

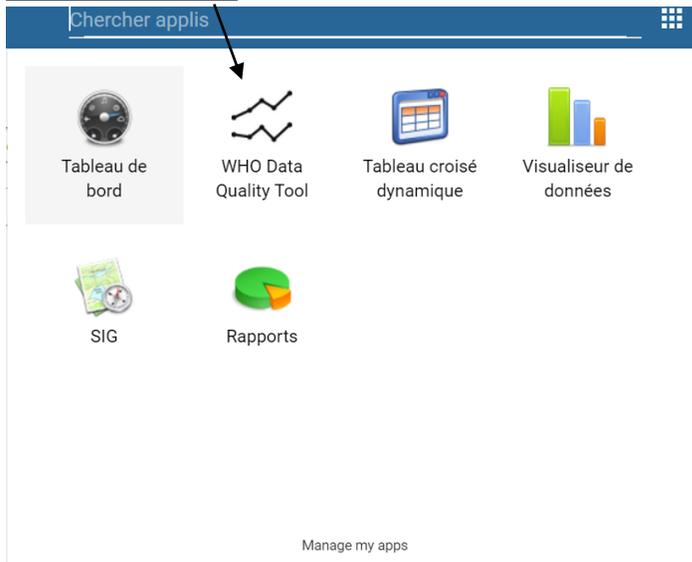
Introduction to WHO's DHIS2 Data Quality Tool

1. Log onto the DHIS2 instance: <https://who.dhis2.net/dq>

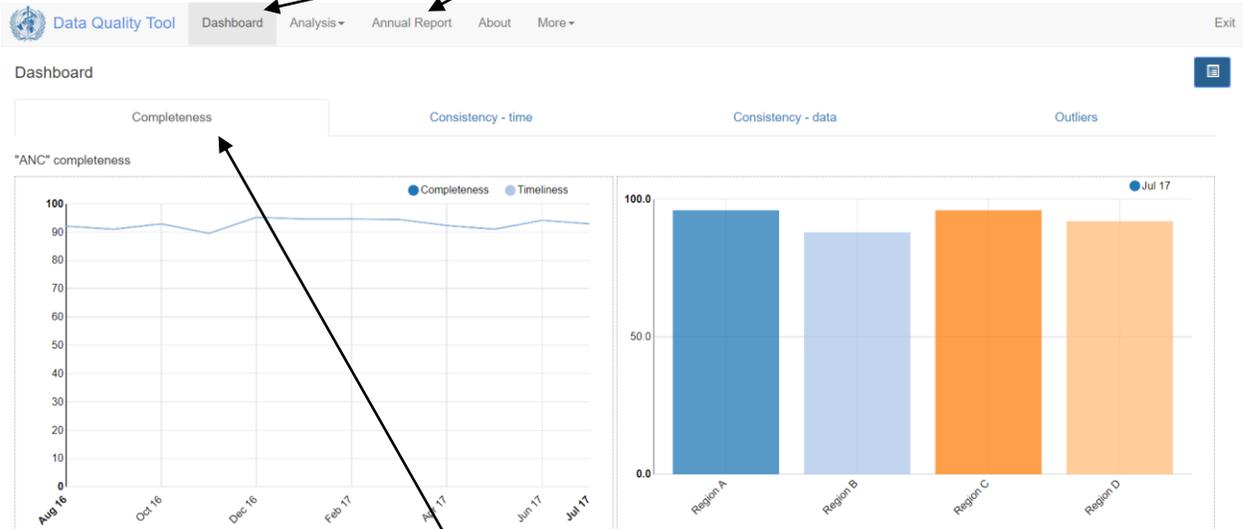
Username: demo Password: UGANDA 2016



2. Click on the menu icon in the upper right of the screen () then click on **WHO Data Quality Tool**.



3. **Quickly review the main menu of the DQ Tool:** After you launch the DQ Tool, you will see the main menu at the top of the page. This is used to select between several different functionalities. The app permits you to quickly generate an Annual Report that is very similar to the report produced by the Excel-based DQR tool. But before we do that, however, let's explore some other features of the app. Notice that the word "Dashboard" is highlighted. And the Dashboard function is selected by default.



