Data Quality Tool.**

3. Quickly review the main menu of the DQ Tool. After launching the DQ Tool, five tabs appear at the top of the page. These are used to select between several different functions. Note that the word "Dashboard" is highlighted and the dashboard function is selected by default.

4. **Quickly review the dashboard menu:** The dashboard has its own set of four tabs. The Completeness tab is selected by default.
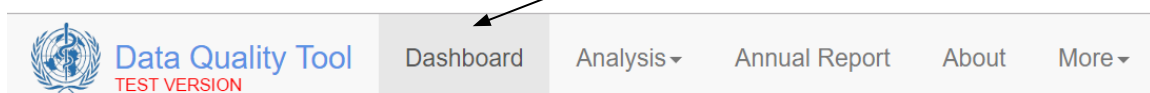
Dashboard

| Completeness | Consistency - time | Consistency - data | Outliers |
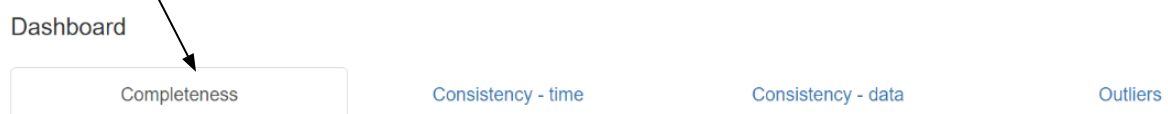
Review the Consistency - time dashboard

Click on the tab for the Consistency - time" dashboard.

1. Review the settings window. If necessary, click on the settings icon to open the window.

2. Click on the rectangle beneath "Data" and view the drop-down list. These are the various indicator groups that can be reviewed using the dashboard of the WHO DQ Tool. Leave Data set to [Core].

3. Set Period to January of the current year. Note that when the settings are changed, a bar appears with moving stripes. This shows that data are being loaded into the DQ Tool. Wait until this bar disappears to see the results. If the internet connectivity is poor or if the DHIS2 instance contains a very large amount of data, this bar may "spin" for a long time.

4. Click on the rectangle beneath "Disaggregation". Select "District".

    Review the chart at the top left. Move the cursor over the lines and read the information in the box which pops up.
    Question 1: The chart shows total values for which level of the health system? (National? Regional? District? Facility?)
    Question 2: The dark blue line shows the trend over which months? The light blue line shows the trend over which months? The orange line shows the trend over which months?
    Question 3: Has ANC 1 had a consistent value from year to year and from month to month? If not, describe any inconsistency seen. For which month(s) were the values not consistent?

Data

[Core]

Period

‹	2020	›
January	February	March
April	May	June
July	August	September
October	November	December

Data for 12 months up to and including the selected month will be used.

Organisation unit

Boundary

| National | Other |

Disaggregation

District

Selected

National > District

45

ANC 1 visits
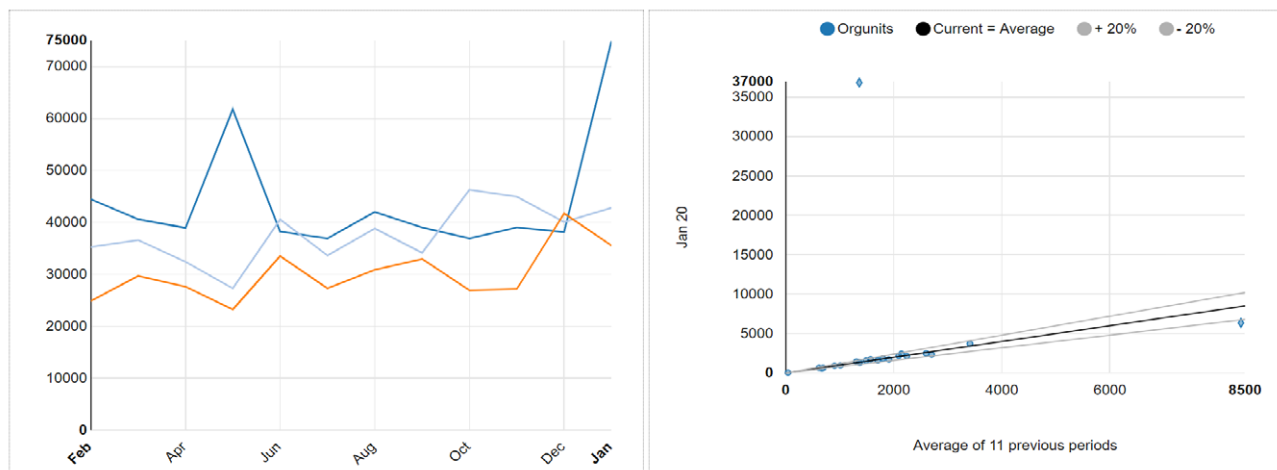
5. Review the chart on the right for ANC 1. This chart is a scatterplot with each dot representing one district. The position of the dot on the vertical axis shows the value of ANC 1 for one district for January of this year. The position of the dot on the horizontal axis shows the average value of ANC 1 for the same district during the 11 months preceding January (i.e. February of last year[1] to December of last year). The dots that lie above the upper grey line or below the lower grey line are suspicious. Note that they are shaped like a diamond rather than a circle. Move the cursor over each of the dots shaped like a diamond and read the information in the box that pops up.

Question 4: There are two dots that are shaped like a diamond: one is below the lower grey line and one is above the upper grey line. These dots represent which two districts?
Question 5: Use the settings window to change Period to May of last year (the other month with a suspicious rise shown in the chart on the left). With Period set to May of last year, the chart on the right has one diamond-shaped dot above the upper grey line. This dot represents which district? Compare the value reported by that district for May of last year with the average value reported for that district during the preceding 11 months (June of the year before last to April of last year).
Question 6: Re-set Period to January of this year. Note again that there is a diamond-shaped dot below the lower grey line of the chart on the right. This dot represents which district? Explain why the dot for this district is below the lower grey line.

Before continuing, ensure that the dashboard is set to Period = January of this year and that Disaggregation is set to District.

Review Consistency - data dashboard

Click on the tab for the Consistency - data dashboard. Each chart on this dashboard assesses the relationship between two related data elements/indicators.

Review the chart assessing Penta dropout rates. Move the cursor over the vertical bars and read the information in the windows that pop up.

Penta 1 to Penta 3 dropout - Feb 19 to Jan 20

District A-3
-4.54% dropout
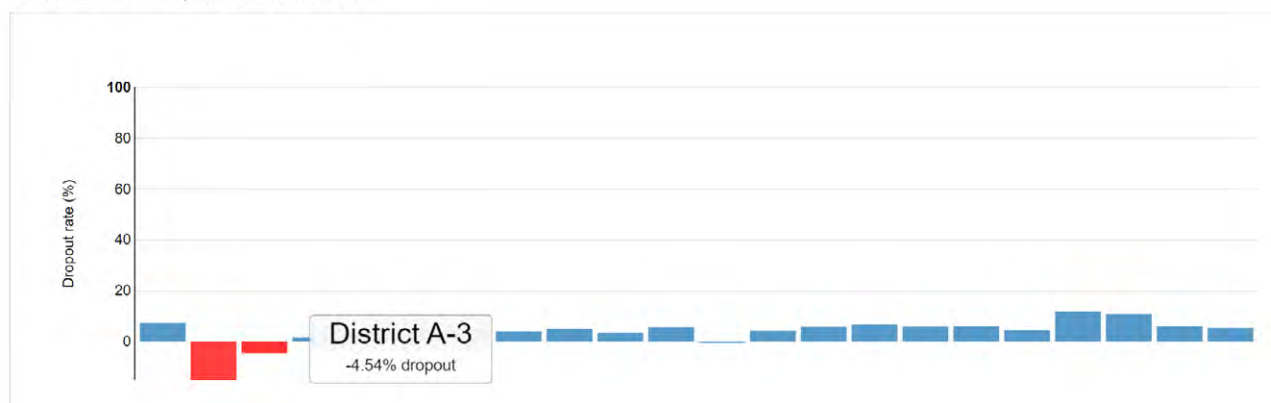
Question 7: How is "dropout rate" defined and how is it calculated?

Question 8: Which districts have a negative Penta 1 to Penta 3 dropout rate? These two districts reported a higher number of Penta 3 doses that Penta 1 doses for the 12-month period. To better understand why Penta 3 is higher than Penta 1 for this district, the Analysis - Consistency function will now be used to investigate further.

Use Analysis - Consistency to assess DPT dropout rates

Click on the Analysis tab, then select "Consistency" from the drop-down menu. Under Analysis Type, click on each of the following: "Between Indicators", "Expected result" and "Dropout".

Click on Data to open the Data settings window. Set Data as shown in the figure to the right:

> Under "Select data element/indicators", click on "Data element";
> Click in the box for "Data element group", then find and click on "Immunisation";
> Click in the box for "Select data element …", then find and click on "Penta 1 given < 1";
> Under "Select data element/indicator for comparison", click on "Data element";
> Click in the box for "Data element group", then find and click on "Immunisation";
> Click in the box for "Select data element …", then find and click on "Penta 3 given < 1".

Click on Period to open the Period settings window. Set Period to Year = last year.
Click on Orgunit to open the Organisation unit settings window. Leave Boundary set to "National". Re-set Disaggregation to District.

Click the Analyze button.

1. Review the chart and the table (note: the top row of the table will not be shaded yellow until the next step is completed). Move the cursor over the vertical red bars in the chart to view the details. Compare the names of the districts with vertical red bars to the names of the districts in the table with Dropout rate shaded in red.

Penta 1 given < 1 to Penta 3 given < 1 dropout. 2019.

