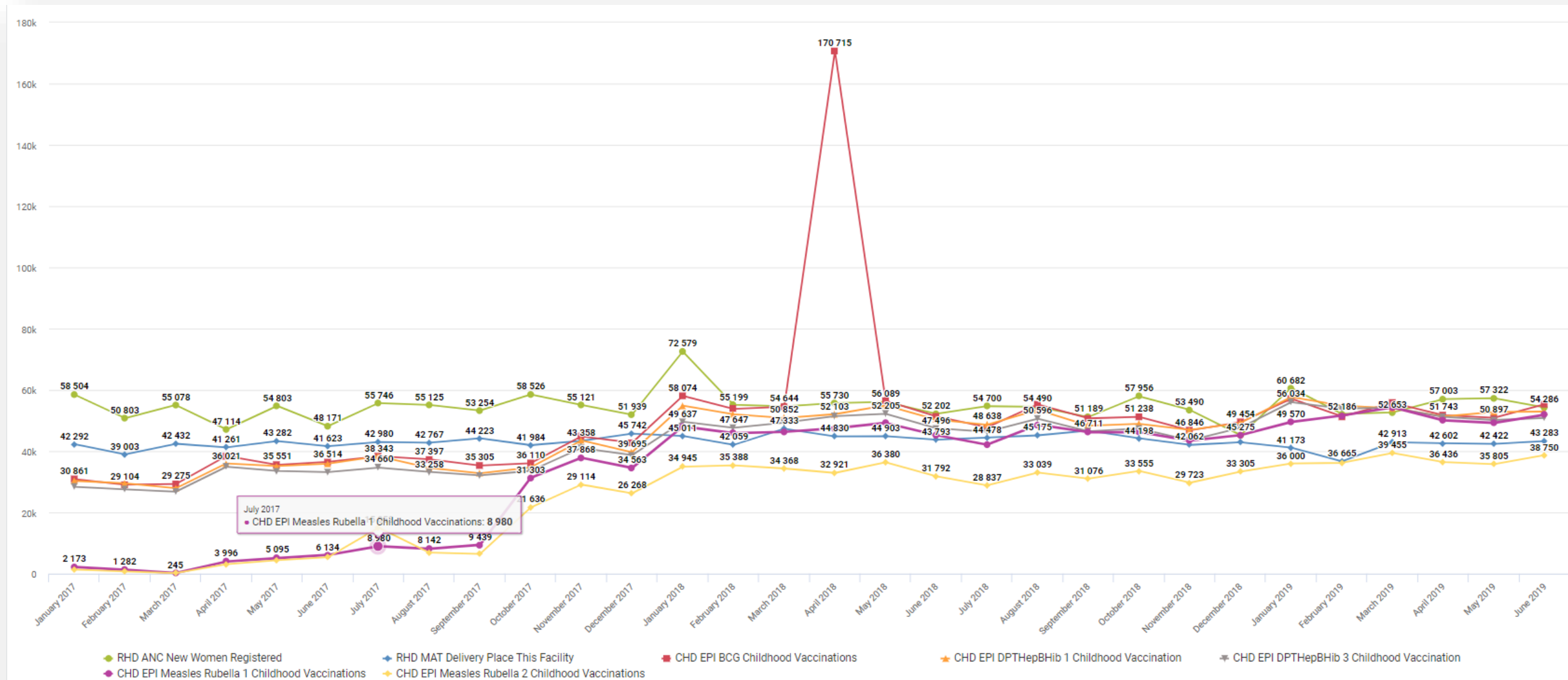
A decorative vertical bar on the left side of the slide, featuring a teal and light blue geometric pattern of overlapping rounded rectangles and circles.

