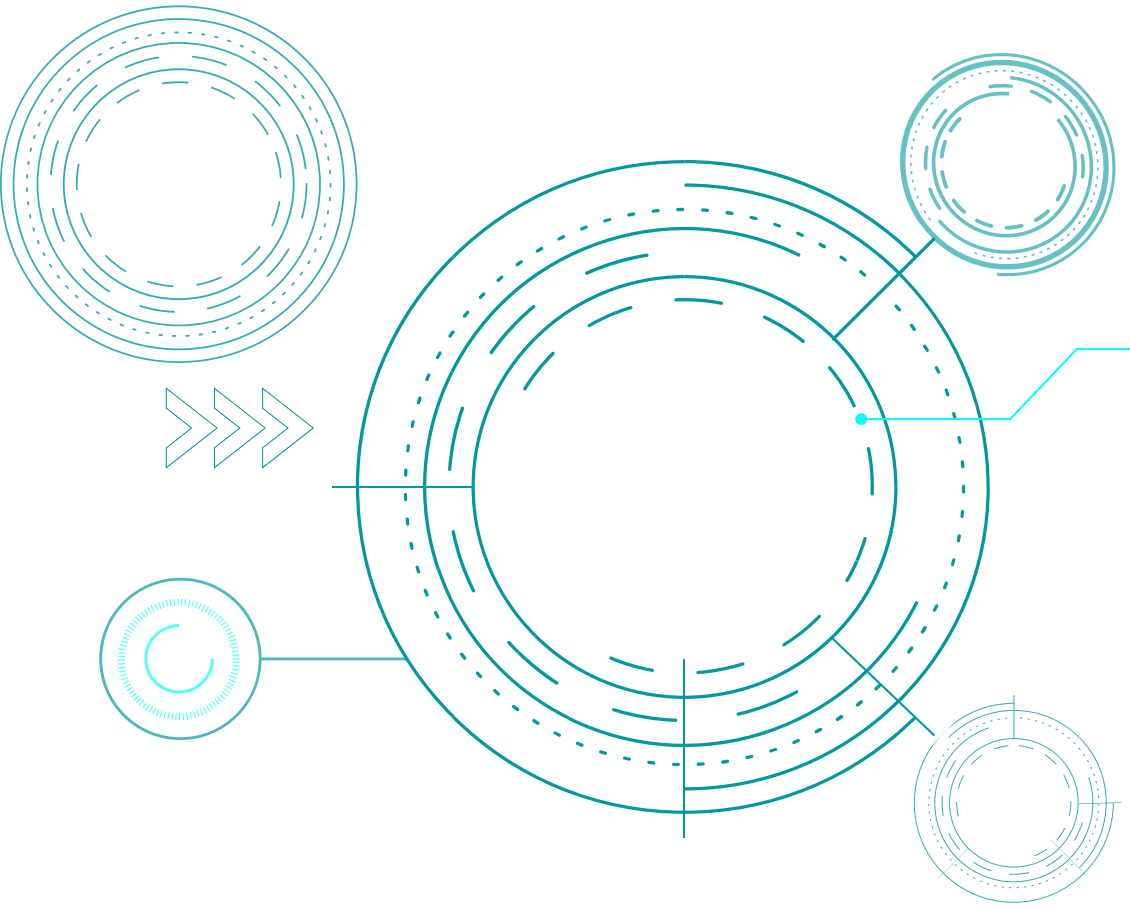
User Manual

HMIS Data Validation Tool



October 2021

This toolbox is developed under project National Data Quality Forum (NDQF) by ICMR-National institute of Medical Statistics and Population Council, India.

For any further clarifications and bug reporting please report at hmisdqtool@gmail.com

Table of Contents

1. Overview

The automated validation check tool is an offline python-based desktop application, which is open-source software, thereby making the tool cost-efficient, easy to use, without having any dependency on the network which often creates a barrier in the remote areas of the country. The automation will help HMIS staff in checking the data inconsistencies at one go without doing any manual checking. The system will be beneficial to the staff present at the national, state, and district levels for undertaking quick data quality assessment checks and providing feedback to data entry operators at facility levels as well as maintain the quality of data. The automated system will enable faster data validation and save time by preventing manual quality checks. This will result in reducing the time lag from data collection to data finalization and analytics.

2. Prerequisites for start

1. Operating system: Windows
2. Compatible on 32/64-bit system.
3. Start with .exe file

4. Background

HMIS gathers, aggregates and analyses routine data pertaining to health service delivery captured at facility levels. However, the reported data are often subjected to inconsistencies affecting the overall quality. National Data Quality Forum (NDQF) in collaboration with HMIS team manually applied existing validation checks to HMIS data to assess the performance of checks at facility level in HMIS 2.0 system. The team mapped data items to aggregate data; modified validation rules, and prepared validation summary reports at facility levels for Delhi, Haryana, Jammu & Kashmir and Mizoram. To reduce human intervention and to reduce the time taken for manually checking the inconsistencies, the NDQF team developed an offline, automated, validation tool. After the first demo given to the HMIS team, the check output was modified between the indicators and categorization of the checks were done as inconsistent, consistent, probable reporting error and blank.

5. Flow Chart

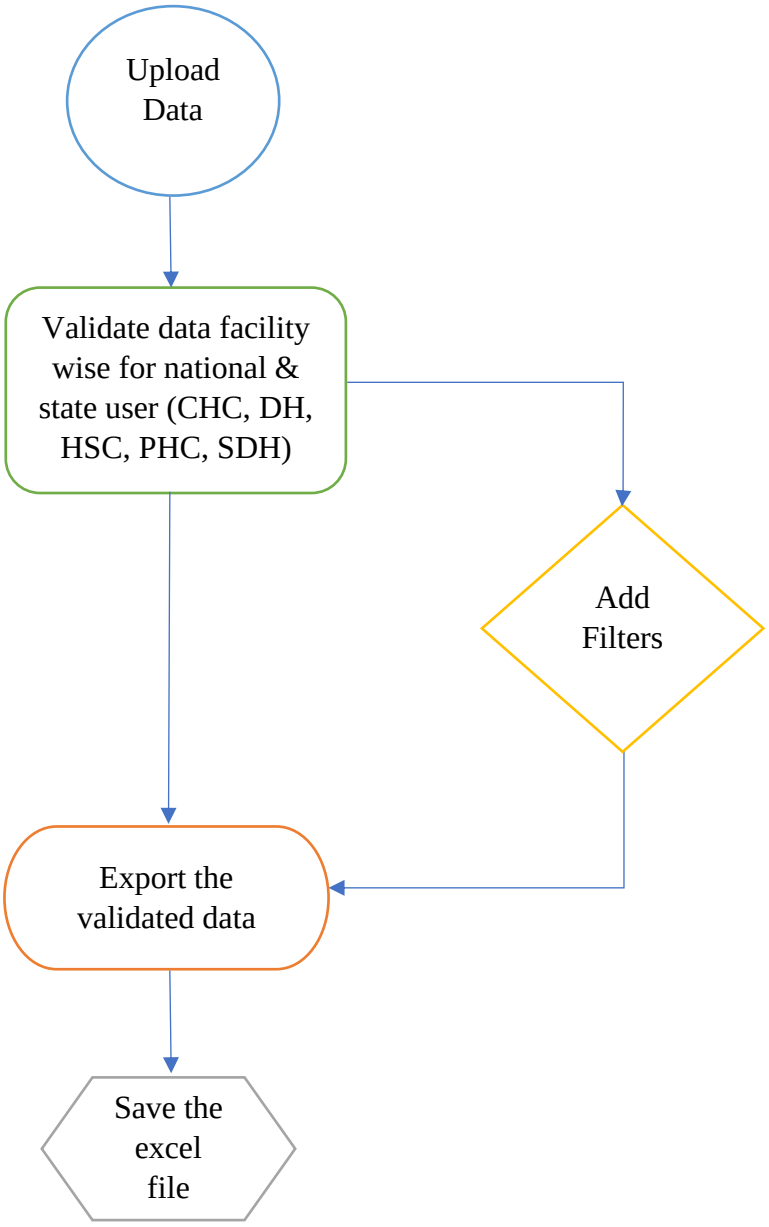


Fig 1: The flow chart of the tool box and the steps to perform validation checks facility wise