4. **Quickly review the Dashboard menu:** The Dashboard has a set of tabs that are used to select from several different pages. The Completeness page is shown by default.
5. **Quickly scroll through the Completeness page:** Scroll down and review the Completeness page of the data quality dashboard. The page shows completeness for each of 5 "core datasets". *Note: for the training instance of DHIS2 the immunization dataset is quite incomplete. It is best to review the quality of the ANC data which are more complete.*
 - a. **Review the graphs on the left side of the page:** The graphs on the left show, for each dataset, the completeness and timeliness by month for each of the last 12 months.
 - b. **Review the graphs on the right side of the page:** The graphs on the right show, for each dataset, the completeness by region for the last month of the period being analyzed. Completeness for the most recent month is actually a measure of timeliness. It is often more

revealing to show completeness for the month before last. Click on the menu icon () in the upper right of the screen to show a menu on the right of the screen.



The reporting period that is charted can be changed by clicking on a different month.

Data
[Core]

Period

<	2017	>
January	February	March
April	May	June
July	August	September
October	November	December

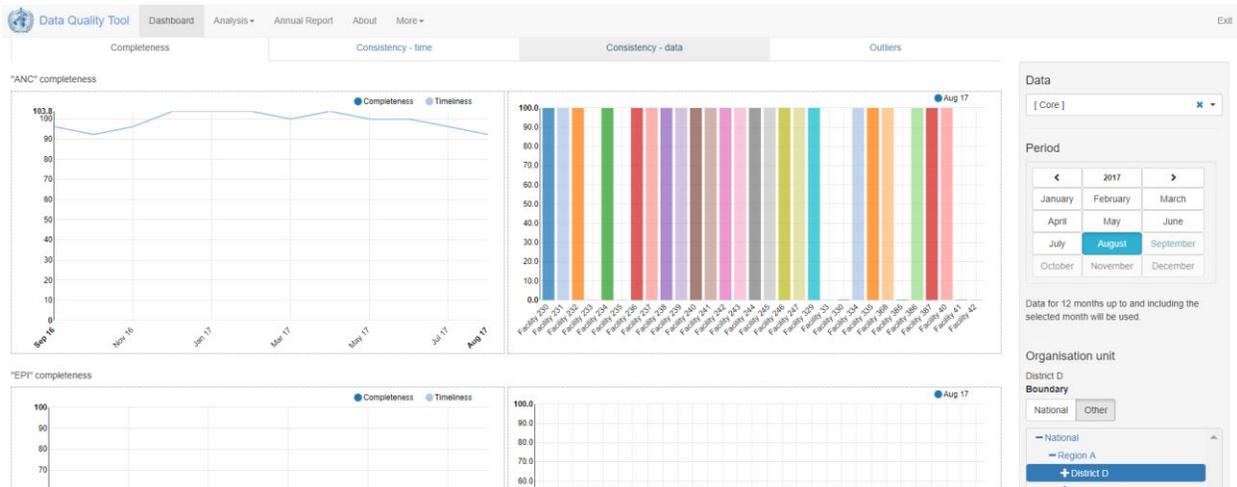
Data for 12 months up to and including the selected month will be used.

Organisation unit
National
Boundary
National Other

Disaggregation
Region

Selected
National > Region

- c. **Change the “Organisation unit” that is being analyzed:** By default, the DQ Tool analyzes all data nationwide, disaggregated by Region. In the “Organization unit” section of the menu at the right, notice that “National” is highlighted. The DQ Tool can also be used at district level with data disaggregated by individual health facility. In the Organisation unit section of the menu, click on and select “District D” of Region A.



The graphs on the left now show the 12 month trend in the reporting completeness of District D. By default, the graphs on the right shows results disaggregated for one level below – facility.
Question: Which individual facilities have not yet reported ANC data for last month?

- 6. Generate the Annual Report:** Go to the menu at the top of the DQ Tool and click on the Annual Report. Use the drop down menus to select “[Core]” under Data -- Group,

select “2016” under Period – Year, and select “District” under Organization unit - Disaggregation.