Detecting Outliers in Dashboards and Validation Notifications

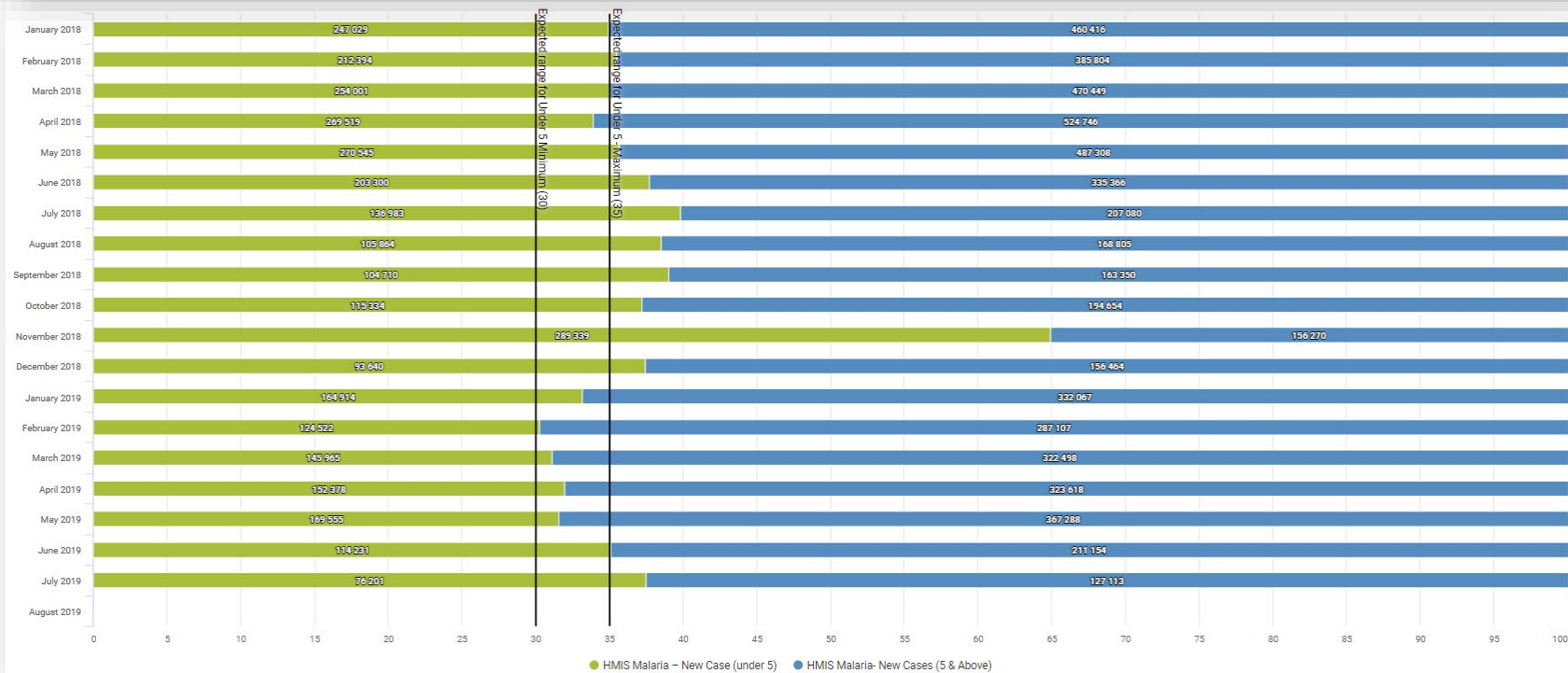
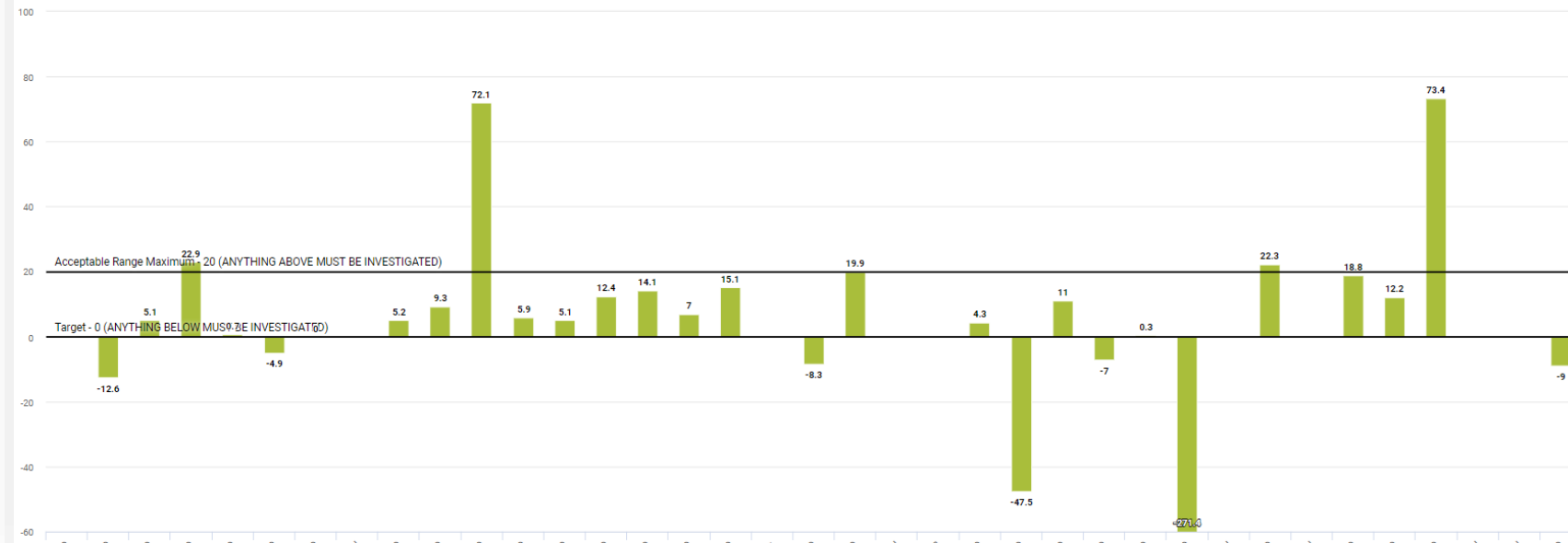
DQ Academy, Day 4 – Oct 22

