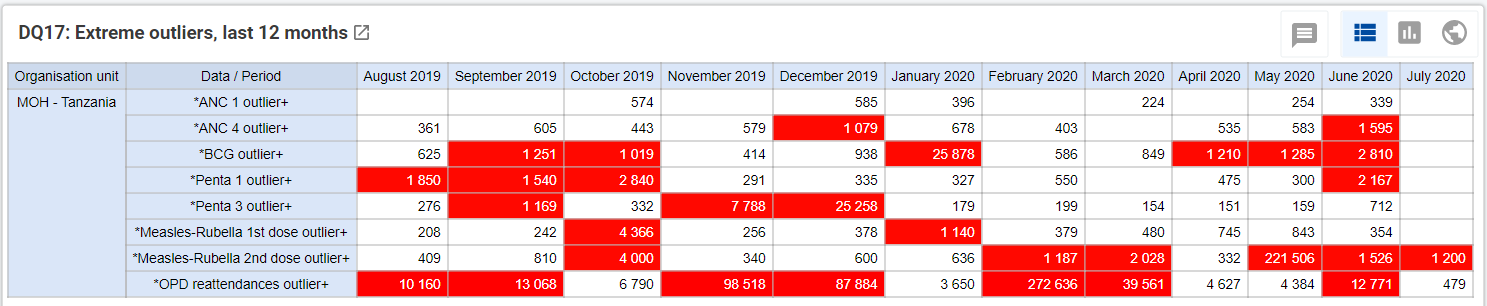
**How to use Predictor rules to visualize outliers**

**Introduction**

* 1. **Why use Predictor to identify extreme outliers** – “Extreme outliers” are values which are highly suspicious and which need to be double checked for accuracy. These are very different from the values normally reported by a health facility. If extreme outliers are found to be erroneous, then they should be edited.

The WHO Data Quality Tool (DQ Tool) does an excellent job of rapidly identifying extreme outliers. However, a user will not be aware that there is an extreme outlier unless they launch the WHO DQ Tool. At present, findings from the WHO DQ Tool cannot be directly presented on a standard dashboard.

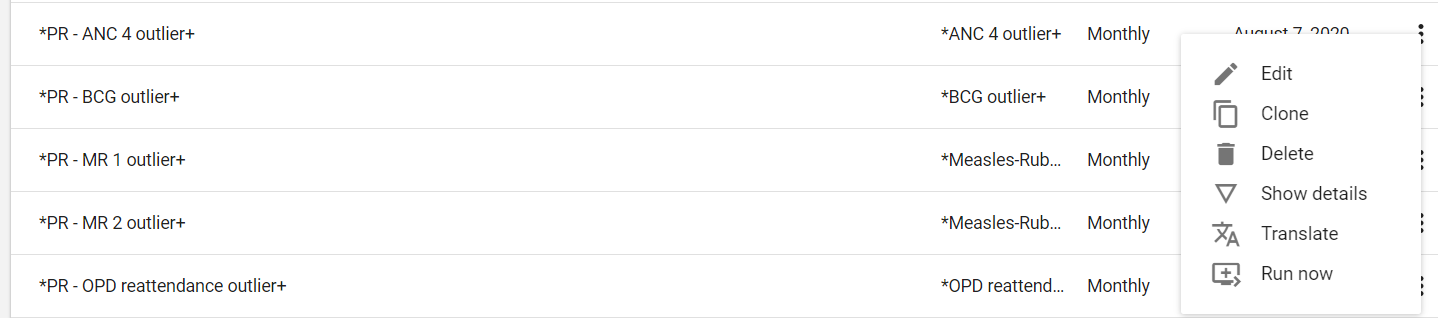
Predictor rules not only identify extreme outliers, but they generate new data on those outliers which can be presented on a standard dashboard. Thus, ***a Predictor rule can make extreme outliers more visible to users***. An example of this is the Pivot table shown below which appears on the dashboard named “\*1. Data quality”. This table shows, for 8 different indicators, the largest extreme outlier reported by any single health facility, during the last 12 months. The cells with the very large outliers are highlighted in red.