Click on “Generate” to view the report. Notice the horizontal bar that appears to track progress with “Loading data”. This means that DHIS2 data are being loaded into the app to be analyzed for data quality. Depending upon your server, this process of loading can be either very quick or quite slow – a lot of patience may be required.

- 7. Quickly review the SUMMARY of the Annual Report:** The report begins with a SUMMARY section followed by a more detailed report. Quickly scroll through the SUMMARY and note the sub-sections devoted to Domain 1 (Completeness), Domain 2 (Internal consistency), Domain 3 (External comparisons) and Domain 4 (Consistency of population data).
- 8. Review the detailed report on 1a: Completeness of facility reporting:** The detailed report begins about 1/3rd of the way through the Annual Report.

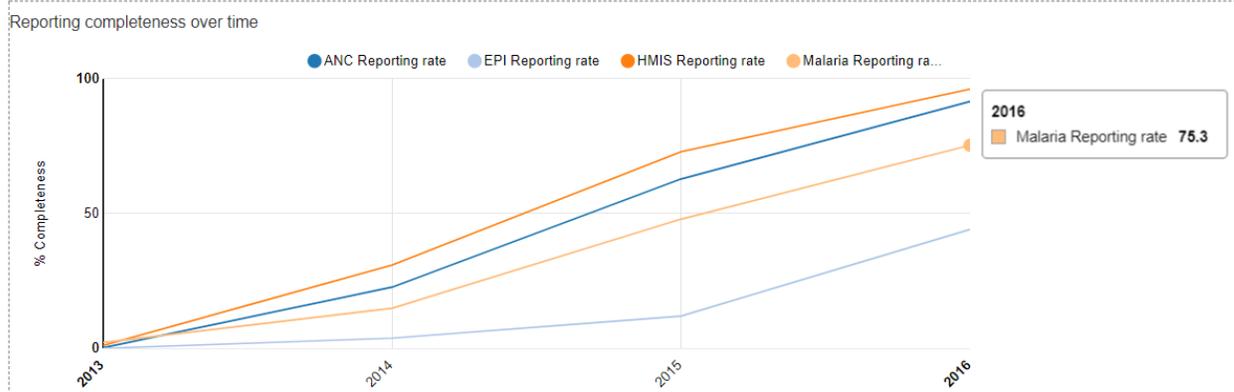
DOMAIN 1: COMPLETENESS OF REPORTING					
1a: Completeness of facility reporting					
<i>The percentage of expected reports that have been entered and completed.</i>					
Data set	Quality threshold	Overall score	District with divergent score		
			Number	Percent	Name
ANC Reporting rate	90%	91.5%	7	31.8%	District B, District E, District M, District O, District P, District Q, District S
EPI Reporting rate	90%	44.1%	18	81.8%	District A, District B, District D, District E, District F, District G, District H, District I, District K, C
HMIS Reporting rate	90%	96.1%	3	13.6%	District B, District E, District O
Malaria Reporting rate	90%	75.3%	14	63.6%	District B, District D, District E, District F, District H, District I, District J, District K, District L, D

Interpretations

Notice that table 1a shows an “Overall score” for each of the core datasets. This is the nationwide completeness of reporting for the dataset, during 2016. The column labeled “Percent” shows the percent of these districts which had completeness below the threshold. **Question:** What is the threshold for ANC Reporting rate? How many and what percentage of districts had a completeness below this threshold? Which specific districts had completeness below this threshold?

- 9. Type something in the box for Interpretation:** Note the box where interpretations and notes on follow-up actions can be typed. Type something into this box [e.g. “The following districts have facility reporting completeness below 90%”]

10. **Review the graph of Reporting completeness over time:** Scroll down to the graph showing “Reporting completeness over time”. If you hover the cursor over a point on these graphs, a box appears to tell you the exact value.