Part 2

Outliers -



July 2019



Example

- <https://academy.demos.dhis2.org/dq/dhis-web-dashboard/#/>

Steps for identifying outliers from trend-line charts.

1. Isolate the period. – Usually just one month
2. Change the chart type to a bar chart
3. Isolate the data element
4. Turn on facility level in your org unit hierarchy.
5. Change the layout so org units is in category, period in filter, and data in series.

Now you should see where the outlier is!

Outliers are a fact of life

Outliers throw off national statistics - They must be caught and corrected quickly

But why should you have to dig through the data to find the outlier?
They should come to you.

From: "Rwanda HMIS Message [No reply]" <trainmutali@gmail.com>
Date: June 6, 2019 at 10:00:00 PM PDT
To: Undisclosed recipients;
Subject: [Rwanda HMIS] Validation violations as of 2019-06-07T07:00:00

Violations: High 23, medium 0, low 0

A violation of the Penta 3 outlier data validation rule was detected at Batsinda HP for 201905 on 20
Please confirm that the value is correct of Penta 3

A violation of the Penta 3 outlier data validation rule was detected at Bumbogo (ex-Gikomero I) CS f
Please confirm that the value is correct of Penta 3

A violation of the Penta 3 outlier data validation rule was detected at Cyuga HP for 201905 on 2019-
Please confirm that the value is correct of Penta 3

A violation of the Penta 3 outlier data validation rule was detected at Cyumba CS for 201905 on 2019
Please confirm that the value is correct of Penta 3

A violation of the Penta 3 outlier data validation rule was detected at Gatsata CS for 201905 on 201
Please confirm that the value is correct of Penta 3

A violation of the Penta 3 outlier data validation rule was detected at Gikore CS for 201905 on 2019
Please confirm that the value is correct of Penta 3

A violation of the Penta 3 outlier data validation rule was detected at Gitarama CS for 201905 on 20
Please confirm that the value is correct of Penta 3

A violation of the Penta 3 outlier data validation rule was detected at Gitwe CS for 201905 on 2019-
Please confirm that the value is correct of Penta 3

A violation of the Penta 3 outlier data validation rule was detected at Harmony Clin for 201905 on 2
Please confirm that the value is correct of Penta 3

A violation of the Penta 3 outlier data validation rule was detected at Jenda (nyabihu) HP for 20190
Please confirm that the value is correct of Penta 3

A violation of the Penta 3 outlier data validation rule was detected at Kirambi CS for 201905 on 201
Please confirm that the value is correct of Penta 3

A violation of the Penta 3 outlier data validation rule was detected at Kirarambogo CS for 201905 on
Please confirm that the value is correct of Penta 3

A violation of the Penta 3 outlier data validation rule was detected at Ntoma CS for 201905 on 2019-
Please confirm that the value is correct of Penta 3

A violation of the Penta 3 outlier data validation rule was detected at Nyarugunga CS for 201905 on
Please confirm that the value is correct of Penta 3

A violation of the Penta 3 outlier data validation rule was detected at Rubengera CS for 201905 on 2
Please confirm that the value is correct of Penta 3

A violation of the Penta 3 outlier data validation rule was detected at Ruheru CS for 201905 on 2019

Using Predictors in Validation Rules

- A predictor is a value calculated from previously reported values in DHIS2.28 – DHIS2.34. This can be done using indicators in DHIS2.35+
 - EX: Average Malaria incidence over the last 6 months.
 - EX: Average monthly ART consumption over the last 3 months.
 - EX: Average ANC 1 Visits with 3 standard deviation for the last 12 months – ANC 1 Outlier Threshold
 - If this value is 0 then count 1 – Count the number of facilities that reported some value.

Warning!