Unit	Penta 1 given < 1	Penta 3 given < 1	Dropout rate (%) ❶	Weight ❶ ⬇
National	479411	451771	5.8%	
District A-2	16532	18986	-14.8%	2454
District A-3	10714	11219	-4.7%	505
District D-4	30699	29000	5.5%	0
District D-3	23025	21676	5.9%	0
District D-2	24519	21930	10.6%	0
District D-1	89320	78435	12.2%	0
District C-6	14302	13648	4.6%	0

Question 9: Which district has the most negative dropout rate? Find it in the chart and find it in the table. For that district, for last year, what was the value of Penta 1 and what was the value of Penta 3 (see table)? Use these numbers to calculate the dropout rate.

2. Click on the row of the table that shows data from the district with the most negative dropout rate. A new table and chart should appear at the bottom of the page.

Selected

Unit	Penta 1 given < 1	Penta 3 given < 1	Dropout rate (%)	
District A-2	16532	18986	-14.8%	▤

● Grouped   ○ Stacked   ● Penta 1 given < 1   ● Penta 3 given < 1
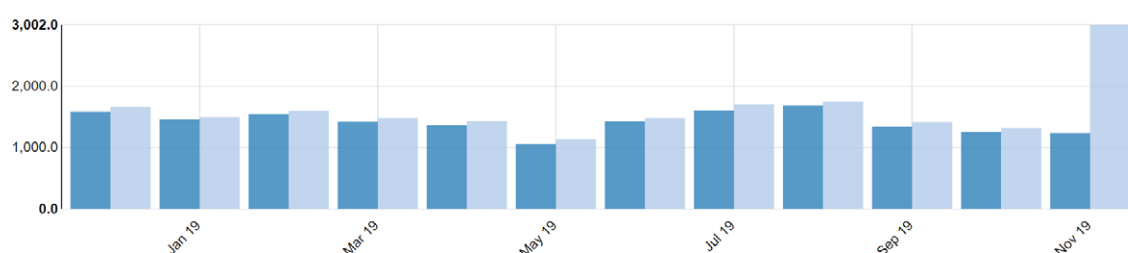
Question 10: Is the negative dropout rate for District A-2 due to an extremely large Penta 3 value for a single month or is it due to higher values for multiple months?

3. Click on the menu icon at the right end of the small table and select "Drill down". A new chart and table should appear showing the dropout rates for last year for each health facility in District A-2.
   Question 11: Is the negative dropout rate for the district due to a value from a single facility or is it due to values from multiple facilities? Does it appear as if the selected district has a systematic problem with over-reporting of Penta 3 (for multiple months and multiple subdistricts)?

Penta 1 given < 1 to Penta 3 given < 1 dropout. 2019.   Download   Previous

Unit	Penta 1 given < 1	Penta 3 given < 1	Dropout rate (%) ❶	Weight ❶ ↓
District A-2	16532	18986	-14.8%	
Facility 197	2320	4151	-78.9%	1831
Facility 463	747	836	-11.9%	89
Facility 469	953	1007	-5.7%	54
Facility 202	573	627	-9.4%	54
Facility 474	634	683	-7.7%	49
Facility 468	930	975	-4.8%	45
Facility 466	616	651	-5.7%	35

4. Click on the row of the table that shows data for the facility with the most negative dropout rate. A new table and chart should appear at the bottom of the page.
   Question 12: Describe what this table shows. Is the negative dropout rate for Facility 197 due to a single very large outlier value of Penta 3 or is it due to Facility 197 repeatedly reporting more Penta 3 doses than Penta 1 doses?

Use the Outliers dashboard to identify important outliers

Click on the tab for the Outliers dashboard. Review the settings window to confirm that Data is set to District, Period is set to January of the current year and Disaggregation is set to District. Click on the settings icon to hide the settings window and free up space on the screen. Click on the Options button then check the box for "Include Z-score" and click on "Outliers", "Modified Z-Score" and "Extreme".

Region	Unit	Data	Feb 19	Mar 19	Apr 19	May 19	Jun 19	Jul 19	Aug 19	Sep 19	Oct 19	Nov 19	Dec 19	Jan 20	Z-score Standard	Modified	Weight Outlier ↓F
Region B	District B-2	ANC 1 visits	1372.0	1311.0	1374.0	1289.0	1413.0	1330.0	1513.0	1335.0	1322.0	1434.0	1304.0	38835.0	3.18	520.27	35472
Region D	District D-1	ANC 1 visits	7519.0	5927.0	6613.0	31494.0	4866.0	6016.0	6803.0	5280.0	5794.0	6036.0	6349.0	6359.0	3.16	41.60	25353
Region C	District C-2	HIV tests done	9275.0	5877.0	5199.0	5560.0	3836.0	4337.0	26652.0	4743.0	4181.0	7089.0	5632.0	9120.0	3.04	10.24	20657
Region D	District D-3	ANC 1 visits	3809.0	1794.0	1652.0	1592.0	1681.0	1489.0	2011.0	1818.0	1722.0	1791.0	1672.0	1723.0	3.11	19.82	2087
Region A	District A-2	Penta 3 given < 1	1599.0	1483.0	1434.0	1130.0	1483.0	1701.0	1751.0	1417.0	1310.0	3002.0	1166.0	1240.0	2.92	5.95	1571
Region B	District B-5	ANC 4 visits	250.0	267.0	239.0	255.0	258.0	224.0	1384.0	377.0	343.0	250.0	264.0	250.0	3.15	84.5	1352
Region C	District C-6	HIV tests done	2291.0	2612.0	2241.0	1883.0	2165.0	2256.0	3050.0	2396.0	2334.0	2429.0	2557.0	3607.0	2.54	5.64	1304
Region B	District B-5	Penta 3 given < 1	1543.0	1497.0	1154.0	1395.0	1162.0	1447.0	1580.0	1528.0	1812.0	1430.0	2573.0	1344.0	2.9	8.03	1146
Region C	District C-4	ANC 1 visits	2030.0	2114.0	2084.0	1947.0	2214.0	2361.0	2226.0	2170.0	1578.0	2161.0	2123.0	2101.0	2.69	5.85	568
Region A	District A-6	ANC 1 visits	812.0	721.0	734.0	662.0	725.0	703.0	750.0	751.0	705.0	674.0	359.0	648.0	2.66	5.65	318
Region A	District A-4	HIV tests done	40.0	72.0	170.0	71.0	77.0	52.0	70.0	88.0	60.0	88.0	100.0	93.0	2.76	5.03	105
Region A	District A-6	Under 5 OPD malaria as a percentage of total OPD malaria (%)	40.9	35.8	31.8	32.2	30.4	30.8	64.4	95.9	23.7	28.3	35.1	36.6	2.75	13.54	95
Region B	District B-3	ANC 4 visits	189.0	250.0	192.0	216.0	181.0	152.0	189.0	213.0	193.0	193.0	187.0	164.0	2.26	6.17	19
Region A	District A-4	ANC 1 visits	43.0	67.0	41.0	60.0	42.0	45.0	46.0	40.0	55.0	9.0	40.0	47.0	2.52	6.74	16

Review the Result table.

Question 13: Each row shows data for which level of the health system? (National? Regional? etc.)
Question 14: Compare the values highlighted in red with the other values in the same row. For the first row, take the value highlighted in red and subtract from this an approximate average of the other values in the same row. How does the result compare to the "weight" for the row? What does the weight represent?
Question 15: Visualize the data. Click on the menu icon at the end of the first row of the table and select "Visualize". Describe the chart. After viewing the chart, click "Close".
Question 16: Drill down to investigate the outlier. Click again on the menu icon at the end of the first row of the table and select "Drill down". A new "Result" table should appear with each row showing 12 months of data for an individual health facility in the selected district. Review the first row of this new table. The extreme outlier was reported by which facility? Is the outlier is due to an error? Try to drill down further. What happens and why?

Use Analysis - Outliers to identify extreme outliers of additional data elements

Click on the tab for "Analysis", then select "Outliers and Missing Data". Select any data element or indicator. Then set Period to last year and set Orgunit to National with Disaggregation by District. Click on Analyze, then proceed to use the Tool to identify extreme outliers using the same steps as with the Outliers dashboard.

Question 17: What are the advantages and disadvantages of the Analysis - Outliers function compared with the Outliers dashboard?
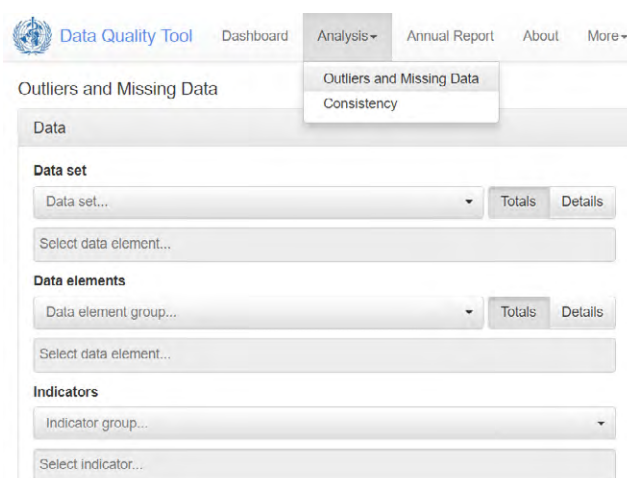
Discuss the steps to take when an error is suspected

Question 18: List the measures that data managers should take if they identify a value that is very suspicious and so large that it probably has a significant effect on a nationwide statistic.

Question 19: Who has authority to edit the DHIS2 data? Do the SOPs specify how to document changes that are made to the data?