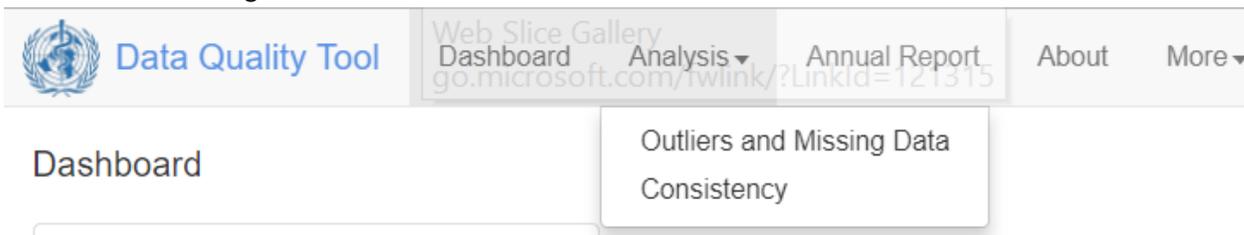
For robust analysis, a dataset should be at least 75% complete. If the completeness of a dataset changes by more than 5%, then the data should be adjusted before analyzing for a year-to-year trend. **Questions:** 1) For which datasets can robust analysis of 2016 data be performed? 2) To assess for trend from 2015 to 2016, should the ANC data be adjusted for a change in completeness?

11. **Review table 2a: Extreme outliers:**

DOMAIN 2: INTERNAL CONSISTENCY OF REPORTED DATA					
2a: Extreme outliers					
<i>Extreme outliers, using the standard method. Threshold denotes the number of standard deviations from the mean. District are counted as divergent if they have one or more extreme outliers for an indicator.</i>					
Indicator	Threshold	Overall score (%)	District with divergent score		
			Number	Percent	Names
ANC 1 visits	3 SD	0.8%	2	9.1%	District E, District K
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	0.8%	2	9.1%	District M, District N
TB cases notified	3 SD	0.8%	2	9.1%	District B, District S

The table shows the Threshold for an extreme outlier -- a monthly value that is more than 3 SD (standard deviations) from the mean monthly value for the period being review. **Question:** For how many districts was there an extreme outlier in the 2016 ANC data?

12. **Use the Analysis function to generate a list of outliers and missing values:** Let’s investigate these extreme outliers in the ANC 1 data. To do that, let’s temporarily exit the Annual Report and use the Analysis function of the tool. Click on “Analysis” in the DQ Tool menu at the top, then select “Outliers and Missing Data”.



A menu appears for setting the outlier analysis function:

Outliers and Missing Data

Data

Data set
 Data set... [Dropdown] Totals Details

Select data element... [Text Box]

Data elements
 ANC [Dropdown] Totals Details

ANC 1 visits [Text Box]

Indicators
 Indicator group... [Dropdown]

Period

Orgunit

Under Data, find the section for Data elements (skip the section for Data set), then click on the down arrow next to the “Data element group” to make a drop down menu appear. Scroll down to find and select “ANC”. This narrows down the data elements to those appearing on the ANC form. Next, click in the empty box (with grey text stating “Select data element”) just below the box you have just filled. A drop down menu will appear. Find and select “ANC 1 visits” to analyze only the data for that one data element. If you wanted to, you could again

click in this cell then select and add more data elements to analyze their data as well. However, for this exercise we will analyze the data for only a single data element. Under Period, click on “By year” then select 2016. Under Orgunit, set Disaggregation to District. Now click “Analyze”. After a delay, a table of Results will appear.