6. Introduction

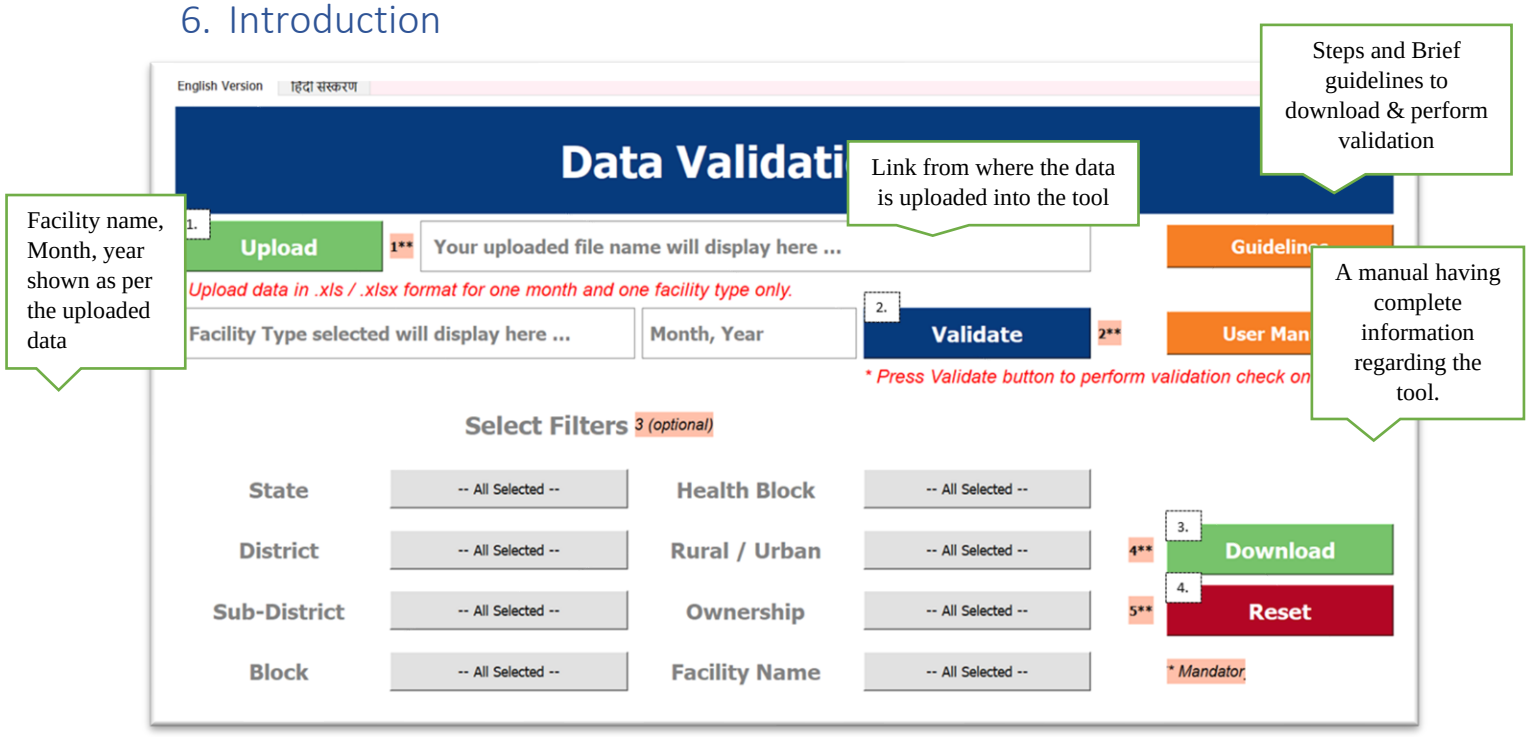


Fig 2: The toolbox window to perform validation checks

- Toolbox performs validation checks for all facilities (HSC, PHC, CHC, DH, SDH) and generate detailed reports to highlight the error.
- Data uploading
- Filters available for facility type, month, state, district, facility name
- Four outputs integrated into one sheet in Excel format
    - Validated data for each observation
    - Indicator wise summary sheet
    - Facility specific summary sheet
    - Top 10 facilities requiring attention

7. Getting started

Start with the data validation by pressing the upload button to upload the raw data downloaded from the HMIS portal for one month facility wise

7.1 Validation without filters

Step-1: Upload the data in the tool.

Step-2: The facility name will be reflected over the window and then press validation button to validate the data.

Step-3: Press Export button to export the validated data in the excel format.

Step-4: Press reset button to start again with the validation.

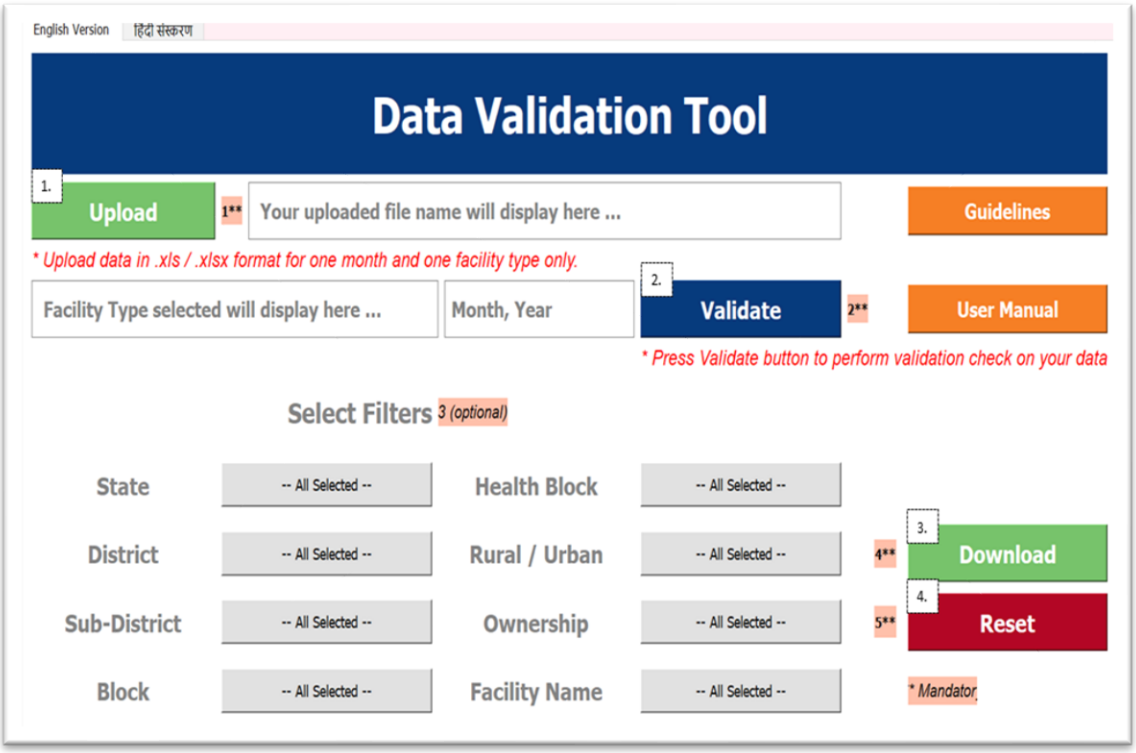


Fig 3: The toolbox window to perform validation checks step wise

7.2 Validation with filters

Step-1: Upload the data in the tool.

Step-2: The facility name will be reflected over the window and then press validation button to validate the data.

Step-3: Select filter (District, Rural/Urban, Ownership, facility name) in the chronological order.

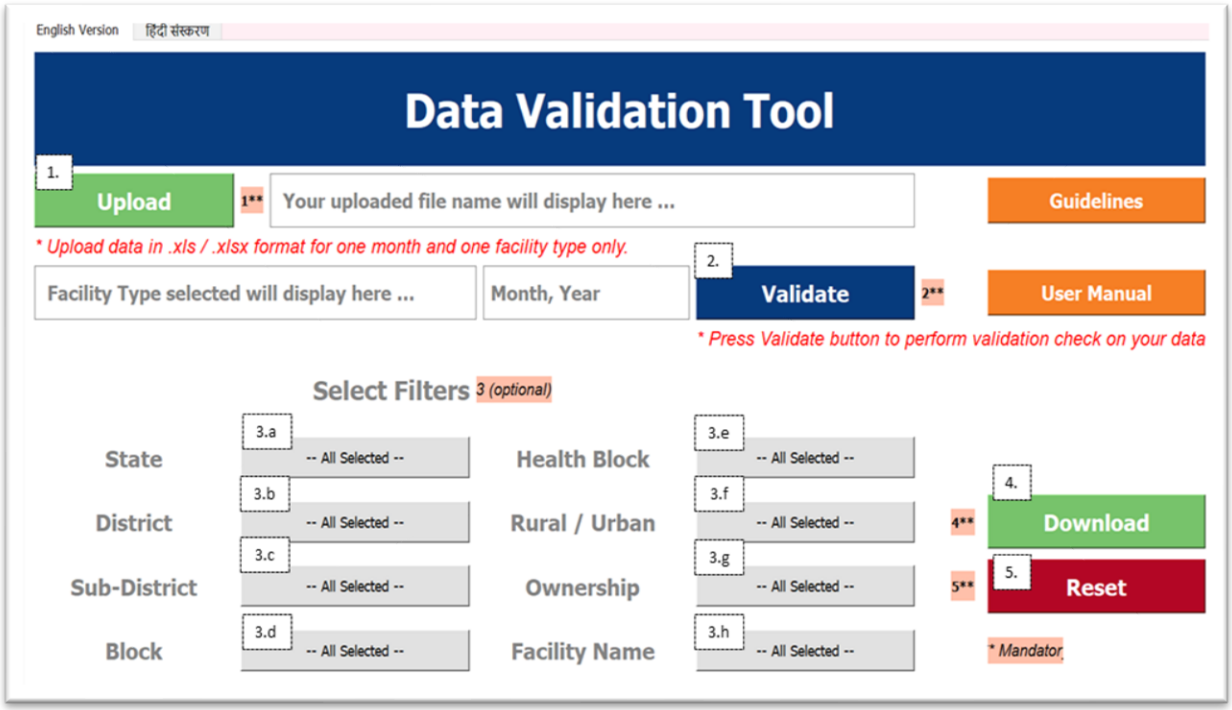


Fig 4: The toolbox window to perform validation checks with filters step wise

The filters are further shown as 3.a, 3.b, 3.c, 3.d, 3.e and are explained as follows:

3.a. It shows the state drop down for the uploaded Datasets if the data set is uploaded for the national level at the centre level for one particular facility.

