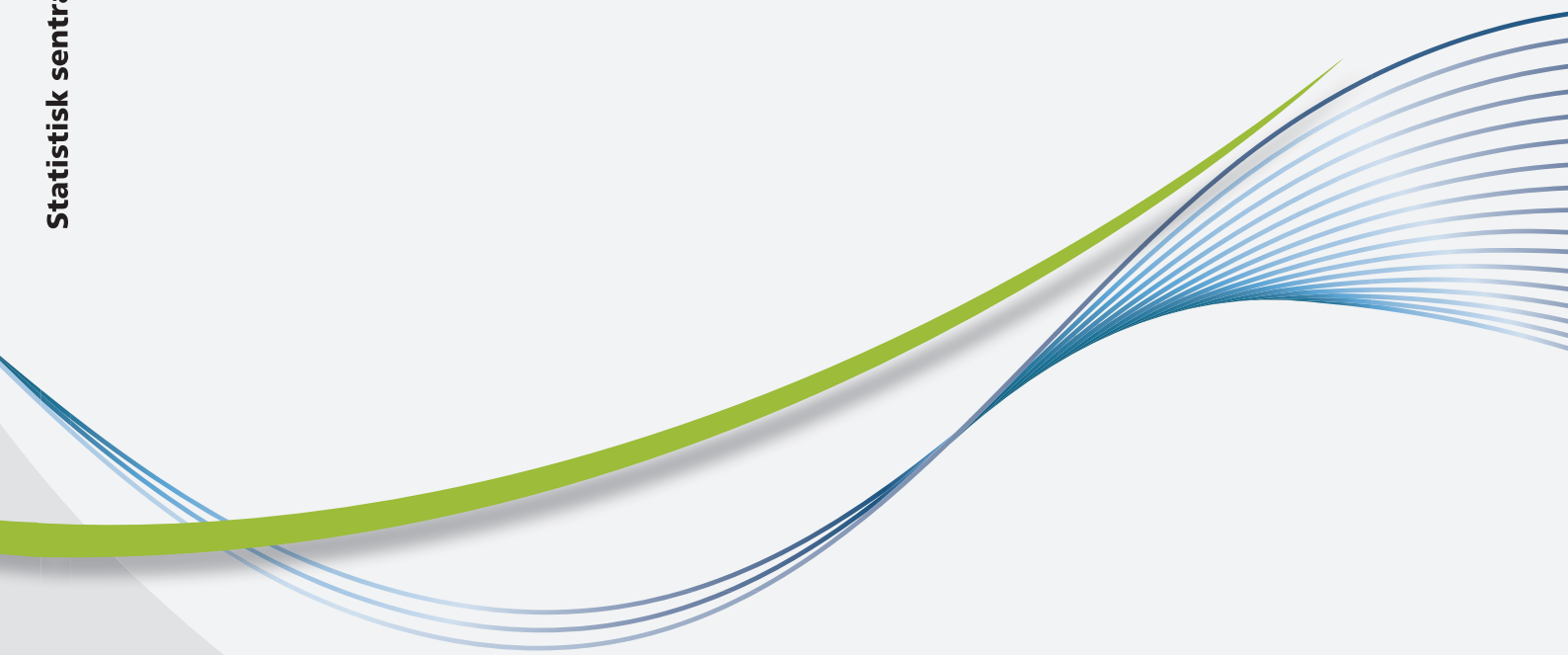
Notater
Documents
2017/4

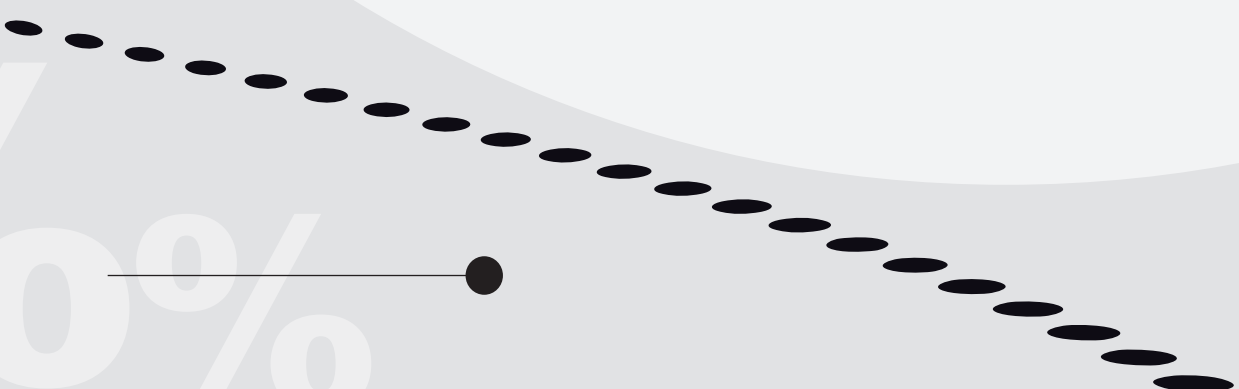
Statistics Norway
Statistisk sentralbyrå

John Åge Haugen, Geir Hjemås and Olav Poppe

Manual for the DHIS2 quality tool

Understanding the basics of improving data quality





Documents 2017/4

John Åge Haugen, Geir Hjemås
and Olav Poppe (WHO)

Manual for the DHIS2 quality tool

Understanding the basics of improving data quality

Statistisk sentralbyrå • Statistics Norway
Oslo–Kongsvinger

In the series Documents, documentation, method descriptions, model descriptions and standards are published.

Published 18 January 2017

ISBN 978-82-537-9477-8 (electronic)

Symbols in tables	Symbol
Category not applicable	.
Data not available	..
Data not yet available	…
Not for publication	:
Nil	-
Less than 0.5 of unit employed	0
Less than 0.05 of unit employed	0.0
Provisional or preliminary figure	*
Break in the homogeneity of a vertical series	—
Break in the homogeneity of a horizontal series	|
Decimal punctuation mark	.

Preface

A Health Management Information System (HMIS) is important for a country's capacity to monitor and improve the delivery of health-care services and programs. The District Health Information System (DHIS2) is a software and a system used in almost 50 countries to collect and process health data. A new quality tool has been developed for this software and this document is the manual for the DHIS2 quality tool.

The manual is a joint effort by the World Health Organization (WHO), the Department of Informatics at the University of Oslo (DoI) and Statistics Norway (SN). Project activities were carried out in 2016.

Thanks to the Norwegian Agency for Development Cooperation (NORAD) for funding of Statistics Norway's contribution to this report.

Statistisk sentralbyrå, 16 December 2016.

Bjørnar Gundersen

Abstract

A Health Management Information System (HMIS) is important for a country's capacity to monitor health, and for evaluating and improving the delivery of health-care services and programs. Many developing countries struggle with data quality problems in their HMIS systems resulting in incomplete, inaccurate and untimely information. As a result, the systems do not provide a good basis for knowledge-based decision-making on health.

The newly developed DHIS2 quality tool is a contribution to a practical approach for improvement of HMIS data quality. By using the quality tool, potential errors in the data are identified. This knowledge can be used to take appropriate action for improving data quality. In many cases, this will be to either edit the data or to improve the data collection system. Implementing modern statistical methods and technology, like the DHIS2 quality tool, are important factors in achieving good quality in data/statistics.

This is a manual for basic use of the DHIS2 quality tool. The manual is aimed at explaining the possibilities of the quality tool, but also to be an input to capacity building and workshops. The data quality tool does not only display potential data errors, but also contributes to improved understanding of the data quality.

The manual is a meant as a practical guide to use the DHIS2 quality tool. Actual screenshots from DHIS2 and a realistic test data set are used to explain and visualize (which is available on the DHIS2 training server). The DHIS2 Data Quality Tool requires an initial configuration for much of the functionality, and this initial setup is the best way to get started.

Contents

1. Introduction

This Manual is part of a NORAD funded project at Statistics Norway (SN) focused on improving data quality in Health Management Information Systems (HMIS). Further, the manual is a result of a joint effort from the World Health Organization (WHO), the Department of Informatics at the University of Oslo (DoI) and Statistics Norway (SN). Project activities on the manual were carried out in 2016. WHO has developed the DHIS2 Data quality tool (software) with SN and DoI as discussion partners (this activity was initiated a few years back).

The DHIS2 quality tool is one step on the ladder for better HMIS data quality. It is important to bear in mind that the tool itself will not improve data, but it can be used to identify potential errors in the data. This knowledge must then be used to take appropriate action; in many cases this will be to either edit the data or to improve the data collection system. Implementing modern statistical methods and technology, as DHIS2, are important factors in achieving good quality in statistics. However, as the report "*Improving health data quality: Recommendations and guidelines"*[1] has detailed, it also depends on other factors; "*As important as modern statistical methods and technology is the role of institutional factors"*[1]. The use of statistical methods for editing in parts depends on having data of adequate quality. Sufficient quality it often dependent on what can be called an institutional framework supporting the data collection system as a whole.

2. Using the DHIS2 quality tool

In this manual the focus is on the basic use of the DHIS2 quality tool. The main goal for using such a tool is to improve data quality. Data improvement usually involves editing data or improving the data collection system. The latter can require more resources and time. Generally most datasets are affected by human errors and misunderstandings. These errors can be identified and should be corrected. The ideal approach is to verify against primary data sources, for instance contacting the local health facillities. If this is not possible, data could be edited using standard statistical methods. It is recommended to read both the report "*Improving health data quality: Recommendations and guidelines*[1] and "*The Health Management Information system in Malawi: Assessment of data quality and methods for improvements*[2]*"* which provides a thorough presentation of tools and techniques to improve HMIS data. The reports are based on the experience of using DHIS2 as the HMIS in Malawi.

3. Using the DHIS2 quality tool manual

This manual is a practical guide to use the DHIS2 quality tool. It is structured around the actual screenshots that are used in DHIS2 and a realistic test data set (which is available on the DHIS2 training server).

The DHIS2 Data Quality Tool requires an initial configuration for much of the functionality, and should be the first step to get started using the quality tool. In chapter 11 there is a step by step guide for this procedure.

After the Quality tool is set up it is recommended to follow the whole manual chronologically, from start to end. It is designed to explain the most basic features first, before moving more in to depts. The chapters can also be looked at independently if one is only interested in a particular subject. Along the way there are also questions that should be discussed.

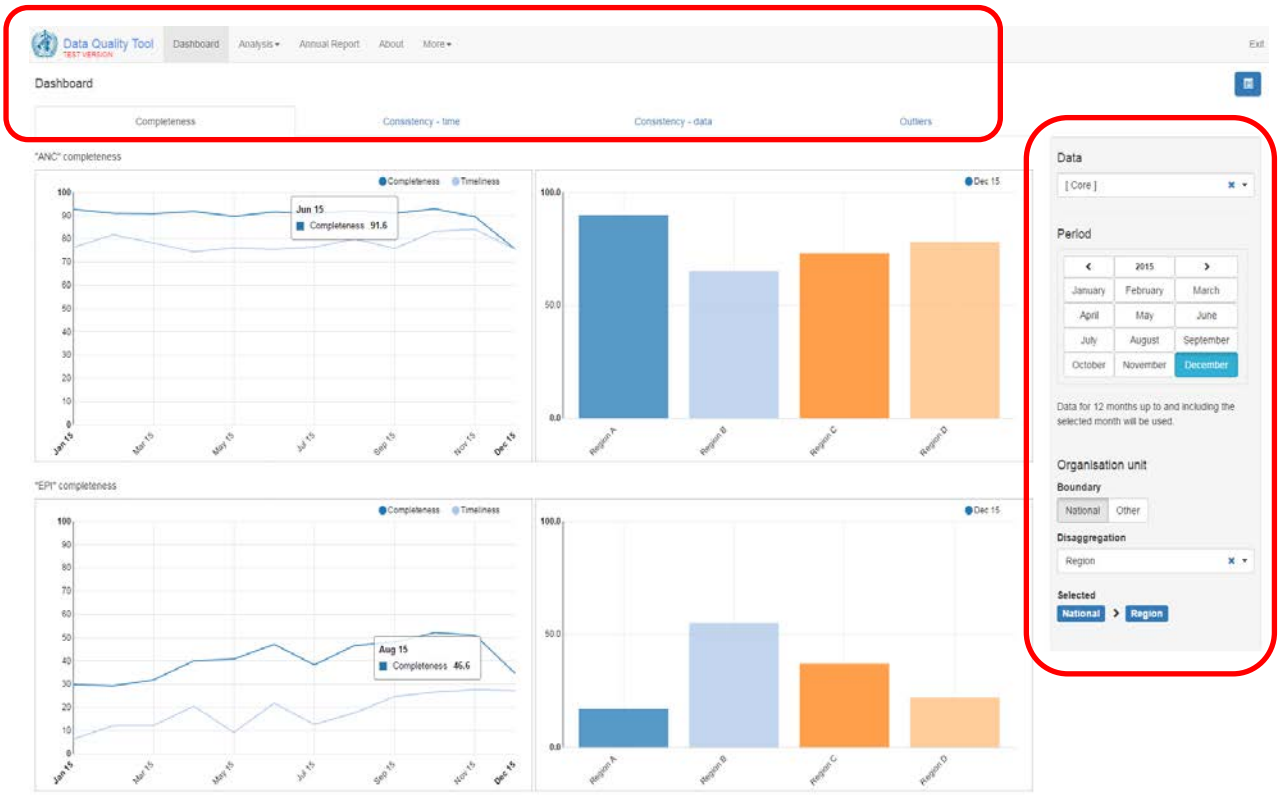
[1] Statistics Norway. Documents 2017/xx. *Improving health data quality: Recommendations and guidelines. Based on the case of the Health Management Information System in Malawi and DHIS2*

[2] Statistics Norway. Documents 2017/xx. *The Health Management Information system in Malawi: Assessment of data quality and methods for improvements*

4. Basics of the DHIS2 Data Quality Tool

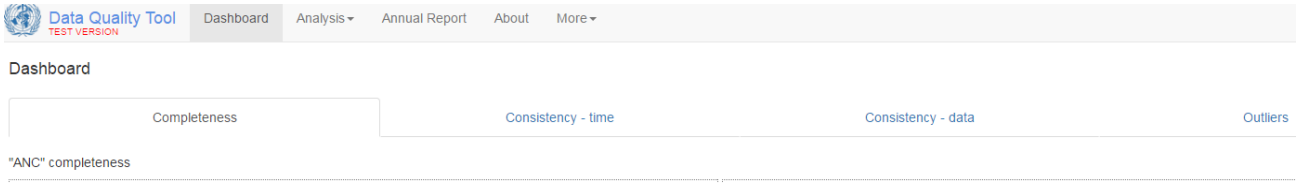
Figure 1 is on a typical first page in the Data quality tool. This is called the Dashboard.

Figure 1 Screenshot from DHIS2 quality tool



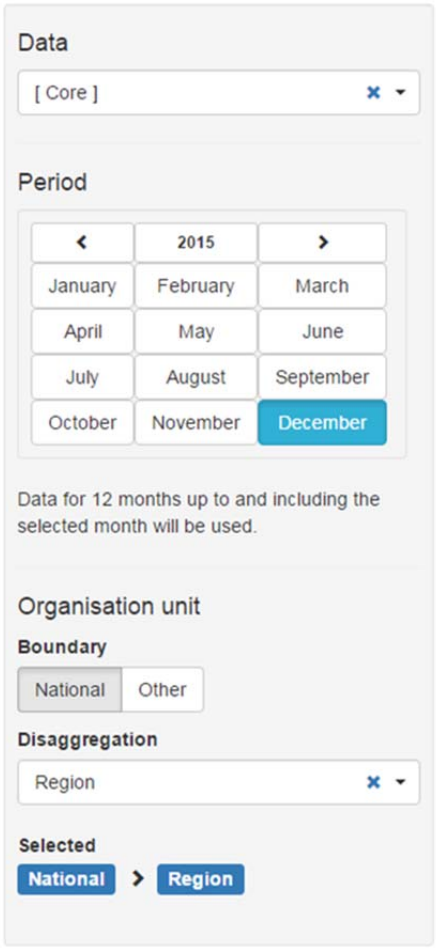
The red circle in the top of Figure 1 shows two important menus (see Figure 1a below). The uppermost menu is the “main menu”. Here the “Dashboard” tab is highlighted (in grey), indicating the current location. In the manual we will cover all areas (except “about”) in the “main menu”. Below the “main menu” the details for the current location is on display. On the “Dashboard” there are four main tabs to choose from, all which will provide different types of analysis. In Figure 1a “Completeness” has been chosen.

Figure 1a Screenshot from DHIS2 quality tool



On the right side of the screen of Figure ,1 inside the second red circle, another important menu ca be observed (See Figure 1b on next page). In this menu one can choose “Data”, “Period”, “Organisation unit”, “Boundary” and “Disaggregation”. A more thorough examination of this will follow in the next chapters.

Figure 1b Screenshot from DHIS2 quality tool



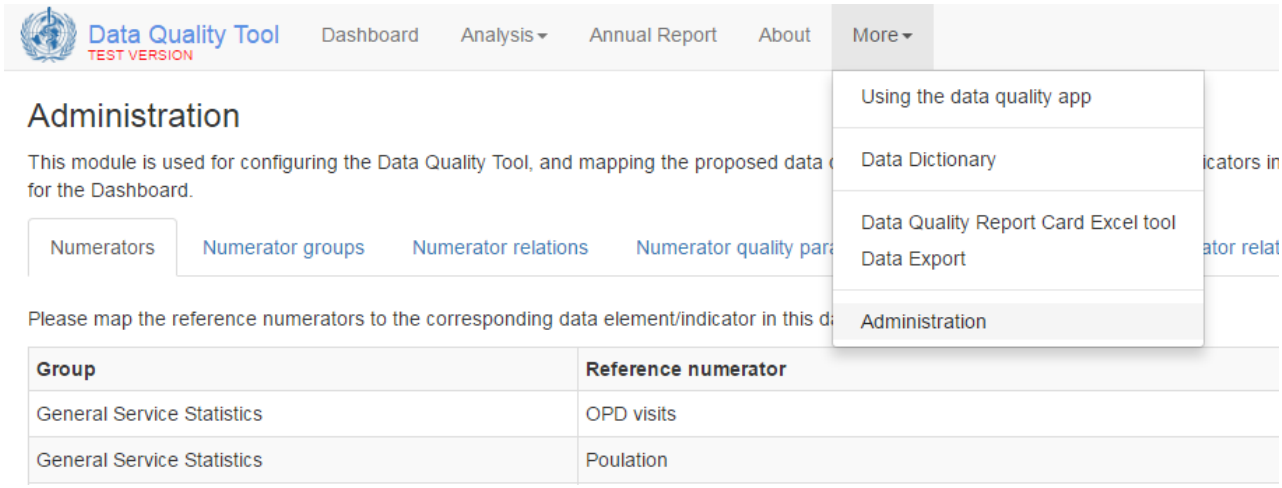
5. Dashboard – completeness

One of the most important factors when evaluating data quality is completeness. A low degree of completeness means that many facilities have missing data in their reported data. Increasing completeness is an important factor for increasing data quality. A low completeness rate can provide a wrong impression of the actual situation. If the completeness is low, actions should be set in place to increase it.

5.1. Step by step setup of Dashboard – completeeness

Completeness compares the proportion of collected data against the potential of “100% complete”. Completeness is measured as the presence of non-blank values. The completeness is shown for the entire data set, which can consist of several data elements and is done on national, region/zone, district and facility level. If an “Orgunit” has reported on 3 out of 4 data elements it will have 75 percent on completeness. An “Orgunit” with more than 10 percent zero/missing values will be flagged in the annual report. The threshold is set in the “Administration” menu (see figure 1c), which can be accessed by the “more” tab at the uppermost part of the screen.

Figure 1c Screenshot from DHIS2 quality tool



Timeliness

Timeliness measures the discrepancy between the time of reporting and the time the event occurred. A high rate of timeliness means that sufficient data is available at the required point in time. Timeliness is in this context measured as proportion of reports received on deadline.

Figure 2 Screenshot from DHIS2 quality tool

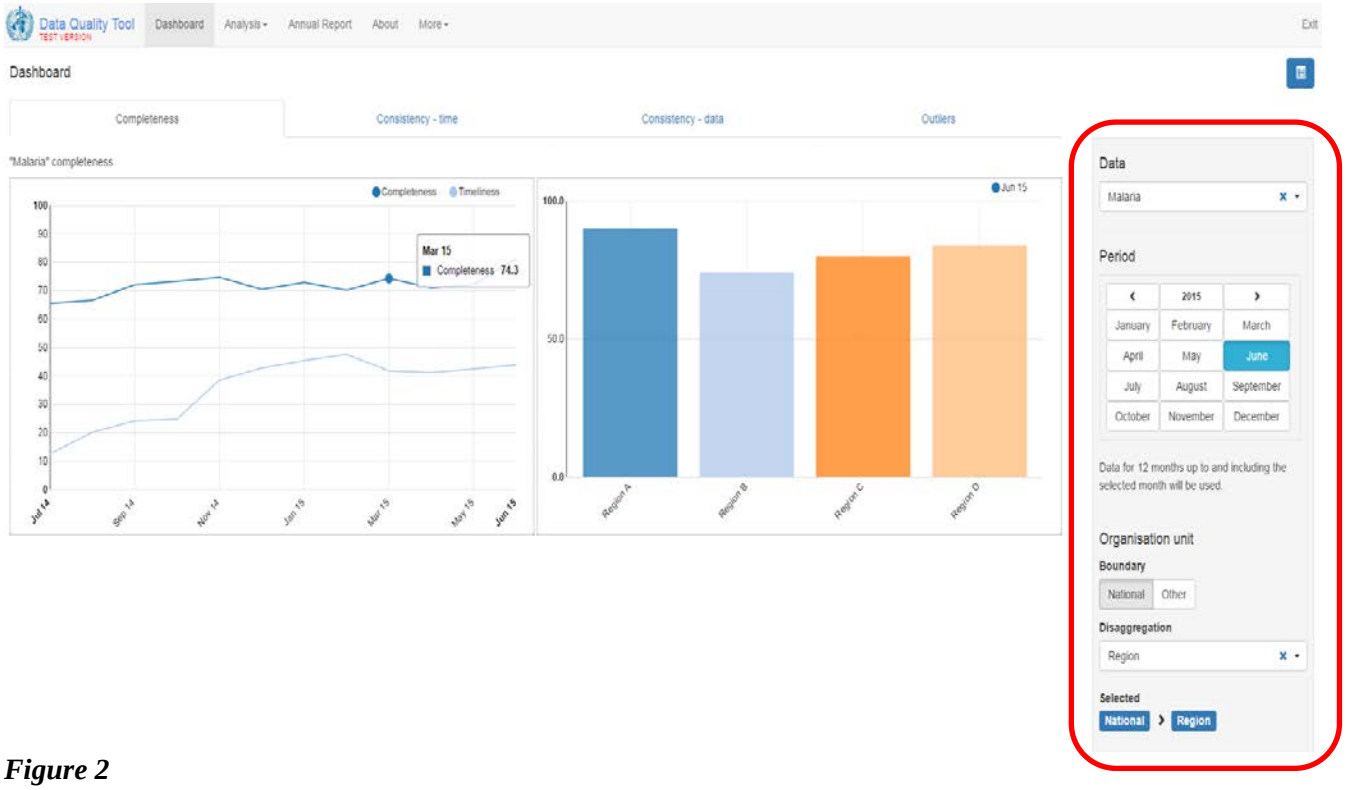


Figure 2

1. Select the “Dashboard” tab (“main menu”)
2. Select the “Completeness” tab under the headline” Dashboard”
3. Select data group in dropdown menu under the headline “Data” (see Figure 2 and inside red circle). Choose “malaria” in the drop down menu
4. Under the headline “Period” select year and month (see Figure 2 and inside red circle), choose “2015” and “June”. By picking June 2015 the months from July 2014 to June 2015 will be analyzed.
5. Hovering with the cursor over the chart will give additional information where this is available.