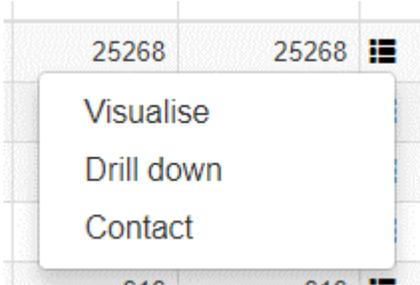
Result

Options Download Previous

Hierarchy		Data	Dec 15	Jan 16	Feb 16	Mar 16	Apr 16	May 16	Jun 16	Jul 16	Aug 16	Sep 16	Oct 16	Nov 16	Weight		
Region	Unit														Missing	Outlier	Total
Region B	District E	ANC 1 visits	6858.0	6780.0	7519.0	5927.0	6613.0	31494.0	4866.0	6016.0	6803.0	5280.0	5794.0	6026.0	0	25268	25268
Region C	District S	ANC 1 visits	3723.0	5632.0	3758.0	4016.0	3377.0	3251.0	3222.0	3039.0	3323.0	3751.0	3108.0	3296.0	0	2190	2190
Region B	District K	ANC 1 visits	1830.0	1852.0	3809.0	1794.0	1652.0	1592.0	1681.0	1489.0	2011.0	1818.0	1722.0	1791.0	0	2061	2061
Region C	District C	ANC 1 visits	2611.0	2789.0	2677.0	2629.0	2648.0	2509.0	3329.0	2325.0	2973.0	2707.0	2761.0	2672.0	0	665	665
Region D	District B	ANC 1 visits	1989.0	1855.0	1964.0	1774.0	1492.0	1044.0	1440.0	1243.0	1611.0	1639.0	1617.0	1633.0	0	616	616
Region A	District D	ANC 1 visits	2350.0	2487.0	2603.0	2418.0	2234.0	1731.0	2214.0	2341.0	2143.0	2497.0	2292.0	2058.0	0	600	600
Region C	District L	ANC 1 visits	2038.0	1976.0	2030.0	2114.0	2084.0	1947.0	2214.0	2361.0	2226.0	2170.0	1578.0	2161.0	0	542	542
Region D	District G	ANC 1 visits	929.0	987.0	1096.0	966.0	1058.0	993.0	970.0	1053.0	1183.0	941.0	975.0	1057.0	0	181	181

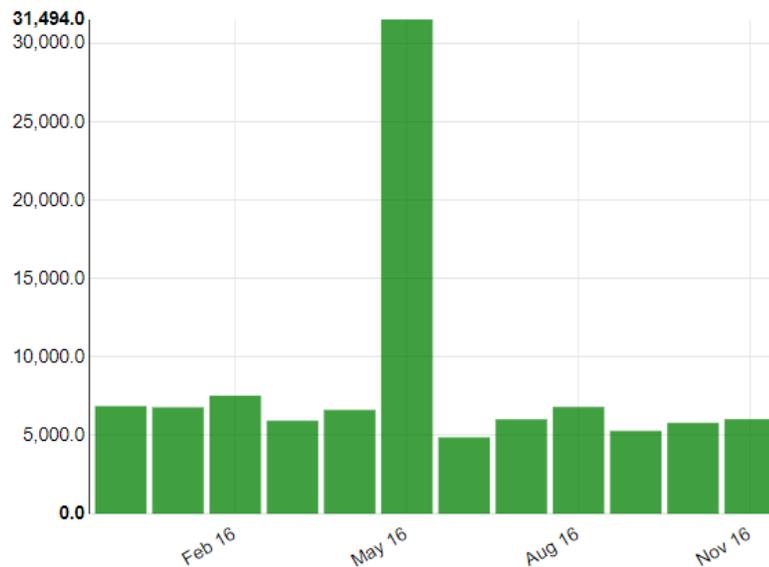
- Review the table of outliers and missing values:** In the table of Results, outliers and missing values are highlighted in red. The outliers and missing values are sorted in order of their “Total Weight” which is shown in the column entitled “Total”. The Total Weight is the difference between an abnormal monthly value and the average monthly value. There are separate rows for each district which reported an outlier.
- Use the DQ Tool to investigate an outlier:** Consider the value (31,494) highlighted in red in the first row of the table (for District E of Region B). Compare this value with the other values in the row (between 4,800 and 7,500). **Question:** Do you think this outlier is the result of a real change in the number of ANC 1st visits? Let us use the DQ Tool to investigate this outlier:

- a. To investigate the problem further, click on the menu icon () at the end of the row and select “Visualize” from the drop down menu.



A bar chart will appear.

District E
ANC 1 visits



Question: what does the bar chart show? Click “Close” to close the chart.

- b. For further investigation of the same outlier, click again on the same menu icon and then select “Drill down”. A notice of “Loading data...” appears above the Result table. Wait while the data is loaded and analyzed.