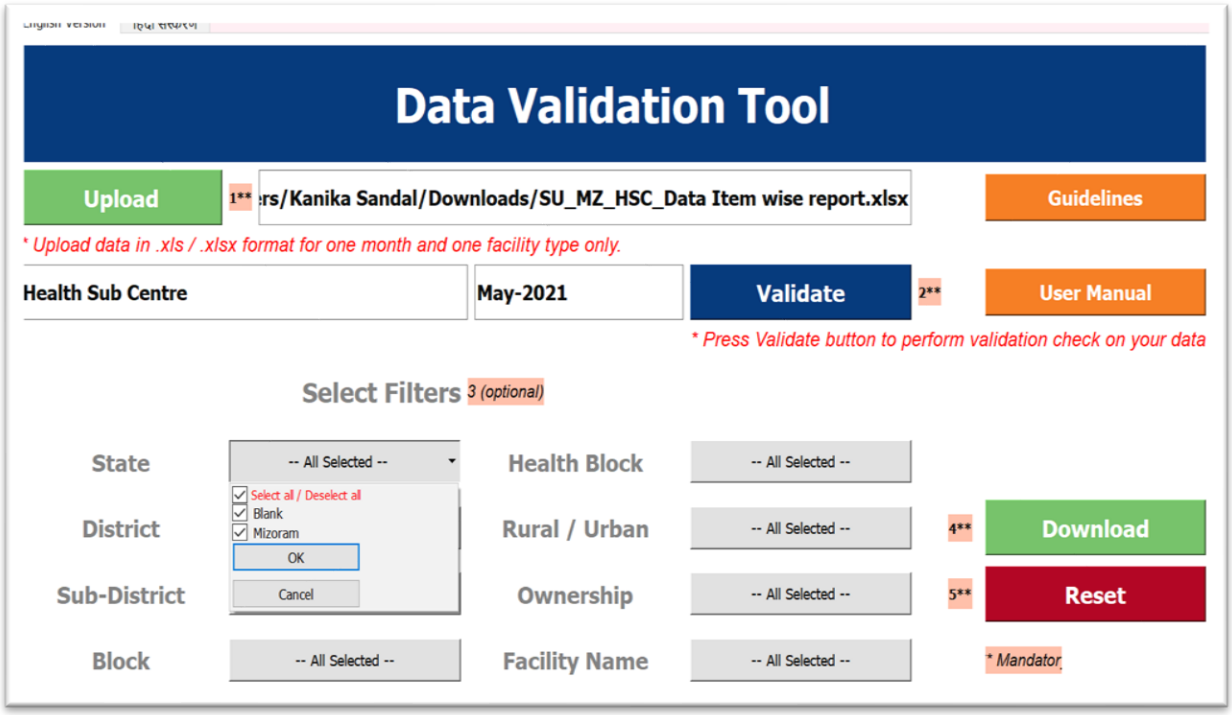


Fig 5: The toolbox window showcasing the first filter of the state category.

3.b. It shows the district drop down for the uploaded Datasets

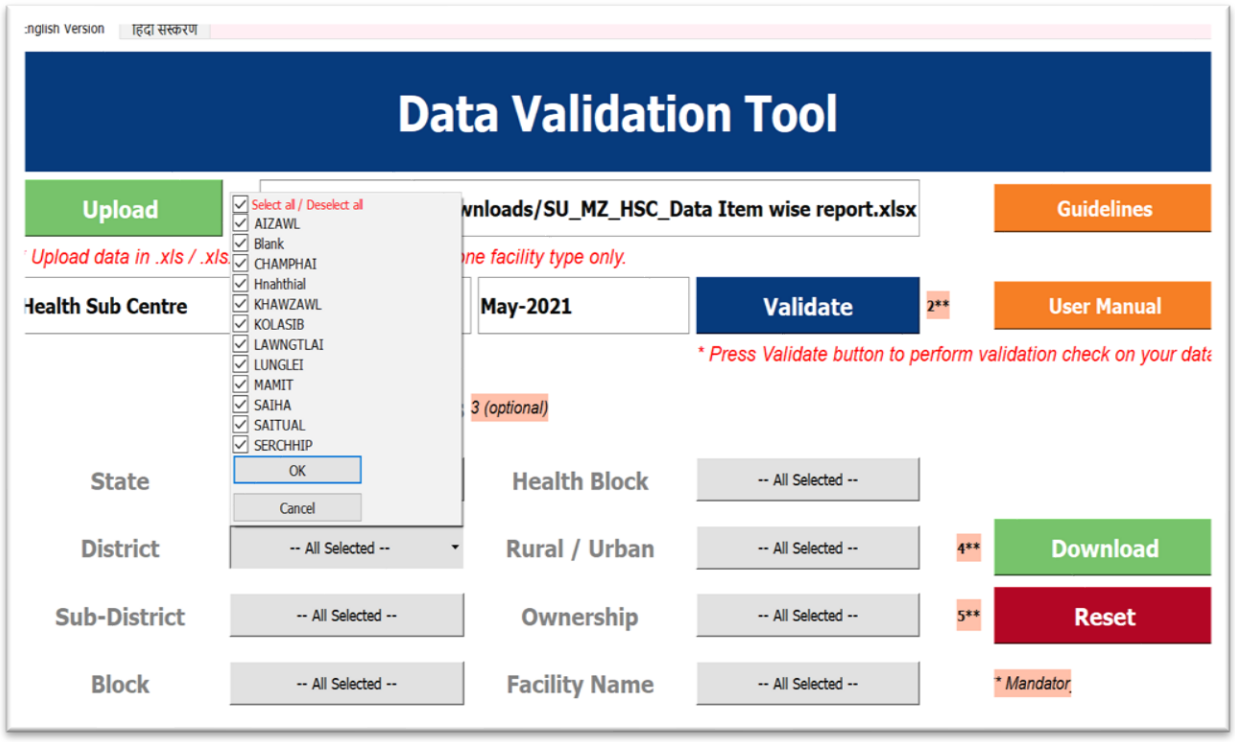


Fig 6: The toolbox window showcasing the district filter.

3.c. It shows the sub district drop down for the uploaded Datasets

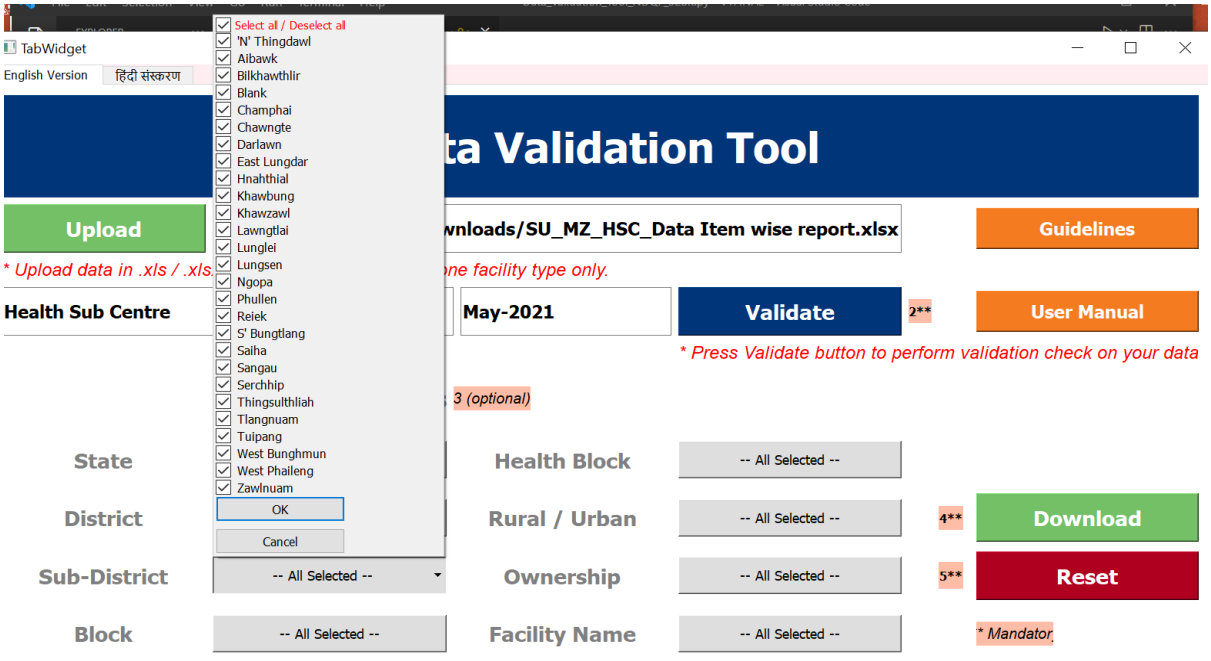


Fig 7: The toolbox window showcasing the sub district filter.