Questions: to the chart on the left:

- Is the completeness rate good and is there a trend?
- Are there many fluctuations in the completeness rate through the period?
- Are there specific months which stand out?
- Is the timeliness rate good and is there a trend?
- Are there many fluctuations in the timeliness rate through the period?
- Are there specific months which stand out?
- Are the orgunits equal in size? If not is completeness on 80 percent actually 80 percent of the total value?

Figure 3 Screensho from DHIS2 quality tool

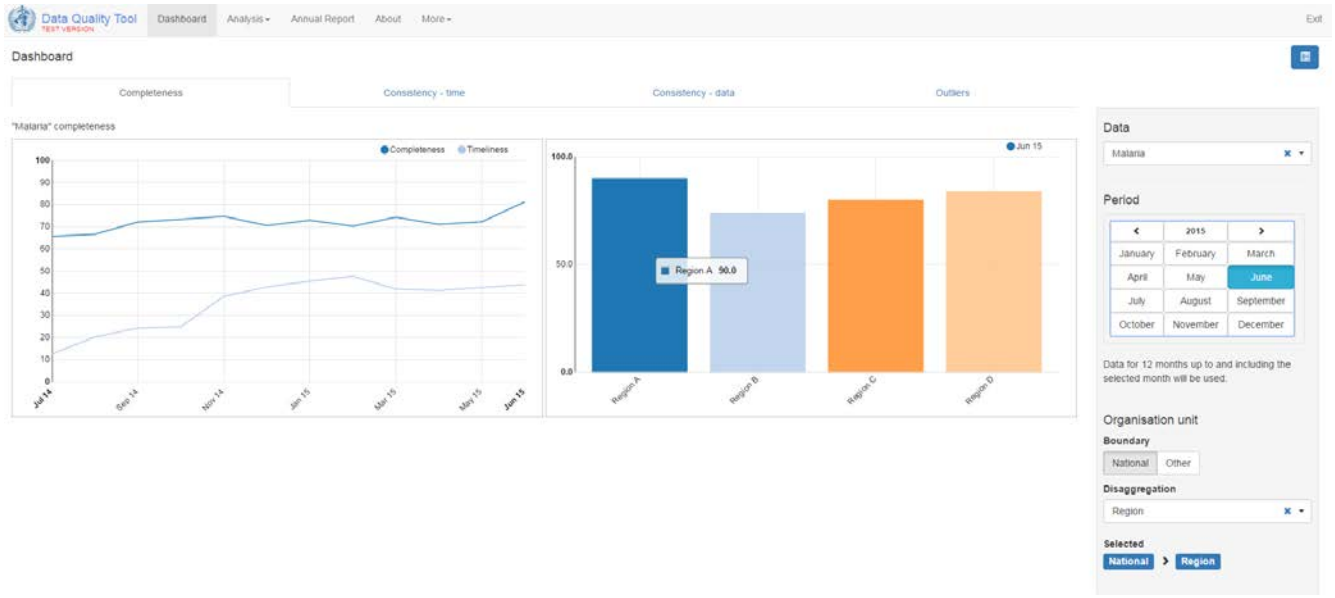


Figure 3

The bar graph to the right shows completeness distributed on regions, and gives an indication of in which regions there might be a data quality challenge. The bars show the completeness for June 2015.

Questions: to the bar chart on the right:

- Is the completeness equally distributed among the regions?
- Is there a region which stands out?

Figure 4 Screenshot from DHIS2 quality tool

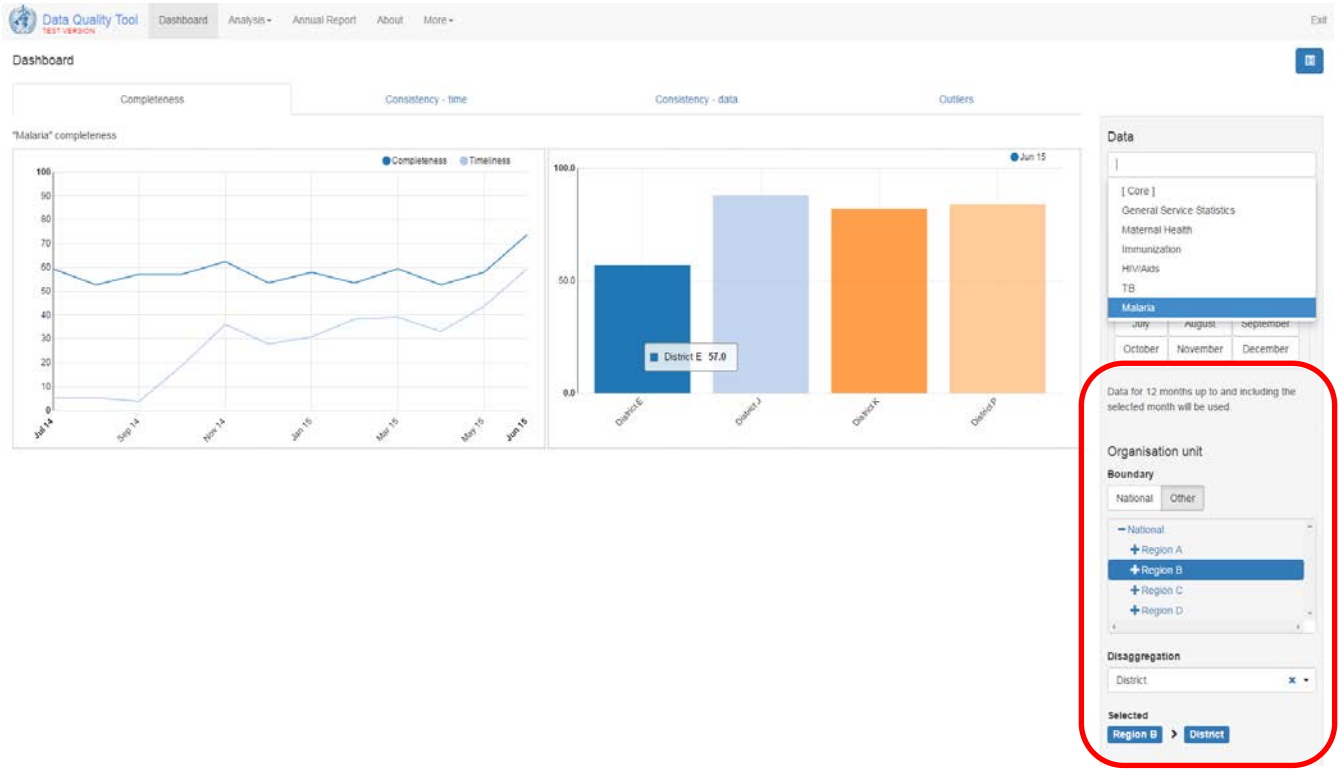


Figure 4

Under the headline "Organisation unit" press "Region B" to drill down[3] (see Figure 4 and inside red circle). The chart to the left shows the completeness and timeliness rate for "Region B" and the chart to the right the distribution on completeness between the districts.

Questions: to the chart on the left:

- Is the completeness rate good and is there a trend?
- Are there many fluctuations in the completeness rate through the period?
- Is there a specific month which stands out?
- Is the timeliness rate good and is there a trend?
- Are there many fluctuations in the timeliness rate through the period?
- Are there specific months which stand out?

Questions: to the bar chart on the right:

- Is the completeness equally distributed among the districts?
- Is there a district which stands out?

[3] To" drill down" means to move down one level, for instance from National level to regional level

Figure 5 Screenshot from DHIS2 quality tool

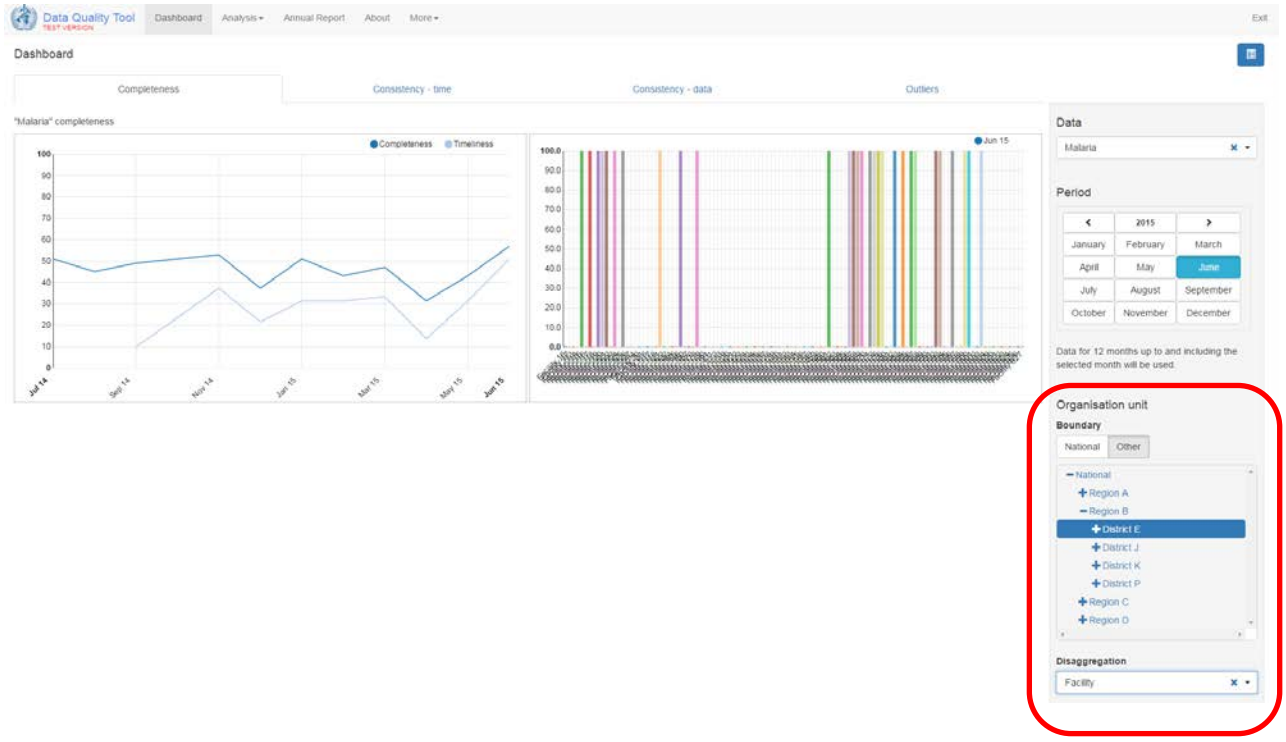


Figure 5

Under the Headline "Orgunit" press District E to drill down (See figure 5 and inside red circle). The chart to the right, even though it is difficult to see all facilities, indicates that just above half of the facilities have reported. If District E had fewer facilities the chart would give a clearer indication.

Questions: to the chart on the left:

- Is the completeness rate good and is there a trend?
- Are there many fluctuations in the completeness rate through the period?
- Are there specific months which stand out?
- Is the timeliness rate good and is there a trend?
- Are there many fluctuations in the timeliness rate through the period?
- Are there specific months which stand out?

6. Dashboard Consistency over time

Comparing reported values for a unit over time is an efficient method to get a better understanding and knowledge of the data. It can also provide valuable insight on issues regarding data quality and data errors. Data values can fluctuate due to seasonal fluctuations, like for instance malaria. The "consistency-time" tool on the "Dashboard" can provide a first impression on fluctuating data. This is useful as a guidance to where to focus quality improving activities. More details of the method follow in this chapter.

6.1. Step by step setup of Dashboard Consistency over time

Figure 6 Screenshot from DHIS2 quality tool

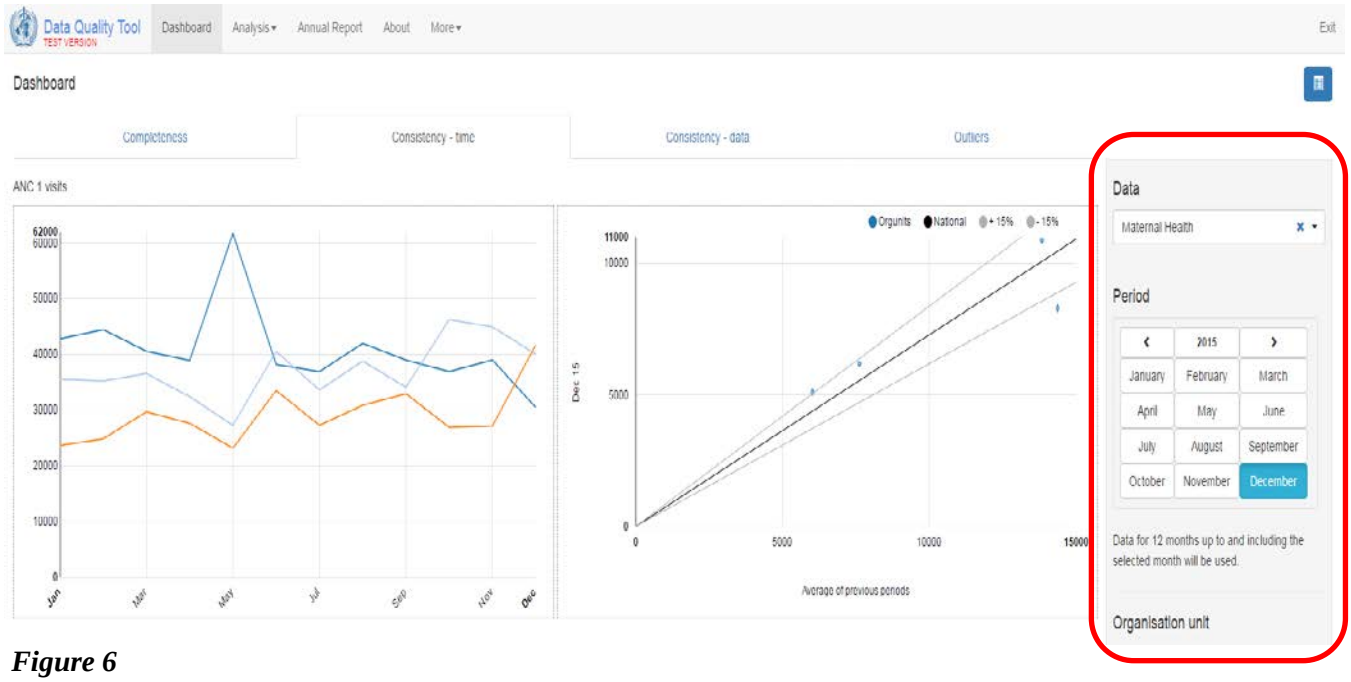


Figure 6

1. Select the "Dashboard" tab ("main menu")
2. Under the headline "Dashboard" (upper left corner) choose the tab "Consistency-time"
3. Select data group in dropdown menu under the headline "Data" (see Figure 6 and inside red circle). Choose "maternal health" in the drop down menu
4. Under the headline "Period" select year and month (see Figure 6 and inside red circle), choose "2015" and "December". By picking December 2015 the months from January 2015 to December 2015 will be analyzed.
5. Hovering with the cursor over the chart will give additional information where this is available.

Questions: to the chart on the left:

- Does the consistency follow any trends and is it comparable to previous years?
- Are there many fluctuations in the consistency through the period?
- Are there specific months which stand out?

Figure 7 Screenshot from DHIS2 quality tool

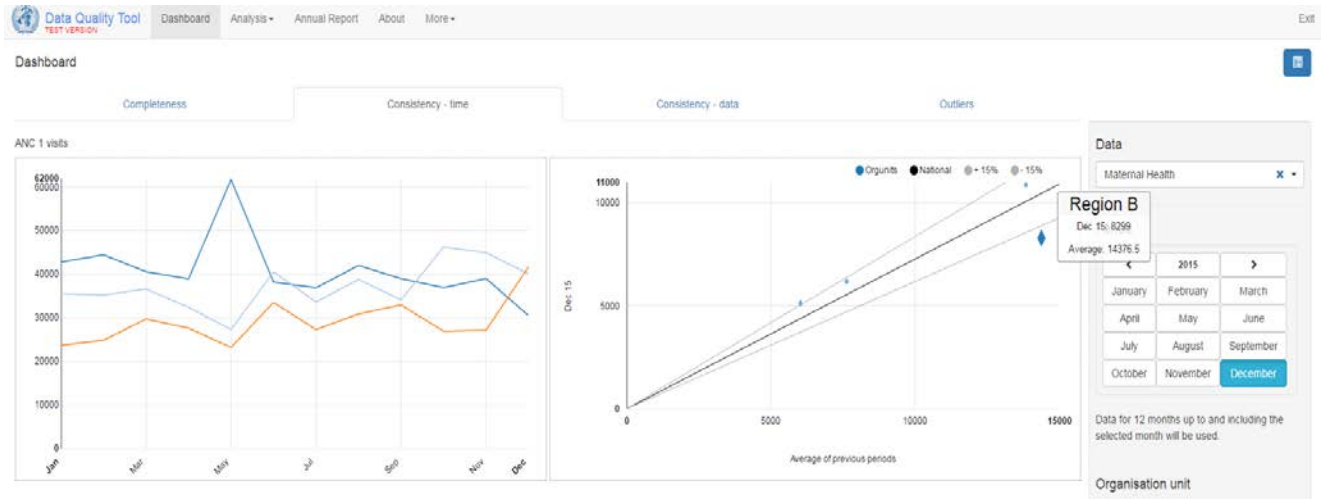


Figure 7

The chart to the right shows how the regions are distributed, and gives an indication in which region there might be a challenge or data errors. The scatter plot shows December 2015 compared to the average for the 12 preceding months.

Questions: to the chart on the right:

- Is the consistency equally distributed among the regions?
- Are there specific regions which stand out?

Figure 8 Screenshot from DHIS2 quality tool

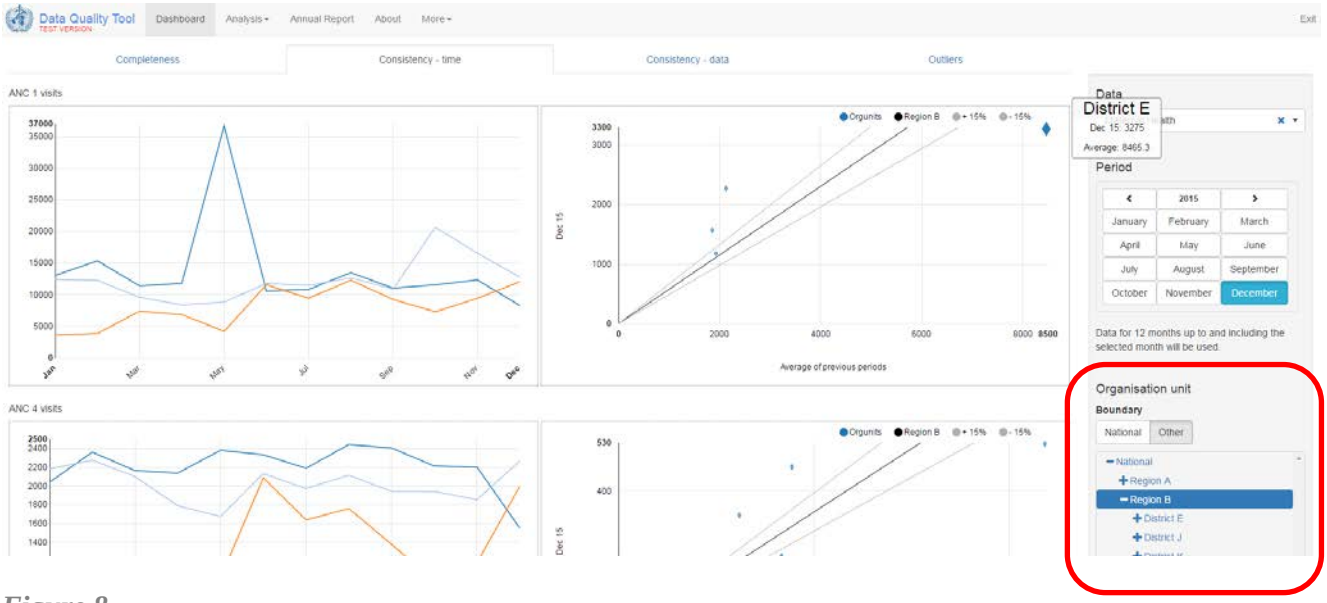


Figure 8

To look more in to detail for “Region B” proceed with the following steps: Under the headline “Organisation unit” and “Boundary” press the button “Other”. Then press region “Region B” to drill down from National to Regional level (see Figure 8 and inside red circle). The chart to the left shows the consistency for Region B and the chart to the right the distribution on consistency between the districts.

Questions: to the chart on the left:

- Does the consistency follow any trends and is it comparable to previous years?
- Are there many fluctuations in the consistency through the period?
- Are there specific months which stand out?

Figure 9 Screenshot from DHIS2 quality tool

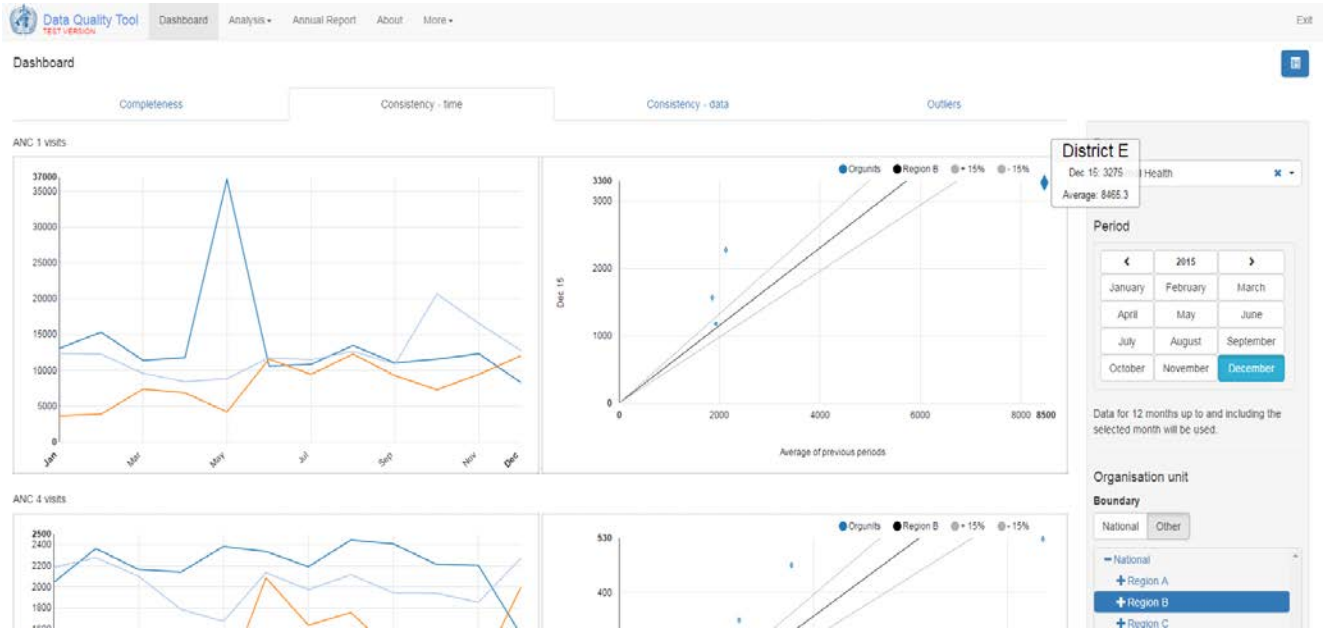


Figure 9

The chart to the right compares reported figures for three years for ANC 1 visits, broken down on months. To the right there is a scatter plot which shows how the regions are distributed, and this gives an indication of in which regions there might be a quality issue. The scatter plot shows December 2015 compared to the average for the 12 preceding months.