- Predictors are a scheduled job that consume a lot of server resources when run.
 - Do not run predictors during peak server usage hours
 - Do not use predictors if you have any server issues at all
 - Do not use predictors if you have limited storage, CPUs, or RAM

How to make a predictor

1. Make a data element to store the predicted value
2. Make the predictor
3. Put the predictor in a predictor group
4. Schedule the predictor group to run
5. Put the predictor in a validation rule

Create outlier threshold data element

- Name – **ANC 1 Outlier Threshold – *Your Name***
- Short name – Same
- Domain type – Aggregate
- Value type – Number
- Aggregation type - Sum



This object will be created with public edit and view rights

Name (*)

ANC1 Outlier Threshold

Short name (*)

ANC1 Outlier Threshold

Code

Color

SELECT COLOR

Icon

ADD ICON

Description

Form name

Domain type (*)

Aggregate

Value type (*)

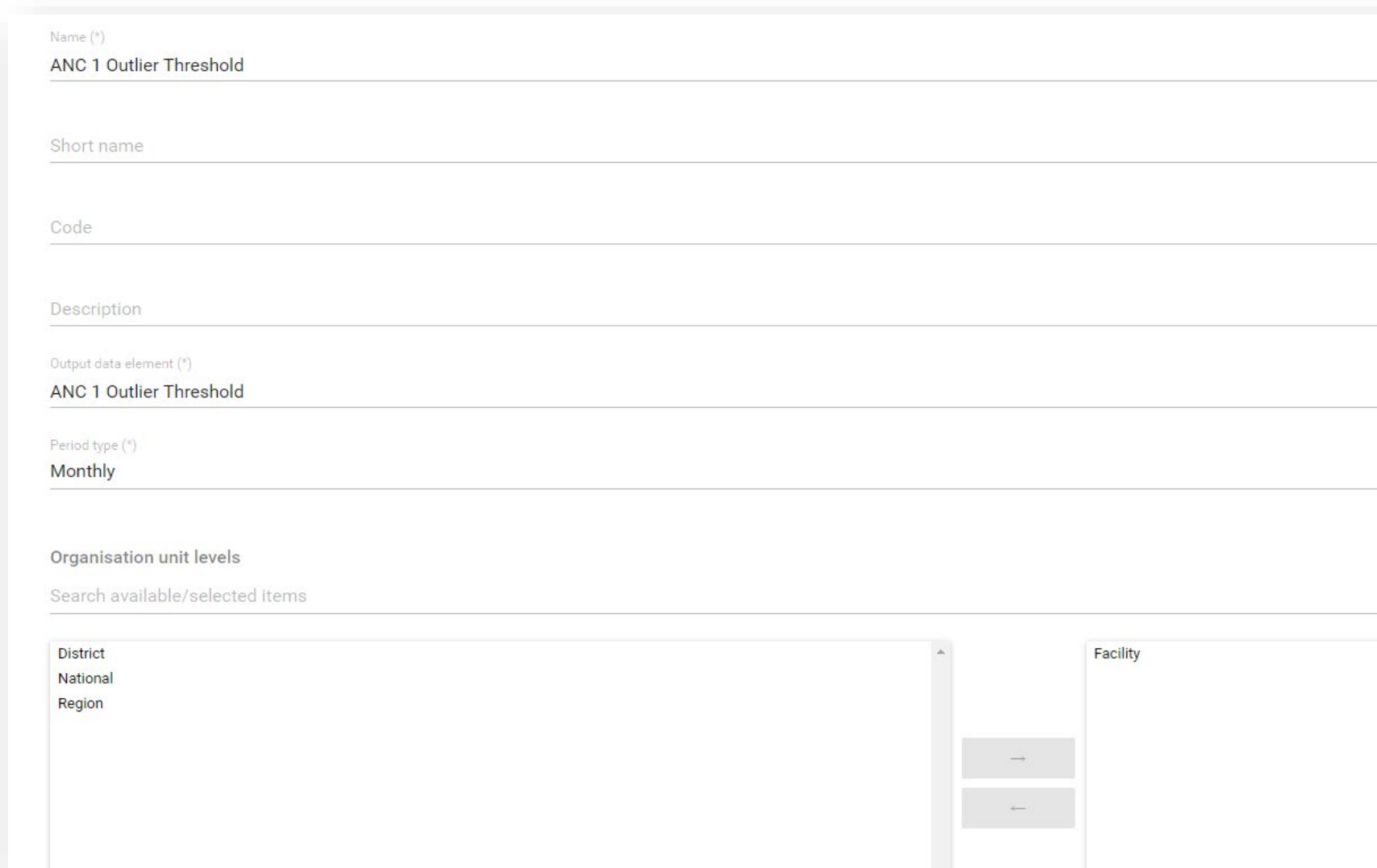
Number

Aggregation type (*)

Sum

Make the predictor

1. Name – **ANC 1 Outlier Threshold**
2. Define output DE – **ANC 1 Outlier Threshold**
3. Select Org unit level - **Facility**



The screenshot shows a web form for configuring a predictor in DHIS2. The form has several sections with labels and input fields:

- Name (*)**: The input field contains "ANC 1 Outlier Threshold".
- Short name**: An empty input field.
- Code**: An empty input field.
- Description**: An empty input field.
- Output data element (*)**: The input field contains "ANC 1 Outlier Threshold".
- Period type (*)**: The input field contains "Monthly".
- Organisation unit levels**: This section includes a search bar labeled "Search available/selected items". Below it, there are two lists of organisation units. The left list, titled "District", contains "National" and "Region". The right list, titled "Facility", is currently empty. Between the two lists are two arrows (one pointing right, one pointing left) for moving items between the lists.