3.d. It shows the block drop down for the uploaded Datasets.

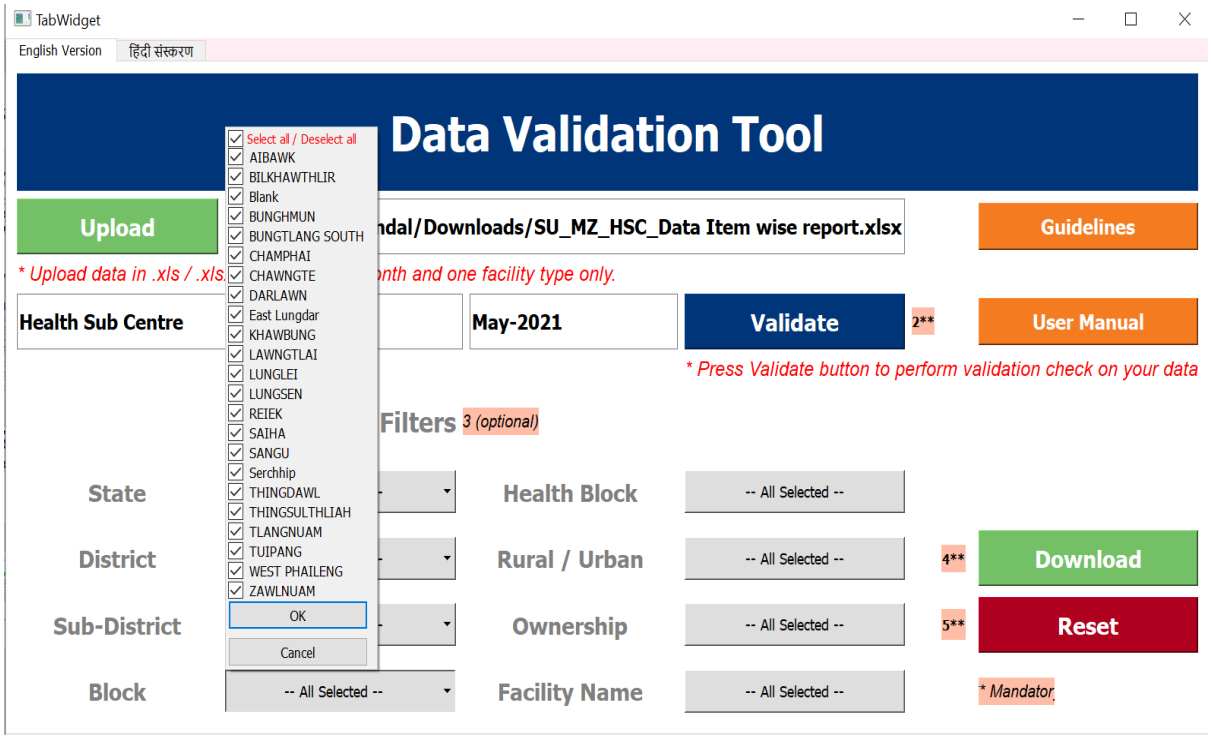


Fig 8: The toolbox window showcasing the block filter.

3.e. It shows the health block drop down for the uploaded Dataset.

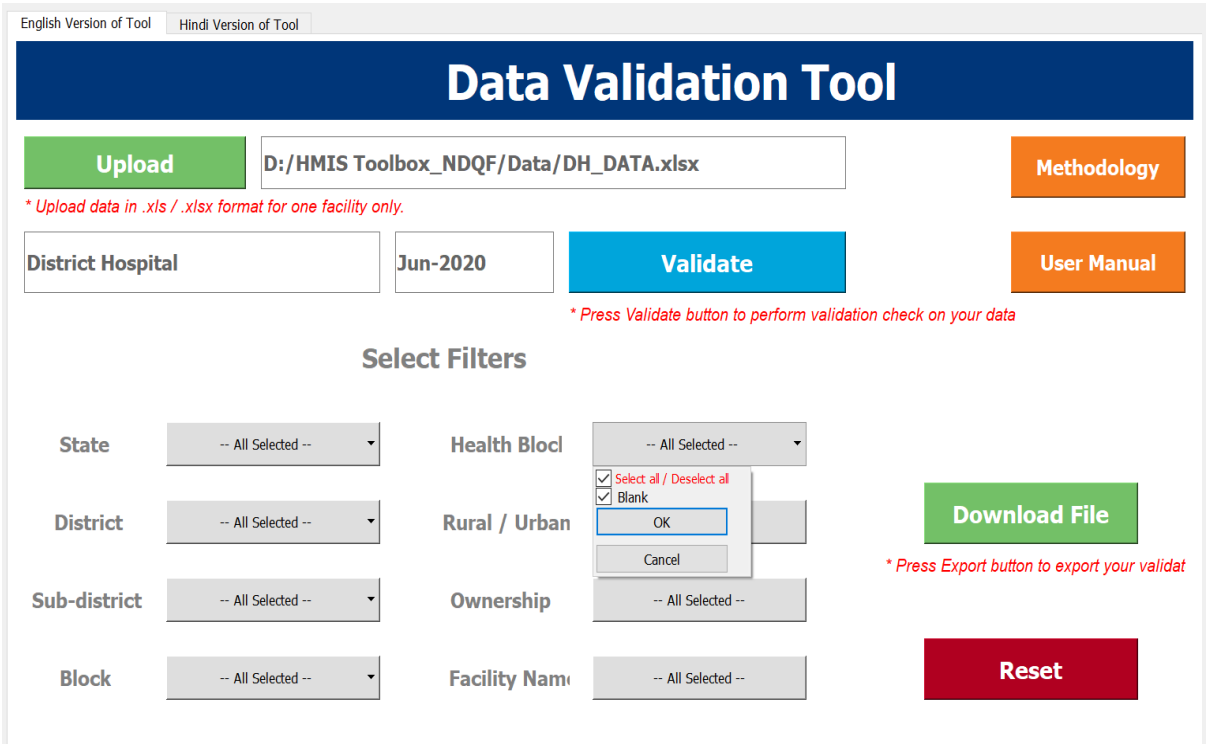


Fig 9: The toolbox window showcasing the health block filter.

3.f. It shows the rural/urban drop down for the uploaded Dataset.

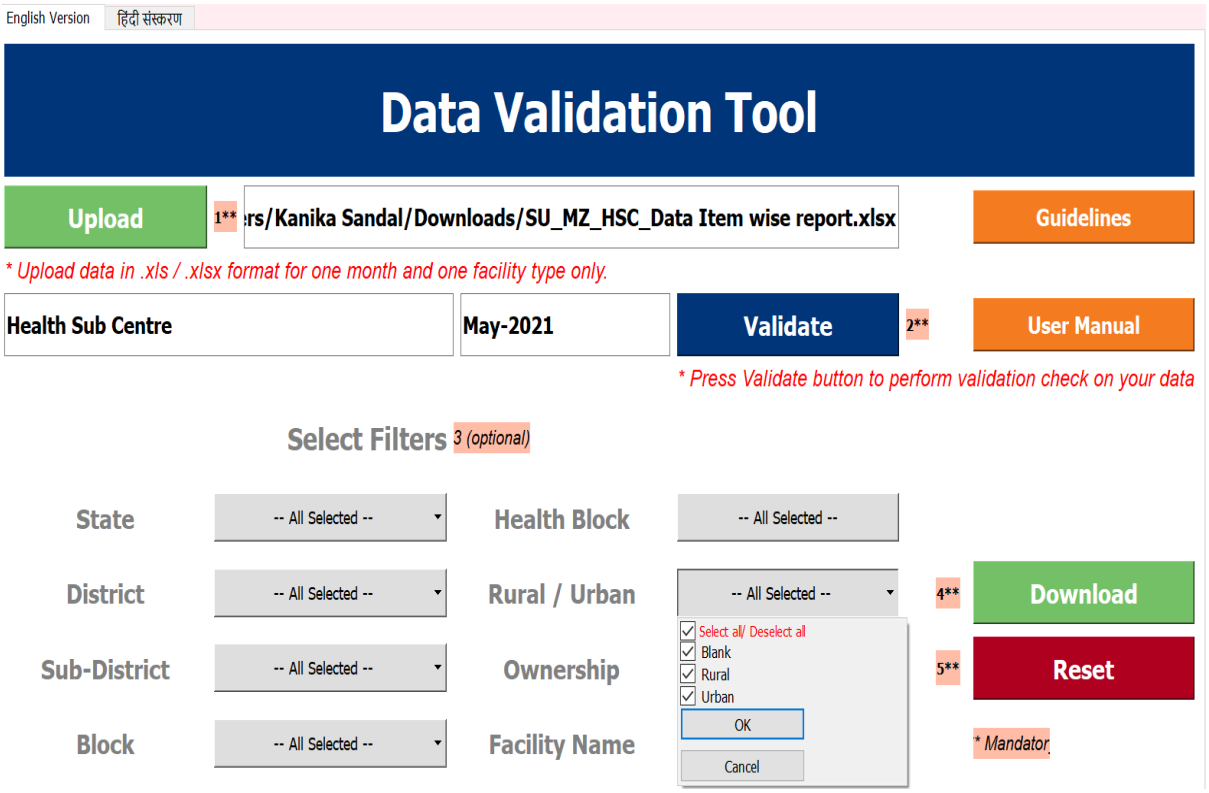


Fig 10: The toolbox window showcasing the rural/urban filter.