Questions: to the chart on the right:

- Is the consistency equally distributed among the districts?
- Are there specific districts which stand out?

Figure 10 Screenshot from DHIS2 quality tool

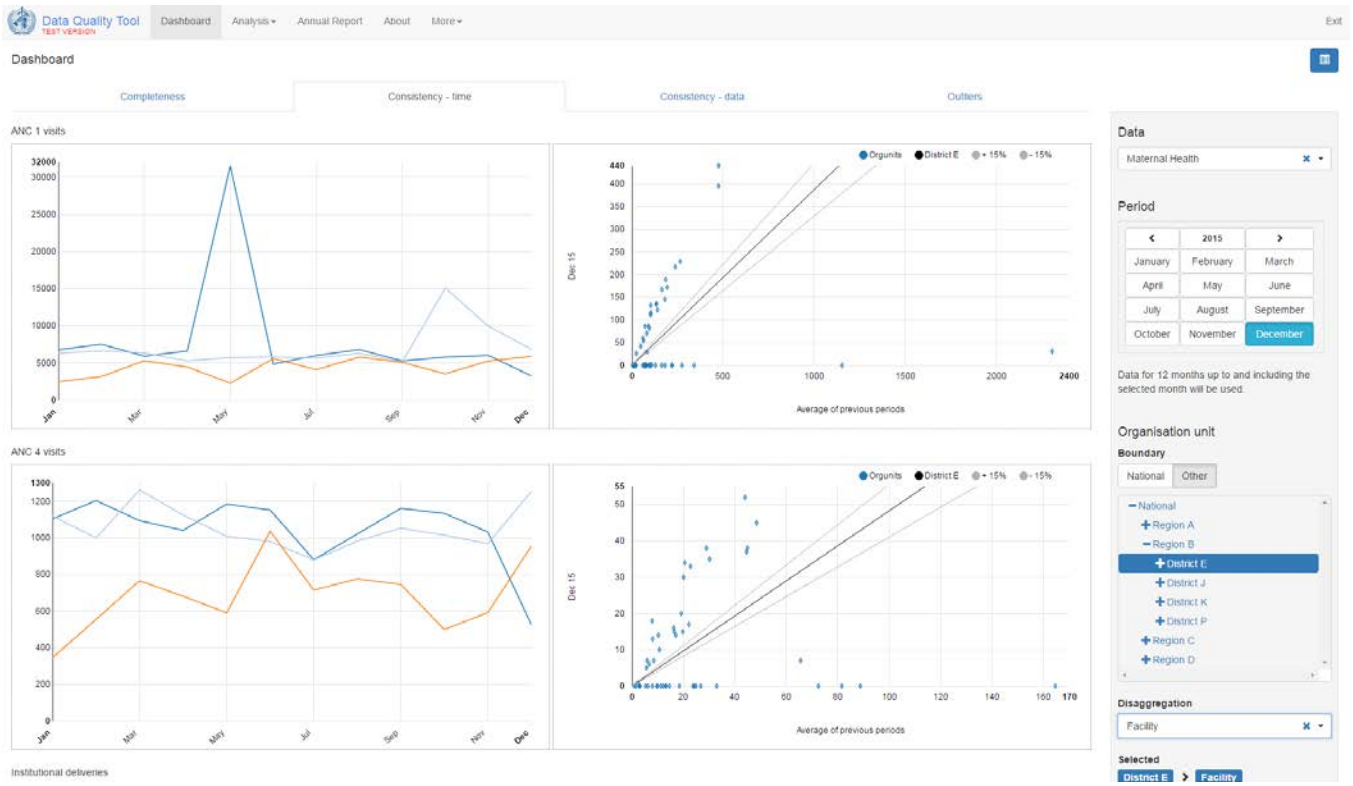


Figure 10

To look more in to detail for District E press “District E” under the headline “Orgunit” to drill down. The chart to the left shows the consistency over time for District E and the chart to the right a scatter plot comparing December 2015 against the 12 preceding months on facility level.

Questions: to the chart on the top left:

- Does the consistency follow any trends and is it comparable to previous years?
- Are there many fluctuations in the consistency through the period?
- Are there specific months which stand out?

Questions: to the chart on the top right:

- Is the consistency equally distributed among the facilities?
- Are there specific facilities which stand out?

7. Analysis of data quality – consistency between indicators

There are many techniques and methods to analyze the quality or search for possible errors in a dataset. One of the more common approaches is for instance to compare the relationship between two variables (using ratio). Further, it is also possible to analyze if this ratio has changed over time.

7.1. Step by step setup of analysis consitency between indicators

Figure 11 Screenshot from DHIS2 quality tool

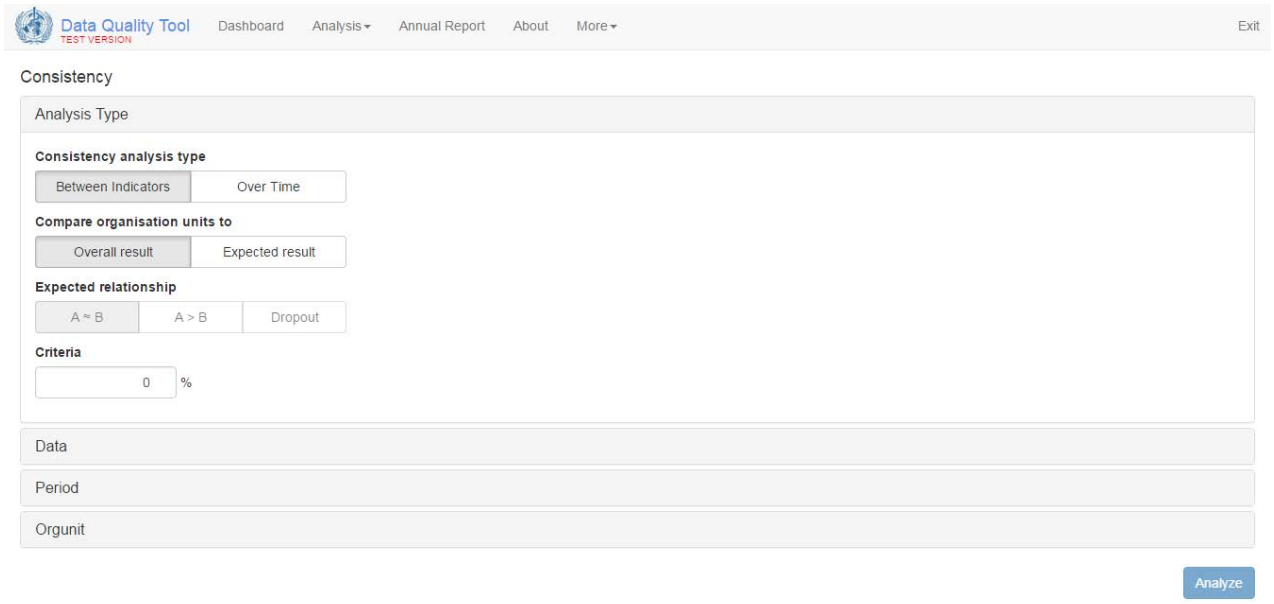
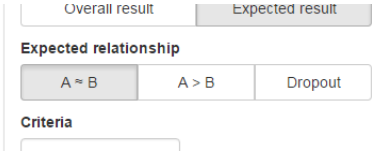


Figure 11 and 12:

1. Press the tab “Analysis” on the “main menu” and choose “consistency” (see figure 11)
2. Under the headline “Consistency” choose “Analysis type” select the button “Between indicators” under the headline “Consistency analysis type” (see figure 11)
3. Under the headline “comparing organization units to”, select the button “Overall result”. “Overall results” means that the “Orgunit” is compared with the national score. “Expected results” means the “Orgunit” is compared to an estimated expected result. When “expected” results” is chosen also choose one of the three Expected relationships (see figure 12)

Figure 12



4. Figure 12 Explanation of “expected results”[4]:
   “A ≈ B”: the two chosen data elements should have almost equal score. This means that the weight is measured against ratio 1. This means that it will not give meaning to run the number of birth and total population together, because we do not expect these values to be equal. The number of birth and delivered by skilled personnel will be a better suit for “Expected results”. The two data elements should be roughly equal in size.
   “A > B”: The data element A should be higher than B
5. Choose “Overall results” for this example.
6. Under the headline “Criteria” the box should be set to “0” (see figure 11)

[4] See further explanations in Figure 60 and 84

Figure 13 Screenshot from DHIS2 quality tool

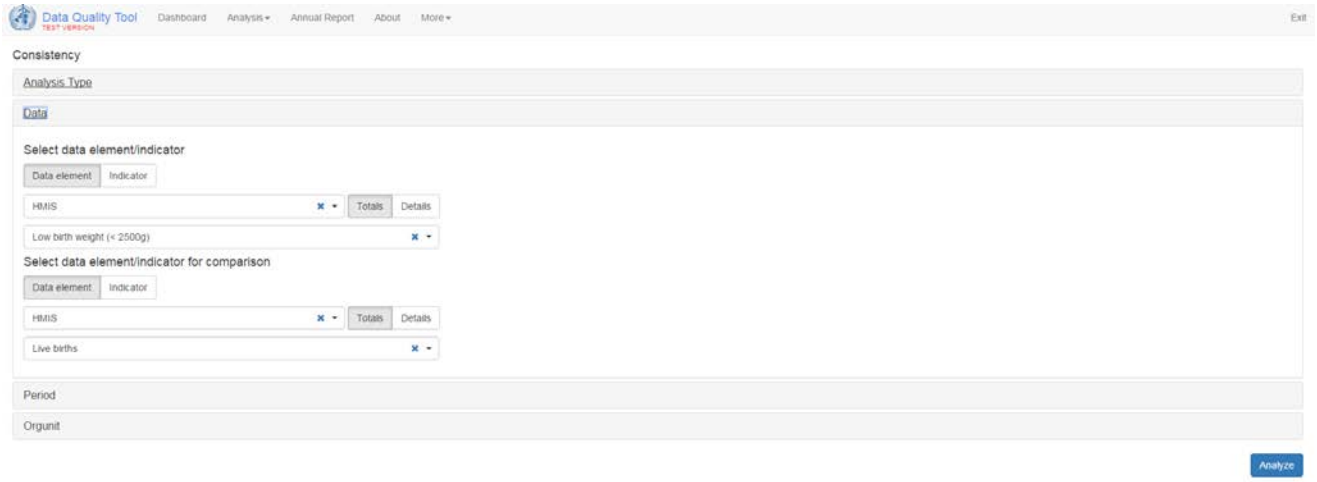


Figure 13:

1. Under the headline “Consistency” choose “Data”
2. Press the button “Data elements” under the headline “Select data element/indicator” (see figure 13)
3. Select data element group “HMIS” and press the button “Totals”
4. Select data element “Low birth weight (< 2500g)”
5. Under the headline “Select data element/indicator for comparison” press “data element”
6. Select data element group “HMIS” and press the button “Totals”
7. Select data element “Live births”

Figure 14 Screenshot from DHIS2 quality tool

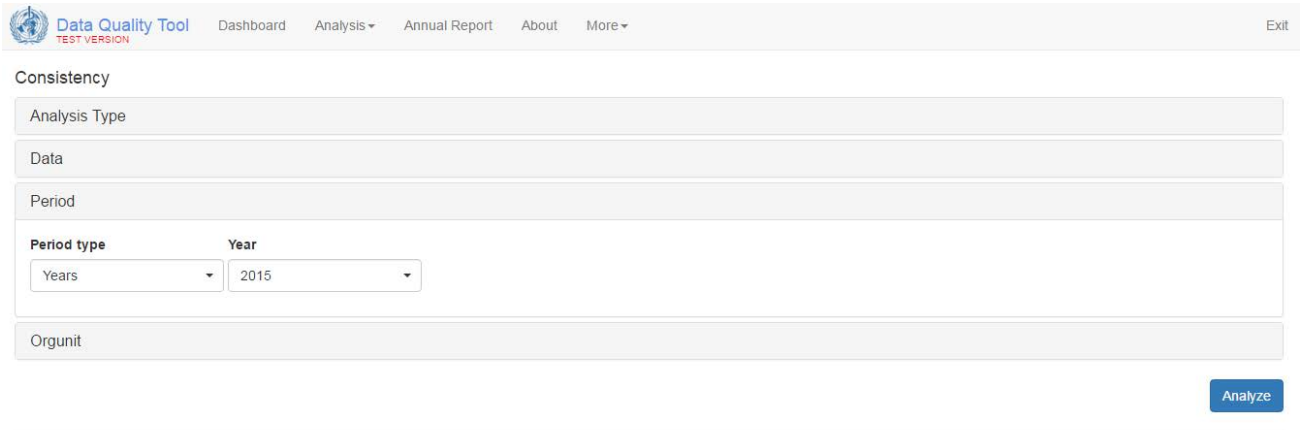


Figure 14 and 15:

1. Under the headline “Consistency” choose “Period”
2. Under the headline “Period type” select “Years” from the drop down menu.
3. Under the headline “Year” choose “2015”
4. If any other period than year is chosen under “Period type” this will have to specified period under “Period” (see figure 15)
5. Select the number of preceding periods. The number of periods which are chosen will depend on the Character trait of the data element.

Figure 15 Screenshot from DHIS2 quality tool

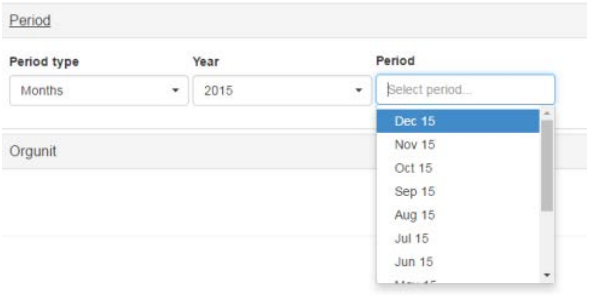


Figure 16 Screenshot from DHIS2 quality tool

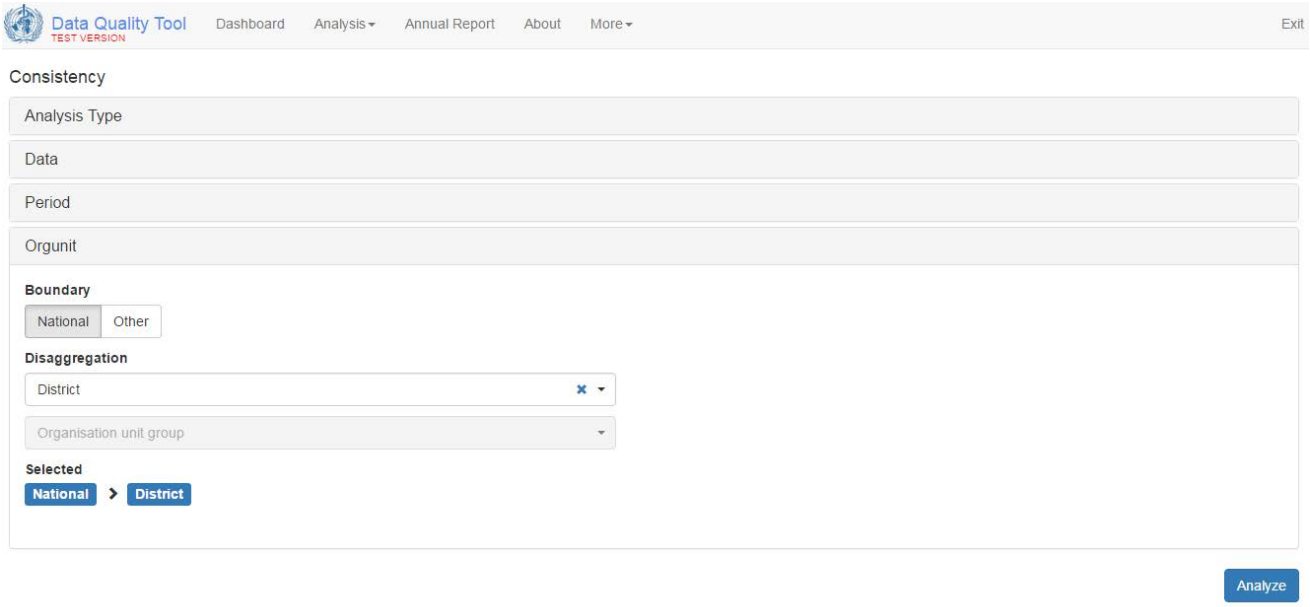
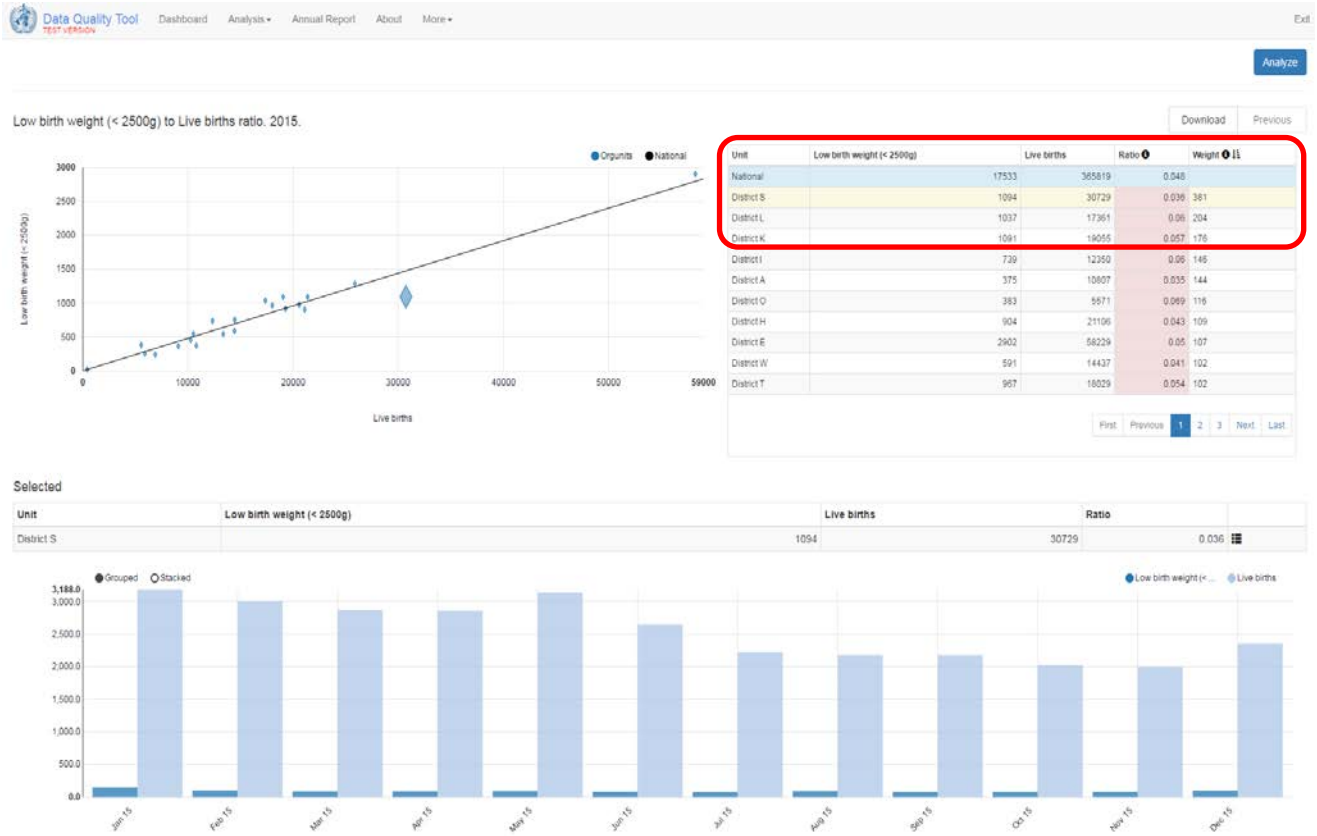


Figure 16:

1. Under the headline “Consistency” choose “Orgunit”
2. Under the headline “Boundary” press the button “national”
3. Under the headline “Disaggregation” choose “District” from the drop down menu
4. Press “Analyze” in the lower right corner when finished. If the analyze-button is grey something is not filled correctly, for instance period or data element.

Figure 17 Screenshot from DHIS2 quality tool



Explanations: Figure 17:
Scatter plot

- The plot, to the left, shows the relation between the two data elements. The black line shows where the relationship is equal to the national level (in this case 0,048). If a dot is above than the black line "Live birth less than 2500g" is higher compared to the national level. Dots close to the black line are similar to the national level.

Table

- **Ratio** is the relationship between the two data elements. The ratio is estimated using the following method: the first data element/the second data element. Usually when assessing the ratios between different data elements there is no "correct" ratio. Unlike what normally is the situation when it comes to ratios based on time series. The ratio at national level for the variables **"Low birth weight (< 2500g)" and "Live births" is: 17 533/ 365 819 = 0,048.**

- **Weight**[5] is an indication of how much figures reported by a health Districts influences national figures (See figure 17 inside red circle). District S has a ratio at (1 094/30 729) 0,036. This is lower than the ratio at national level which is 0,048. Since District S has a lower ratio than the national, the variable "Low birth weight (< 2500g)" will need 381 to reach the national ratio, and this is defined as the weight. The method is as follows: **"Low birth weight (< 2500g)" + Weight / "Live births" = National Ratio**. With actual figures for District S this formula will be: ((1 094+381)/30 729= 0,048). If the ratio is over the national ratio the weight needs to be deducted from the first data element. The weight does not necessary pin point which data element that is erroneous, but it will give an indication of which data elements that should be looked more into. The table (to the right in figure 17 and 18) is originally sorted by weight. This means that a potential erroneous data element with the highest influence on the national level is placed first.