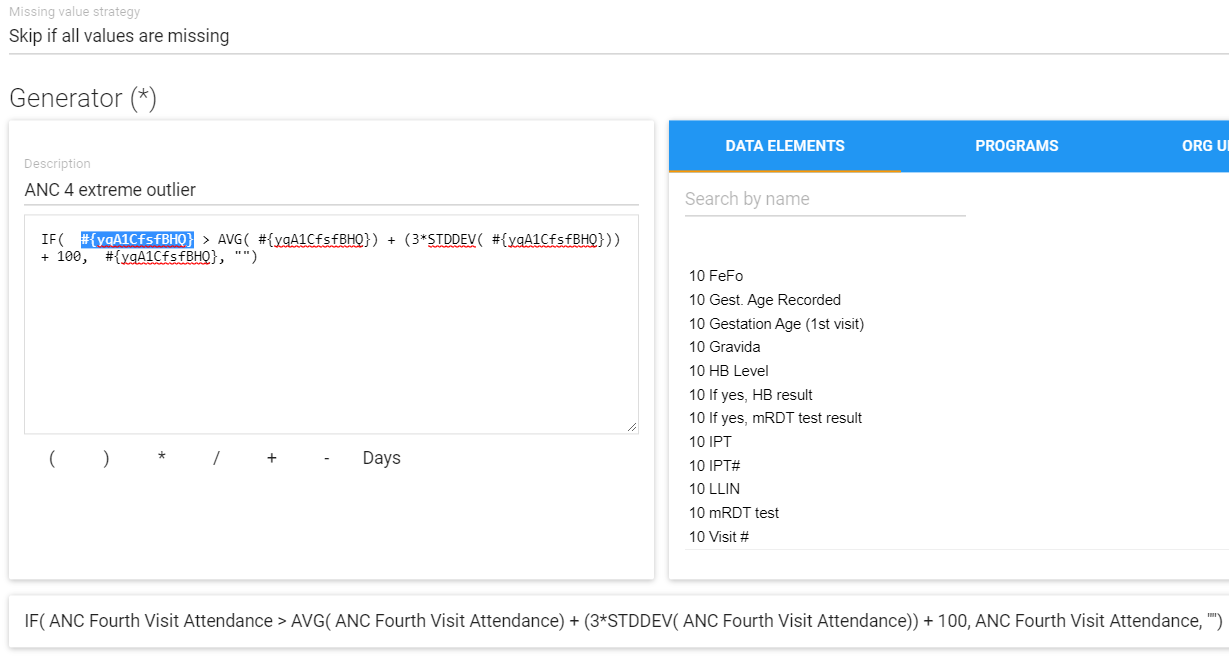
* 1. **The steps involved**
     1. Configure a new data element and a new dataset
     2. Configure a new Predictor rule
     3. Configure Scheduler
     4. Use Data Export to confirm that new data have been generated
     5. Update the Analytics tables
     6. Configure the pivot table to visualize the outlier values
  2. **The key to successful use of Predictor** -- For success, the following instructions must be followed with great attention to the details. At key stages of the process (e.g. after configuration of the Predictor rule; after configuration of the Data Validation rule), it is advisable to pause to verify that the configuration is working correctly. Instructions are provided on how to do this.

More than an hour may be required to run a Predictor rule. In addition, even after the Predictor rule has generated new data, these data may not be visible until the Analytics have been updated – another step for which more than an hour may be required. As a result, considerable patience and several prolonged pauses in the workflow are likely to be required the first time the guidance is followed. However, once you have successfully configured for one data element/indicator and confirmed that the resulting outputs are correct, the configuration process for each subsequent data elements/indicator can involve cloning and should take much less time.

1. **Configuring new data elements and a new dataset**. Before a Predictor rule can be configured, a new data element must be configured which will be used to store the value of the outlier threshold. Use Maintenance – Data element to define a new data element
   1. **Configure a data element for outliers from the last 12 month** -- Give your new data element a name such as “\*ANC 4 outliers+”. Set Domain type to “Aggregate”. Set Value type to “Number”. Set Aggregation type to “Max”. Set category combination to “None”. Leave Aggregation levels blank. By setting Aggregation type to “Max” it is possible to configure a pivot table which will show the largest outlier reported by a single health facility in a single month.
   2. **Configure additional data elements depending upon how you want to visualize the suspicious data** -- Predictor rules can also be used to visualize outliers in other ways besides the one described in this document: validation rules, maps showing the locations of health facilities reporting extreme outliers, etc … Refer to separate documents for instructions on how to do this.
   3. **Add the new data element to the dataset named “\*Outliers (from PRs) for data export”** – Use the maintenance app to edit the dataset named “\*Outliers (from PRs) for data export”. Delete any data elements which were previously included in the dataset. Add the new data element and save the edited dataset. The Data Export app can now be used to export the values of the new data element to confirm that the outlier data have been generated.
2. **Configure Predictor rules** 
   1. Go to Maintenance – Other – Predictor
   2. Click on the three dots of the “Actions” icon at the end of the row for “PR – ANC 4 outlier+” and select “Clone