Question 20: In your health system, how should the DQ Tool be used to identify extreme outliers – at district level as well as at national level – each month or quarter as well as each year?

Tutorial 2 exercise: Using a data quality dashboard

Introduction

This exercise provides practice in using a data quality dashboard. This is a dashboard that is visible as soon as the user logs into the DHIS2 instance. It is not necessary to launch any application to view the dashboard. The dashboard permits users to quickly assess the completeness and consistency of DHIS2 data.

The exercise provides practice in how to use a data quality dashboard at national level. The screenshots show what national-level users will see if they log into https://who-demos.dhis2.org/dq.

As https://who-demos.dhis2.org/dq is a live DHIS2 instance, the months shown in some charts and tables update each month to include data from the current "last month". For this reason, there will be differences between the items on the instance and some of the screenshots included in this document.

The dashboard is structured according to the four dimensions of the WHO Data Quality Assurance framework: completeness; internal consistency of data (do DHIS2 data values agree with each other); external consistency of data (do DHIS2 data agree with statistics from other sources such as a health facility survey); and consistency of DHIS2 denominator/target estimates.

1. **Log into the DHIS2 instance:** https://who-demos.dhis2.org/dq
   Username: demo
   Password: District1#
   In the dashboard ribbon at the top of the screen, click on the button for **Data quality.**
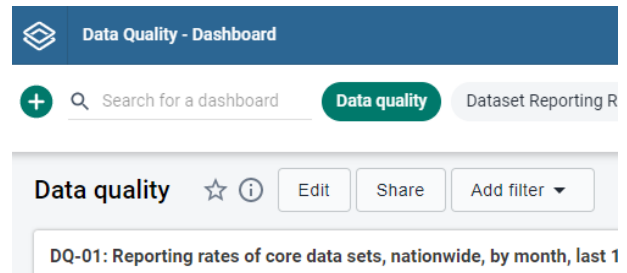
2. **Quickly review the data quality dashboard:**
   Question 21: Which items on the dashboard are used to assess each of the following aspects of data quality: a) completeness of reporting for key datasets; b) month-to-month consistency; c) consistency between related indicators; d) consistency between survey and routine coverage estimates; e) consistency of targets/denominators?

3. **Practise using the cursor** to interact with and simplify DQ-01. Place the cursor on the orange line so that the other lines become faint. Next, click on the indicator names at the bottom of the chart so that only the orange line is visible.
   Question 22: In December of last year, what percentage of health facilities submitted their report on the Malaria dataset?
   Question 23: The points on the right end of each line show the reporting rates for last month. If DQ-01 is viewed at the beginning of a month, all of these points for last month may be at zero or near to zero. Why?
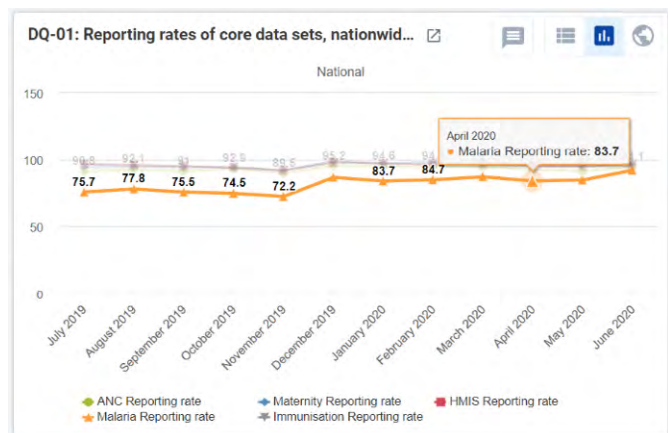
4. **Review DQ-02:**
   Question 24: Which types of health facility have had the lowest reporting rates for the Malaria dataset?

5. **Review DQ-03:**
   Question 25: The table shows the reporting rates for which period? Click once on the header for Malaria Reporting rate, then review the reporting rates that appear at the top. Click a second time on the header for Malaria Reporting rate.
   Question 26: Which districts had the lowest Malaria Reporting rate for the last 3 months?

6. **Review DQ-04:** Place the cursor on one line, then another, to view the trend for each indicator.
   Question 27: For which indicators is the trend almost like a straight, horizontal line?
   Question 28: For which indicators and which months do the trend lines show a suspicious rise?
   Question 29: Explain why the trend line for most indicators of maternal health and immunization services is almost like a straight, horizontal line.

7. **Review DQ-05:**
   Question 30: Is the number of malaria cases roughly the same each month? Compared with other indicators, is it easier or is it more difficult to detect errors in malaria data?

8. **Review DQ-05b:**

DQ-05b: Under five OPD malaria as a percentage of total ...
Under 5 OPD malaria as a percentage of total OPD malaria (%)
May 2020 • District D-4: **93.4**
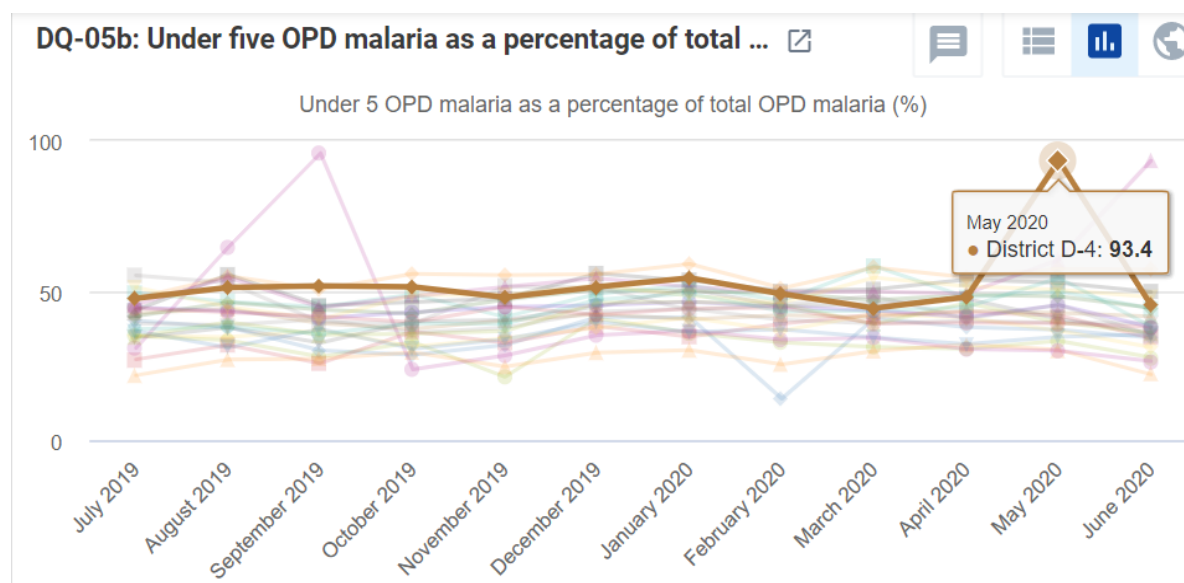
Question 31: What is the indicator presented in this chart? How is it calculated? What is the numerator and what is the denominator?
Question 32: For a sample of lines, place the cursor on one, line then another, to focus on the trend. For most lines, does the value of the indicator change a lot from month to month or is the trend almost a straight, horizontal line?

Question 33: For which districts are the lines not straight? For which months?

Question 34: Place the cursor on either September of last year for the purple line (District A-6) or May of this year for the brown line (District D-4). For the district and the month selected, malaria in children less than 5 years of age was what percentage of all reported malaria cases? Do you think this value suggests that there is an error in the data on malaria cases?

9. **Review DQ-06:** This table shows very large outliers for five different indicators. These are the largest single suspicious values reported by any facility for any month during the last 12 months.

    Question 35: Explain how the extreme outlier values shown in DQ-06 help to explain the suspicious values revealed in DQ-04.

    Question 36: DQ-06 shows that some extreme outlier values were reported for Penta 1 and Penta 3. Why are these extreme outliers not evident from DQ-04?

10. **DQ-06b** can be used to find out which facility reported each of the very large outliers shown in DQ-06. The first column (on the left) lists individual facilities. There is a very large extreme outlier for each of the columns immediately following this first column (ANC 1 outliers > 1 000; etc.) To identify the facility that reported the very large outlier, click on the cell specifying the month, beneath the indicator. The rows of the table will then be sorted so that the very large outlier appears at the top.

    Question 37: Which health facility reported each of the following very large outliers?
    a) ANC 1 = 35 888 in January 2020; b) Penta 1 = 1 155 in February 2020; c) Penta 3 = 1 117 in December 2019; d) OPD presumed malaria = 653 536 in September 2019.

DQ-06b: Very large outliers, by health facility and by month, last 12+ months ⧉

	ANC 1 outliers > 1,000					Penta 1 outliers > 1,000	
	January 2019	February 2019	May 2019	January 2020	September 2020	February 2020	August 2020
National / Region A / District A-3 / Facility 251							
National / Region A / District A-3 / Facility 258							
National / Region A / District A-5 / Facility 177							
National / Region A / District A-6 / Facility 73							
National / Region A / District A-6 / Facility 79							
National / Region A / District A-7 / Facility 185							
National / Region B / District B-1 / Facility 238							
National / Region B / District B-1 / Facility 240							1 323
National / Region B / District B-2 / Facility 147				35 888			

(DQ-06c will be used during the next session when learning how to make use of a data quality dashboard at district level. At national level, DQ-06c may take very long to appear on the screen. Do not wait for it to appear at national level.)