Figure 18 Screenshot from DHIS2 quality tool

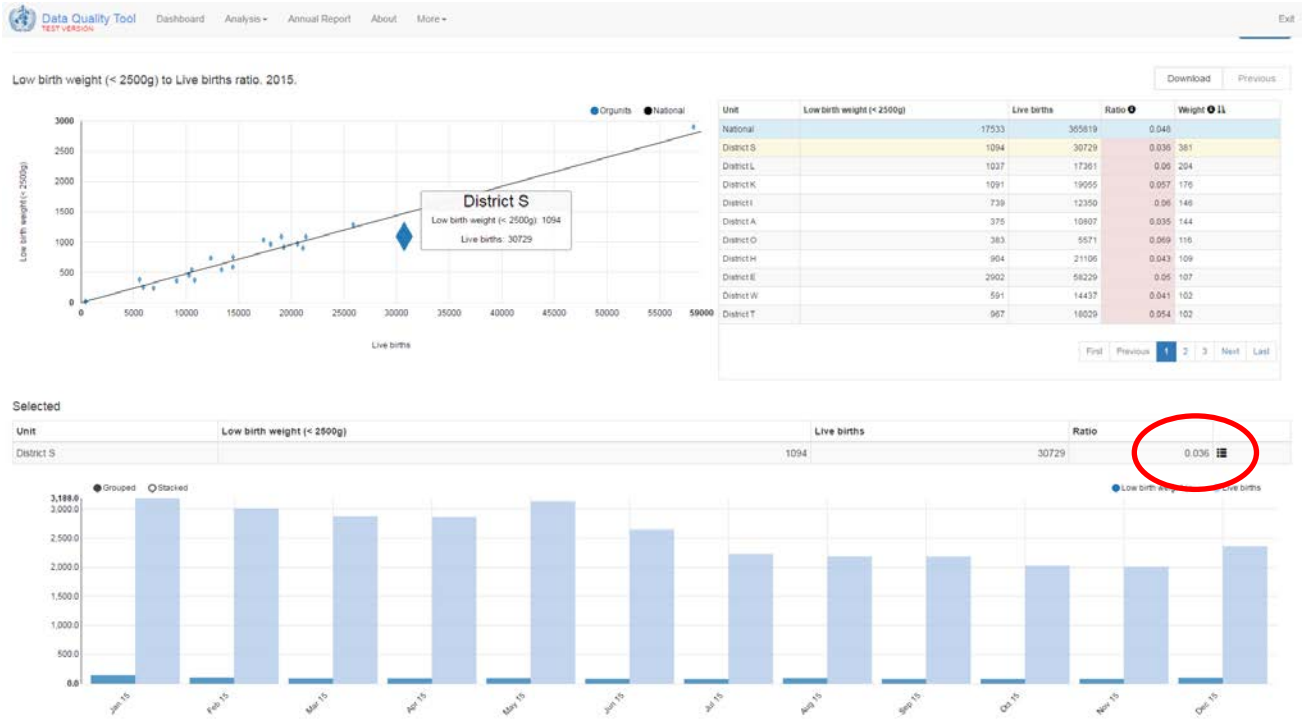


Figure 18:
Manoeuvring in the scatter plot and table

- Additional information will pop up by hovering the cursor over the plot (See figure 18 and box in the scatter plot)

[5] In general Weigh is here an indication of influence. In this manual more than one method estimating weight will be used.

- By clicking on a dot in the scatter plot or in the table the "Orgunit" will be highlighted both in the table and in the scatter plot. A bar chart will also appear below the scatter plot. The chart shows the relationship between the data elements for the last twelve months (depending on choice in period menu).
- It is also possible to click on a district in the table and it will be highlighted in the plotter chart
- Press the black bars on the right just above the bar chart (see Figure 18 and inside red circle) and choose "drill down" from the drop down menu (see results in Figure 19).
- "Drill down" means to go downwards in the data hierarchy, this can for instance be from National level, to District level and further to facility level. The purpose of this is to identify the unit causing a data value in need for further investigation.

Questions related to figure 18:

- Are the districts in general close to the national level?
- Are there any districts that should be looked more into?
- District O has the highest ratio, but not the highest weight. Why?
- When looking at District S and the bars at the bottom, is there any pattern? And could it have a logical explanation?
- How can you investigate District S at facility level?

Figure 19 Screenshot from DHIS2 quality tool

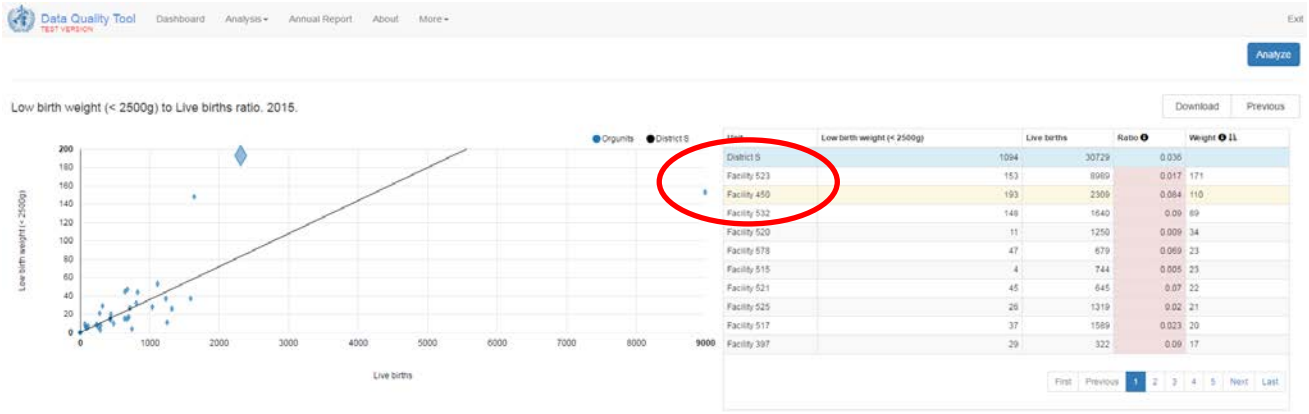
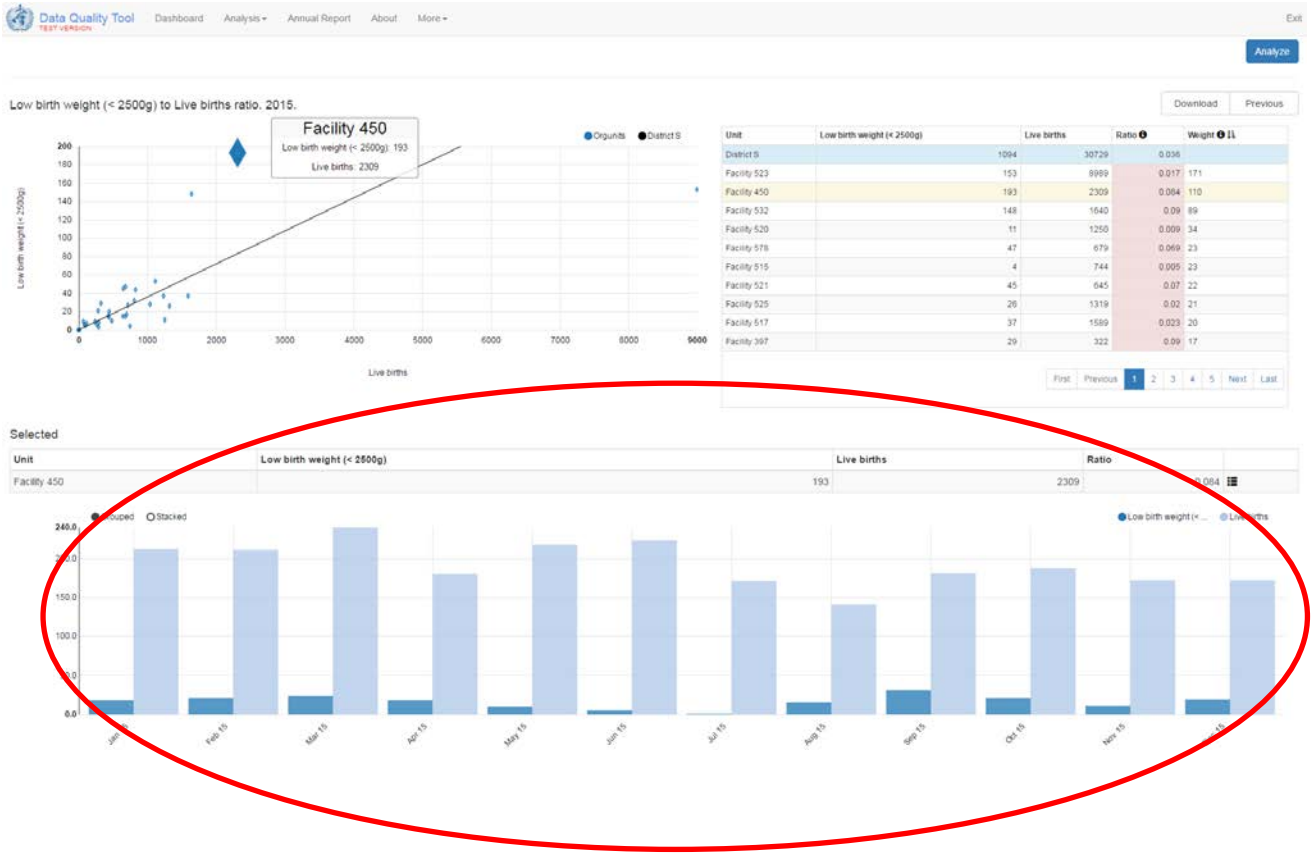


Figure 19 and 20:
Drill down:

- Drilling down in District S will expose the facilities which this district consists of (see figure 19). The units in the scatter plot are now facilities.
- Press "facility 450" in table to the right, to see more details on this facility (see red circle in figure 19)
- By pressing "facility 450" it is both highlighted in the table and in the scatterplot (see figure 19).
- By pressing "facility 450" a new diagram is also produced (see figure 20 and red circle). The Diagram shows details for the chosen facility. Here it is reports for the last 12 months'.
- To get back to the previous view press "Previous" top right

Figure 20 Screenshot from DHIS2 quality tool



Questions related to figure 19 and 20:

- What is the minimum number of facilities that should be investigated further?

8. Analysis of data quality – consistency over time

Comparing data over time is probably the most common way to look for errors in a dataset. In many cases the figures reported one year or month should be somewhat similar to the previous year and month, but of course dramatic changes can be explained by local contexts and actual incidents. The reason for these changes should be documented. If there is no logical reason for a marked increase or decrease from one month to the other, it can be advisable to look more in to the issue and look for the potential reason. For more details on how to use the "consistency over time" approach see explanation of Figure 27.

8.1. Step by step setup of consistency over time analysis

Figure 21 Screenshot from DHIS2 quality tool

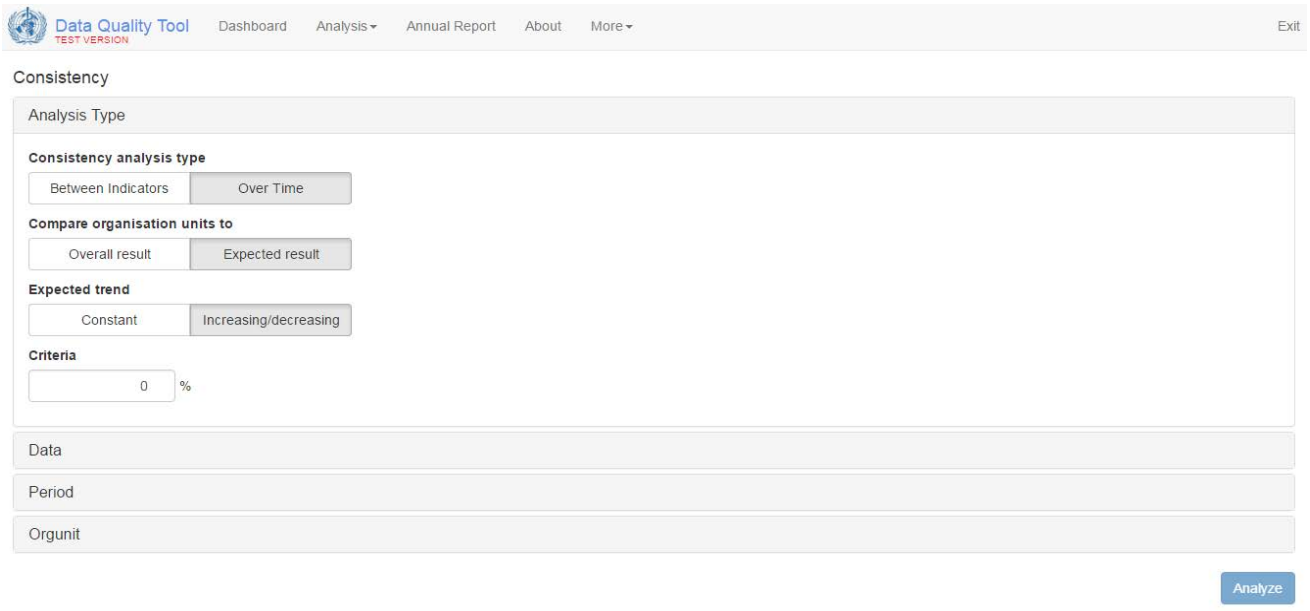


Figure 21:

1. Press the tab “Analysis” in the “main menu” and choose “consistency” (see Figure 21)
2. Under the headline “Consistency” choose “Analysis type” select the button “Over time” under the headline “Consistency analysis type” (see Figure 21)
3. Under the headline “Compare organisation units to” select the button “Expected result”
4. Under the headline “Expected trend” select the button “Increasing/decreasing”.
5. Depending on what you expect, you can choose “constant” or “increasing/decreasing”. The nature of the data element will give an indication on what to choose. The number of births is likely to be relatively constant. If you were to compare December with an average for the preceding 4 months you would expect the results to be fairly similar. Malaria on the other hand is seasonal, so an average for the preceding months would probably give a result that either is to low or too high depending on the month of choice.
6. Under the headline “Criteria” the value in the box should be set to “0” (see figure 21)

Figure 22 Screenshot from DHIS2 quality tool

Figure 22:

1. Under the headline “Consistency” choose “Data”
2. Under the headline “select data element/indicator” press the button “Data element” (see figure 22)
3. Select data element group “HMIS” and press the button “Details”
4. Select data element “Malaria < 5 years (HMIS)”

Figure 23 Screenshot from DHIS2 quality tool

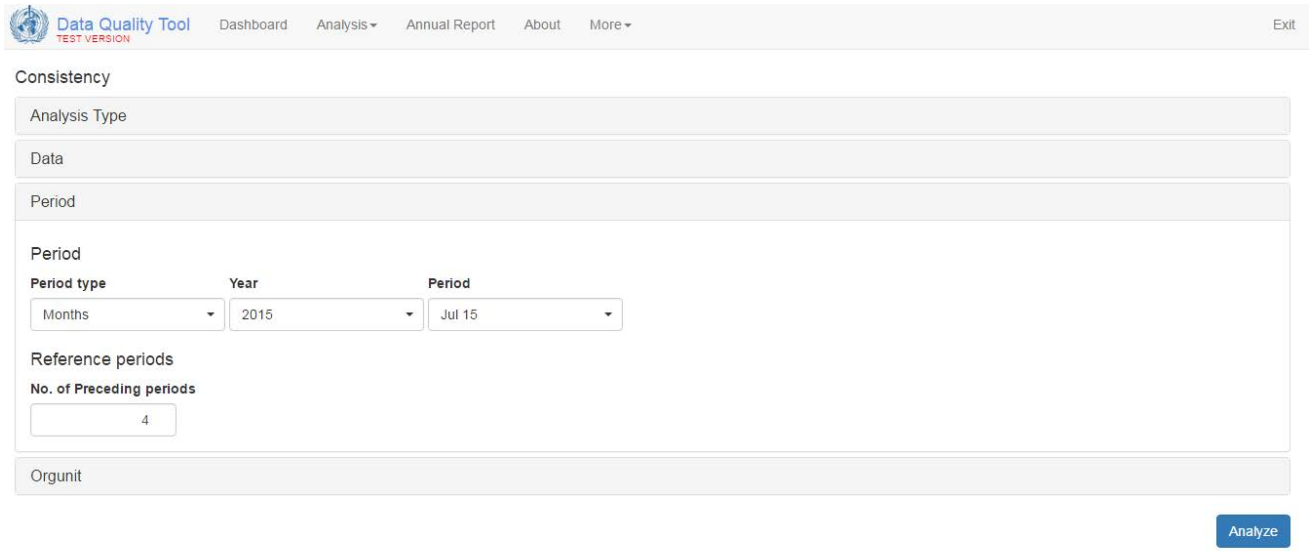


Figure 23:

1. Under the headline “Consistency” choose “Period”
2. Under the headline “Period type” select “Months” from the drop down menu.
3. Under the headline “Year” choose 2015. If anything else than “year”, for instance “months”, is chosen under “Period type” you have to specify period under “Period”.
4. Under the headline “Period” choose “Jul 15”
5. Under the headline “Reference periods” set “No. of preceding periods” to “4”. The number periods chosen should depend on the character trait of the data element.

Figure 24 Screenshot from DHIS2 quality tool

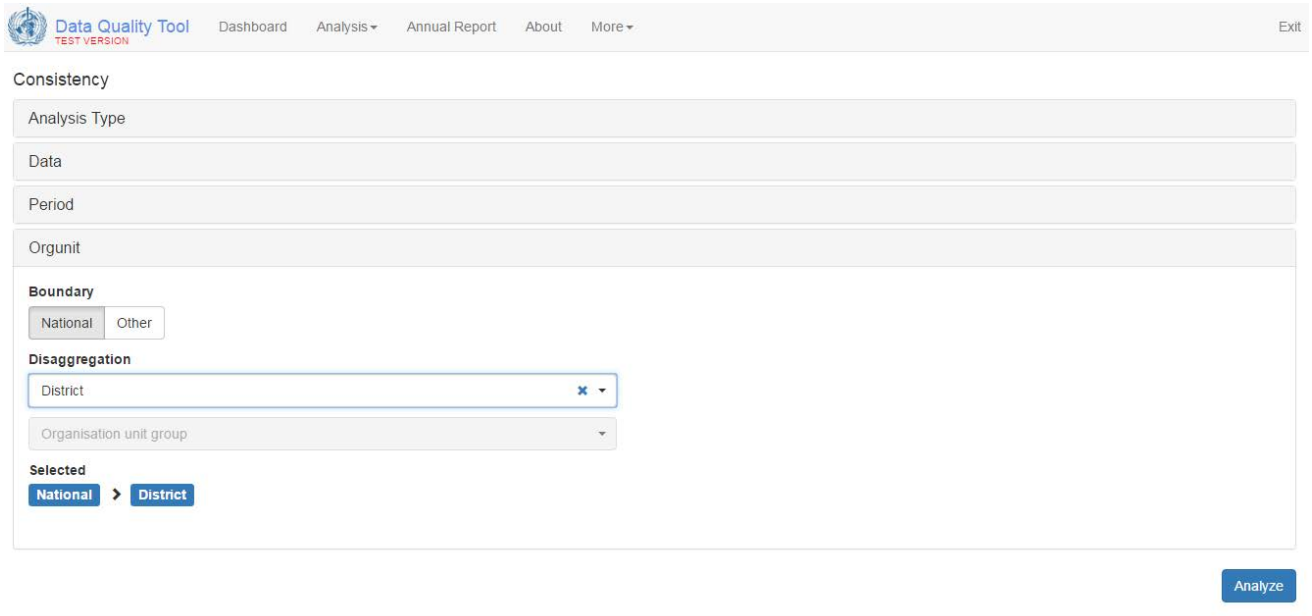
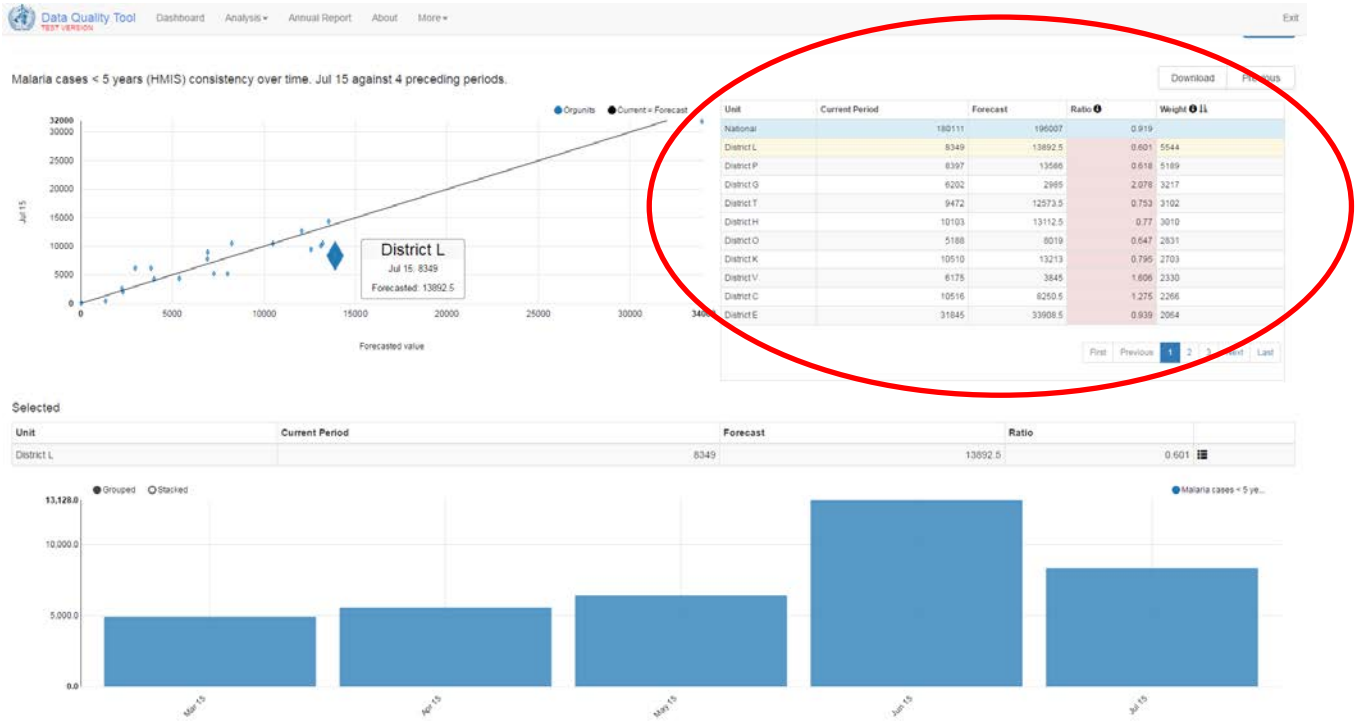


Figure 24:

1. Under the headline “Consistency” choose “Orgunit”
2. Under the headline “Boundary” press the button “national”
3. Under the headline “Disaggregation” choose “District” from the drop down menu
4. Press “Analyze” when finished. If the analyze-button is grey something is not filled correctly, for instance period or data element.

Figure 25 Screenshot from DHIS2 quality tool



Explanation Figure 25:
Scatter plot

- The plot shows the relation between the current period and the predicted forecast. The black line shows where the relationship is equal to 1. If the dot is higher than the black line the current period is higher than the prediction. The further away from the black line the worse the prediction between the current period and the forecast are.

Table (see Figure 25 on page 23 inside red circle))

- **Current period** shows the number of new malaria cases for those who are < 5 years and for July 2015.
- **Forecast** is the prediction for July 2015 based on the 4 preceding periods.
- **Ratio** is the relationship between current period and the forecast. The ratio is calculated as (current period/forecast). If the ratio is 1 there is no difference between the current period and forecast. The closer to 1 the better, provided that the data are of good quality on both sides of the fraction line.
- **Weight** is an indication of influence of one district on the national result. District L has 8 349 cases in July, but the forecast was 13 892. This gives a ratio (8 349/ 13 892.5) at 0,601 and a weight at 5544. When the ratio is below 1 the weight is the figure that has to be added to the current period to achieve ratio 1 ((8 349+5 544/13892.5) = 1). If the ratio is over 1 the weight should be deducted from current period. The table is originally sorted by weight. This means that the potential error with the most influence on the national level is placed first. The most important factor in the sorting is the size of the "Orgunit" and distance from ratio 1.

Manoeuvring in plot and table

- Additional information will pop up by hovering the cursor over the plot (See figure 25 and box in the scatter plot)
- By clicking on a dot in the scatter plot or in the table the "Orgunit" will be highlighted both in the table and in the scatter plot.
- A bar chart will also appear below the scatter plot. The chart shows the preceding periods and current period for the "Orgunit", and gives an indication on why the prediction of the forecast was good or bad.