3.g. It shows the ownership drop down for the uploaded Dataset.

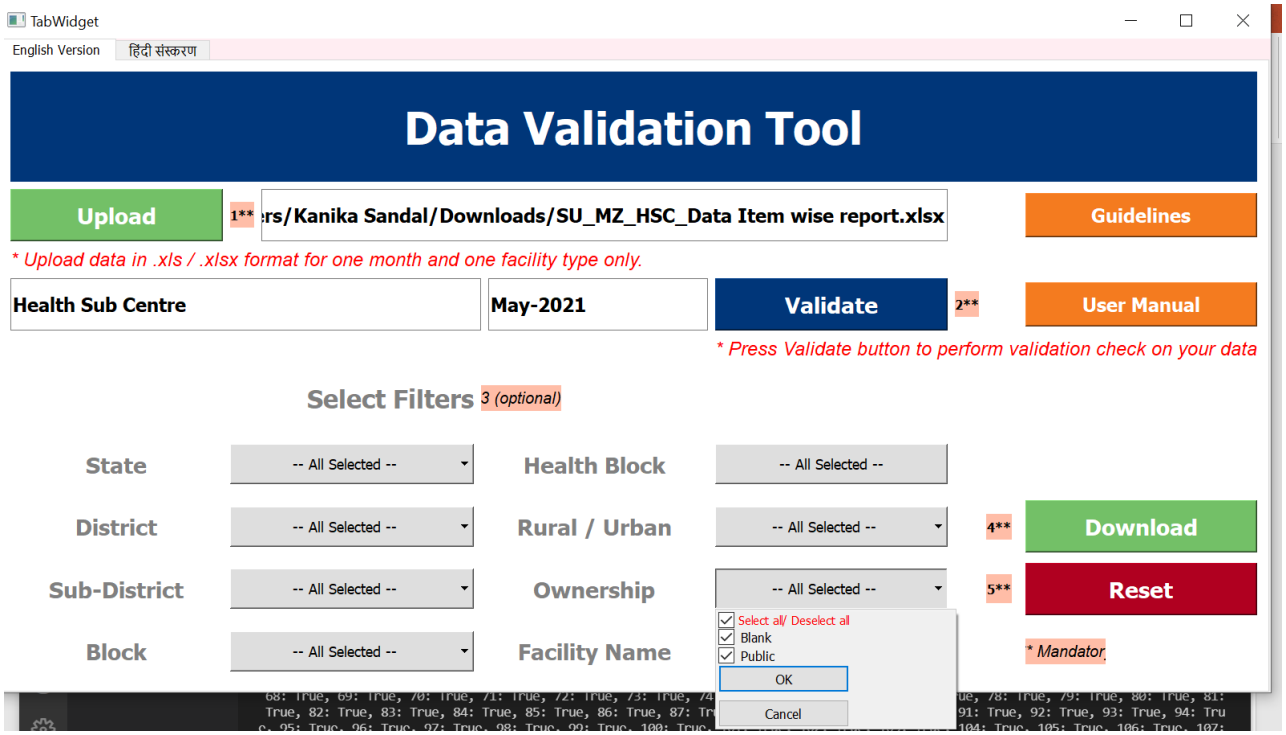


Fig 11: The toolbox window showcasing the ownership filter.

3.h. It shows the facility name drop down for the uploaded Datasets in the alphabetical order.

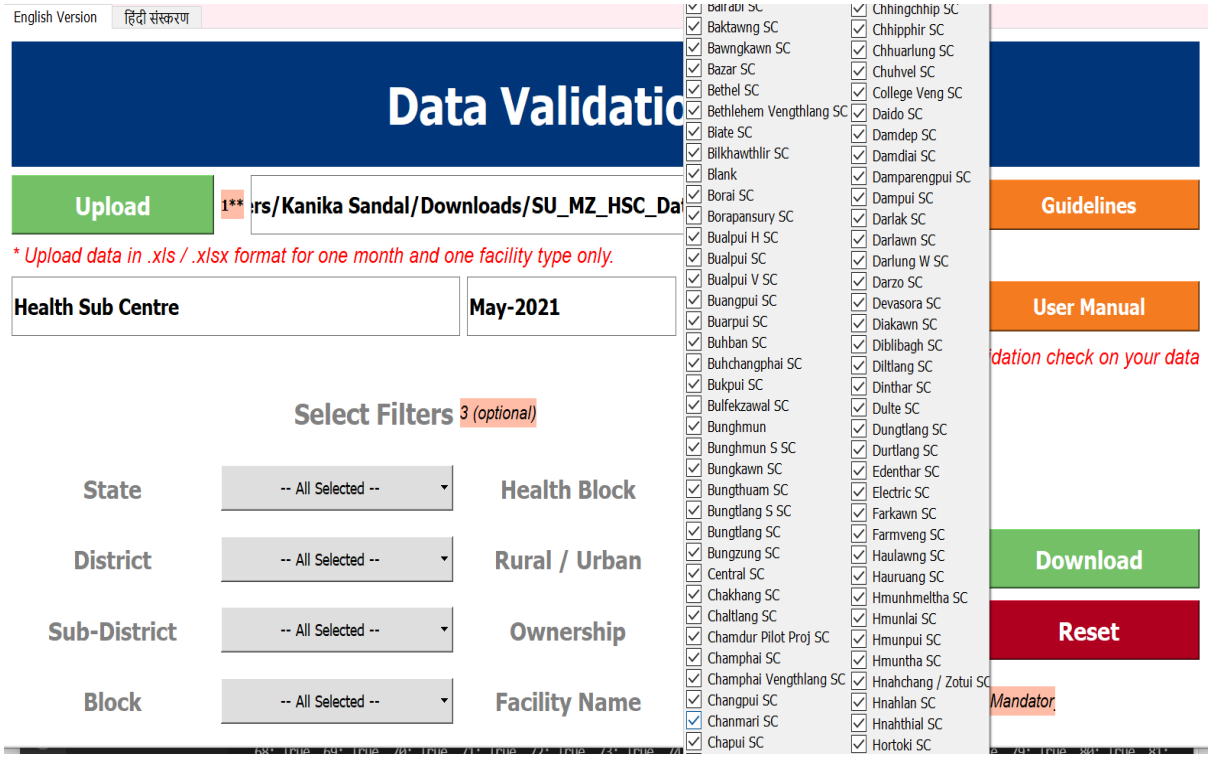


Fig 12: The toolbox window showcasing the facility name filter.

Step-4: Press Export button to export the validated data in the excel format.

Step-5: Save the excel sheet in your system.

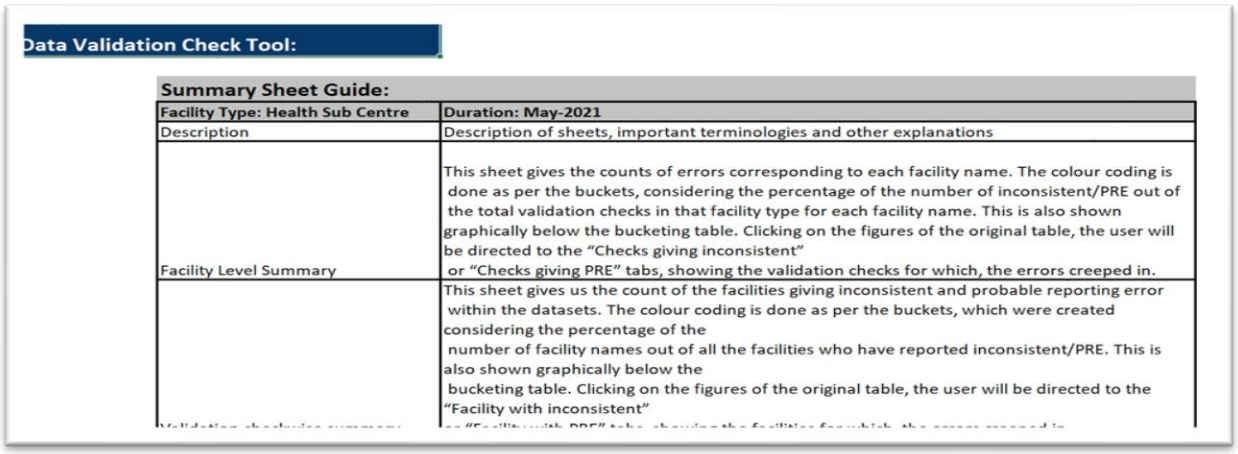
Step-6: Press reset button to start again with the validation

8. Summary Sheet (Downloaded File)

The validation tool generates the excel sheet (.xls) which gives detailed summary sheets for the validation checks focus on the error like Inconsistencies and Probable reporting error within the datasets.