11. **Review DQ-07:** This is an example of how to compare the values of related indicators.
Question 38: The chart shows the "Penta 1 to Penta 3 dropout rate". How is this indicator defined? How is it calculated? What is the numerator? What is the denominator?
Question 39: Which districts have a negative Penta 1 to Penta 3 dropout rate over the last 12 months?
Question 40: If a district has a negative Penta 1 to Penta 3 dropout rate for a full 12 months, does this suggest that the district has a problem with data quality?

12. **Review DQ-08:**
Question 41: For most regions, is the survey estimate of Penta 3 coverage consistent with the Penta 3 coverage calculated from routine health facility data? For which region(s) is the survey estimate of coverage not consistent with the routine estimate of coverage?
Question 42: List some possible explanations for any inconsistency shown by DQ-08.
Question 43: What are some possible reasons why a survey estimate may not be reliable?

13. **Question 44:** Review the chart below. The chart shows several types of inconsistencies in the denominators used to calculate coverage with maternal and infant health services. Describe at least three inconsistencies. (Hint: there are inconsistencies in the year-to-year trend as well as inconsistencies between related denominators.)

National

	2015	2016	2017	2018	2019
Estimated pregnancies	339 604	354 214	346 809	378 513	386 082
Estimated live births				340 632	347 441
Estimated population < 1 year	313 722	306 983	307 826		
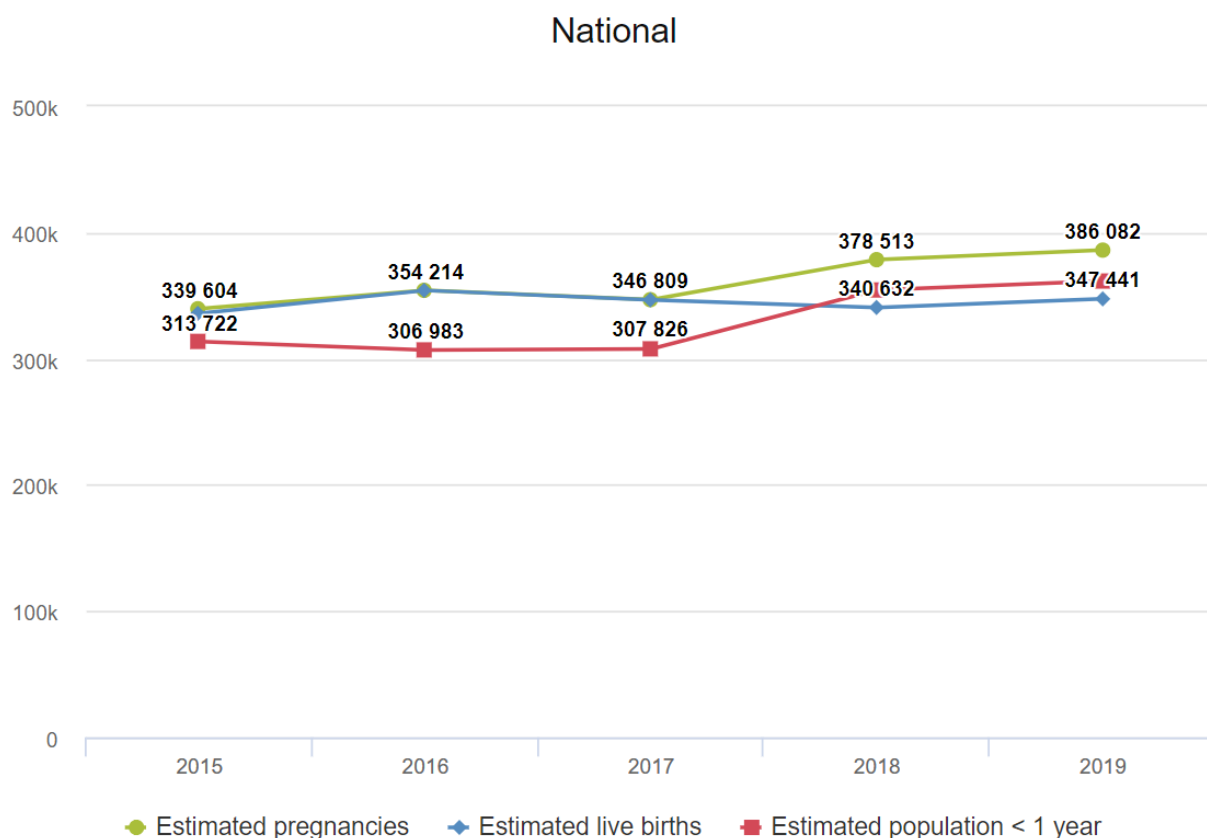
Tutorial 3 exercise: Monthly use of data quality tools at district level

Introduction

The best way to assure the quality of DHIS2 data is for data managers at district and facility level to check the data each month, soon after they have entered the data for most health facilities. This tutorial walks the user through the steps to use these DHIS2 data quality tools to rapidly identify "outliers" (suspicious data values) and important missing data values.

Review the data quality dashboard

Log into DHIS2 and review the data quality dashboard: URL: who-dev.dhis2.org/dq
Username: demo   Password: District1#

A data quality dashboard is visible as soon as the user logs into DHIS2. It is not necessary to launch any application to view the dashboard.
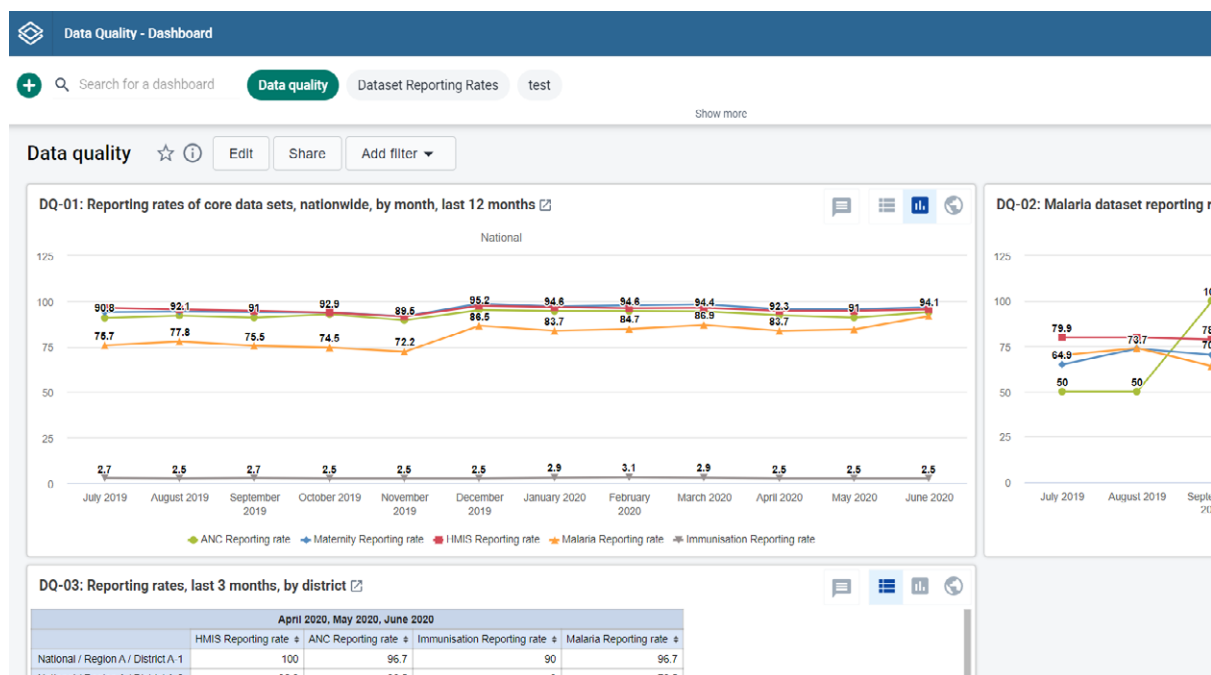
When a user logs on at national level, they see charts and tables for assessing the data of the entire country.
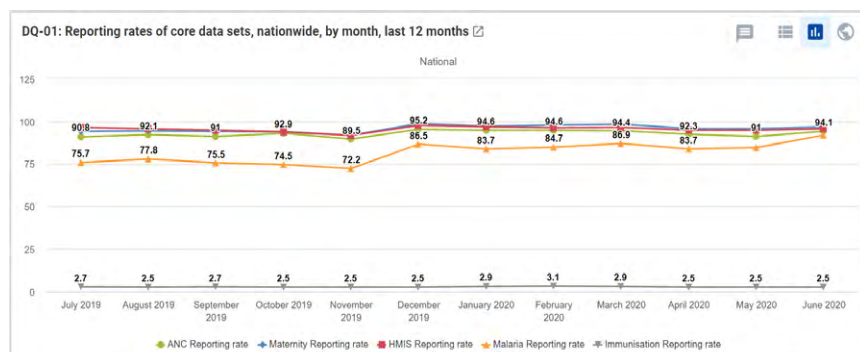
When a district data manager logs on, however, they usually see data only for their specific district.
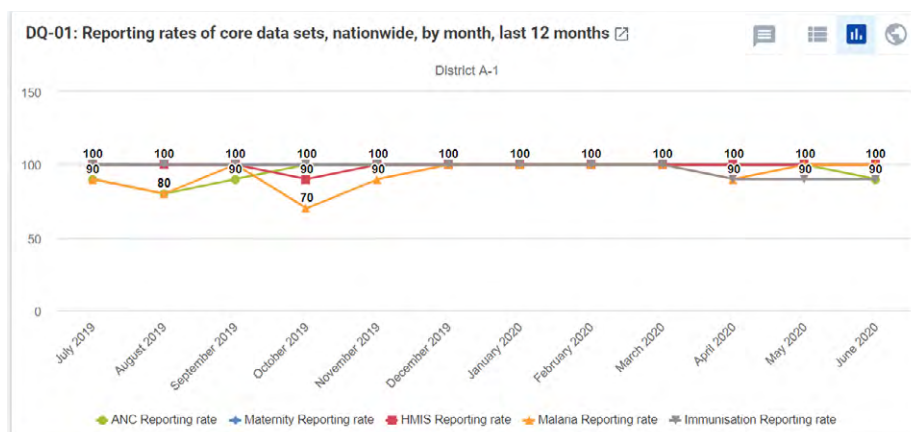
Assess the completeness of reporting

The first four items on the data quality dashboard can be used to assess the "reporting rates" (also known as "completeness") for key datasets. **DQ-01** shows the trend, over the last 12 months, in the district's reporting rates for multiple datasets.