Figure 26 Screenshot from DHIS2 quality tool

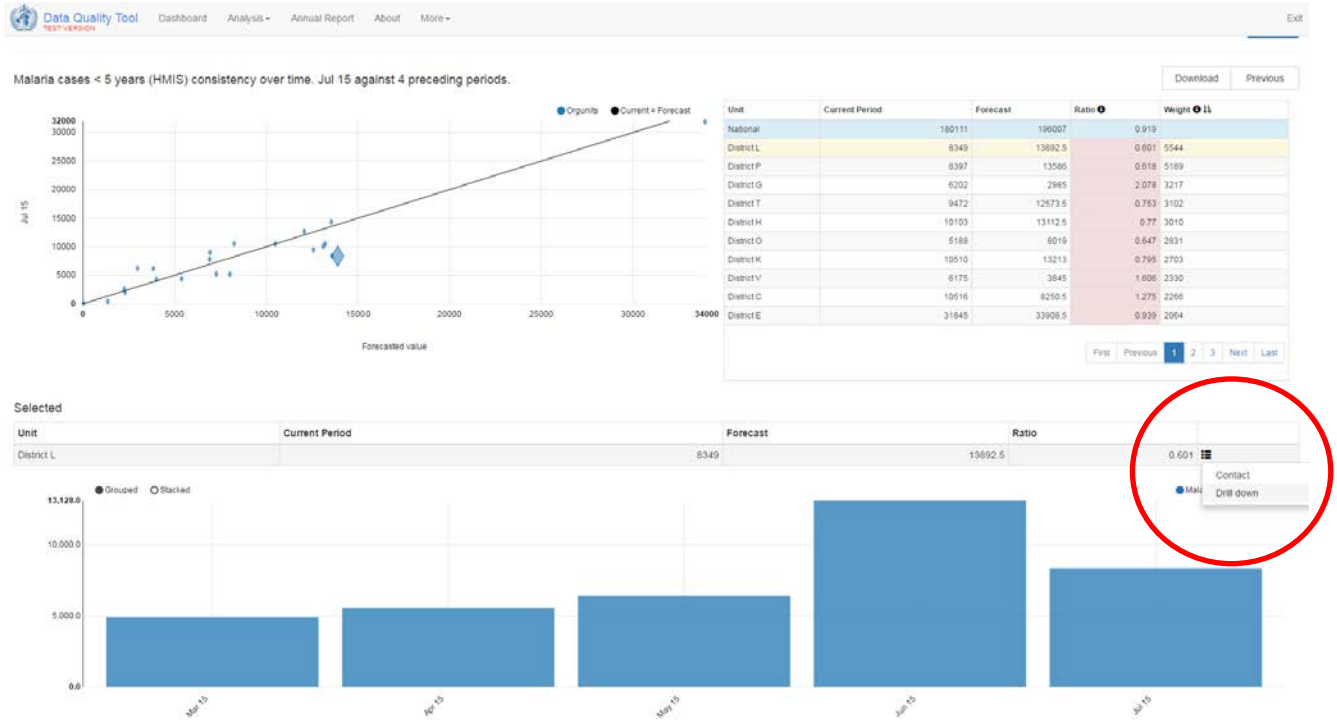


Figure 26:

- Press the black bars on the right just above the bar chart (see red circle in figure 26) and choose “drill down” from the drop down menu and observe results in Figure 27.

Questions related to figure 26:

- Does the forecast at national level come close to the current period?
- Are there any districts that should be investigated further?
- District G (on the first page of the table) has the highest ratio, but not the highest weight. Why?
- Looking at District L and the bars at the bottom, which month looks most suspicious? Could it have a logical explanation?
- How can you investigate District L at facility level?

Figure 27 Screenshot from DHIS2 quality tool

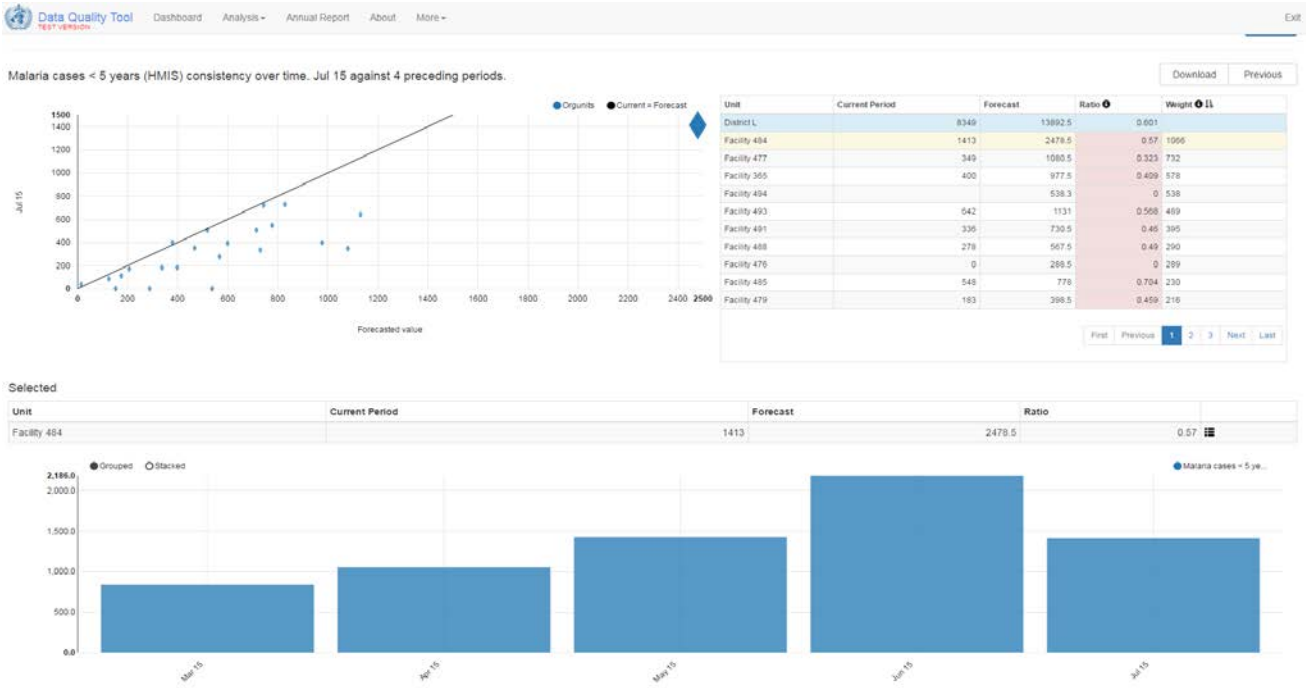


Figure 27:

- The scatter plot and table shows the facilities for District L.
- To get back to the previous view press “Previous” top right.

Questions related to figure 27:

- Why are almost all the dots in the scatter plot under the black line?
- Which facility is potential most problematic?

Additional question:

- In figure 21 we chose “increasing/decreasing” for expected trend under “analyze type”. Change it to “constant”. Is the prediction better at national level better?
- Can you explain the difference in ratio between the “increasing/decreasing” and “constant” trend for District E and L?

9. Analysis of data quality – outliers and missing data

This feature should only be used on data with sufficient data quality and when data is not affected by seasonal changes (e.g. data on malaria). If the data vary significantly between periods, the DHIS2 quality tools ability to efficiently identify outliers is vastly reduced. Additionally, the previous periods should already be checked and verified, otherwise the ability to identify outliers in the final period will be reduced.

9.1. Step by step setup of outliers and missing data analysis

Figure 28 Screenshot from DHIS2 quality tool

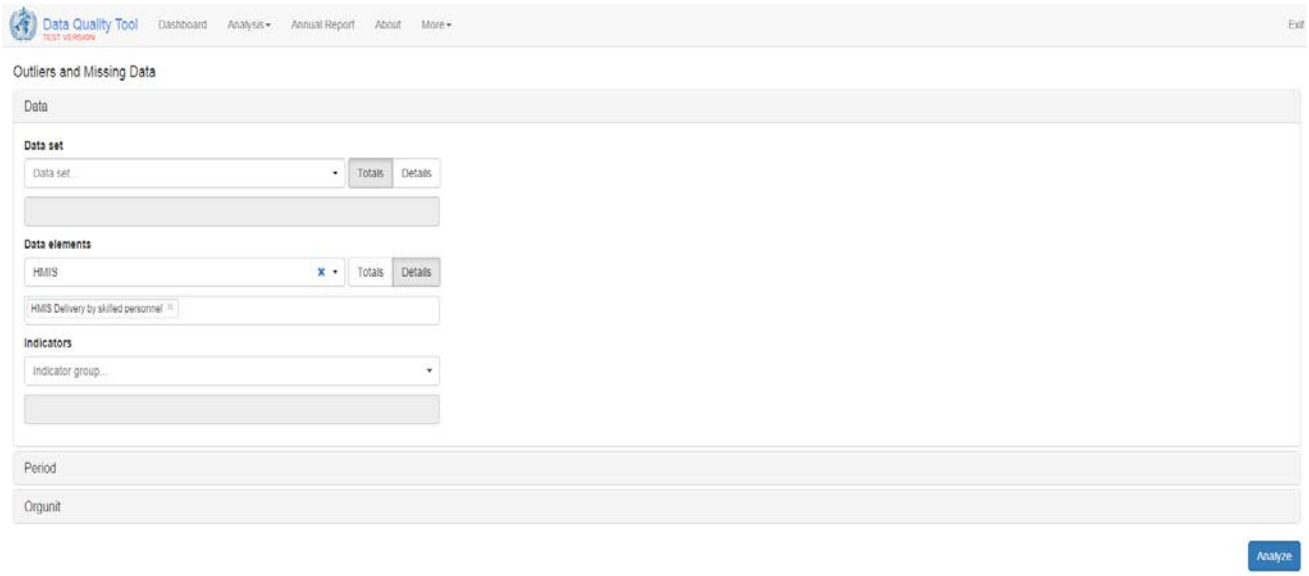


Figure 28

1. Press the tab “Analysis” in the “main menu” and choose “outliers and missing data” (see Figure 28)
2. Under the headline “Outliers and missing data” choose “Data” and select data element group “HMIS” under the headline “Data elements” and then select data element “HMIS Delivery by skilled personnel” underneath. There are three headlines: “Data set”, “Data elements” and “Indicators”, but it is only possible to choose one at a time.

Figure 29 Screenshot from DHIS2 quality tool

Figure 29

1. Under the headline “Outliers and missing data” choose “Period”
2. Under the headline “Period” Press the button “By year”
3. Underneath select “months” and “2015” from the drop down menu.

Figure 30 Screenshot from DHIS2 quality tool

Figure 30:

1. Under the headline “Outliers and missing data” choose “Orgunit”
2. Under the headline “Boundary” press the button “national”
3. Under the headline “Disaggregation” choose “District” from the drop down menu
4. Press “Analyze” when finished. If the analyze-button is grey something is not filled correctly, for instance period or data element.

Figure 31 Screenshot from DHIS2 quality tool

Explanations Figure 31:

- As mention in the introduction to chapter 9 the main focus should be on the last month. Values for previous months should already be checked and approved.
- A value with red background is an outlier and missing is marked with a yellow background (see figure 31, inside red circle).
- The column “Total” indicates the outliers impact, higher figure means higher impact on aggregated figures.
- An example using District E in Region B the value in the (see Figure 31): The “Total” column is an estimation of the figure that should be added to reach an estimated “correct” value. For District E in Region B the “correct” value for December is estimated to be 4 963 (2 185 + 2 788 = Value for Dec 15 + Total). The same value in column “Total” for District V in Region C is 201. The table is by default sorted by the Total column, high to low. Therefore, it provides a prioritized list of which District that should be checked first.
- The estimation becomes difficult when there is more than one outlier in the period, especially if the “errors” goes both ways. If one value is +1000 and the other is -1000 then the value in column “Total” will be zero. So this column should be used with caution.

Figure 32 Screenshot from DHIS2 quality tool

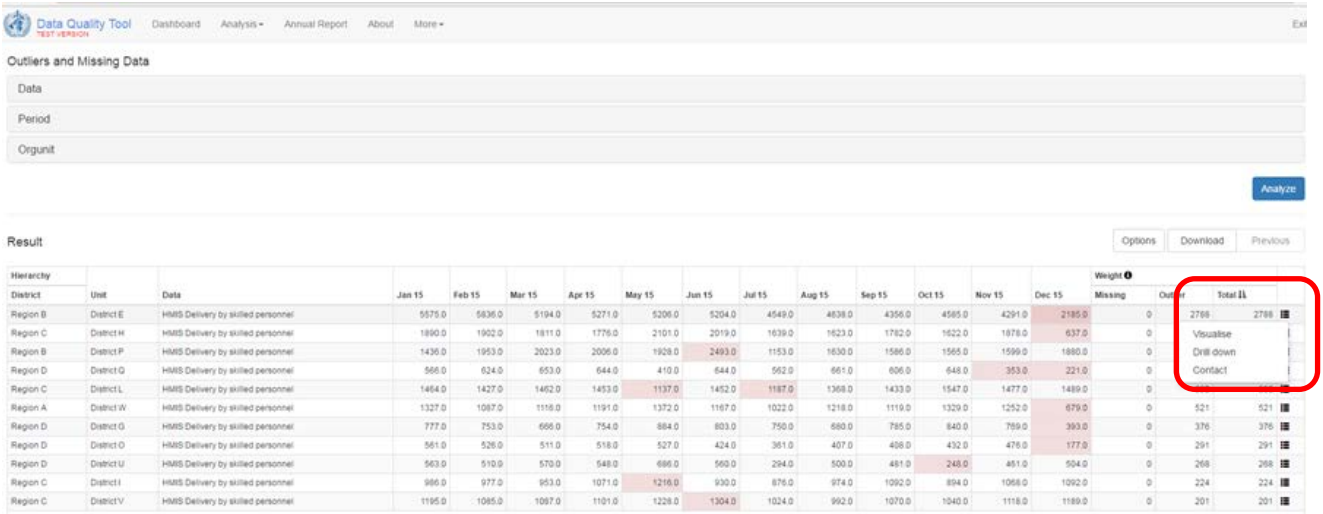


Figure 32:

1. Press the black symbol on the far right (See Figure 32, inside red circle)
2. By choosing “Visualize” from the drop down menu, a bar chart will of district E will be displayed. See Figure 33 for results.
3. By choosing “Drill down”, the facilities which District E consist of will be displayed. See figure 34 for results.

Figure 33 Screenshot from DHIS2 quality tool

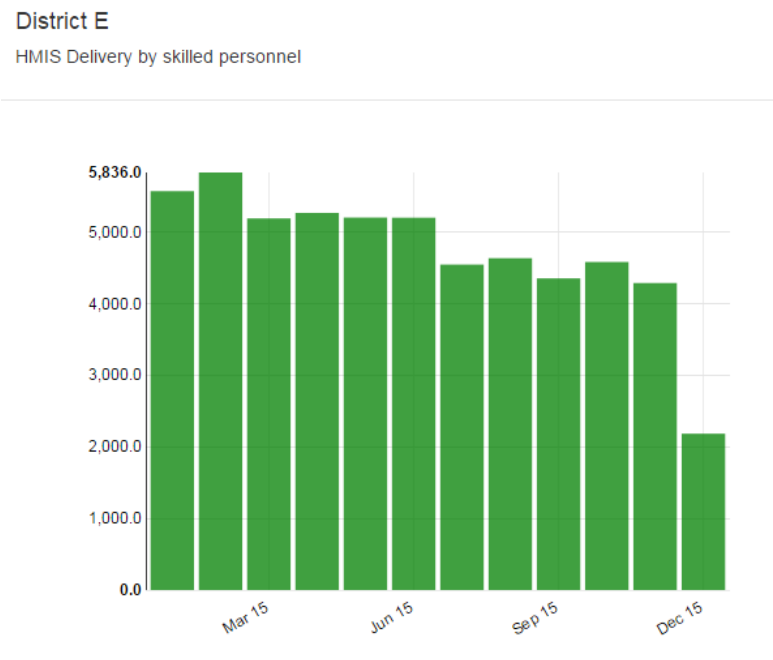


Figure 34 Screenshot from DHIS2 quality tool

Hierarchy																				
District	Facility	Unit	Data	Jan 15	Feb 15	Mar 15	Apr 15	May 15	Jun 15	Jul 15	Aug 15	Sep 15	Oct 15	Nov 15	Dec 15	Weight	Missing	Outlier	Total
Region B	District E	Facility 222	HMIS Delivery by skilled personnel	1283.0	1383.0	1003.0	1243.0	1161.0	1110.0	1080.0	1086.0	1006.0	1122.0	1293.0			1122	0	1122
Region B	District E	Facility 561	HMIS Delivery by skilled personnel	430.0	365.0	392.0	327.0	336.0	518.0	273.0		307.0	323.0	350.0			686	173	859
Region B	District E	Facility 223	HMIS Delivery by skilled personnel	269.0	296.0	259.0	282.0	300.0	286.0	229.0	261.0		260.0	270.0			551	49	600
Region B	District E	Facility 563	HMIS Delivery by skilled personnel	264.0	282.0	357.0	327.0	322.0	260.0	208.0	212.0	215.0	246.0				524	0	524
Region B	District E	Facility 569	HMIS Delivery by skilled personnel	115.0	137.0	113.0	127.0	149.0	156.0	145.0	154.0	115.0					411	0	411
Region B	District E	Facility 449	HMIS Delivery by skilled personnel	88.0	97.0	103.0	110.0	81.0			79.0	63.0	75.0				342	0	342
Region B	District E	Facility 590	HMIS Delivery by skilled personnel	327.0	237.0	238.0	207.0	276.0	209.0	192.0	240.0	217.0	234.0	182.0			234	104	338
Region B	District E	Facility 609	HMIS Delivery by skilled personnel	53.0	44.0	33.0	15.0	30.0	27.0	97.0	53.0						154	61	215
Region B	District E	Facility 575	HMIS Delivery by skilled personnel	46.0	250.0	74.0	104.0	24.0	82.0	68.0	78.0	86.0	62.0	75.0	75.0		0	179	179
Region B	District E	Facility 221	HMIS Delivery by skilled personnel	16.0	18.0	21.0	23.0					17.0					126	0	126
Region B	District E	Facility 581	HMIS Delivery by skilled personnel	25.0	31.0	24.0	25.0	17.0	32.0	11.0							125	0	125
Region B	District E	Facility 566	HMIS Delivery by skilled personnel	46.0	57.0	9.0	14.0	3.0	12.0	10.0	14.0		0.0		0.0		22	88	110
Region B	District E	Facility 593	HMIS Delivery by skilled personnel	66.0	79.0	104.0	87.0	72.0	81.0	63.0	52.0	83.0	62.0	69.0			72	33	105
Region B	District E	Facility 567	HMIS Delivery by skilled personnel	95.0	108.0	84.0	101.0	119.0	136.0	81.0	118.0	103.0	82.0		93.0		101	0	101
Region B	District E	Facility 604	HMIS Delivery by skilled personnel		0.0	0.0	0.0	0.0		0.0	0.0				97.0		0	97	97

Explanation: Figure 34:

- From the table in Figure 34 the reason for the outlier in District E becomes clear. Several facilities have not reported any activity for December, and this result in a low figure.
- Press the tab “Previous” to go back to district level.

10. The Annual Report module

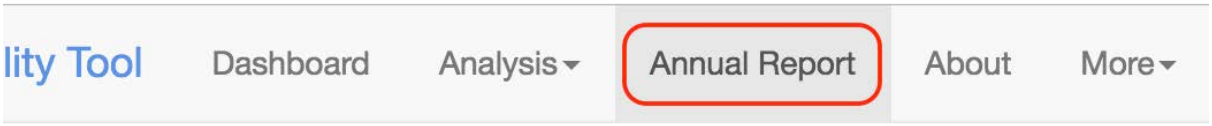
The Annual Report module provides an overview of the key data quality metrics for one year, looking at data quality in four domains: completeness, internal consistency, external comparison, and consistency of population data (denominators). Some of the metrics of the Annual Report may only be available at the national level, such as the denominator consistency and external comparison checks; however, the report can still be useful when run at e.g. province or district level.

The report is based on a selection of key indicators, either a set of cross-cutting indicators from different areas, or a more focused set of indicators for e.g. one disease Programme. The groups of indicators are based on the "numerator groups" that are described in the section on "Administration". Other parameters are also based on the configuration done in the "Administration" section, such as thresholds for completeness, outliers, consistency over time etc.

10.1. Accessing the Annual Report

The Annual Report module is access directly from the main menu of the tool.

Figure 35 Screenshot from DHIS2 quality tool



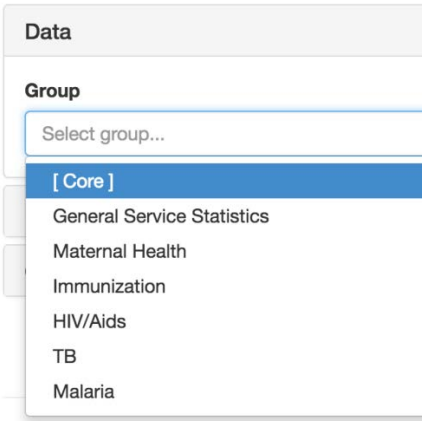
10.2. Running the Annual Report

Three selections are required for running the Annual Report: Data, period and organisation unit.

Figure 36 Screenshot from DHIS2 quality tool

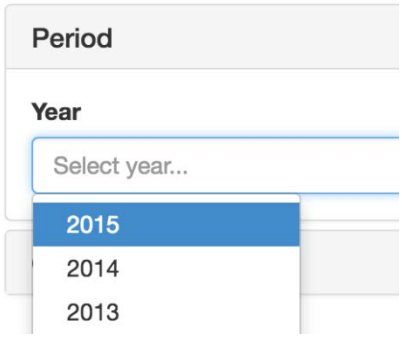
In the "Data" section, users can choose between different groups of indicators (based on "Numerator groups"), including defined set of core indicators.

Figure 37 Screenshot from DHIS2 quality tool



In the "Period" section, users choose the year to run the report for. Note that some of the quality checks look at data for the selected year compared to the three previous years, and they might not work properly when selecting a year for which there is not data in previous periods.

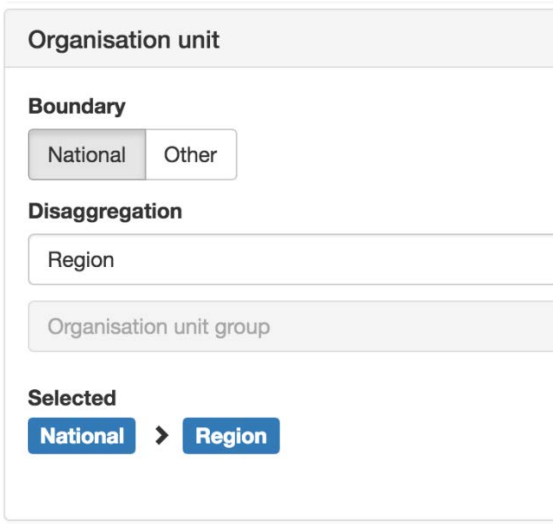
Figure 38 Screenshot from DHIS2 quality tool



In the “Organisation unit” section, users choose the “boundary” for the report, and disaggregation either by level or organisation unit group. By default, the boundary is set to the organisation unit the user is assigned to, and the disaggregation is set to the level below that organisation unit.