Define Generator

1. Select missing value strategy: Skip if all values are missing
 - **Default is skip is any is missing?** Is this good for this?
 - No, we want the calculation even if there are missing values

Missing value strategy

Skip if all values are missing

Generator (*)

Description

ANC 1 Outlier Threshold

```
avg(#{RvArfOFKdXe}) + ( 3 * stddev(#{RvArfOFKdXe}))
```

() * / + - Days

```
avg(ANC 1 visits) + ( 3 * stddev(ANC 1 visits))
```

Valid

DATA ELEMENTS

Search by name

Albendazole 1 dose at ANC
Albendazole 1 dose at ANC expected
Albendazole 1 dose at ANC present
ANC 1 coverage (DHS 2014)
ANC 1 outliers > 1,000
ANC 1 outliers > 1,000 - 3 step
ANC 1 outliers > 1,000, last 12 months (Y/N)
ANC 1 Outlier Threshold
ANC 1 Outlier Threshold >= 1,000
ANC 1 visits
ANC 1 visits expected
ANC 1 visits present

Define Generator

2. Name – **ANC 1
Outlier Threshold**
3. Formula –
 $\text{avg}(\text{ANC1 UID}) + (3 * \text{STDDEV}(\text{ANC1 UID}))$

Missing value strategy
Skip if all values are missing

Generator (*)

Description
ANC 1 Outlier Threshold

```
avg({RvArfQFKdXe}) + ( 3 * stddev({RvArfQFKdXe}))
```

() * / + - Days

avg(ANC 1 visits) + (3 * stddev(ANC 1 visits))

Valid

DATA ELEMENTS

Search by name

- Albendazole 1 dose at ANC
- Albendazole 1 dose at ANC expected
- Albendazole 1 dose at ANC present
- ANC 1 coverage (DHS 2014)
- ANC 1 outliers > 1,000
- ANC 1 outliers > 1,000 - 3 step
- ANC 1 outliers > 1,000, last 12 months (Y/N)
- ANC 1 Outlier Threshold
- ANC 1 Outlier Threshold >= 1,000
- ANC 1 visits
- ANC 1 visits expected
- ANC 1 visits present

Other functions for predictors

Aggregate function	Means
AVG	Average (mean) value
COUNT	Count of the data values
MAX	Maximum value
MEDIAN	Median value
MIN	Minimum value
STDDEV	Standard deviation
SUM	Sum of the values

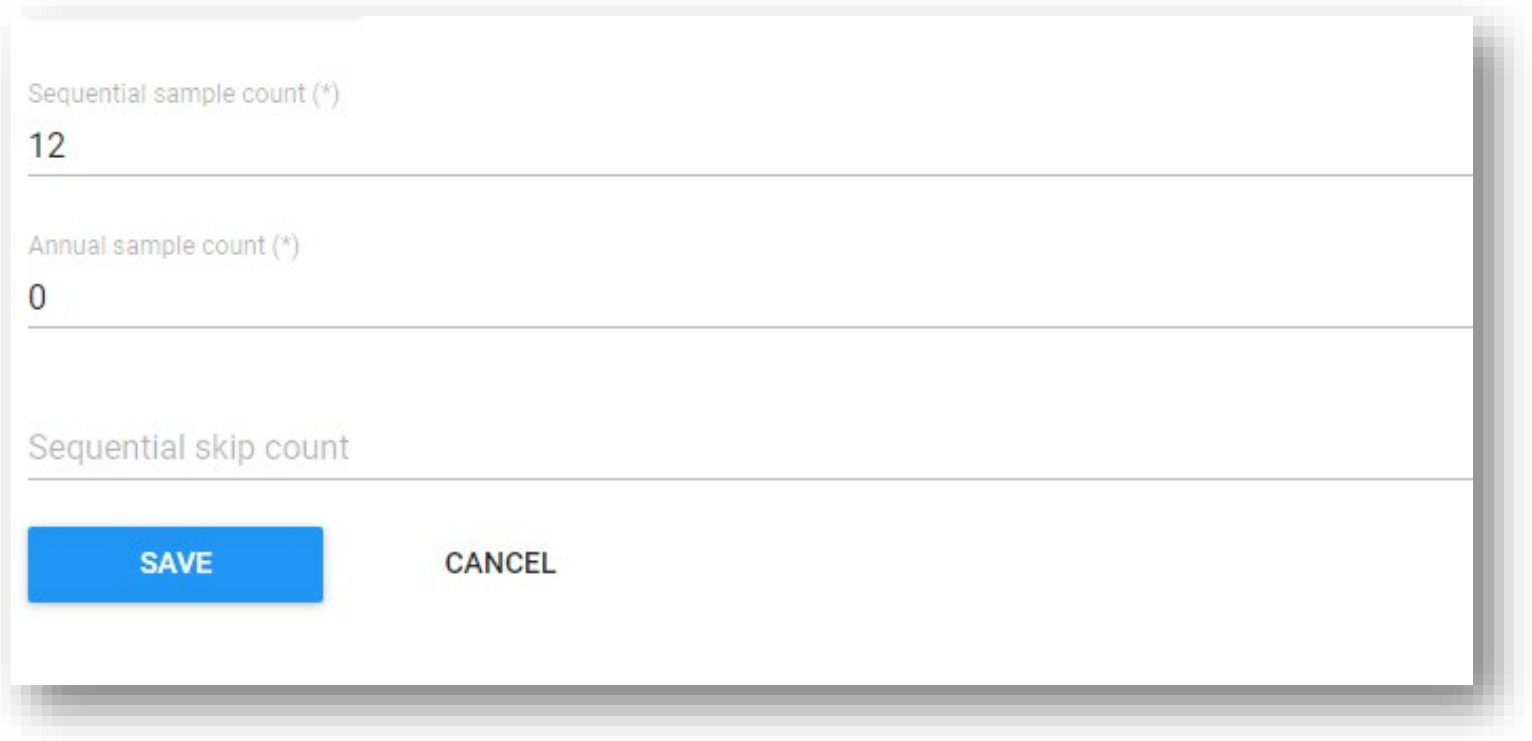
BE AWARE: the functions are case sensitivie.