Question 45: When DQ-01 is viewed at the beginning of a month, the reporting rates for last month (the month on the right) will be very low or zero. Explain why.

A chart with multiple lines can sometimes be difficult to read. To simplify the chart, click on the names of the indicators in the legend at the bottom of the chart. This will hide the lines for those indicators and the names will become light grey.

Question 46: Which dataset appears to have had the lowest reporting rate over the last 12 months?
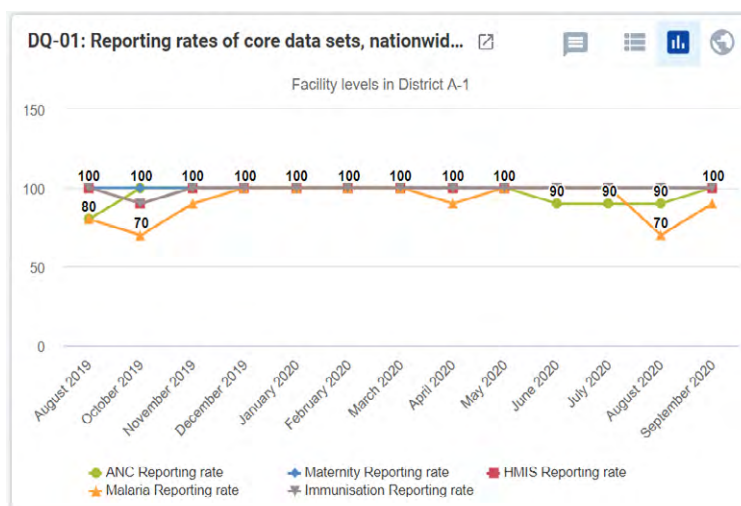
DQ-02 shows that, using the DHIS2 data visualizer app, it is possible to make a chart that *disaggregates* a reporting rate *by type of facility*.

Question 47: Which type of facility had the lowest reporting rates for the Malaria dataset?

DQ-03 presents reporting rates for multiple datasets over the last
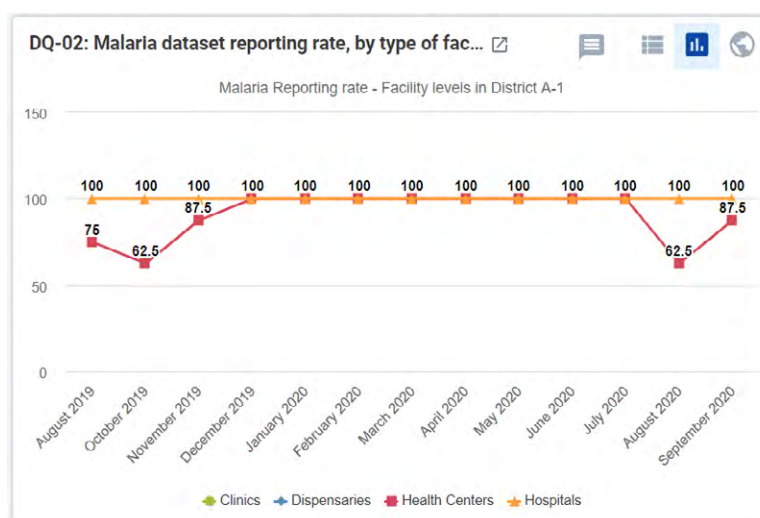
3 months for each of the health facilities in the district. If necessary, click once on the column header for a dataset to sort the rows from high to low.

Question 48: When this table is viewed at the beginning of a month, the numbers will be lower than when the table is viewed at the end of a month. Why?

Question 49: How many facilities have so far submitted all three Malaria reports from the last 3 months? How many have submitted two of the three Malaria reports? How many have submitted only one of the three Malaria reports? How many have not submitted any of their monthly reports?

Findings related to the internal consistency of the data

The next three items on the dashboard can be used to assess the consistency over time of the data. The **DQ-04** presents the district totals, over the last 12 to 24 months, for several related maternal health and immunization indicators.

DQ-03: Reporting rates, last 3 months, by district ⧉

July 2020, August 2020, September 2020				
	HMIS Reporting rate	ANC Reporting rate	Immunisation Reporting rate	Malaria Reporting rate
National / Region A / District A-1 / A-1 District Hospital	100	100	100	100
National / Region A / District A-1 / A-1 NGO Hospital	100	100	100	100
National / Region A / District A-1 / Facility 313	100	33.3	100	100
National / Region A / District A-1 / Facility 323	100	100	100	100
National / Region A / District A-1 / Facility 325	100	100	100	100
National / Region A / District A-1 / Facility 326	100	100	100	33.3
National / Region A / District A-1 / Facility 327	100	100	100	100
National / Region A / District A-1 / Facility 328	100	100	100	66.7
National / Region A / District A-1 / Facility 340	100	100	100	100
National / Region A / District A-1 / Facility 610	100	100	100	66.7

Question 50: Are there any suspicious values? If yes, for which indicator(s) and which month(s) are outliers seen?

Question 51: Which tool is more sensitive for identifying suspicious values: DQ-04 or the outlier function of the WHO Data Quality Tool?

DQ-06c summarizes some of the extreme outliers reported in the last 12 months for facilities in the district. Extreme outliers which are large (those highlighted in red) are especially important to investigate because they can significantly distort even the district annual total value and result in over-estimation of coverage and other indicators.

Question 52: Which of the following large outliers in table DQ-06c are apparent in DQ-04: a) ANC 1 = 341 in May of the current year; b) ANC 4 = 130 in October of last year; c) Penta 1 = 325 and Penta 1 = 177 in March of the current year; d) Penta 3 = 330 in June of the current year.

Question 53: Discuss how you would go about investigating these values.

When assessing data quality at district level and below, the last two charts on the data quality dashboard are not used.

Review the WHO Data Quality Tool

Log into the DHIS2 instance: URL: https://who-demos.dhis2.org/dq
Username: demo   Password: District1#

Type "WHO" in the search apps line in the upper right of the DHIS2 Home page. Click on **WHO Data Quality Tool.**

Quickly review the menus of the DQ Tool

After the DQ Tool is launched, five tabs will appear at the top of the page. These are used to select between several different functions. Notice that the word "Dashboard" is highlighted and the Dashboard function is selected by default.
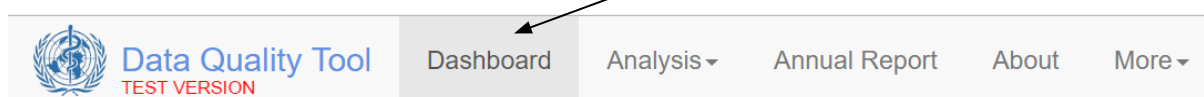
The Dashboard has its own set of four tabs. The Completeness tab is selected by default.

Do not spend long reviewing the Completeness dashboard. (The completeness tab of the DQ Tool shows the reporting rate for select datasets. However, the same information is presented more comprehensively with the charts and tables on the data quality dashboard.)

Explore the settings window of the DQ Tool dashboard

Click on the settings icon ( ) in the upper right of the screen to show a settings window on the right of the screen. If the settings window appears briefly but then disappears, click again until the window stays on the screen. Under "Data", there is a white space where [Core] appears. Click on the down arrow to the far right of this space and a drop-down menu will appear, showing the various groups of indicators that can be reviewed on the DQ Tool dashboard: [Core], General Service Statistics, Maternal Health, HIV/Aids, Malaria, District and Immunisation.
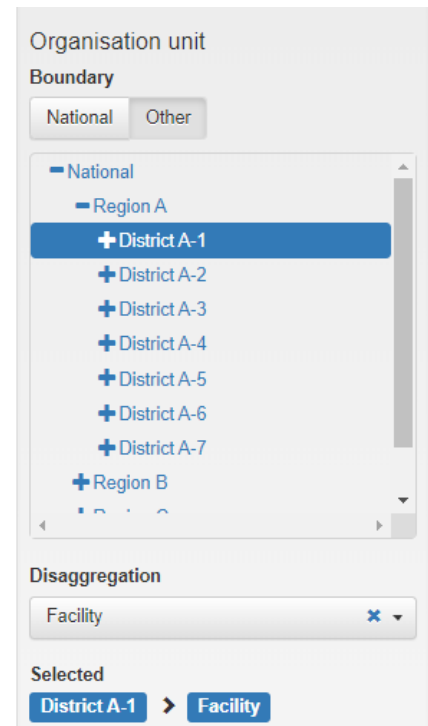
Change "[Core]" to "District". The reporting period that is shown on the dashboard can be changed by clicking on a different month. After experimenting with changing the **Period**, reset it **January of the current year.**
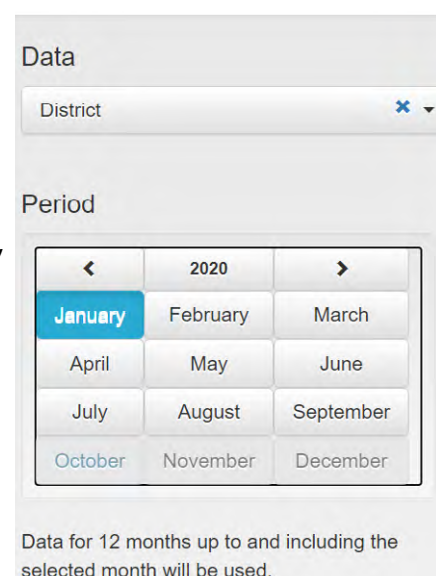
Change the "Organisation unit" that is being analysed

Under Boundary, click on "Other", then click on the "+" sign to the left of "National". The list of regions will then appear. Click on the "+" sign to the left of "Region A" and the list of districts in Region A will appear.