Result

Options Download Previous

Hierarchy		Unit	Data	Dec 15	Jan 16	Feb 16	Mar 16	Apr 16	May 16	Jun 16	Jul 16	Aug 16	Sep 16	Oct 16	Nov 16	Weight		
Region	District															Missing	Outlier	Total
Region B	District E	Facility 599	ANC 1 visits	24.0	28.0	28.0	34.0	28.0	25125.0	22.0	18.0	23.0	16.0	24.0	23.0	0	25101	25101
Region B	District F	Facility 777	ANC 1 visits	1087.0	1248.0	1306.0	1204.0	1080.0	964.0	1128.0	1247.0	1124.0	1076.0	2252	0	2252		

Questions: How has the table changed? What does each row represent? What do you think is the cause of the extreme outlier?

- How to further investigate an outlier and take action to correct any errors:** **Question:** What would you do if you found an obvious error in your data? The DQ Tool offers a way to quickly contact staff at district level. Hit the “Previous” button to return to a table that shows one row for each district. Use the menu at the end of the first row to select “Contact”.
- Interpret the overall findings on extreme outliers:** Return to the Annual report with results disaggregated by district. Return to the detailed report on 2a: Extreme outliers (about 1/2 of the way through the Annual report). Review the findings in table 2a.

DOMAIN 2: INTERNAL CONSISTENCY OF REPORTED DATA

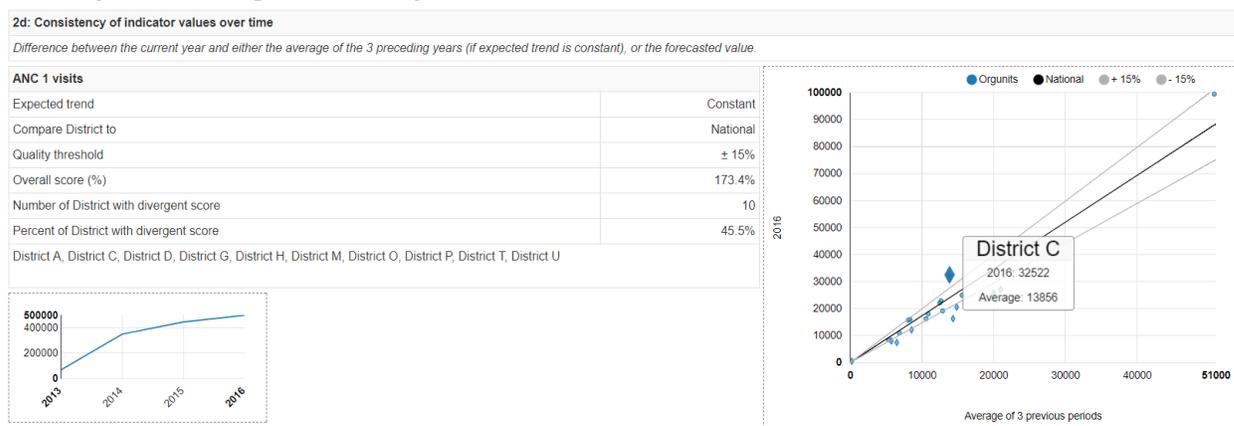
2a: Extreme outliers

Extreme outliers, using the standard method. Threshold denotes the number of standard deviations from the mean.

Indicator	Threshold	Overall score (%)	District with divergent score		
			Number	Percent	Names
ANC 1 visits	3 SD	0.8%	2	9.1%	District E, District K
Confirmed malaria cases	3 SD	0%	0	0%	

Question: the “Overall score” for ANC 1 visits was 0.8%. What does this mean? How was it calculated? What is meant by the figure of 9.1%?