THEY MUST BE IN LOWER CASE/SMALL LETTERS

EX: avg, count, max, median, min, stddev

Define counts

- Sequential sample count – 12
- Annual sample count 0
- Sequential skip count 0



Sequential sample count (*)

12

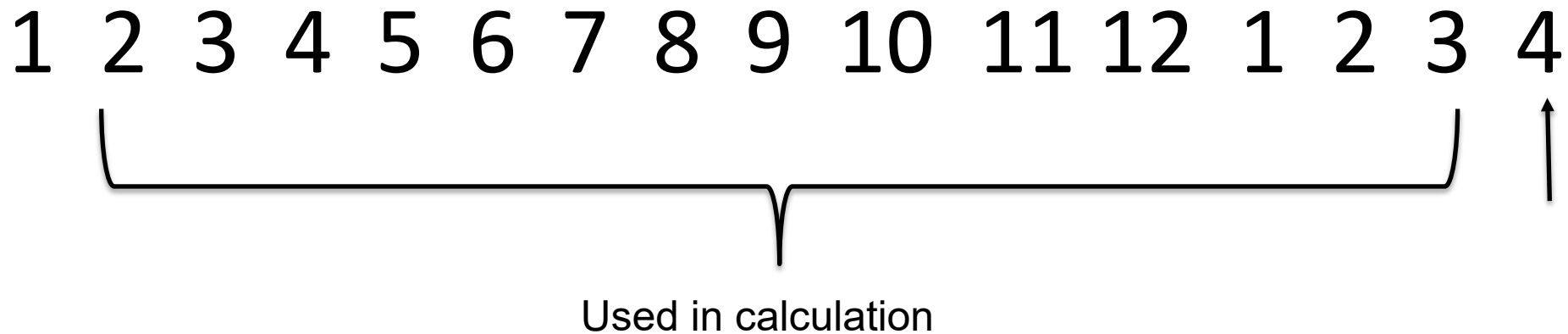
Annual sample count (*)