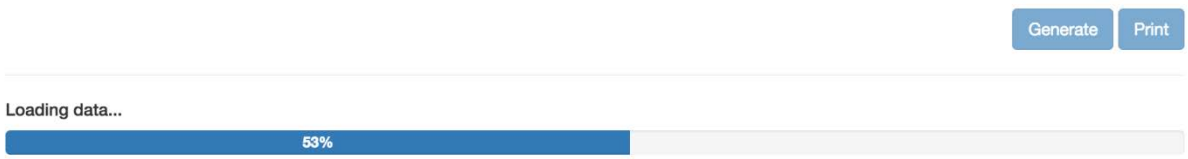
Note: throughout the rest of this chapter, the terms *boundary* and *subunits* will be used as generic terms for the selected organisation units. In the screenshots, you will see “National” as the name of the boundary, and “Region” as the name for the disaggregation.

Figure 39 Screenshot from DHIS2 quality tool



Once “Data”, “Period” and “Organisation unit” has been selected, the Annual Report can be produced by clicking the “Generate” button. A progress indicator will be displayed. Note that the process can take anything from a few seconds to 5-10 minutes depending on the selections that have been made, number of organisation units to be checked, the internet connection etc. During this time, the progress indicator might stop at a certain percentage for some time before continuing.

Figure 40 Screenshot from DHIS2 quality tool



10.3. The Annual Report Results

The result of the annual report is divided into four main sections, each with a number of sub-sections:

1) Completeness of reporting
2) Internal consistency of reported data
3) External comparison
4) Consistency of population data

These four main sections are preceded by a Summary section, which includes the key outputs from each main section.

Each of the four main sections will be described below. The summary section will not be described separately, as it has only the same content as the main section with details left out.

10.4. Completeness of reporting

1a – completeness of facility reporting. This table shows the percentage of expected data sets (reports, forms) that have been entered and completed.

Figure 41 Screenshot from DHIS2 quality tool

1a: Completeness of facility reporting

The percentage of expected reports that have been entered and completed.

Dataset	Quality threshold	Overall score	Region with divergent score		
			Number	Percent	Name
ANC Reporting rate	90%	90.1%	2	50%	Region B, Region D
EPI Reporting rate	90%	40.9%	4	100%	Region A, Region B, Region C, Region D
HMIS Reporting rate	90%	94.7%	0	0%	
Malaria Reporting rate	90%	73.5%	4	100%	Region A, Region B, Region C, Region D

Figure 42 Explanation of variables for annual report (see figure 41)

Column	Description
Dataset	The dataset (report, form) that is checked for completeness. Data sets are included based on being the source for the selected indicators.
Quality threshold	The minimum completeness that is expected.
Overall score	The overall completeness for the selected boundary organisation unit.
Subunits with divergent score	- Number: number of subunits below the threshold - Percent: percent of subunits below the threshold - Name: name of subunits below the threshold

1b – timeliness of facility reporting. This table shows the percentage of expected data sets (reports, forms) that have been entered and completed *on time*. This metric is only included when the tool is used with DHIS version 2.23 and later. The criteria for on time is specified in DHIS2 per dataset, and may vary accordingly. It is by default 15 days after the end of the reporting period. Refer to 1b for a detailed explanation of the table, with the only difference being that 1b refers to timeliness rather than completeness.

1c – completeness of indicator data. This table shows the completeness of the indicator (numerator). In cases where the numerator can be disaggregated, one of the disaggregation's is used as a proxy for the variable completeness. If zeros are not stored, zeros are counted as missing.

Figure 43 Screenshot from DHIS2 quality tool

1c: Completeness of indicator data

Reports where values are not missing. If zeros are not stored, zeros are counted as missing.

Indicator	Quality threshold	Values		Overall score	Region with divergent score		
		Expected	Actual		Number	Percent	Name
ANC 1 or more visits default	90%	5724	5091	88.9%	2	50%	Region B, Region D
DPT-HepB-Hib 3 static Age < 1 year default	90%	6252	2575	41.2%	4	100%	Region A, Region B, Region C, Region D
Malaria cases confirmed by RDT 5+ years default	90%	6108	4552	74.5%	4	100%	Region A, Region B, Region C, Region D
PLHIV receiving ART default	90%	6120	5497	89.8%	2	50%	Region B, Region D
TB cases notified default	90%	6120	5363	87.6%	3	75%	Region A, Region C, Region D

Figure 44 Explanation of columns for annual report (see figure 43)

Column	Description
Indicator	The variable that is checked for completeness.
Quality threshold	The minimum completeness that is expected.
Values	- Expected: the number of values expected to be reported. This is calculated based on the data set (form) the variable is reported through, i.e. reports per year multiplied with the number of facilities expected to report - Actual: the number of values that have been reported.
Overall score	The percentage of actual values out of the expected values, i.e. the completeness.
Subunits with divergent score	- Number: number of subunits below the threshold - Percent: percent of subunits below the threshold - Name: name of subunits below the threshold

1d – consistency of dataset completeness over time. This table and accompanying chart shows the consistency of reporting completeness over time, i.e. between the selected year and the three preceding years.

Figure 45 Screenshot from DHIS2 quality tool

1d: Consistency of dataset completeness over time

Completeness of datasets in 2015 compared to previous three years.

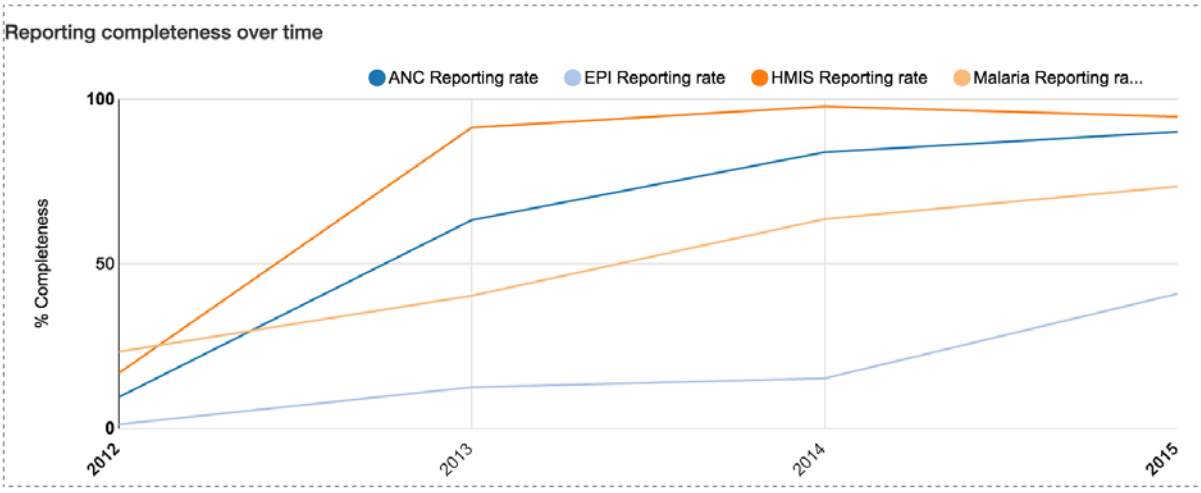
Data set	Expected trend	Compare Region to	Quality threshold	Overall score	Region with divergent score		
					Number	Percent	Name
ANC Reporting rate	Constant	National	± 33%	172.5%	0	0%	
EPI Reporting rate	Increasing	Current vs forecast	± 33%	173.1%	3	75%	Region A, Region B, Region C
HMIS Reporting rate	Constant	National	± 33%	137.9%	0	0%	
Malaria Reporting rate	Constant	National	± 33%	173.3%	1	25%	Region C

Figure 46 Explanation of columns for annual report (see figure 44)

Column	Description
Data set	The variable that is checked for consistency.
Expected trend	Whether completeness is expected to be constant or increase/decrease. If the expected trend is constant, the completeness of the selected year is compared with the *average* of the three previous years. If the expected trend is increasing or decreasing, the completeness of the selected year is compared with a forecast based on the trend in the three previous years.
Compare subunits to	Whether the consistency of reporting for subunits is compared to the boundary (e.g. national) consistency ratio, or to the expected consistency between current and previous periods.
Quality threshold	The consistency (+/-) of completeness that is expected.
Overall score	The consistency between current year and the average or forecast of the three previous years, i.e. value for the current year divided by the average or forecast for 3 previous years, expressed as a percentage.
Subunits with divergent score	Number: number of subunits below the threshold Percent: percent of subunits below the threshold Name: name of subunits below the threshold

The chart shows the trend of reporting completeness for the selected year and the 3 previous ones.

Figure 47 Screenshot from DHIS2 quality tool



10.5. Internal consistency of reported data

2a – extreme outlier. This table shows the percentage of values that are extreme outliers, using the standard method. The threshold for what is defined as an extreme outlier can be specified per indicator (numerator), with the default value being 3 standard deviations from the mean.

Figure 48 Screenshot from DHIS2 quality tool

2a: Extreme outliers

Extreme outliers, using the standard method. Threshold denotes the number of standard deviations from the mean. Regions are counted as divergent if they have one or more extreme outliers for an indicator.

Indicator	Threshold	Overall score (%)	Region with divergent score		
			Number	Percent	Names
ANC 1 or more visits	3 SD	2.1%	1	25%	Region B
Confirmed malaria cases	3 SD	0%	0	0%	
DPT-HepB-Hib 3 doses given < 1 year	3 SD	0%	0	0%	
PLHIV receiving ART	3 SD	0%	0	0%	
TB cases notified	3 SD	4.2%	2	50%	Region A, Region D

Figure 49 Explanation of columns for annual report (see figure 48)

Column	Description
Indicator	The variable (numerator) that is checked for outliers
Threshold	The number of standard deviations from the mean a value must be to be counted as an outlier.
Overall score	The percentage of values that are extreme outliers.
Subunits with divergent score	- Number: number of subunits with extreme outliers - Percent: percent of subunits with extreme outliers - Name: name of subunits with extreme outliers

2b – moderate outlier. This table shows the percentage of values that are moderate outliers, using the standard method. The threshold for what is defined as a moderate outlier can be specified per indicator (numerator), with the default value being 2 standard deviations from the mean.

Figure 50 Screenshot from DHIS2 quality tool

2b: Moderate outliers.

Moderate outliers, using the standard method. Threshold denotes the number of standard deviations from the mean. Regions are counted as divergent if they have two or more moderate outliers for an indicator.

Indicator	Threshold	Overall score (%)	Region with divergent score		
			Number	Percent	Names
ANC 1 or more visits	2 SD	4.2%	0	0%	
Confirmed malaria cases	Not applicable				
DPT-HepB-Hib 3 doses given < 1 year	2 SD	4.2%	0	0%	
PLHIV receiving ART	2 SD	2.1%	0	0%	
TB cases notified	2 SD	4.2%	0	0%	

Figure 51 Explanation of columns for annual report (see figure 50)

Column	Description
Indicator	The variable (numerator) that is checked for outliers
Threshold	The number of standard deviations from the mean needed by a value to be defined as an outlier.
Overall score	The percentage of values defined as moderate outliers.
Subunits with divergent score	- Number: number of subunits with two or more moderate outliers - Percent: percent of subunits with two or more moderate outliers - Name: name of subunits with two or more moderate outliers

2c – moderate outlier. This table shows the percentage of values that are moderate outliers, using the modified Z score method.

Figure 52 Screenshot from DHIS2 quality tool

2c: Moderate outliers

Moderate outliers, based on median (modified Z score). Threshold denotes the number of standard deviations from the mean. Regions are counted as divergent if they have two or more moderate outliers for an indicator.

Indicator	Threshold	Overall score (%)	Region with divergent score		
			Number	Percent	Names
ANC 1 or more visits	3.5	0%	0	0%	
Confirmed malaria cases	Not applicable				
DPT-HepB-Hib 3 doses given < 1 year	3.5	0%	0	0%	
PLHIV receiving ART	3.5	0%	0	0%	
TB cases notified	3.5	0%	0	0%	

Figure 53 Explanation of columns for annual report (see figure 52)

Column	Description
Indicator	The variable (numerator) that is checked for outliers
Threshold	The number of standard deviations from the mean needed by a value to be defined as an outlier.
Overall score	The percentage of values defined as moderate outliers.
Subunits with divergent score	- Number: number of subunits with two or more moderate outliers - Percent: percent of subunits with two or more moderate outliers - Name: name of subunits with two or more moderate outliers

2d – consistency of indicator values over time. This table and charts show the consistency of the indicator (numerator) values over time. For indicators that are expected to be constant, the selected year is compared with the average of the three preceding years. For indicators that are expected to increase or decrease, the selected year is compared with a forecast based on the three preceding years. For subunits (e.g. regions, districts), the consistency of an indicator is either compared to the overall consistency, or the expected ratio, e.g. current to average or forecast from previous periods.

Figure 54 Screenshot from DHIS2 quality tool

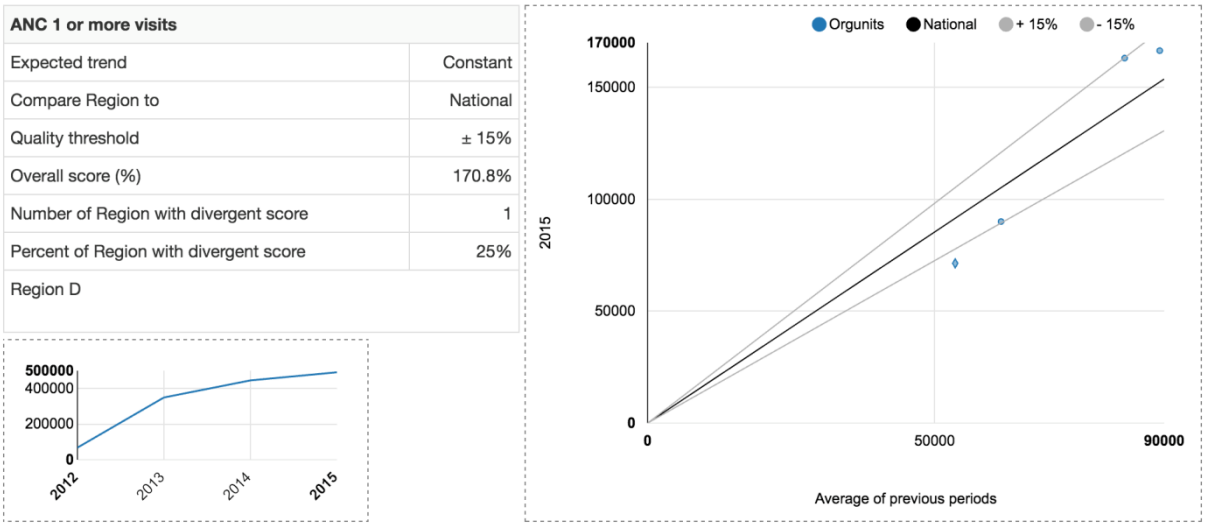
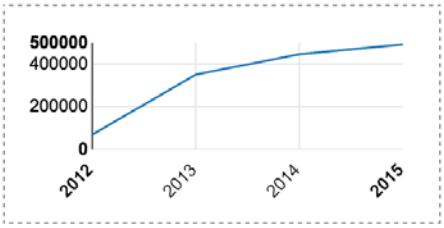


Figure 55 Explanation of fields for annual report (see figure 54)

Field	Description
Expected trend	Whether reporting is expected to be constant or increase/decrease. If the expected trend is constant, the indicator value of the selected year is compared with the *average* of the three previous years. If the expected trend is increasing or decreasing, the indicator value of the selected year is compared with a forecast based on the trend in the three previous years.
Compare subunits to	Whether the consistency for subunits is compared to the overall (e.g. national) consistency, or to the expected consistency between current and previous periods.
Quality threshold	The consistency (+/-) of completeness that is expected.
Overall score	The consistency between current year and the average or forecast of the three previous years, i.e. value for the current year divided by the average or forecast for 3 previous years, expressed as a percentage.
Number of Subunits with divergent score	Number of subunits with a consistency outside the threshold.
Percent of Subunits with divergent score	Percent of subunits with a consistency outside the threshold.
[Last row]	Names of subunits with a consistency outside the threshold.

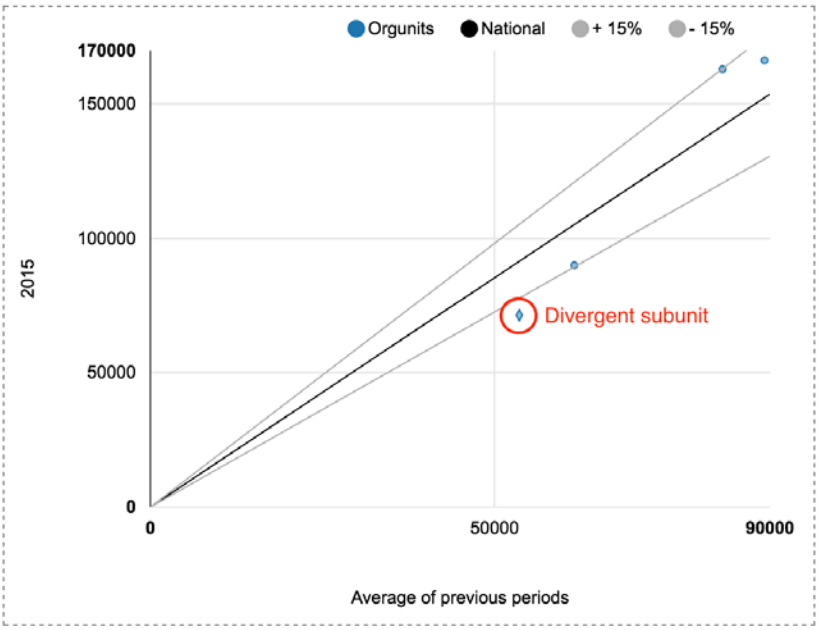
The small line chart shows the reported value for the selected year, and the three preceding ones.

Figure 56 Screenshot from DHIS2 quality tool



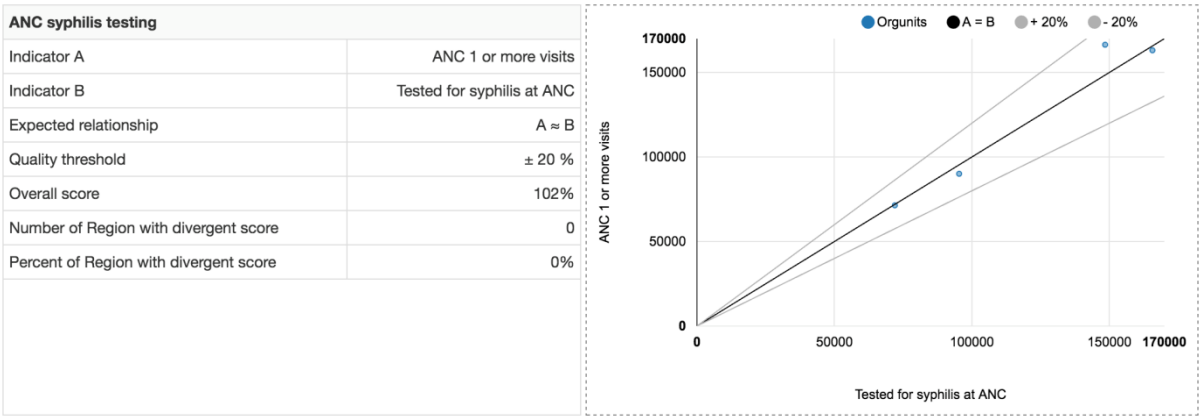
The scatterplot shows the subunits (e.g. regions) as points, along with three solid lines. The solid black line is the ratio that the subunits are compared to. This could be either the boundary (e.g. national) ratio, or the expected ratio (e.g. current period against the average of previous periods). The threshold values are shown as light grey lines. Subunits outside these lines are counted as “divergent”, and they get a different symbol in the scatterplot.

Figure 57 Screenshot from DHIS2 quality tool



2d – consistency between related indicators. This table and charts show the consistency of related the indicators (numerators) for the selected period.

Figure 58 Screenshot from DHIS2 quality tool



ANC syphilis testing	
Indicator A	ANC 1 or more visits
Indicator B	Tested for syphilis at ANC
Expected relationship	A ≈ B
Quality threshold	± 20 %
Overall score	102%
Number of Region with divergent score	0
Percent of Region with divergent score	0%

Figure 59 Explanation of fields for annual report (see figure 58)

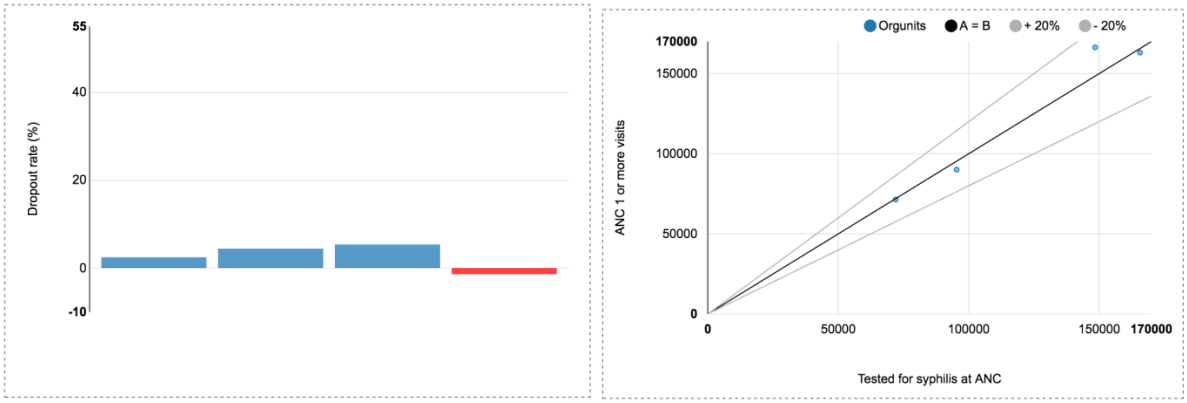
Field	**Description**
Indicator A	The first indicator (numerator) in the comparison.
Indicator B	The second indicator (numerator) in the comparison.
Expected relationship	The expected relationship between the indicators, presented in a separate table below.
Quality threshold	The consistency (+/-) of completeness that is expected.
Overall score	The ratio between the two indicators (numerators), expressed as a percentage.
Number of Subunits with divergent score	Number of subunits with a ratio outside the threshold.
Percent of Subunits with divergent score	Percent of subunits with a ratio outside the threshold.
[Last row]	Names of subunits with a ratio outside the threshold.

Figure 60 Overview of The different types of comparisons between indicators are outlined in the table

Type	Description	Threshold	Example
A ≈ B	A and B is expected to be roughly equal.	The threshold denotes the percentage difference that is allowed between the two.	BCG ≈ DOT 1
A > B	A should be greater than B.	The threshold denotes the percentage B is permitted to be greater than A. If the threshold is 0, the comparison is strictly A > B.	Malaria cases tested > Malaria cases confirmed
Dropout rate	Dropout rate from A to B	Calculates the dropout rate from A to B, which should not be negative.	ANC 1 to ANC 4
Equal across organisation units	Ratio A/B	Compares the ratio of the numerators between "subunits" to their "parent", e.g. district to national level.	Any numerators where a consistent ratio is expected

There are two different chart types, depending on the type of the expected relationship. Dropout rates are presented as a bar chart, with subunits (e.g. regions) with negative dropouts coloured red. For other types of relationships, the results are presented as a scatterplot showing the subunits (e.g. regions) as points, along with two or three solid lines. The solid black line is the ratio that the subunits are compared to. This could be either the boundary (e.g. national) ratio, or the expected ratio (i.e. 1 if A ≈ B or A > B). The threshold values are shown as light grey lines. Subunits outside these lines are counted as "divergent", and they get a different symbol in the scatterplot.