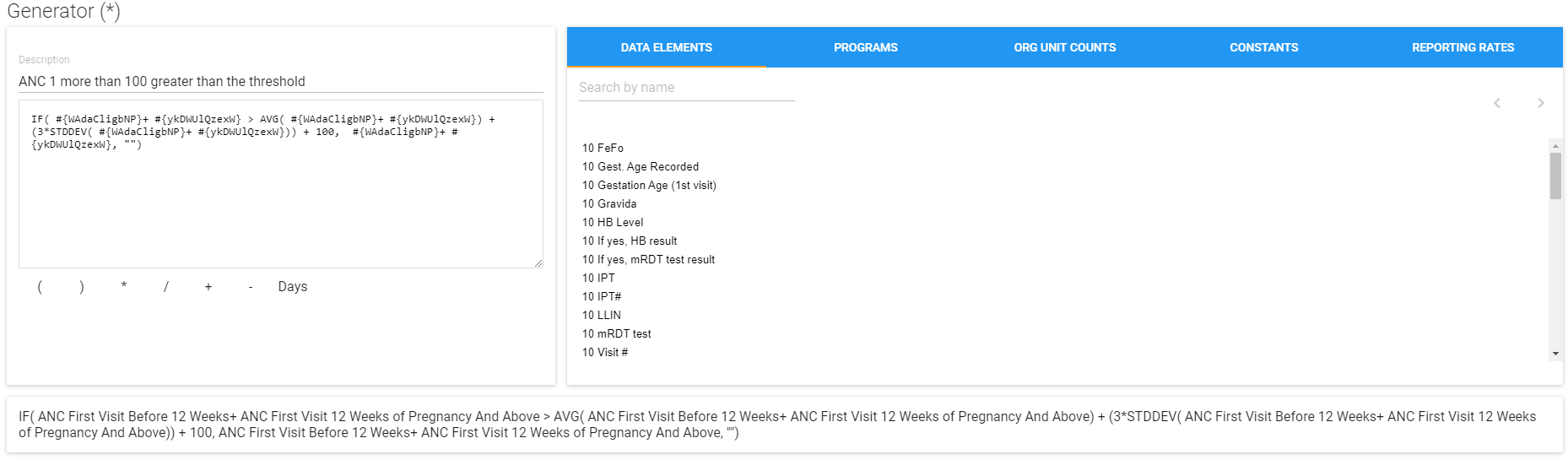
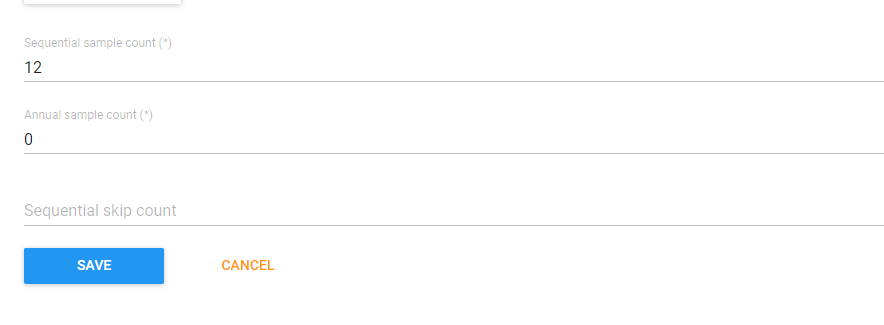
* 1. Click on “Output data element (\*)” and a window will appear with the names of all data existing data elements. On the “filter list” line type the exact name of the new data element. Find and click on the new data element in the dropdown list that appears. The name of the new data element should now appear on the line beneath “Output data element (\*)”.
  2. Leave Period type to “Monthly” and Organisation unit level to set to “Facilities”.
  3. Click on Generator (\*). Here is where you enter the formula for generating the data for the new data element. Lease Missing value strategy set to “Skip if all values are missing”.