0

Sequential skip count

SAVE CANCEL

Sequential – 12, Annual – 0, Skip – 0



Sample count for seasonal data – Malaria

Sequential sample count is = 4

Annual sample count is = 2

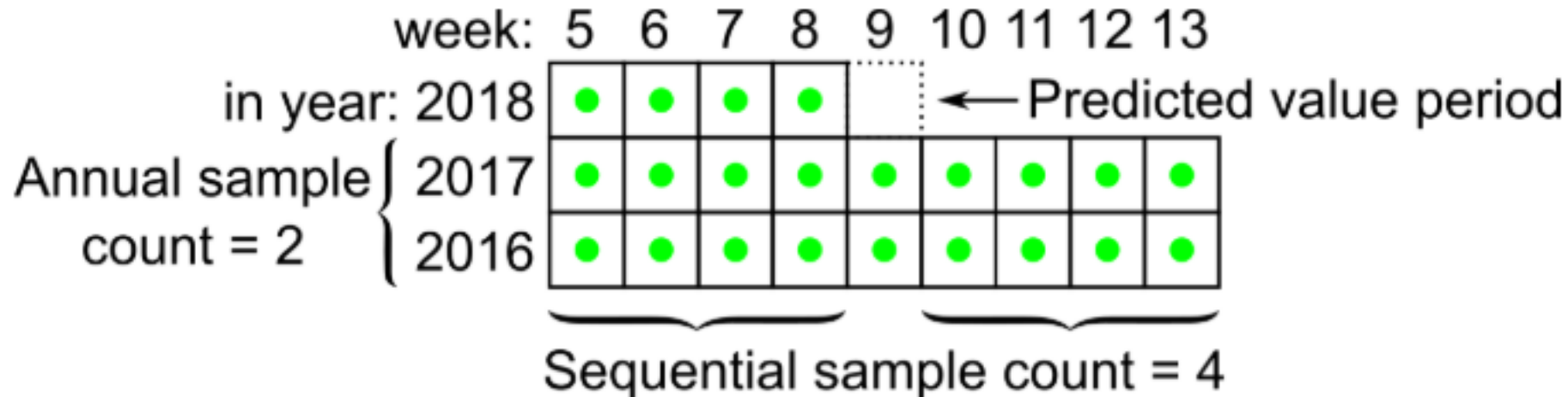
Total sampled period is 22 =

4 previous weeks

+ 8 weeks from the previous weeks over the last two years

+ 8 weeks from the period weeks over the last two years

+ 2 week 9 from the last two years



Output

Period type

Last updated



ANC 1 Outlier Thresh...