Click on and select District A-1. When staff of District A-1 first launch the DQ Tool, they do not need to reset Organisation unit. It will automatically show the name of their district. By default, Disaggregaton will be set to "Facility".

Use the Outliers page of the dashboard of the DQ Tool

Use the Outliers dashboard to identify suspicious values:
Click on the tab for "Outliers". At first there may be a bar with moving diagonal stripes that appears in the upper right of the screen, to the left of the settings icon. This indicates that data are being loaded into the Outlier detection tool. Wait for this bar to disappear. To free up space on the page, click on the settings icon to make the settings window disappear.

Each row is for one health facility. The name of the facility is shown in the column headed "Unit". Each row contains one core indicator reported by that facility. The name of the indicator is shown in the column headed "Data". Each row has the 12 values for that indicator reported by that facility during a 12-month period.

Each row has one or more values highlighted in red and/or one or more cells shaded yellow.
Question 54: How are the values highlighted in red different from the other values in the same row?
Question 55: How are the cells highlighted in yellow different from the other cells?
Question 56: Are the rows of the table sorted in any order? Review the table, then describe how you think the rows are sorted.
Question 57: How do you think "Outlier weight" is calculated?
Question 58: How do you think the "Missing weight" is calculated? How is "Total weight" calculated?

Use the Options window to filter the Result table and see the most important outliers and missing values. A good way to find the most suspicious values is to use their "Z-score". To use this method, click on the Options button to open the Options window. To see the Z-score for each row, click on the box "Include Z-score". Review the modified Z-score values in the Result table. They range from zero (not suspicious) to more than 5 (extremely suspicious). Next, click on the following buttons: 1) Outliers; 2) Modified Z-Score; and 3) Extreme. Note how the Result table reduces to only a few rows.

Result   Options   Download   Previous

Display columns

| All | Missing data | Outliers |

☑ Include Z-score

Outlier filter

| Off | Standard Score | Modified Z-Score |

| Moderate | Extreme |

Regions	Districts	Unit	Data	Jul 19	Aug 19	Sep 19	Oct 19	Nov 19	Dec 19	Jan 20	Feb 20	Mar 20	Apr 20	May 20	Jun 20	Standard	Modified	Outlier
Arusha Region	Arusha City Council	Lovolosi Health Center	*Measles/Rubella 2nd doses given	25.0	20.0	19.0	2410.0	33.0		21.0	25.0	48.0				2.47	292.49	2383
Arusha Region	Arusha City Council	Daraja II Health Center	ANC Fourth Visit Attendance	90.0	69.0	570.0	128.0	97.0	86.0	88.0	101.0	139.0	165.0	135.0	131.0	3.11	12.54	458
Arusha Region	Arusha City Council	Kaloleni Health Center	ANC 1st visit total	416.0	226.0	260.0	240.0	246.0	246.0	248.0	251.0	291.0	228.0	218.0	268.0	2.94	7.12	169
Arusha Region	Arusha City Council	Muriet Health Center	*Measles/Rubella 2nd doses given	28.0	29.0	10.0	38.0		36.0	5.0	32.0				173.0	2.41	14.79	148
Arusha Region	Arusha City Council	Lovolosi Health Center	*Measles/Rubella 1st doses given	44.0	40.0	41.0	200.0	45.0		79.0	55.0	70.0				2.39	10.65	147
Arusha Region	Arusha City Council	Muriet Health Center	*Measles/Rubella 1st doses given	28.0	41.0	27.0	38.0	81.0	53.0	12.0	34.0			184.0		2.48	8.95	145
Arusha Region	Arusha City Council	Lovolosi Health Center	Children Voucher For Mosquito Nets	31.0	60.0		200.0	45.0		79.0	55.0	70.0				2.18	6.3	143
Arusha Region	Arusha City Council	LCC Huruma Dispensary	*Penta 3 doses given before 12 months	88.0	40.0	88.0	74.0	80.0		100.0	90.0	24.0	35.0	0.0	90.0	1.92	5.4	128
Arusha Region	Arusha City Council	Cogi Daraja 11 Dispensary	*Measles/Rubella 1st doses given	17.0	11.0	23.0	142.0	17.0		14.0	17.0	22.0		16.0	11.0	2.83	21.08	126
Arusha Region	Arusha City Council	Kaloleni Health Center	*Measles/Rubella 1st doses given	84.0	86.0	76.0	73.0	53.0			84.0	66.0	58.0	34.0	105.0	2.6	7.1	117
Arusha Region	Arusha City Council	Muriet Health Center	ANC 1st visit total	86.0	82.0	102.0	104.0	77.0	90.0	103.0	110.0	101.0	111.0	207.0	110.0	2.98	8.81	109
Arusha Region	Arusha City Council	Arusha Lutheran Center Referral Hospital	*Penta 1 doses given before 12 months	86.0	39.0	56.0	39.0	41.0	38.0	63.0		40.0	40.0	26.0	73.0	2.04	15.51	107
Arusha Region	Arusha City Council	Muriet Health Center	Children Voucher For Mosquito Nets	28.0	41.0	27.0	38.0	39.0	52.0	12.0				139.0		2.36	6.16	105
Arusha Region	Arusha City Council	Daraja II Health Center	*Penta 1 doses given before 12 months	145.0	124.0	129.0	111.0	135.0		112.0	122.0	122.0	121.0	36.0	138.0	2.82	5.8	90
Arusha Region	Arusha City Council	AICC Hospital at District Level	ANC Fourth Visit Attendance	8.0	6.0	10.0	10.0	31.0	20.0	11.0	19.0	17.0	7.0	7.0	92.0	3.02	13.74	79

Question 59: What is the smallest value of Modified Z-score that you now see in the table?
Inspect each row. Click on the menu icon at the end of the first row and select "Visualize" from the drop-down menu.
Question 60: Is the red value suspicious? After reviewing the bar chart, click on Close.

Use the "Outliers" dashboard to identify important missing values: To review the missing values, return to the Options window and click on "Missing Data" and "Off" (to turn off the Outlier filter).

Question 61: Are the rows of the table sorted in any order? Review the table, then describe how you think the rows are sorted.

Question 62: Which rows are most important to focus on? Which rows show the health facilities which have the largest amount of missing data?

Download the data to a spreadsheet and use the spreadsheet to view the rows

Click on the "Download" button.

When the Download window appears, leave the default setting as they are and simply click on "Download".

The file will appear in the bottom left corner of the screen. Click on the file name once to open the file.
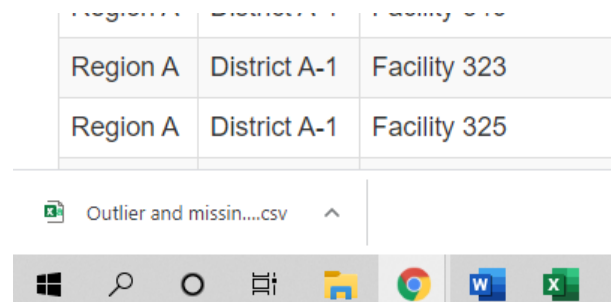
Outliers and missing values will **not** be highlighted on this CSV file, so you will have to visually scan the numbers. To identify health facilities with the most suspicious values, sort the rows based on "Max Modified Z-score":

Using Excel, select "Data" then select "Sort".
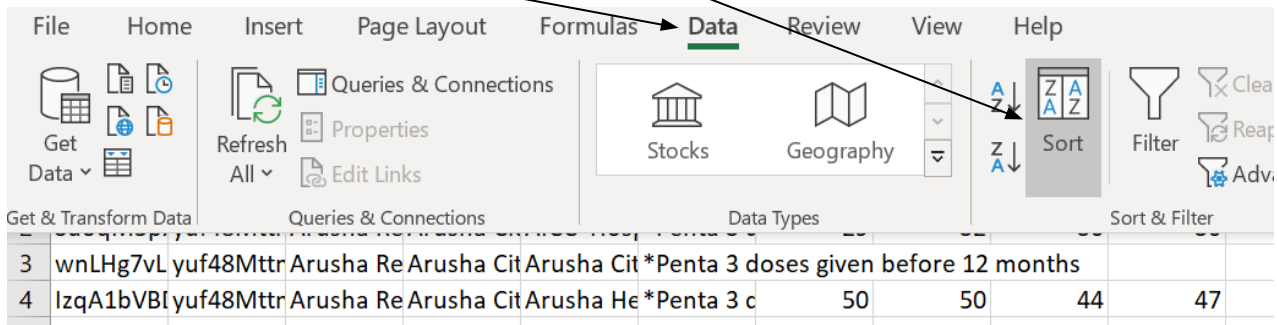
Set the Sort by function to "Max modified Z-score" (scroll down to find it). Set Order to "Largest to Smallest". Click "OK".
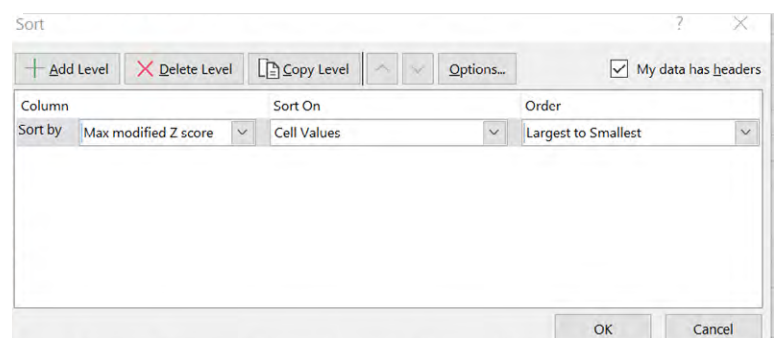
Review each of the top three rows. Identify the suspicious value in each of these rows. For each of these top three rows, identify the cell with the most suspicious value and highlight it in yellow (Home – Fill colour: yellow).