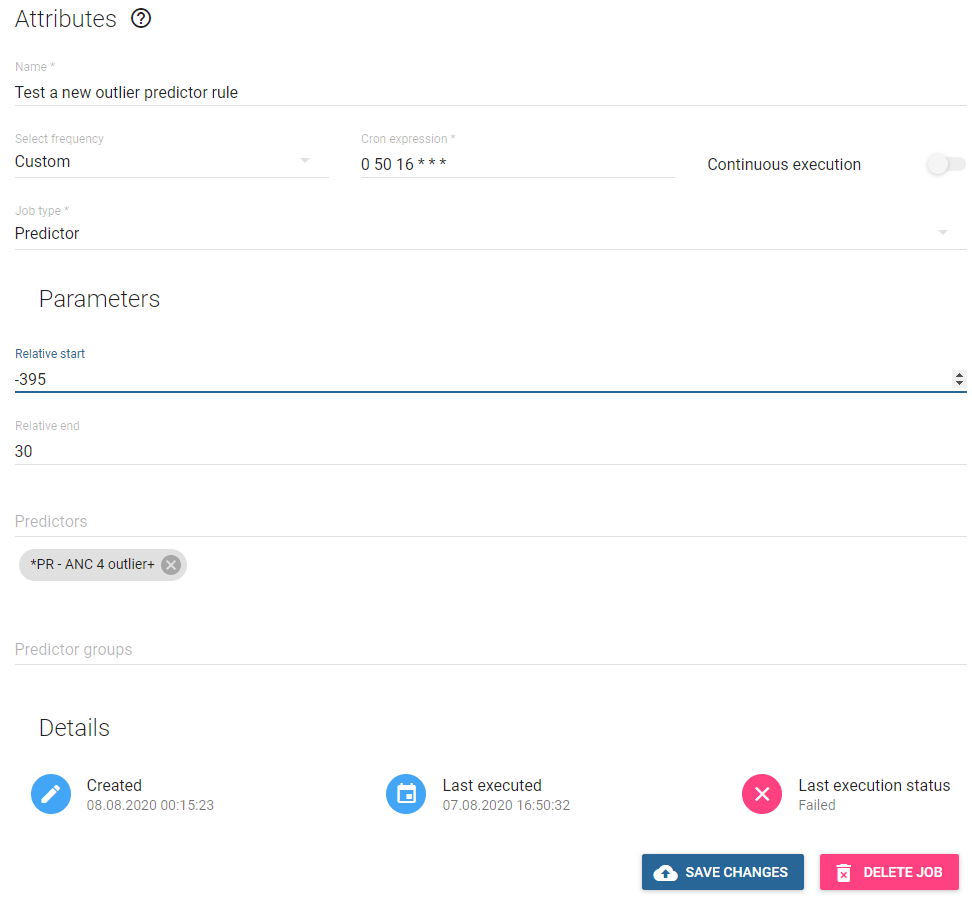
The formula has been configured properly to identify extreme outlier values of the indicator ANC Fourth Visit Attendance (“ANC Hudhurio la nne wajawazito wote “; “Wajawazito Hudhurio la nne”; data element id = #{yqA1CfsfBHQ} ). The Generator expression can be interpreted as follows:

* The threshold for an extreme outlier is calculated by taking the average value of the indicator (i.e. ANC 4) then adding 3 times the standard deviation of the indicator then adding 100.
* If, for a given month, the value of an indicator is greater than the threshold then the value is an extreme outlier and the generator will give the new data element a value for that month equal to the value of the indicator.
* If, however the value of the indicator is less than or equal to the threshold then the generator will give the new data element a blank or missing value for that month.
  1. To configure the Generator expression for the new data element, the id for the old indicator (i.e. #{yqA1CfsfBHQ}) needs to be replaced with the id or formula for the new indicator.
  2. A good way to do this is as follows:
     1. type a “+” sign after the existing Generator formula;
     2. enter the id or formula for the new indicator;
     3. select and copy (with CTrl+C) the id or formula for the new indicator;
     4. select each instance of #{yqA1CfsfBHQ} then replace (with Ctrl+V) ot with the id or formula for the new indicator;
     5. Delect the “+” sign then check at the bottom of the screen to see if the expression is well-formed.

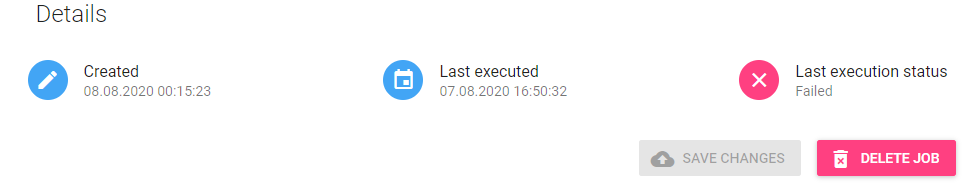
**NOTE: All statements (EX: “if” or “stddev” in 2.34 must be lower case. In versions <2.34 it can be in eather capital letters or lowercase.**