8.1 Description Sheet (sheet -1)

This is the first sheet of the generated summary sheet from the validation toolbox that gives a glimpse and introduction to the user of the particular sheet uploaded by the user. The sheet also provides you with the type of the data set and the facility type of the sheet uploaded by the user. It also provides the examples with some of the examples of the check implemented.

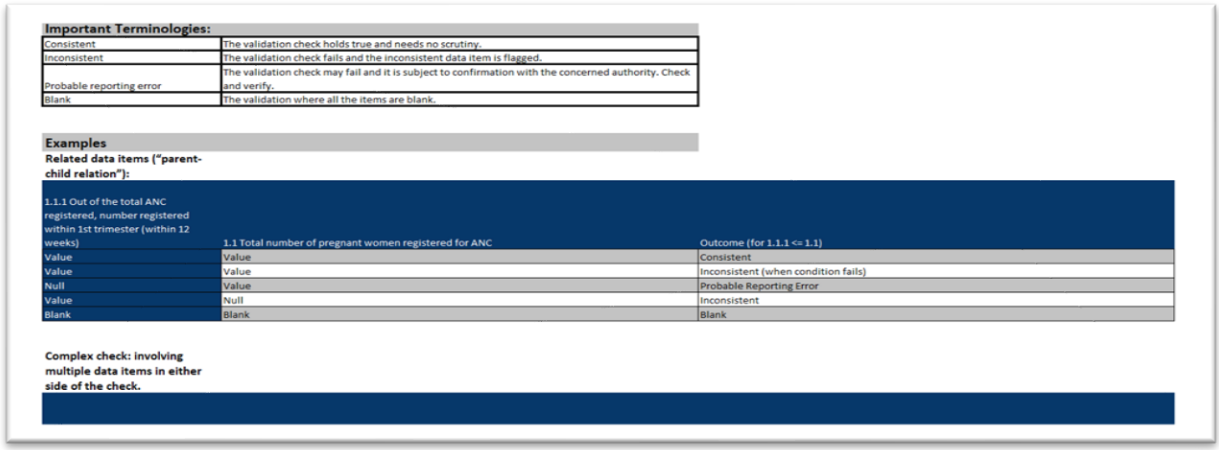


Data Validation Check Tool:

Summary Sheet Guide:

Facility Type: Health Sub Centre	Duration: May-2021
Description	Description of sheets, important terminologies and other explanations
Facility Level Summary	This sheet gives the counts of errors corresponding to each facility name. The colour coding is done as per the buckets, considering the percentage of the number of inconsistent/PRE out of the total validation checks in that facility type for each facility name. This is also shown graphically below the bucketing table. Clicking on the figures of the original table, the user will be directed to the "Checks giving inconsistent" or "Checks giving PRE" tabs, showing the validation checks for which, the errors creeped in.
	This sheet gives us the count of the facilities giving inconsistent and probable reporting error within the datasets. The colour coding is done as per the buckets, which were created considering the percentage of the number of facility names out of all the facilities who have reported inconsistent/PRE. This is also shown graphically below the bucketing table. Clicking on the figures of the original table, the user will be directed to the "Facility with inconsistent"

Fig 13: The first sheet of the downloaded file from the validation tool box giving brief information regarding the dataset uploaded and sheet description.



Important Terminologies:

Consistent	The validation check holds true and needs no scrutiny.
Inconsistent	The validation check fails and the inconsistent data item is flagged.
Probable reporting error	The validation check may fail and it is subject to confirmation with the concerned authority. Check and verify.
Blank	The validation where all the items are blank.

Examples

Related data items ("parent-child relation"):

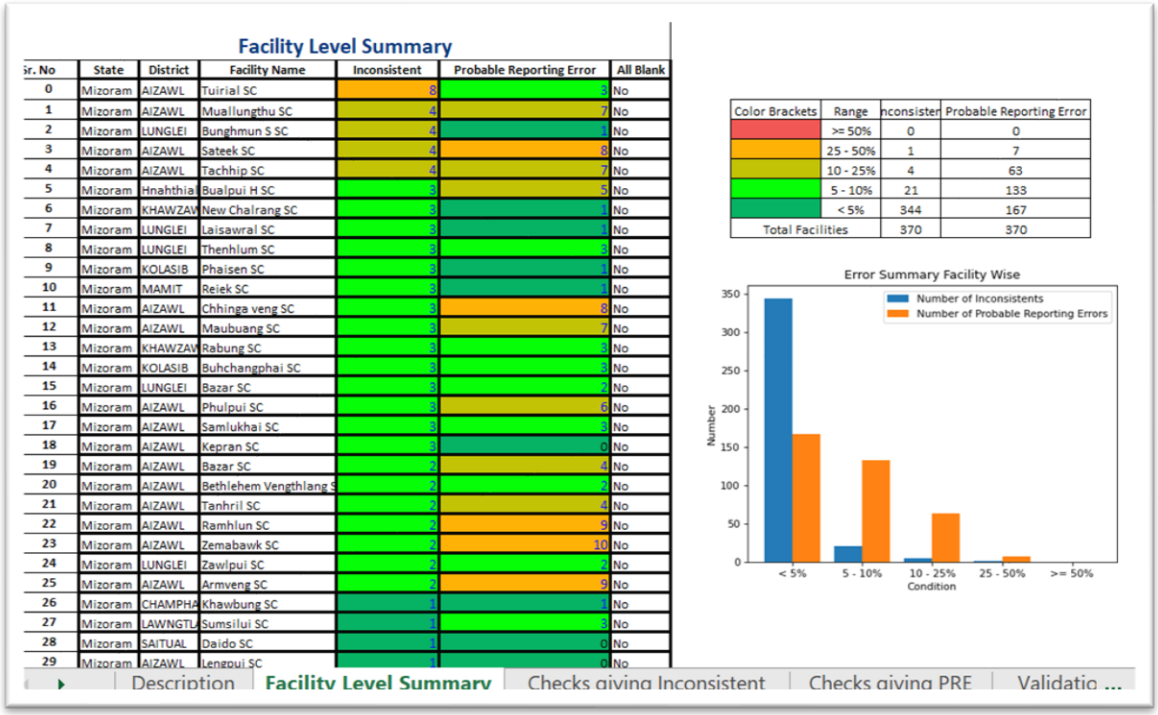
1.1.1 Out of the total ANC registered, number registered within 1st trimester (within 12 weeks)	1.1 Total number of pregnant women registered for ANC	Outcome (for 1.1.1 <= 1.1)
Value	Value	Consistent
Value	Value	Inconsistent (when condition fails)
Null	Value	Probable Reporting Error
Value	Null	Inconsistent
Blank	Blank	Blank

Complex check: involving multiple data items in either side of the check.

Fig 14: The first sheet of the downloaded file from the validation tool box giving brief information regarding important terminologies and related examples.

8.2 Facility Level Summary (Sheet -2)

It cabinets the count of the validation checks giving inconsistent and probable reporting error with reference to the facility involved and also gives the information about the facilities which are having all blank entry for the indicators of the particular facility.



Facility Level Summary

Sr. No	State	District	Facility Name	Inconsistent	Probable Reporting Error	All Blank
0	Mizoram	AIZAWL	Tuirial SC	8	3	No
1	Mizoram	AIZAWL	Muallungthu SC	4	7	No
2	Mizoram	LUNGLEI	Bunghmun S SC	4	1	No
3	Mizoram	AIZAWL	Sateek SC	4	8	No
4	Mizoram	AIZAWL	Tachhip SC	4	7	No
5	Mizoram	Hnahthial	Bualpui H SC	3	5	No
6	Mizoram	KHAWZAW	New Chalrang SC	3	1	No
7	Mizoram	LUNGLEI	Laisawral SC	3	1	No
8	Mizoram	LUNGLEI	Thenhlum SC	3	3	No
9	Mizoram	KOLASIB	Phaisen SC	3	1	No
10	Mizoram	MAMIT	Reiek SC	3	1	No
11	Mizoram	AIZAWL	Chhinga veng SC	3	8	No
12	Mizoram	AIZAWL	Maubuang SC	3	7	No
13	Mizoram	KHAWZAW	Rabung SC	3	3	No
14	Mizoram	KOLASIB	Buhchangphai SC	3	3	No
15	Mizoram	LUNGLEI	Bazar SC	3	2	No
16	Mizoram	AIZAWL	Phulpui SC	3	6	No
17	Mizoram	AIZAWL	Samlukhai SC	3	3	No
18	Mizoram	AIZAWL	Kepran SC	3	0	No
19	Mizoram	AIZAWL	Bazar SC	2	4	No
20	Mizoram	AIZAWL	Bethlehem Vengthlang S	2	2	No
21	Mizoram	AIZAWL	Tanhril SC	2	4	No
22	Mizoram	AIZAWL	Ramhlun SC	2	9	No
23	Mizoram	AIZAWL	Zemabawk SC	2	10	No
24	Mizoram	LUNGLEI	Zawlpui SC	2	2	No
25	Mizoram	AIZAWL	Armveng SC	2	9	No
26	Mizoram	CHAMPHA	Khawbung SC	1	1	No
27	Mizoram	LAWNGTL	Sumsilui SC	1	3	No
28	Mizoram	SAITUAL	Daido SC	1	0	No
29	Mizoram	AIZAWL	Lengpui SC	1	0	No

Color Brackets	Range	Inconsistent	Probable Reporting Error
	>= 50%	0	0
	25 - 50%	1	7
	10 - 25%	4	63
	5 - 10%	21	133
	< 5%	344	167
Total Facilities		370	370

Description | Facility Level Summary | Checks giving Inconsistent | Checks giving PRE | Validatio ...