Return to the outlier dashboard and, under Options, click again on Outliers, Modified Z-Score and Extreme. Sort the result table by clicking once on the column header for Modified Z-Score.
Question 63: How does the Result table of the Outlier dashboard compare with your CSV file?
Question 64: How do you sort the table so that the top rows show health facilities with the largest amounts of missing data?

Answers to exercise questions

Tutorial 1 answers

1. The chart shows total values nationwide.
2. The dark blue line shows values for February of last year to January of this year. The light blue line shows values for February of the year before last to January of last year. The orange line shows values for February of 2 years ago to January of the year before last.
3. The values of ANC 1 have been roughly consistent from year to year and month to month, except for abrupt rises in May of last year and January of this year.
4. The diamond-shaped dot below the lower grey line represents District D-1. The diamond-shaped dot above the upper grey line represents District B-2.
5. With Period set to May of last year, the chart on the right shows a diamond-shaped dot above the upper grey line for District D-1. For May of last year, District D-1 reported 31 494. For the previous 11 months (June of the year before last) District D-1 reported an average of only 7 426.9.
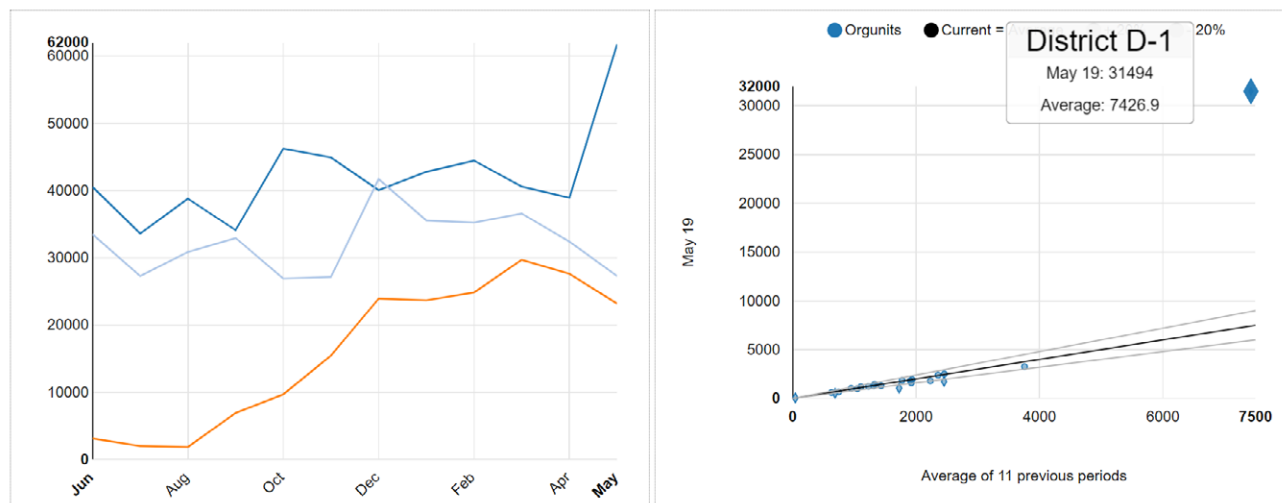
6. With Period set to January of the current year, the chart on the right shows a diamond-shaped dot below the lower grey line for District D-1. We have just learned that District D-1 had a suspiciously high value for May of last year. Due to that suspiciously high value, the average value for the previous 11 months (February of last year to December of last year) is raised to 8 426.1. This average value is more than 20% greater than the value reported for District D-1 for January 2020. For this reason, the dot falls below the lower grey line and is diamond-shaped.

7. Dropout rate is defined as the percentage of patients that have a first visit but fail to come for a subsequent visit.
Penta 1 to Penta 3 dropout rate (%) = ((Penta 1 minus Penta 3) x 100)/ Penta 1

8. District A-2 and district A-3 had negative dropout rates.

9. District A-2 has a dropout rate of -14.2%. For last year, the district reported 16 532 Penta 1 doses and 18 986 Penta 3 doses. Penta 1 to 3 dropout rate = (16 532 – 18 896)/16 532 = - 14.8%.

10. This is a trick question. The negative dropout rate is due to both an extremely high value for December of last year and higher values for each of the other months.

11. For last year, 27 out of 31 health facilities in District A-2 reported a higher number of Penta 3 doses than Penta 1 doses. The facilities in this district appear to have a systematic problem with over-reporting of Penta 3. It is possible that the staff in this district are misclassifying doses of Penta 1 and Penta 2 and reporting them as Penta 3 doses.

12. The chart at the bottom of the page shows month-to-month data for Facility 197. The chart shows that an extreme outlier value of Penta 3 was reported for Facility 197 for November of last year. However, the chart shows that, also for most of the other months, the number of Penta 3 doses was greater than Penta 1 doses. Thus, this is another trick question. A single large and erroneous value for November of last year is largely responsible for the negative dropout rate. However, even without this clearly erroneous value, Facility 197 would have a negative dropout rate for last year.

Unit	Penta 1 given < 1	Penta 3 given < 1	Dropout rate (%)
Facility 197	2320	4151	-78.9%

● Grouped ○ Stacked   ● Penta 1 given < 1 ● Penta 3 given < 1
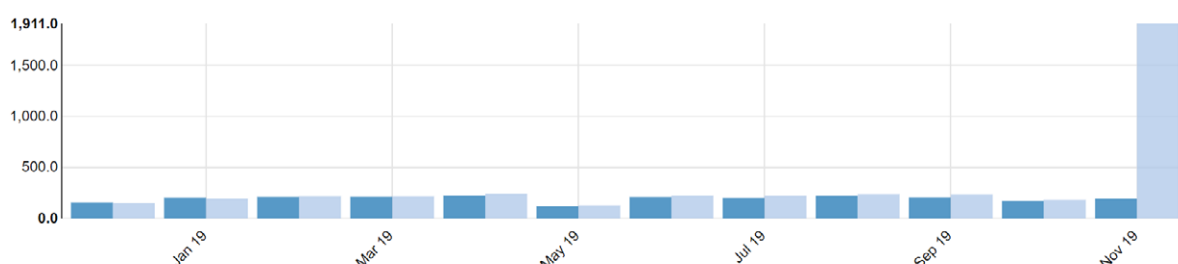
13. District level: each row of the Result table now shows 12 monthly values for one district. Recall that the dashboard was set to Disaggregation = District.

14. The "Weight" is roughly equal to the value of the outlier (highlighted in red) minus the average value that was reported for other months. If we assume that the outlier value is due to an error and that the actual value is roughly the same as the average value reported in other months, then the weight is the amount which has erroneously been added to the data.

15. The Visualize chart shows that the value of ANC 1 visits for District B-2 for January of this year was very different from the values for other months. Values such as this, which are extremely different from the values reported in other months, are usually due to an error.

16. The extreme outlier value of ANC 1 visits for January of this year was reported by Facility 147. This is almost certainly an error. Upon trying to drill down, a message appears: "Not possible to drill down, Facility 147 has no children." This means that there are no data for sub-units below Facility 147.

17. The user must configure the Analysis function. This is both an advantage (greater flexibility) and a disadvantage (it is not always easy to find and select the most appropriate data element or indicator unless the user is quite familiar with the DHIS2 datasets and indictors). An example of the flexibility of the Analysis - Outliers function is that the user can assess any data element or indicator for outliers and missing data – not only those that the data administrator has selected for the dashboard. Another example of flexibility is that the Analysis - Outliers function can assess for outliers using 12 quarterly values (e.g. for tuberculosis data) rather than 12 monthly values.

18. Possible steps include:
    a. Conduct further desk review. For example, use DHIS2 to compare the suspicious value with the value for related data elements (e.g. ANC 4, ANC testing for HIV, etc.) for the same facility and same month.
    b. Use DHIS2 to send a message to the person responsible for the data.
    c. At district level, the paper forms can be inspected.
    d. Telephone the person in charge at the health facility and ask them to check the paper form.
    e. A staff member with appropriate authority can edit the value.

19. In most health systems, it is the policy that the data are owned by the district and only staff at district level have permission to edit the data. In some health systems, the policy is to "lock down" the data a few weeks after the deadline for submitting the reports. In this case, district staff must first request permission from a higher level before they can edit data from several months ago. Whenever data are edited, it is very important that the changes are carefully documented, specifying the previous value, the new value, the person that made the change and the reason.