* 1. What is meant by “formula for the new indicator”? In the case of ANC Fourth Visit Attendance, the indicator is a single data element so the formula includes a single data element id ( #{yqA1CfsfBHQ} ). However, in the case of “ANC 1st visits” the indicator must be calculated from the sum of two data elements: ANC First Visit Before 12 Weeks+ ANC First Visit 12 Weeks of Pregnancy And Above. So each instance of #{yqA1CfsfBHQ} in the Generator expression must be replaced with #{WAdaCligbNP}+ #{ykDWUlQzexW}
  2. Once the expression is well-formed, click on Submit.
  3. Ignore the button for “Sample skip test”
  4. Leave Sequential sample count (\*) set to 12. This is the number of months of data which are used for the average and the standard deviation of the Generator formula. If last month is, for example, July 2020, then, to calculate the threshold, the Generator formula takes the average of the values for the same health facility for July 2019 to June 2020 then adds 3 times the standard deviation of the values for July 2019 to June 2020 then adds 100.
  5. Leave Annual sample count (\*) set to 0
  6. Leave Sequential skip count blank.
  7. Save the new Predictor rule.

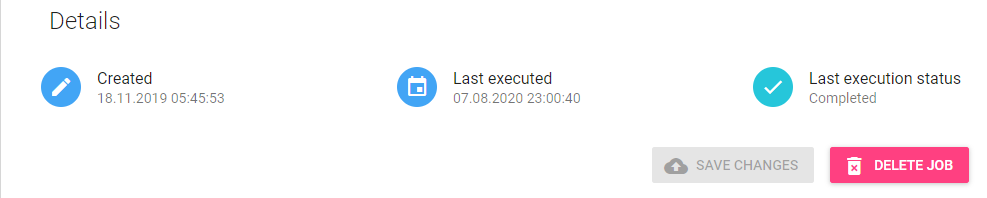
1. **Configure Scheduler**
   1. Launch the Scheduler app.
   2. To test a new Predictor rule, edit the existing job named “Test a new outlier predictor rule”. Click on the name of this job and a window will open.



* 1. Leave Name as it is.
  2. Leave Select frequency set to Custom
  3. The Cron expression determines when the job will start. Set this to 00 XX YY \* \* \* where XX is 5 minutes after the current minutes and YY is the current hour. Make sure that there is a single space between each part of this cron expression: 00 and XX and YY and \* and \* and \*. With this expression, the job will run 5 minutes after you set the cron expression.
  4. LEAVE “Continuous execution” SET TO OFF>
  5. Leave Job type\* set to Predictor
  6. Leave Relative start set to -395.
  7. Leave Relative end set to 30. With this Relative start and Relative end, the Predictor job will identify any extreme outliers reported in the last 12 months.
  8. Leave Predictor groups blank
  9. Click on SAVE CHANGES.
  10. The name of your new or modified Predictor job should now appear under “Scheduled jobs”. The Type should be shown as Predictor. The Status should be Scheduled. Next execution should say something like “07:08:2020 18:10” (if the job were configured in August of 2020). Enabled should be set to on.
  11. Wait 15 minutes, then refresh the screen. If the Predictor is still running the Scheduler will show the Status of the job as Running. If the Predictor has finished running, the status will shows as Scheduled, but the date that it is scheduled will be tomorrow.
  12. Once the Predictor has finished running, click on it and examine the “Last execution status” at the bottom of the screen. The status will either be Failed



or it will be Completed.



1. **Check that your new predictor rule is running correctly**. The best way to determine whether the PR has run correctly is to use the Data Export app to export data for the new outlier data element. Otherwise, you may have to wait until the Analytics tables have been updated overnight. Only then can the outlier threshold data be visualized with the Pivot table or Data Visualizer apps. Meanwhile, use the following steps to check your predictor rule:
   1. Launch the Data Export app. Under Organisation unit, select the national level. Under Children select Include descendants of organization unit. Under Datasets, select the “\*Outliers (from PRs) for data export” dataset. Set START DATE to the first day of the month 12 months previously. Set END DATE to today’s date. Set FORMAT to CSV. Set COMPRESSION to Uncompressed. Click on EXPORT and wait for the icon for the CSV file to appear in the lower left of the screen.
   2. Open the CSV file and check to see whether any outliers have been identified.
   3. To verify further, compare the outliers identified in the CSV file to the outliers identified with the WHO Data Quality Tool.
2. **Add the new data element to DQ17** (the pivot table of outliers).