Fig 15: The second sheet of the downloaded file providing error information facility wise.

To analyse the level of error the count is divided into the colour brackets which gives a brief summary of number of inconsistent and probable reporting error encountered. The colour bracket is divided into five categories, i. e

1) <5%: Highlighted by dark green colour to showcase the number of facilities falling into this category which is referred to as result that is very good.
2) 5% -10%: Highlighted by light green colour to showcase the number of facilities falling into this category which is referred to as result that is good.
3) 10% -25%: Highlighted by light yellow colour to showcase the number of facilities falling into this category which is referred to as result that is moderate.
4) 25% -50%: Highlighted by orange colour to showcase the number of facilities falling into this category which is referred to as result that is poor.
5) >=50%: Highlighted by orange colour to showcase the number of facilities falling into this category which is referred to as result that is very poor.

Calculation to calculate %:

To segregate the number of consistent and probable reporting error for particular facility into percentage the following formula is used:

$$\%\ of\ Inconsistent = \frac{\text{Number of Inconsistent in particular check}}{\text{Total Number of check}}\ X\ 100$$

$$\%\ of\ PRE = \frac{\text{Number of PRE in particular check}}{\text{Total Number of check}}\ X\ 100$$

After calculation of all the percentages individually the facilities percentages into the colour brackets to give a brief summary of errors in the file is done.

Visualization:

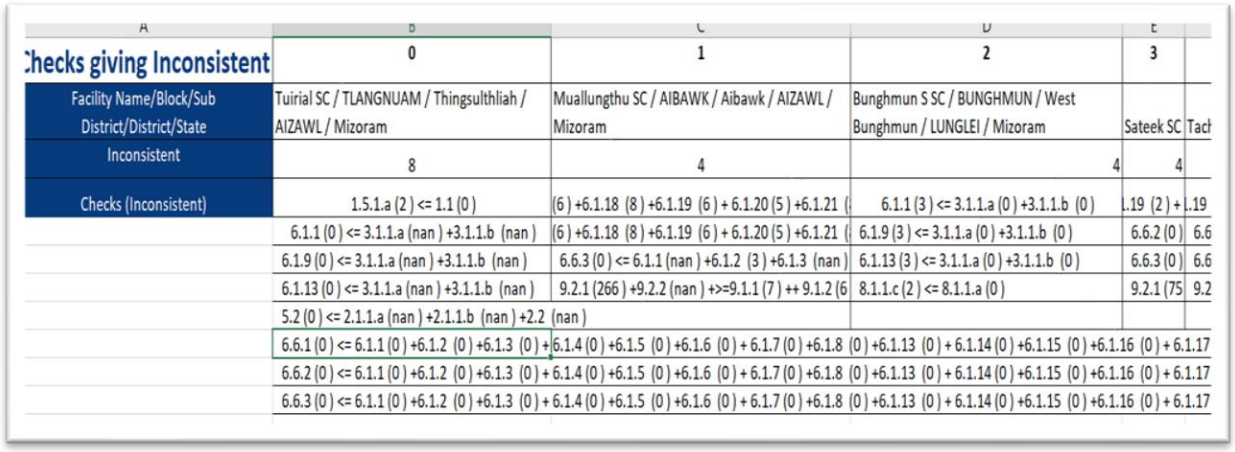
The data showcased in the colour bracket for the total number of inconsistent and probable reporting error is shown in a bar graph to give a better idea of the number of checks. The orange bar represents the Probable reporting error and blue bar represents the inconsistent error.

Hyperlinks:

Hyperlinks are added to both the columns i.e., Number of inconsistent and probable reporting error that takes the user to the next separate sheets for both inconsistent and probable reporting error named as checks giving inconsistent and checks giving probable reporting error. The detailed of these sheets are explained in the F and G section.

8.3 Checks with Inconsistencies (Sheet-3)

This sheet provides the list of the checks give the number of inconsistent errors within a facility and also the input provided by the user in the given indicator of the check. This sheet is directly hyperlinked with the sheet no 2, i. e. facility level summary.

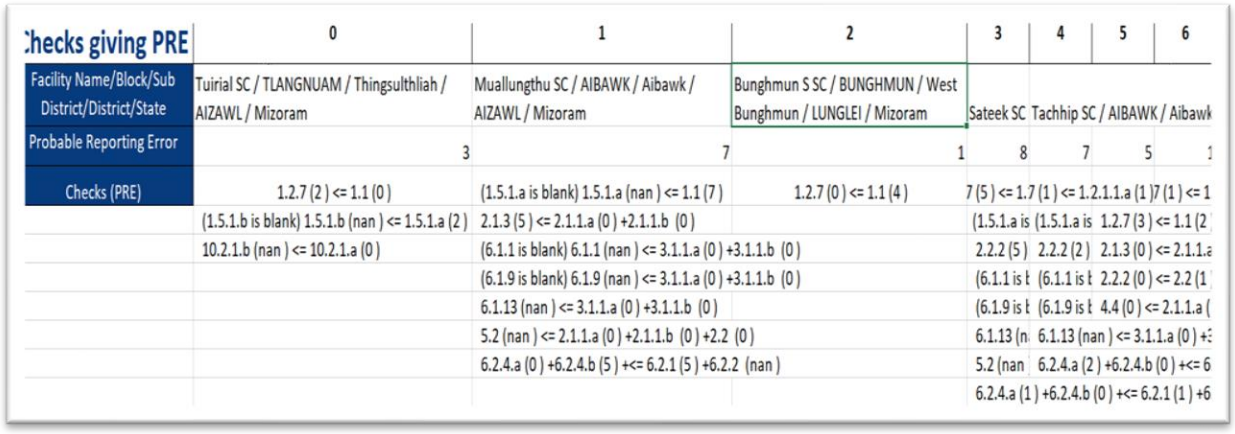


Checks giving Inconsistent	0	1	2	3	
Facility Name/Block/Sub District/District/State	Tuirial SC / TLANGNUAM / Thingsulthliah / AIZAWL / Mizoram	Muallungthu SC / AIBAWK / Aibawk / AIZAWL / Mizoram	Bunghmun S SC / BUNGHMUN / West Bunghmun / LUNGLEI / Mizoram	Sateek SC	Tach
Inconsistent	8	4	4	4	
Checks (Inconsistent)	1.5.1.a (2) <= 1.1 (0)	(6) +6.1.18 (8) +6.1.19 (6) + 6.1.20 (5) +6.1.21 (	6.1.1 (3) <= 3.1.1.a (0) +3.1.1.b (0)	.19 (2) + .19	
	6.1.1 (0) <= 3.1.1.a (nan) +3.1.1.b (nan)	(6) +6.1.18 (8) +6.1.19 (6) + 6.1.20 (5) +6.1.21 (	6.1.9 (3) <= 3.1.1.a (0) +3.1.1.b (0)	6.6.2 (0)	6.6
	6.1.9 (0) <= 3.1.1.a (nan) +3.1.1.b (nan)	6.6.3 (0) <= 6.1.1 (nan) +6.1.2 (3) +6.1.3 (nan)	6.1.13 (3) <= 3.1.1.a (0) +3.1.1.b (0)	6.6.3 (0)	6.6
	6.1.13 (0) <= 3.1.1.a (nan) +3.1.1.b (nan)	9.2.1 (266) +9.2.2 (nan) +>=9.1.1 (7) ++ 9.1.2 (6	8.1.1.c (2) <= 8.1.1.a (0)	9.2.1 (75	9.2
	5.2 (0) <= 2.1.1.a (nan) +2.1.1.b (nan) +2.2 (nan)				
	6.6.1 (0) <= 6.1.1 (0) +6.1.2 (0) +6.1.3 (0) +	6.1.4 (0) +6.1.5 (0) +6.1.6 (0) + 6.1.7 (0) +6.1.8 (0) +6.1.13 (0) + 6.1.14 (0) +6.1.15 (0) +6.1.16 (0) + 6.1.17			
	6.6.2 (0) <= 6.1.1 (0) +6.1.2 (0) +6.1.3 (0) + 6.1.4 (0) +6.1.5 (0) +6.1.6 (0) + 6.1.7 (0) +6.1.8 (0) +6.1.13 (0) + 6.1.14 (0) +6.1.15 (0) +6.1.16 (0) + 6.1.17				
	6.6.3 (0) <= 6.1.1 (0) +6.1.2 (0) +6.1.3 (0) + 6.1.4 (0) +6.1.5 (0) +6.1.6 (0) + 6.1.7 (0) +6.1.8 (0) +6.1.13 (0) + 6.1.14 (0) +6.1.15 (0) +6.1.16 (0) + 6.1.17				

Fig 16: The third and the hyperlinked sheet of the downloaded file providing inconsistent detailed information

8.4 Checks with Probable Reporting Error (Sheet-4)

This sheet provides the list of the checks give the number of PRE error within a facility and also the input provided by the user in the given indicator of the check. This sheet is directly hyperlinked with the sheet no 2, i. e. facility level summary.