Monthly

April 1, 2010



Edit



Clone



Delete



Show details



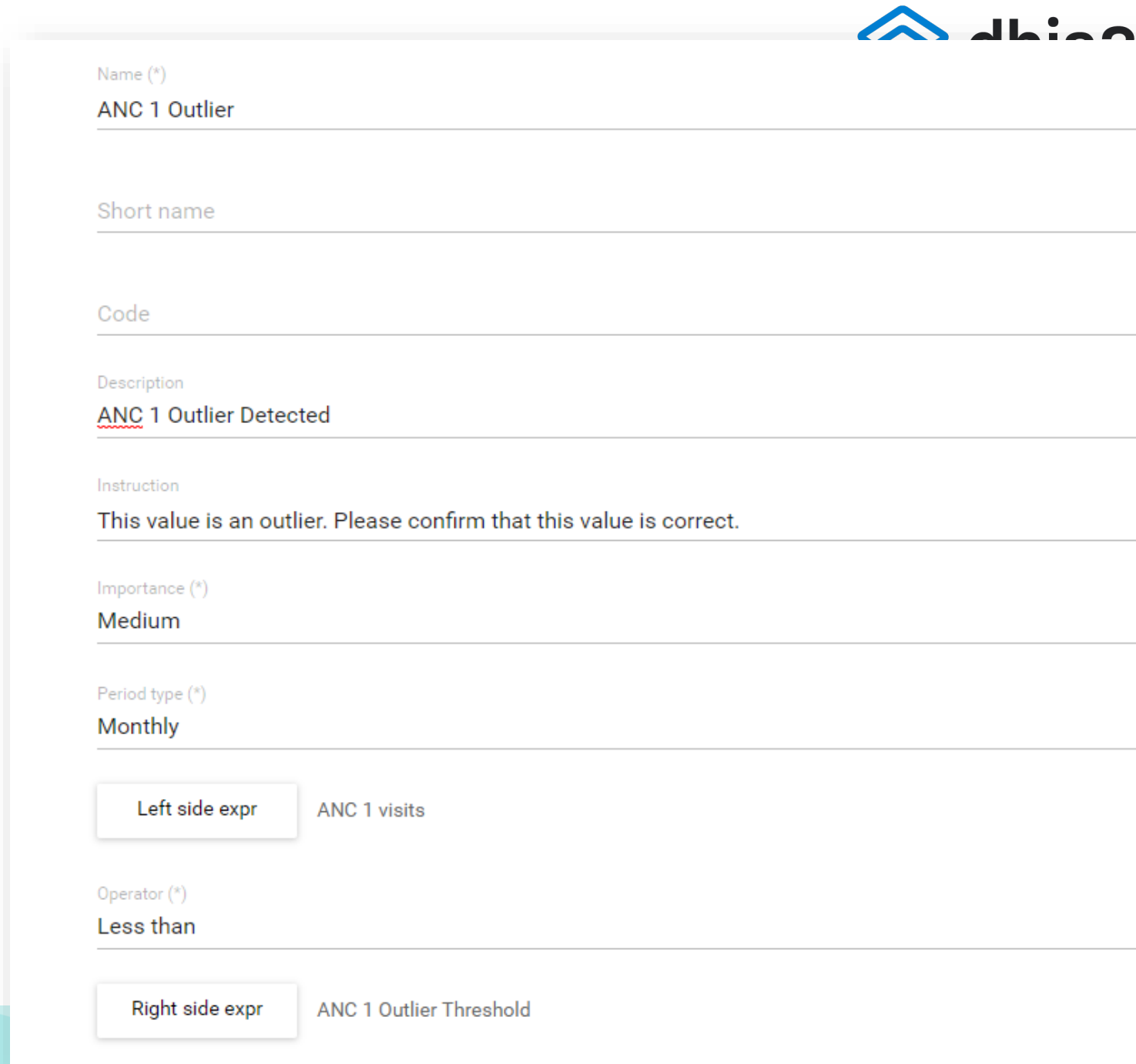
Translate



Run now

Outlier Threshold into validation rule

- Name
- Description
- Instruction
- Importance
- Period type
- Left side – **ANC 1 Visits**
- Operator – **Less than**
- Right side – **ANC 1 Outlier Threshold**



The screenshot shows a web form for configuring a validation rule. The form has several input fields and a final section for defining the rule's logic. The fields are as follows:

- Name (*)**: ANC 1 Outlier
- Short name**: (empty)
- Code**: (empty)
- Description**: ANC 1 Outlier Detected
- Instruction**: This value is an outlier. Please confirm that this value is correct.
- Importance (*)**: Medium
- Period type (*)**: Monthly
- Left side expr**: ANC 1 visits
- Operator (*)**: Less than
- Right side expr**: ANC 1 Outlier Threshold




















The form is styled with a light blue header and a white body. The 'Left side expr' and 'Right side expr' fields are highlighted with a light blue background.

Last Step

- Add it to a validation rule group and run it:

More than 500 values found, please narrow the search to see all

DOWNLOAD AS PDF DOWNLOAD AS XLS DOWNLOAD AS CSV

Organisation Unit	Period	Importance	Validation Rule	Value	Operator	Value	Details
Facility 536	October 2019	MEDIUM	ANC 1 Outlier Det...	0.00	<	0.00	
Facility 537	October 2019	MEDIUM	ANC 1 Outlier Det...	0.00	<	0.00	
Facility 539	October 2019	MEDIUM	ANC 1 Outlier Det...	0.00	<	0.00	
Facility 553	October 2019	MEDIUM	ANC 1 Outlier Det...	0.00	<	0.00	
Facility 556	October 2019	MEDIUM	ANC 1 Outlier Det...	0.00	<	0.00	
Facility 559	October 2019	MEDIUM	ANC 1 Outlier Det...	0.00	<	0.00	
Facility 570	October 2019	MEDIUM	ANC 1 Outlier Det...	0.00	<	0.00	
Facility 573	October 2019	MEDIUM	ANC 1 Outlier Det...	0.00	<	0.00	
Facility 580	October 2019	MEDIUM	ANC 1 Outlier Det...	0.00	<	0.00	
Facility 583	October 2019	MEDIUM	ANC 1 Outlier Det...	126.00	<	117.90	
Facility 585	October 2019	MEDIUM	ANC 1 Outlier Det...	0.00	<	0.00	
Facility 589	October 2019	MEDIUM	ANC 1 Outlier Det...	0.00	<	0.00	
Facility 591	October 2019	MEDIUM	ANC 1 Outlier Det...	0.00	<	0.00	
Facility 594	October 2019	MEDIUM	ANC 1 Outlier Det...	0.00	<	0.00	
Facility 598	October 2019	MEDIUM	ANC 1 Outlier Det...	0.00	<	0.00	
Facility 602	October 2019	MEDIUM	ANC 1 Outlier Det...	0.00	<	0.00	
Facility 603	October 2019	MEDIUM	ANC 1 Outlier Det...	0.00	<	0.00	
Facility 604	October 2019	MEDIUM	ANC 1 Outlier Det...	0.00	<	0.00	
Facility 605	October 2019	MEDIUM	ANC 1 Outlier Det...	0.00	<	0.00	

Counting Org Units Using Predictors

- Example: I want to know all of the org units the recorded a positive value for the number of stock-out days.

Missing value strategy

Generator (*)

Description

Count is stockout days is greater than 0

```
if({#w1uAGVuJ9PV} > 0,1,0)
```

() * / + - Days

if(Stock-out of RDTs (days) > 0,1,0)

Valid

DATA ELEMENTS

PRO

stock

Stock-out of co-trimoxazole more than 1 week

Stock-out of RDTs (days)

Stock-out of SP (days)

Stock out of SP more than 1 week

Stock RDTs issues

Stock RDTs received

Stock SP issued

Stock SP received