- Review 2d: Consistency of indicator values over time:** Scroll down to section 2d (about 60% of the way through the Annual Report). This section compares, for the 5 core indicators, the total value reported in 2016 to the average annual total value for 2013 to 2015. When values of an indicator are consistent from year to year, this is reassuring. If values jump up and down erratically from year to year it can be a sign of real changes in program performance (e.g. as a result of a stock out). However, inconsistency from year to year may be a sign of a data quality problem. **Review the scatterplot showing consistency over time of ANC 1 visits:**



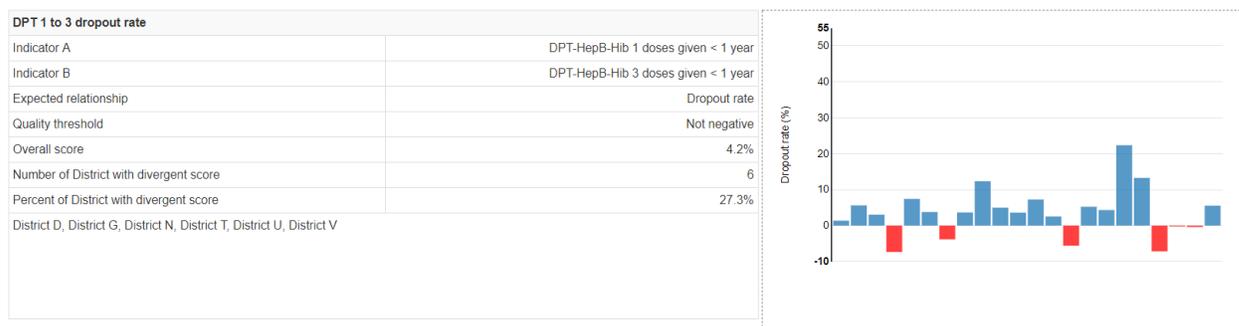
Each dot represents the data from one district. The position of the dot on the vertical axis is the number of 1st ANC visits reported by the district in 2016. The position of the dot on the horizontal axis is the average annual number of 1st ANC visits reported during 2013 to 2015. Place the cursor over the dot that is furthest from the upper gray line. It shows that District C reported 32,522 first ANC visits in 2016 but an average of only 13,856 first ANC visits in the 3 previous years. **Question:** Is District C included in the list of districts with a divergent score?

- i. **Review the diagonal black line**– Based upon the national ratio, the black line indicates the expected relationship between data reported by a District in 2016 and the average values reported by the same district from 2013 to 2016. The line starts at the origin (0,0) and runs diagonally upwards. **Question:** What is the slope of this line?
- ii. **Review the diagonal grey lines** – These lines also start at the origin and also run diagonally upwards. For points in the same position on the horizontal axis, the values on the upper grey lines are 15% greater than values on the black line. The values on the lower grey line are 15% less than the values on the black line. Hence, the grey lines show the thresholds for district values that are within 15% of their expected value. The table notes that the “Number of Districts with a divergent score” was 10. These 10 districts are represented by dots that fall either above the upper grey line or below the lower grey line.
- iii. **Interpret your findings** -- Ask yourself whether there have been any sub-national changes in program operations (e.g. stock outs, new interventions), changes in the health system (e.g. disruptions in clinic operations; splitting of an old district into two new districts) or in changes in reporting completeness that would explain the inconsistency over time in some districts. Type your answer into the box for Interpretations.

18. **Review section 2e: Consistency between related indicators** – In some instances, the value of one indicator is closely linked to the value of another indicator.

A good example of this is the DPT1 to DPT 3 dropout rate. $DPT\ 1\ to\ DPT\ 3\ dropout = (DPT\ 1\ minus\ DPT\ 3) / DPT\ 1$. The third dose of DPT cannot be administered until a child has received the first and second dose of DPT. For this reason, the value of DPT 3 should be less than or at most equal to the value of DPT1. When this is not true, this often indicates a problem with data quality – especially if it is true for a district or higher level and true for a full year.

Review the graph and table in section 2e that is devoted to the DPT1 to 3 dropout rate.



This graphs show a way to study the relationship between two indicators when the first indicator should necessarily be greater than the second indicator. The “Overall score” is the national dropout rate. A lower dropout rate is a better dropout rate – as long as the dropout rate is not negative. **Question:** How can the DPT 1 to DPT3 dropout rate be negative? [A: At district level, when data are compiled for an entire year, DPT 1 should always be greater than DPT 3. When DPT 3 is greater than DPT 1 it almost always a sign of a data quality problem. It may mean that there are incentives for health staff to misclassify first and second doses of DPT as third doses of DPT.] **Question:** Type in your interpretations of the DPT 1 to 3 dropout rate.