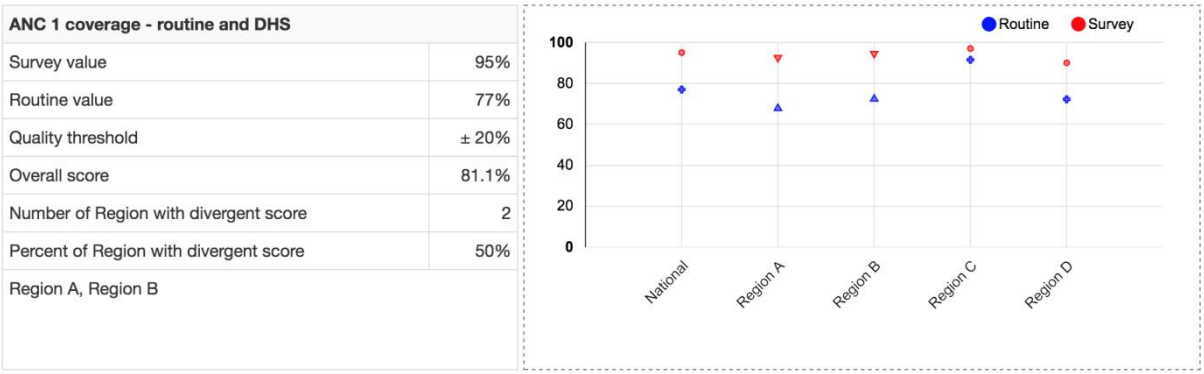
Figure 61 Screenshot from DHIS2 quality tool



External comparison

3a – Comparison with external/survey data. This check compares the routine data, typically collected monthly, with external data, typically from household surveys. The results are presented as a table and a chart.

Figure 62 Screenshot from DHIS2 quality tool



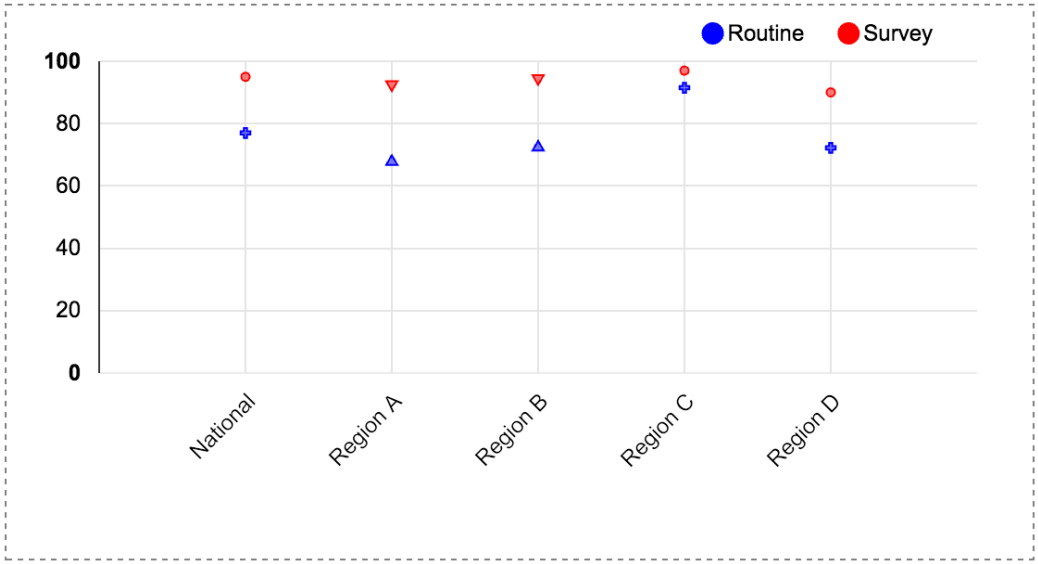
ANC 1 coverage - routine and DHS	
Survey value	95%
Routine value	77%
Quality threshold	± 20%
Overall score	81.1%
Number of Region with divergent score	2
Percent of Region with divergent score	50%
Region A, Region B	

Figure 63 Explanation of fields for annual report (see figure 62)

Field	Description
Survey value	The survey/external value, imported into DHIS2.
Routine value	The routine value, normally the regular DHIS2 data.
Quality threshold	The ratio (+/-) between the routine and survey value, expressed as a percentage.
Overall score	The ratio between the external and routine indicators, expressed as a percentage.
Number of Subunits with divergent score	Number of subunits where the difference between the routine and survey values are greater than the allowed threshold.
Percent of Subunits with divergent score	Percent of subunits where the difference between the routine and survey values are greater than the allowed threshold.
[Last row]	Names of subunits where the difference between the routine and survey values are greater than the allowed threshold.

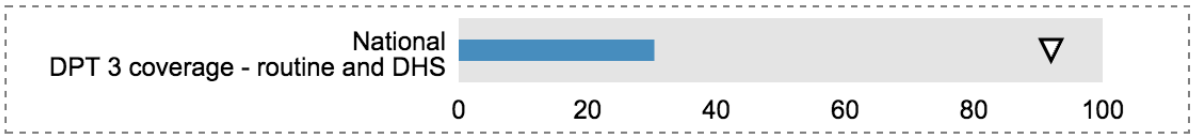
There are two different charts that might be displayed, depending on whether the external data comparison is available only for the boundary (e.g. national) level, or also for each subunit (e.g. region). The chart used when external data is available for subunits has two series, for routine and survey data, with different colours and symbols – routine data as a plus and survey data as a circle. Organisation units with divergent scores are presented with different symbols (triangles).

Figure 64 Screenshot from DHIS2 quality tool



The chart is used when data is only available for the boundary organisation unit is a bullet chart where the routine value is presented as a blue bar, and the survey values as a triangle.

Figure 65 Screenshot from DHIS2 quality tool



10.6. Consistency of population data

4a – consistency with UN population projection. The check shows the ratio between the population figure used within DHIS2, typically census projections, and the UN population projection.

Figure 66 Screenshot from DHIS2 quality tool

4a: Consistency with UN population projection.	
Ratio of population projection from the Bureau of Statistics to a UN projection	1.042

4b – consistency of denominators. This check compares related denominators, e.g. denominator estimates from different sources. The result is presented as a table and chart, similar to the "*2d – consistency between related indicators"* checks described above, only with the comparison type fixed to A ≈ B.

Figure 67 Screenshot from DHIS2 quality tool

Population < 1 year - census vs EPI	
Denominator A	Under 1 Population
Denominator B	Population < 1 year (EPI)
Quality threshold	± 10%
Overall score	104.2%
# Region with poor score	0
% Region with poor score	0%

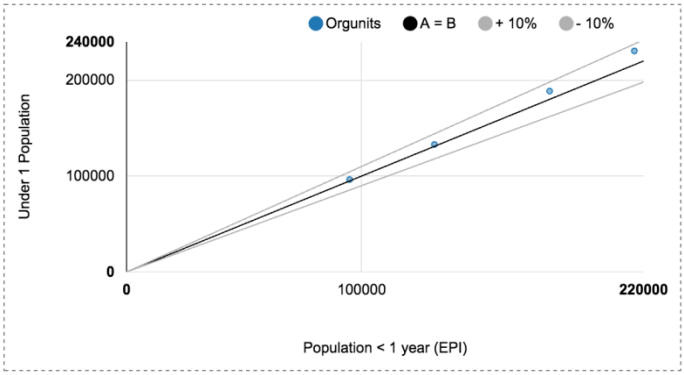
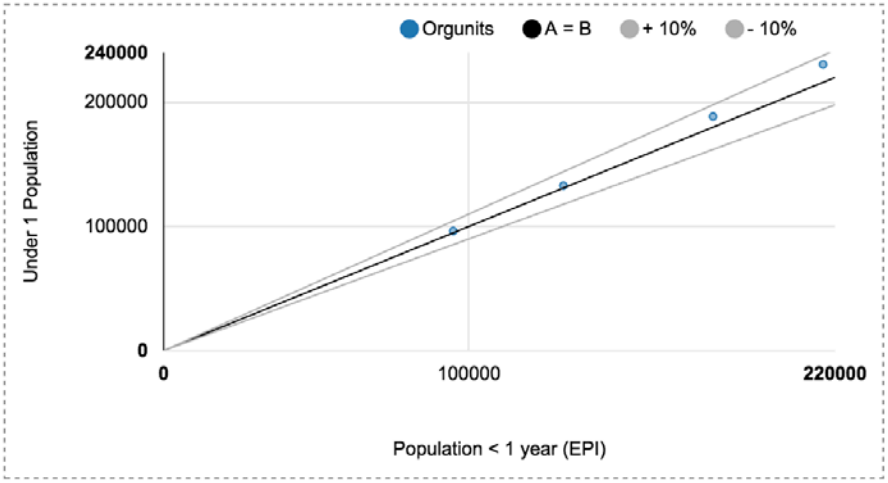


Figure 68 Explanation of fields for annual report (see figure 66)

Field	Description
Denominator A	The first denominator (numerator) in the comparison.
Denominator B	The second denominator (numerator) in the comparison.
Quality threshold	The consistency (+/-) of completeness that is expected.
Overall score	The ratio between the two denominators, expressed as a percentage.
Number of Subunits with divergent score	Number of subunits with a ratio outside the threshold.
Percent of Subunits with divergent score	Percent of subunits with a ratio outside the threshold.
[Last row]	Names of subunits with a ratio outside the threshold.

The chart is a scatterplot showing the subunits (e.g. regions) as points, along with two or three solid lines. The solid black line is equality line where the two denominators are equal, i.e. where the ratio is 1. The threshold values are shown as light grey lines. Subunits outside these lines are counted as "divergent", and they get a different symbol in the scatterplot.

Figure 69 Screenshot from DHIS2 quality tool



10.7. Remarks

At the bottom of the page, below the content of the annual review, a section called “Remarks” may appear. This table shows any errors or exceptions that have come up while the report has been generated, in most cases where data has been missing for certain organisation units or periods and is not included in the results.

Figure 70 Screenshot from DHIS2 quality tool

Remarks

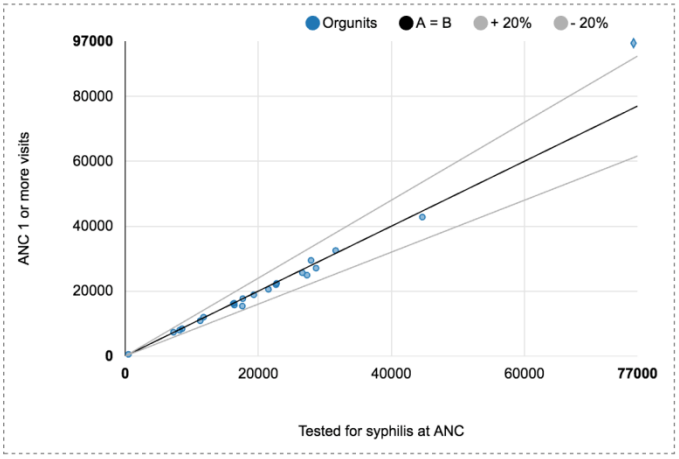
Analysis type	Item	Remarks
Consistency betweeen indicators	Total Population - UN Population projection	Skipped for the following units due to missing data: District A, District B, District C, District D, District E, District F, District G, District H, District I, District J, District K, District L, District M, District N, District O, District P, District Q, District S, District T, District U, District V, District W
Consistency betweeen indicators	DPT-HepB-Hib 1 doses given < 1 year - DPT-HepB-Hib 3 doses given < 1 year	Skipped for the following units due to missing data: District B
Consistency over time	DPT-HepB-Hib 3 doses given < 1 year	Skipped for the following units due to missing data: District B
Consistency over time	EPI Reporting rate	Skipped for the following units due to missing data: District B

10.8. Commenting and printing the annual report

Each of the outputs of the annual report has an associated box where comments and interpretations of the outputs can be written.

Figure 71 Screenshot from DHIS2 quality tool

ANC syphilis testing	
Indicator A	ANC 1 or more visits
Indicator B	Tested for syphilis at ANC
Expected relationship	A ≈ B
Quality threshold	± 20 %
Overall score	102%
Number of District with divergent score	1
Percent of District with divergent score	4.5%
District E	



Interpretations

The ratio between clients registered for ANC and clients tested for syphilis at ANC is generally quite consistent across districts, with an overall ratio of 102%. District E is the only district with a divergent score, and it is notably also by far the biggest district.

The report can be printed using the “Print” button at the top of the report. It is important to use this button rather than a keyboard shortcut, as this triggers a number of adjustments to the report to make it more printer-friendly. A front page will also be added when printing.

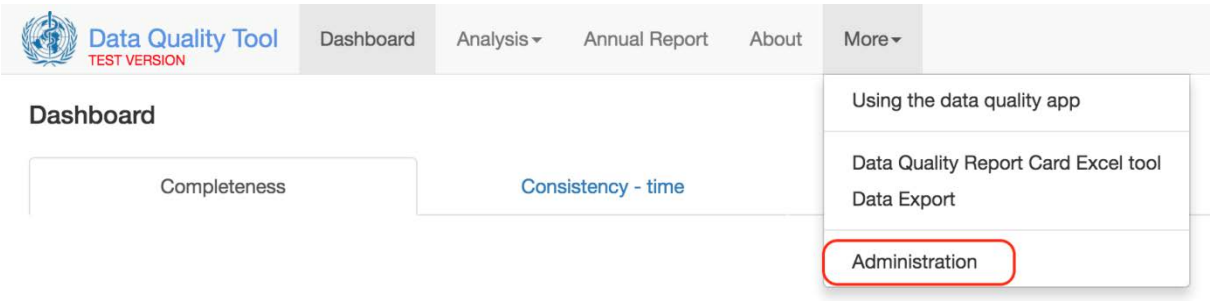
Figure 72 Screenshot from DHIS2 quality tool

Many web browsers, e.g. Google Chrome (https://support.google.com/chrome/answer/1379552), let users generate PDF documents from the print dialog, which is useful for saving an electronic copy of the report (with comments).

11. Setting up the DHIS2 Quality Tool

The Data Quality Tool requires an initial configuration for much of the functionality. This configuration consists primarily of selecting what data elements and indicators should be displayed on the data quality dashboard and in the annual report, as well as specifying the quality parameters for these variables. The configuration is done in the "Administration" section of the tool, accessed under "More" in the main menu.

Figure 73 Screenshot from DHIS2 quality tool



Any user with access to the tool can see the configuration; however, making changes requires having a DHIS2 user role with the "Add/edit public indicator" authority[6].

The WHO Data Quality Review Toolkit suggests a number of *indicators* that should be checked for data quality. However, the data quality checks/metrics are done separately on the numerator and the denominator. When configuring the data quality tool, numerator and denominators are therefore configured separately.

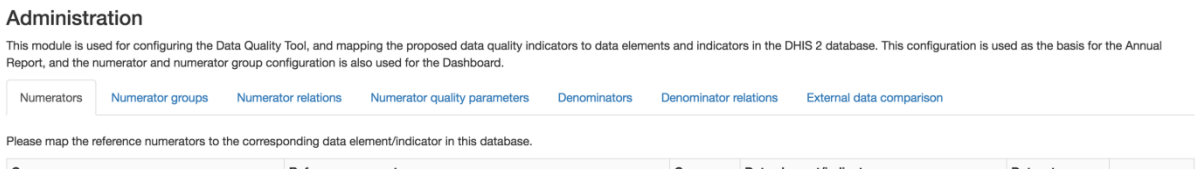
The Administration section of the tool has different 7 tabs/pages, for configuring different aspects of the tool:

- Numerators
- Numerator groups
- Numerator relations
- Numerator quality parameters
- Denominators
- Denominator relations
- External data comparison

[6] Note that this restriction only applies to the user interface. Any user with access to the app could in theory change the configuration through the DHIS2 API.

Each of these will be described below.

Figure 74 Screenshot from DHIS2 quality tool



The 4 first pages for configuring numerators are necessary for both the dashboard and annual report functionality of the tool. The denominators and external data are used in relation to the annual report only.

11.1. Numerators

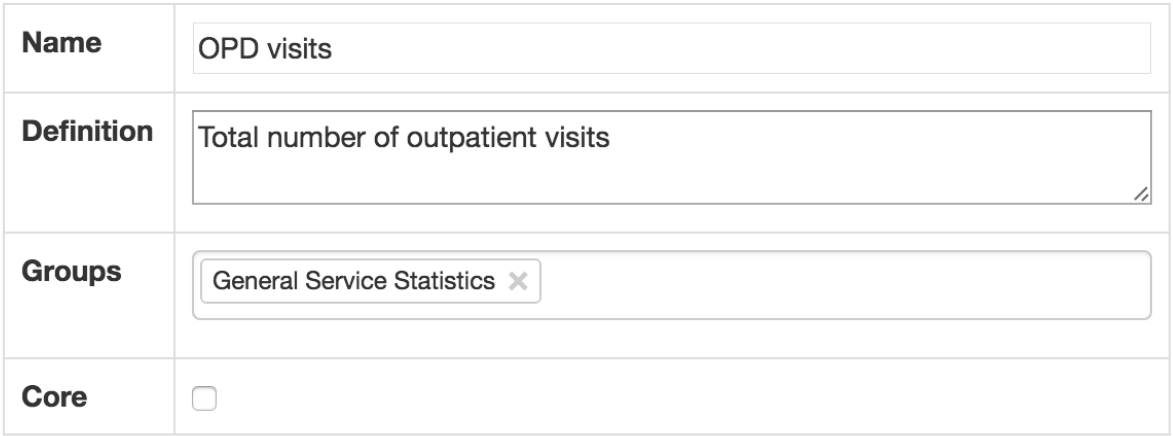
The numerator page is used for configuring the numerators, which is typically disease or service data such as "ANC 1 visits", "Clients tested for HIV", "Malaria cases confirmed" etc. The tool will list a number of numerators, based on indicators recommended in the WHO Data Quality Toolkit. Since each DHIS2 implementation is different, these recommended numerators must be mapped to the correct variable in DHIS2. It is also possible add any number of additional numerators, this will be described below.

To map a recommended indicator to a variable in the DHIS2 database, click the "Edit" button in the last column of the table.

Figure 75 Screenshot from DHIS2 quality tool

A window will appear. The top part of the window shows the name and definition of the numerator, as well as the numerator group(s) it is part of (groups are discussed in the next section). For the default recommended numerators, the name and definition cannot be changed. A checkbox allows administrators to choose whether the numerator should be included as "core" indicator, which means that it will be shown on the data quality dashboard by default for all users of the tool.

Figure 76 Screenshot from DHIS2 quality tool



The second part of the window is the "Data mapping", which is where the numerator is mapped to a specific data element or indicator in the DHIS2 database. The steps for mapping indicators are as follows:

1) *Select whether the variable is a data element or indicator, and for data elements, whether to look at the total or disaggregation ("Details").* To specify an indicator as a numerator can seem counterintuitive, but is necessary in cases where the numerator consists the sum of two or more data

elements, e.g. “Malaria cases confirmed by RDT” + “Malaria cases confirmed by lab test” to get “Malaria cases confirmed”.

2) *Select a data element or indicator group.*
3) *Select a data element or indicator.*

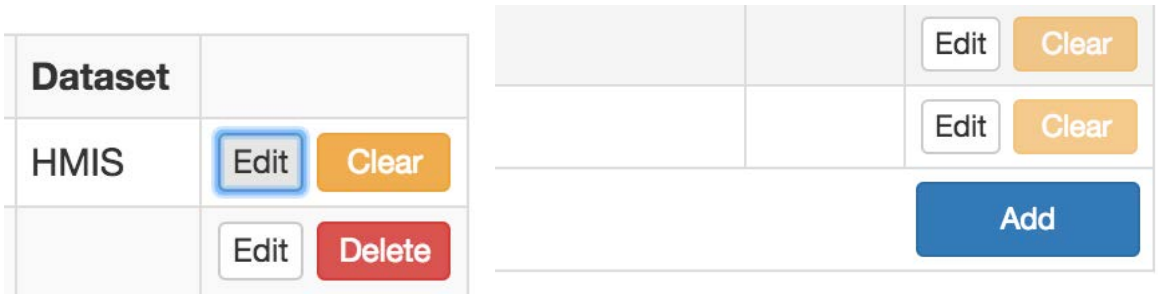
Figure 77 Screenshot from DHIS2 quality tool

Finally, two selections must be made in relation to completeness. For each numerator, the data set (reporting form) to use to estimate completeness and timeliness of reporting must be specified. In most cases, there will only be on option in the drop down list. However, in cases where the same data elements are collected in multiple data sets, or where the numerator is calculated based on data from multiple data sets, a choice must be made for which dataset to use as a proxy. The last field is “Variable for completeness”. If the variable selected as numerator can be disaggregated further, i.e. if it is a data element with categories for disaggregation or an indicator made up of multiple data elements, you must here select what disaggregation to use as a proxy for variable-level completeness. You should normally choose the disaggregation where you expect the highest number of reported values.

Figure 78 Screenshot from DHIS2 quality tool

Click “Save” in the window to save the numerator configuration. For the default recommended numerators, the configuration can be “Cleared”, in which case the mapping is removed and the numerator is hidden from the dashboard and in the annual report. Additional numerators added by the administrators can be removed completely with a “Delete” button.

Figure 79 Screenshot from DHIS2 quality tool



Additional numerators can be added using the “Add” button at the bottom of the table. This brings up a window similar to what is described above, with the only exception that the name and description fields can be edited.

11.2. Numerator groups

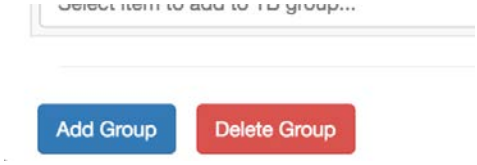
The numerator group page is used for grouping the numerators described in the last section. These groups are used on the dashboard and for the annual report, where data is displayed based on the selection of numerator group.

To add a numerator to a group, select the numerator from the dropdown list at the bottom of the group's table and click "Add". To remove a numerator from a group, click the "Remove" button next to the numerator.

Figure 80 Screenshot from DHIS2 quality tool

Groups can be added or removed using the "Add group" and "Delete group" buttons at the bottom of the page.

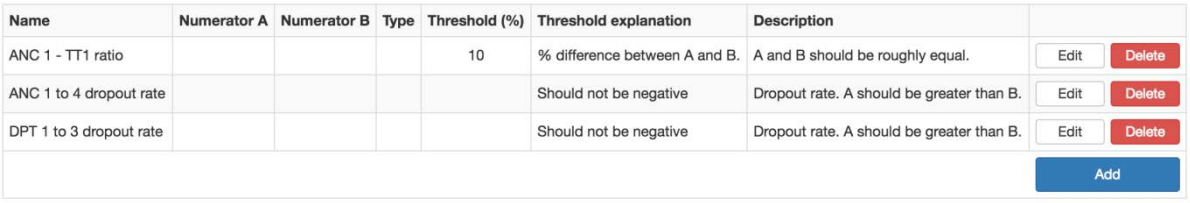
Figure 81 Screenshot from DHIS2 quality tool



11.3. Numerator relations

The numerator relations page is used to configure internal consistency checks, i.e. checks of the relation between related indicators (numerators). The tool comes with a default list of recommended checks, however, these needs to be configured in the same way as the numerators - in the screenshot below, the relations have not been configured, as can be seen from the fact that the "Numerator A" and "Numerator B" columns are empty.