20. This will be discussed in depth in Tutorials 2 and 3.

Tutorial 2 answers

21. a) DQ-01, DQ-02 and DQ-03; b) DQ-04, DQ-05, DQ-06 and DQ-06b; c) DQ-07; d) DQ-08; e) DQ-09.

22. 86.5%.

23. At the beginning of each month, data for last month have not yet been entered in the DHIS2. Hence, to review completeness for last month, it is best to wait until the end of the month. Note that with this training instance of DHIS2 (who-dev.dhis2.org/dq), data for last month as well as for all the months for the remainder of the current year have already been entered. Therefore, in this respect the training instance behaves differently from a real-world instance of DHIS2.

24. The answer will depend on the current month. If the current month is earlier than May, clinics had the lowest reporting rate, followed by dispensaries and health centres. If the current month is May or later, health centres and dispensaries have recently had the lowest reporting rates.

25. DQ-03 shows reporting rates for the last 3 months.

26. The districts with lowest malaria dataset reporting rates depend on the current month in which DQ-03 is viewed: January (D-1, A-2, C-2); February (D-1, A-2, C-2); March (A-2, D-1, D-3); April (D-1, D-3, A-2); May (D-1, C-2, D-3); June (D-1, C-2, D-3); July (D-2, D-3, D-1); August (D-3, D-2, D-1); September (D-3, D-2, D-1); October (C-2, D-3, D-2); November (D-2, C-2, D-3); December (D-2, C-2, D-3).

27. The trend for most indicators is almost like a straight, horizontal line without much variation from month to month.

28. The exceptions are the values of ANC 1 visits for January of the current year and May of last year.

29. Usually, the values of maternal health services (ANC, deliveries) and immunization services (BCG, Penta 1, etc.) do not vary a great deal from one month to the next. The reason for this is that roughly the same number of pregnancies start each month, so there are roughly the same number of deliveries each month (after subtracting abortions/miscarriages) and so there are roughly the same number of children coming for their first dose of Penta vaccine each month (after subtracting neonatal deaths).

30. Each year, the number of malaria cases is much lower during December to May than it is during June to November. This increases the difficulty of detecting values that may be erroneous – especially if a high value is reported during the malaria season.

31. The chart shows the trend, for each district, in under-5 malaria cases as a percentage of all reported malaria cases = (malaria < 5 years)/(malaria < 5 years + malaria ≥ 5 years).

32. Most lines are almost straight, with the value staying roughly constant between 25% and 50%.

33. The trend lines for districts A-6, A-2, D-4 and D-1 each include at least 1 month with a value that is quite different from the average monthly value. For District A-6, the values for August and September of last year are much greater than for the average. For District A-2, the value for February of this year is much less than average. For District D-4, the value for May of this year is higher than average. For District D-1 the value for June of this year is higher than the average.

Data Quality Assurance. Module 2: Desk review of data quality

67

67

34. For District A6, 95.9% of malaria cases were children less than 5 years of age. For District D-4, 93.4% of the reported malaria cases were children less than 5 years of age. This is very different from other months when roughly half the malaria cases were in this age group. A sudden rise in the indicator to more than 90% is very suspicious and suggests that either the number of cases in older patients was under-reported or the number of cases in children was over-reported. (The table here shows the data for May 2020 for Facility 292 of district D-4.)

Malaria microscopy-confirmed outpatient cases < 5 years	51
Malaria microscopy-confirmed outpatient cases >= 5 years	33
Malaria presumed inpatient cases < 5 years	34
Malaria presumed inpatient cases >= 5 years	29
Malaria presumed outpatient cases < 5 years	221217
Malaria presumed outpatient cases >= 5 years	65
Malaria RDT-confirmed outpatient cases < 5 years	522
Malaria RDT-confirmed outpatient cases >= 5 years	644
Malaria RDTs done on outpatient cases < 5 years	729
Malaria RDTs done on outpatient cases >= 5 years	933

35. The extreme outlier values for ANC 1 visits reported in May of last year, January of this year and September of this year, are so large that they appear as sharp increases in the total value at national level.

36. The extreme outlier values for Penta 1 and Penta 3 are not large enough to visibly increase the nationwide total values. This provides an important lesson: most extreme outliers are not apparent when reviewing nationwide totals. This is a further reason to decentralize data quality review: as seen in the next session, a chart like DQ-04 is much more sensitive for identification of extreme outliers if it is viewed at district level.

37. a) Facility 147; b) Facility 567; c) Facility 482; d) Facility 79.

38. The "Penta 1 to Penta 3 dropout rate" is the percentage of children that received their first dose of Penta vaccine, but failed to receive their third dose of this vaccine. It is equal to (Penta 1 minus Penta 3)/(Penta 1).

39. The two districts on the far left of the chart, A2 and A3, had a negative dropout rate over the last 12 months.

40. If the dropout rate is negative, it means that the district reported more third doses of vaccine than first doses. At district level, a negative Penta 1 to Penta 3 dropout rate for a full 12 months suggests that the district probably has a problem with some health facilities sometimes mistakenly reporting a vaccination as a Penta 3 dose when, in fact, it was a Penta 1 or Penta 2 dose. This may be due to a single, very large outlier value reported by a single health facility or it may be due to many health facilities systematically over-reporting Penta 3.

41. The survey estimates of coverage are consistent with the routine estimates of coverage for all regions except Region B.

42. The higher routine estimate of coverage may be due to erroneous over-reporting of the numerator (reported Penta 3 doses) or under-estimation of the denominator (estimated population under 12 months of age).

43. Survey measures can be less reliable for the following reasons: 1) too small a sample size; 2) a sample that is not selected randomly; 3) recall bias – reliance on the memory of caretakers when vaccination cards are not available for a large percentage of children.

44. a. Year-to-year inconsistencies:

    i. From 2016 to 2017 there was a decline in the estimated pregnancies and estimated live births. This is not consistent with most other years when there was an increase from year-to-year.

    ii. From 2017 to 2018 there was a much greater increase in estimated pregnancies and population < 1 year compared with other years.

    b. Inconsistency between related denominators: for 2018 and 2019 the estimated population < 1 year was greater than the estimated number of live births. This is not plausible, because the population < 1 year should be less than the estimated number of live births (as a result of infant mortality).

Tutorial 3 answers

45. When the chart is viewed at the beginning of a month, the monthly reports for the preceding month have not yet been entered in the DHIS2. Over the course of the month, the reporting rates for last month will increase.

46. The answer will vary depending on the month in which you view the dashboard. When viewed in October, the reporting rate was lowest for the malaria dataset followed by the HMIS dataset.

47. Health centres had the lowest reporting rate.

48. When DQ-03 is viewed at the beginning of a month, the data will not have been entered for last month. As a result, the highest reporting rate for any health facility will be only 66.7%. For this reason, it is best to wait until the data for last month have been entered for most health facilities before reviewing the data quality dashboard.

49. The answer depends on the month in which the dashboard is viewed. At the time of writing this document, seven facilities had submitted all of their last 3-monthly reports, one facility had submitted two of their last 3-monthly reports and two facilities had submitted only one of their last 3-monthly reports. No facility had failed to submit any reports at all.

50. The following values appear suspicious: ANC 1 for May of the current year; ANC 4 for October of last year (this is not very obvious when compared with the other outliers, but nevertheless suspicious); Penta 1 for March of this year; Penta 3 for August of last year and June of this year. The values for measles vaccine doses vary a lot and are also quite suspicious.

51. With a month-to-month chart like DQ-04, only the largest and most suspicious values can be identified; the Outlier function of the WHO DQ Tool is much more sensitive for detecting suspicious values.

52. This is a trick question. There is evidence in DQ-04 for all of the large outliers listed in the question.

53. The investigation can begin with further DHIS2-based desk review using the Pivot Table app and other DHIS2 applications. Next, the paper forms should be investigated to see if the outlier was recorded on the paper form or if a mistake was made during entry of the data into DHIS2. Next, the staff of the facility may need to be contacted to investigate further (e.g. by looking at individual registers).

54. The Tool finds the values highlighted in red to be suspicious because they are either significantly higher or significantly lower than the other values in the row.

55. The cells coloured yellow are blank. Data are missing.

56. The rows are sorted in order of "weight". The row with the highest weight appears at the top.

57. The "Outlier weight" is the difference between the suspicious value (the value highlighted in red) and the average of other values in the row.

58. The "Missing Weight" is equal to the average value reported by that facility multiplied by the number of months with missing (yellow) values. "Total Weight" is Outlier Weight plus Missing Weight.

59. After using the options to filter the table, all rows of the table should have a modified Z-score of five or more. These are the most suspicious values and are the most important to investigate.

60. If the bar chart resembles the example shown, the red value is very much greater than the values reported in other months. This means that the red value is very suspicious.

61. The rows are sorted according to the "Missing Weight".

62. The rows at the top show the facilities and the indicators for which the largest total values of data are missing.

63. The sorted Result table of the Outlier dashboard should be identical to the sorted CSV file.

64. Use the Excel Data-Sort steps to sort by "Gap weight" with "Order" from "Largest to Smallest".

Notes

. . . .
. . . .
. . . .
. . . .
. . . .
. . . .
. . . .
. . . .
. . . .
. . . .
. . . .
. . . .
. . . .
. . . .
. . . .
. . . .
. . . .
. . . .
. . . .
. . . .
. . . .
. . . .
. . . .
. . . .
. . . .
. . . .
. . . .
. . . .

Notes

DQA

DATA QUALITY ASSURANCE

9789240036468

9 789240 036468