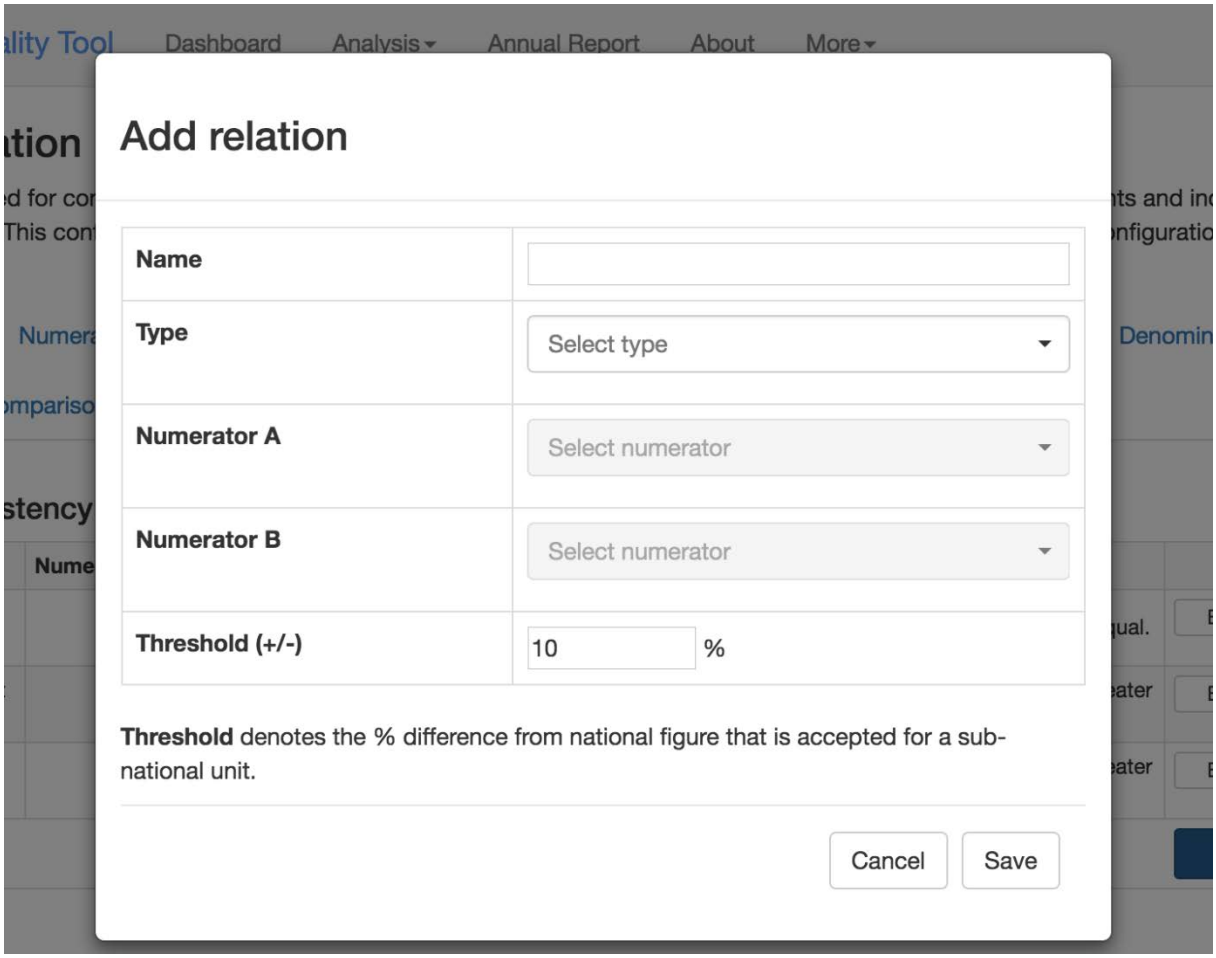
Figure 82 Screenshot from DHIS2 quality tool



Name	Numerator A	Numerator B	Type	Threshold (%)	Threshold explanation	Description	
ANC 1 - TT1 ratio				10	% difference between A and B.	A and B should be roughly equal.	Edit Delete
ANC 1 to 4 dropout rate					Should not be negative	Dropout rate. A should be greater than B.	Edit Delete
DPT 1 to 3 dropout rate					Should not be negative	Dropout rate. A should be greater than B.	Edit Delete
							Add

The table gives an overview of the internal consistency checks. To configure a new or edit an existing numerator relation check, click the "Edit" or "Add" buttons respectively, and a window will appear.

Figure 83 Screenshot from DHIS2 quality tool



First, specify the name of the consistency check, which will appear on the dashboard and in the annual report. Second, select a type of consistency check. The table below explains the different types that are available. Third, select the numerators (A and B) to be compared. And finally, set a threshold for the comparison. Click "Save" to save the numerator relation.

Figure 84 Explanation of relations in figure 82

Type	Description	Threshold	Example
A ≈ B	A and B is expected to be roughly equal.	The threshold denotes the percentage difference that is allowed between the two.	BCG ≈ DPT 1
A > B	A should be greater than B.	The threshold denotes the percentage B is permitted to be greater than A. If the threshold is 0, the comparison is strictly A > B.	Malaria cases tested > Malaria cases confirmed
Dropout rate	Dropout rate from A to B	Calculates the dropout rate from A to B, which should not be negative.	ANC 1 to ANC 4
Equal across organisation units	Ratio A/B	Compares the ratio of the numerators between "subunits" to their "parent", e.g. district to national level.	Any numerators where a consistent ratio is expected

11.4. Numerator quality parameters

The numerator quality parameters page allows administrators to specify the parameters that should be applied to each numerator and dataset when performing the data quality checks. These parameters are used on the dashboard and in the annual review. Only numerators that have been mapped to DHIS2 database variables are displayed, along with the relevant data sets.

The parameters for numerators are described in the table below.

Figure 85 Screenshot from DHIS2 quality tool

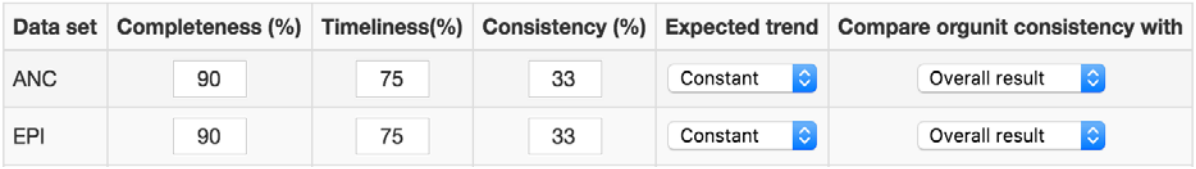
Group	Reference indicator/data element	Local data element/indicator	Moderate outlier (SD)	Extreme outlier (SD)	Consistency (%)	Expected trend	Compare orgunit consistency with	Missing/zero values (%)
General Service Statistics	OPD visits	HMIS OPD total attendance	2	3	33	Constant	Overall result	90
General Service Statistics	Poulation	Expected pregnant women	2	3	33	Constant	Overall result	90

Figure 86 Explanation of Parameters in Figure 84

Parameter	Relevant for…	Description
Moderate outlier	Outliers	Number of standard deviations (SD) from the mean for a values to quality as a moderate outlier.
Extreme outlier		Number of standard deviations (SD) from the mean for a values to quality as a moderate outlier.
Consistency	Consistency over time	Threshold for consistency over time (percentage change over time).
Expected trend		Whether the numerator value is expected to be constant over time or increase/decrease.
Compare orgunit consistency with…		Whether to compare consistency over time across organisation units, or to the expected change (e.g. constant or increasing/decreasing)
Missing/zero values	Completeness	Threshold for missing/zero values for variable completeness. Note: when zero values are not stored for a data element, zeros and missing are not differentiated.

Parameters for dataset completeness are described in the table below.

Figure 87 Screenshot from DHIS2 quality tool



Data set	Completeness (%)	Timeliness(%)	Consistency (%)	Expected trend	Compare orgunit consistency with
ANC	90	75	33	Constant	Overall result
EPI	90	75	33	Constant	Overall result

Figure 88 Explanation of Parameters in Figure 86

Parameter	Description
Completeness	Threshold for completeness of reporting.
Timeliness	Threshold for timeliness of reporting.
Consistency	Threshold for consistency of completeness of reporting over time (percentage change over time).
Expected trend	Whether the completeness is expected to be constant over time or increase/decrease.
Compare orgunit consistency with…	Whether to compare consistency over time across organisation units, or to the expected change (e.g. constant or increasing/decreasing)

After the necessary changes have been made to numerator and dataset completeness parameters, click the" Save changes" button in the bottom left corner of the page.

11.5. Denominators

The denominators page allows administrators to identify and map denominators in the DHIS2 database. Denominators are needed for three different data quality checks:

- Comparison of national figures with UN population projections.
- Comparison of the same denominator from different sources, e.g. in cases where disease programme and the national bureau of statistics produce different estimates for population groups.
- For comparing routine data with external data sources (together with numerators), e.g. household survey results.

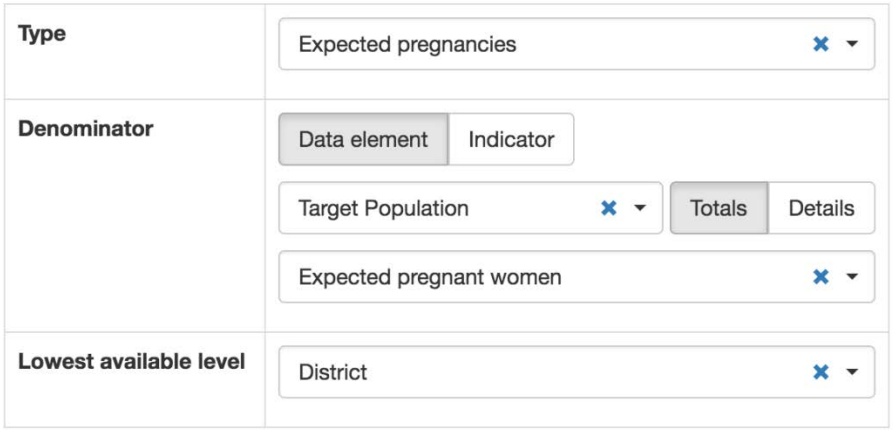
Which denominators are needed will thus depend on what external data is available.

The table of denominators has two columns: “Data element/indicator”, which shows the name of the DHIS2 data element/numerator (once configured), and “Type” which is the type of denominator (e.g. total population, < 1 year population, expected pregnancies, etc.). Use the “Edit” or “Add” buttons to open the add/edit denominator window.

Figure 89 Screenshot from DHIS2 quality tool

To configure the denominator, first select the type of denominator. Second, select the DHIS2 data element or indicator to map to that denominator type. Finally, select the lowest level the denominator data is available for, and click “Save”.

Figure 90 Screenshot from DHIS2 quality tool



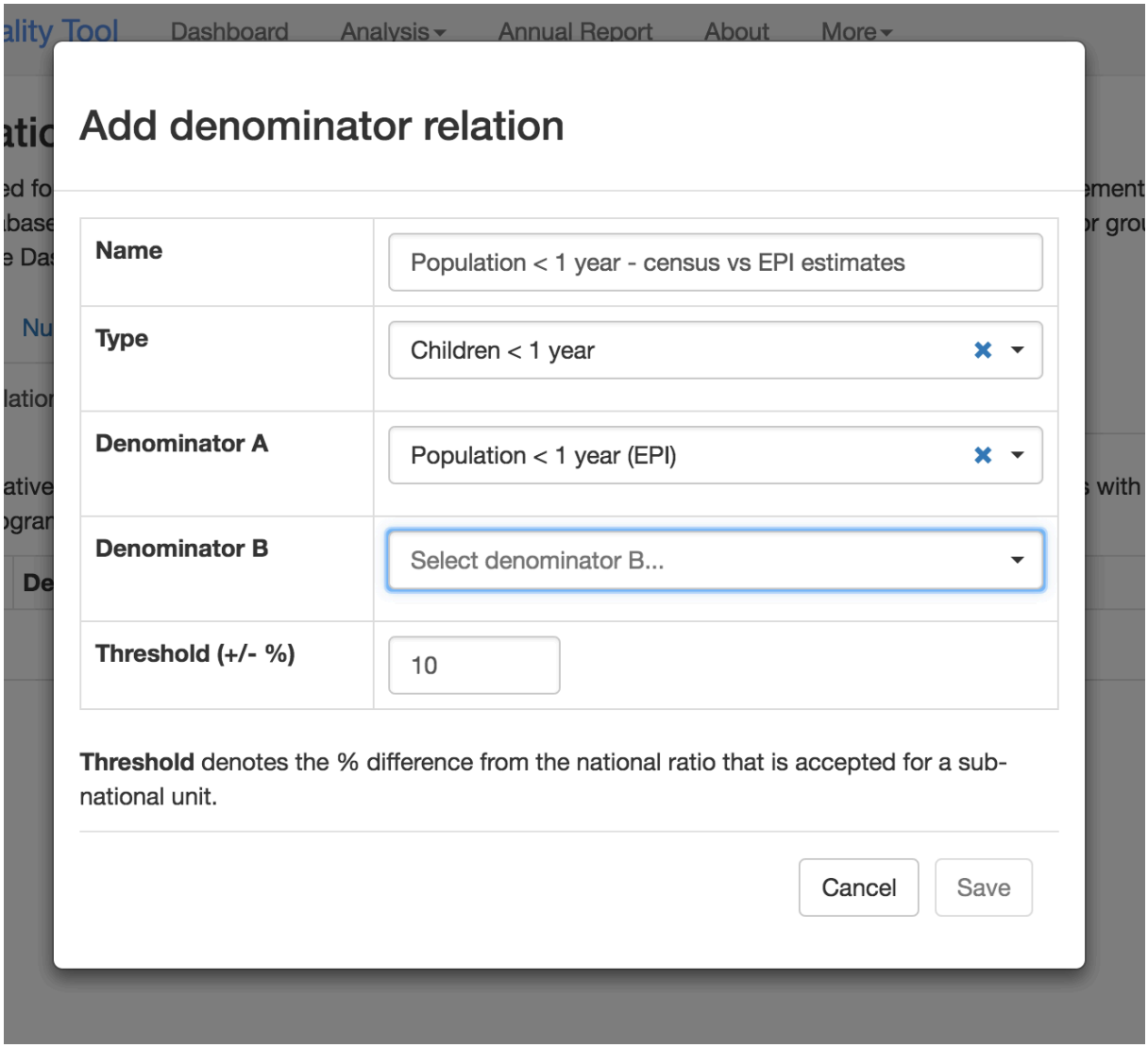
As noted above, one of the data quality metrics is a consistency check of denominators in cases where different estimates for the same denominator are used. In those cases, all the different denominators must be configured.

11.6. Denominator relations

The denominator relations page lets administrators set up internal consistency checks between denominators in cases where there are multiple data sources for the same denominator type, e.g. if disease programmes use estimates that are different from the official population figures, or for comparing national census figures with UN population projections.

Use the “Add” or “Edit” buttons to bring up the window to add/edit denominator relations. First, specify a name for the denominator relation. Second, choose from the drop down the denominator type of the relation, e.g. “Live births”. Third, select the two denominators A and B to compare – only denominators that match the selected denominator type will be displayed. Finally, set the threshold, which denotes the percentage difference allowed between the two denominators.

Figure 91 Screenshot from DHIS2 quality tool

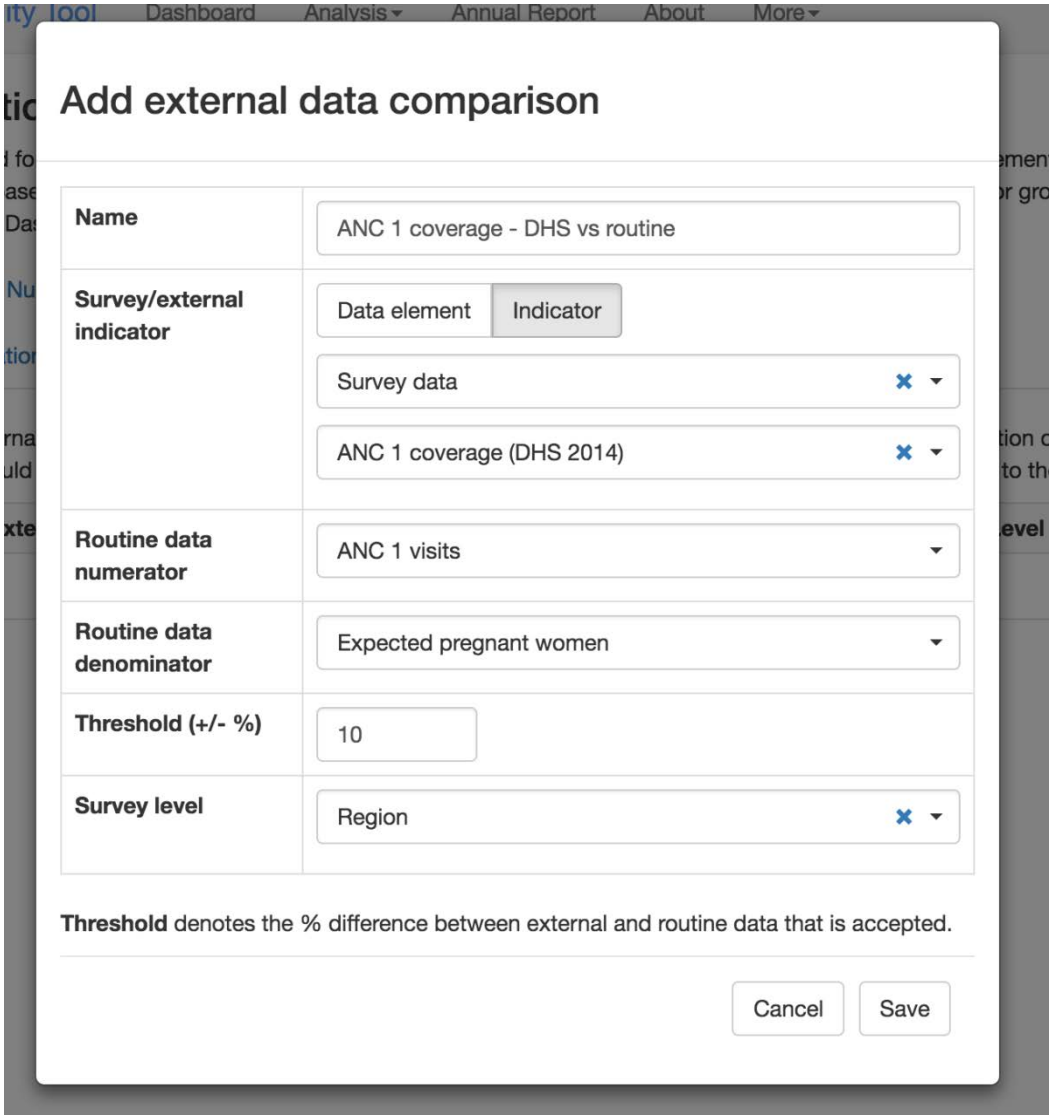


11.7. External data comparison

On the external data comparison page, comparison of indicators between routine data and external data such as household surveys can be specified. This is done by specifying the data element/variable for the external data, a numerator and a denominator. External data comparisons require that the external data has first been imported into DHIS2 and is available as data elements or indicators.

Click “Add” or “Edit” to add/edit external data comparisons. First, give the comparison a name. Second, select the data element or indicator where the external data (e.g. a value from a household survey) is stored. Second, select the numerator and denominator to be used to use for calculating the routine indicator value, which will be compared to the external data. Third, select the threshold used when comparing the external and the routine data. Finally, select the lowest level at which the survey data is available.

Figure 92 Screenshot from DHIS2 quality tool



References

Statistics Norway. Documents 2017/3. *The Health Management Information system in Malawi: Assessment of data quality and methods for improvements*

Statistics Norway. Documents 2017/5. *Improving health data quality: Recommendations and guidelines Based on the case of the Health Management Information System in Malawi and DHIS2*

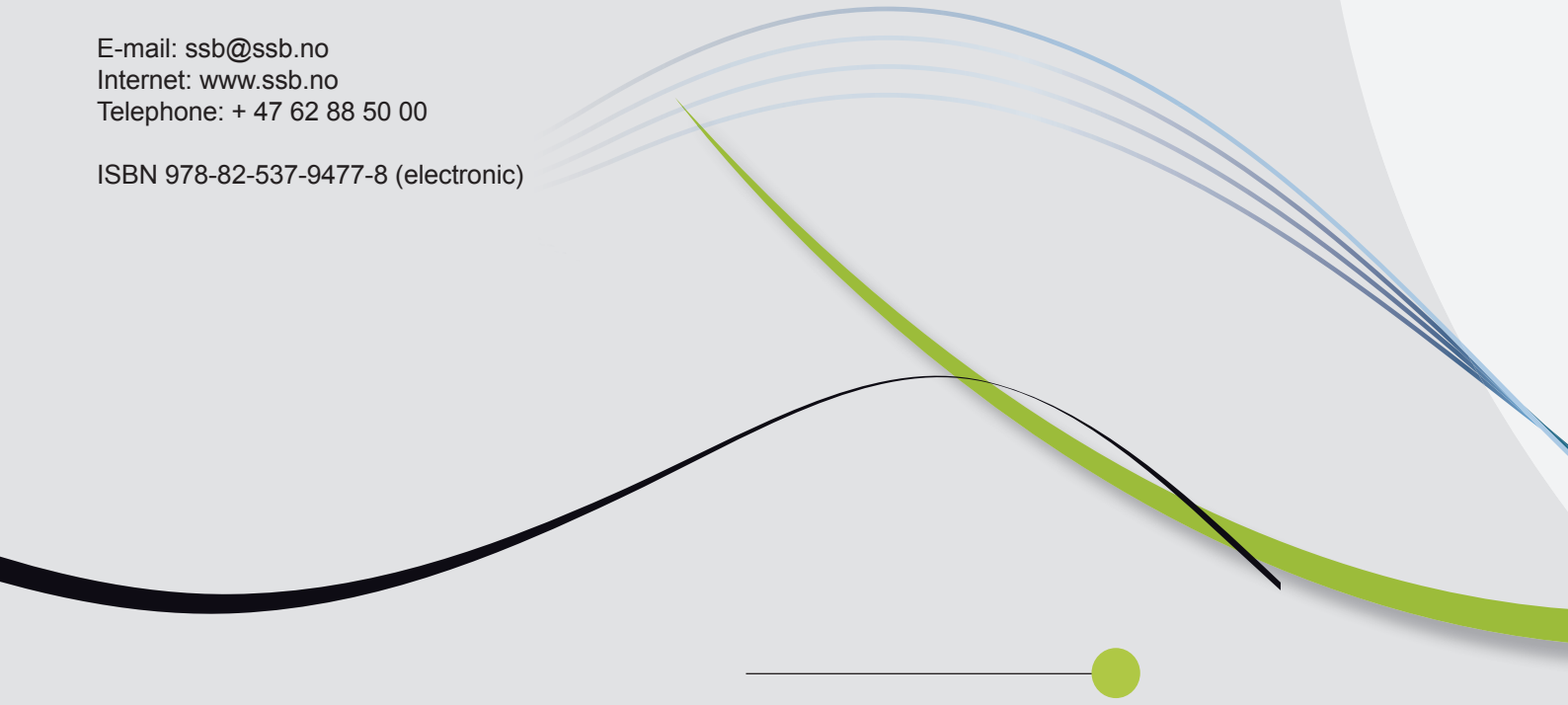
Statistics Norway

Postal address:
PO Box 8131 Dept
NO-0033 Oslo

Office address:
Akersveien 26, Oslo
Oterveien 23, Kongsvinger

E-mail: ssb@ssb.no
Internet: www.ssb.no
Telephone: + 47 62 88 50 00

ISBN 978-82-537-9477-8 (electronic)



Statistisk sentralbyrå
Statistics Norway

Design: Siri Boquist