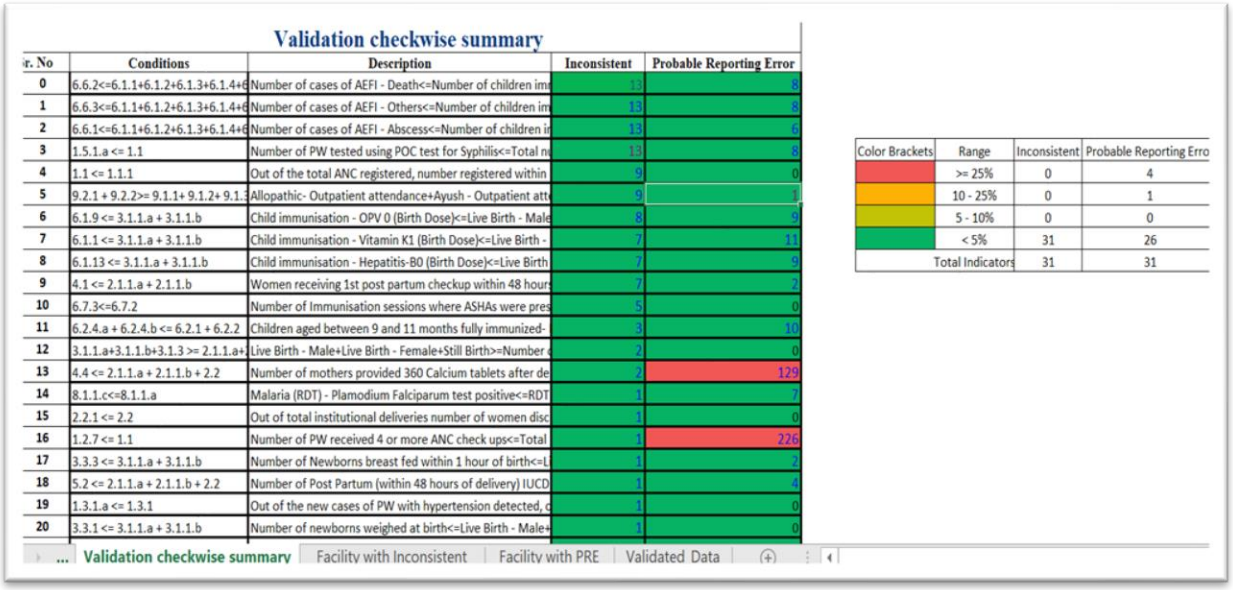


Checks giving PRE	0	1	2	3	4	5	6
Facility Name/Block/Sub District/District/State	Tuirial SC / TLANGNUAM / Thingsulthliah / AIZAWL / Mizoram	Muallungthu SC / AIBAWK / Aibawk / AIZAWL / Mizoram	Bunghmun S SC / BUNGHMUN / West Bunghmun / LUNGLEI / Mizoram	Sateek SC	Tachhip SC / AIBAWK / Aibawk		
Probable Reporting Error	3	7	1	8	7	5	1
Checks (PRE)	1.2.7 (2) <= 1.1 (0)	(1.5.1.a is blank) 1.5.1.a (nan) <= 1.1 (7)	1.2.7 (0) <= 1.1 (4)	7 (5) <= 1.7 (1) <= 1.2.1.1.a (1)	7 (1) <= 1		
	(1.5.1.b is blank) 1.5.1.b (nan) <= 1.5.1.a (2)	2.1.3 (5) <= 2.1.1.a (0) +2.1.1.b (0)		(1.5.1.a is	(1.5.1.a is	1.2.7 (3) <= 1.1 (2	
	10.2.1.b (nan) <= 10.2.1.a (0)	(6.1.1 is blank) 6.1.1 (nan) <= 3.1.1.a (0) +3.1.1.b (0)		2.2.2 (5)	2.2.2 (2)	2.1.3 (0) <= 2.1.1.a	
		(6.1.9 is blank) 6.1.9 (nan) <= 3.1.1.a (0) +3.1.1.b (0)		(6.1.1 is b	(6.1.1 is b	2.2.2 (0) <= 2.2 (1	
		6.1.13 (nan) <= 3.1.1.a (0) +3.1.1.b (0)		(6.1.9 is b	(6.1.9 is b	4.4 (0) <= 2.1.1.a (	
		5.2 (nan) <= 2.1.1.a (0) +2.1.1.b (0) +2.2 (0)		6.1.13 (n	6.1.13 (nan) <= 3.1.1.a (0) +3		
		6.2.4.a (0) +6.2.4.b (5) +<= 6.2.1 (5) +6.2.2 (nan)		5.2 (nan	6.2.4.a (2) +6.2.4.b (0) +<= 6		
				6.2.4.a (1) +6.2.4.b (0) +<= 6.2.1 (1) +6			

Fig 16: The fourth and the hyperlinked sheet of the downloaded file providing PRE detailed information

8.5 Validation Check wise summary (Sheet 5)

This sheet gives us the count of the facilities giving inconsistent and probable reporting error within the datasets.



Validation checkwise summary

Sr. No	Conditions	Description	Inconsistent	Probable Reporting Error
0	6.6.2<=6.1.1+6.1.2+6.1.3+6.1.4+6	Number of cases of AEFI - Death<=Number of children im	13	8
1	6.6.3<=6.1.1+6.1.2+6.1.3+6.1.4+6	Number of cases of AEFI - Others<=Number of children im	13	8
2	6.6.1<=6.1.1+6.1.2+6.1.3+6.1.4+6	Number of cases of AEFI - Abscess<=Number of children i	13	6
3	1.5.1.a <= 1.1	Number of PW tested using POC test for Syphilis<=Total n	13	8
4	1.1 <= 1.1.1	Out of the total ANC registered, number registered within	9	0
5	9.2.1 + 9.2.2>= 9.1.1+ 9.1.2+ 9.1.3	Allopathic- Outpatient attendance+Ayush - Outpatient att	9	1
6	6.1.9 <= 3.1.1.a + 3.1.1.b	Child immunisation - OPV 0 (Birth Dose)<=Live Birth - Male	8	9
7	6.1.1 <= 3.1.1.a + 3.1.1.b	Child immunisation - Vitamin K1 (Birth Dose)<=Live Birth -	7	11
8	6.1.13 <= 3.1.1.a + 3.1.1.b	Child immunisation - Hepatitis-B0 (Birth Dose)<=Live Birth	7	9
9	4.1 <= 2.1.1.a + 2.1.1.b	Women receiving 1st post partum checkup within 48 hour	7	2
10	6.7.3<=6.7.2	Number of Immunisation sessions where ASHAs were pres	5	0
11	6.2.4.a + 6.2.4.b <= 6.2.1 + 6.2.2	Children aged between 9 and 11 months fully immunized-	3	10
12	3.1.1.a+3.1.1.b+3.1.3 >= 2.1.1.a+	Live Birth - Male+Live Birth - Female+Still Birth>=Number	2	0
13	4.4 <= 2.1.1.a + 2.1.1.b + 2.2	Number of mothers provided 360 Calcium tablets after de	2	129
14	8.1.1.c<=8.1.1.a	Malaria (RDT) - Plamodium Falciparum test positive<=RDT	1	7
15	2.2.1 <= 2.2	Out of total institutional deliveries number of women disc	1	0
16	1.2.7 <= 1.1	Number of PW received 4 or more ANC check ups<=Total	1	226
17	3.3.3 <= 3.1.1.a + 3.1.1.b	Number of Newborns breast fed within 1 hour of birth<=L	1	2
18	5.2 <= 2.1.1.a + 2.1.1.b + 2.2	Number of Post Partum (within 48 hours of delivery) IUCD	1	4
19	1.3.1.a <= 1.3.1	Out of the new cases of PW with hypertension detected, c	1	0
20	3.3.1 <= 3.1.1.a + 3.1.1.b	Number of newborns weighed at birth<=Live Birth - Male+	1	0

Color Brackets	Range	Inconsistent	Probable Reporting Erro
	>= 25%	0	4
	10 - 25%	0	1
	5 - 10%	0	0
	< 5%	31	26
	Total Indicators	31	31

... Validation checkwise summary | Facility with Inconsistent | Facility with PRE | Validated Data

Fig 17: The fifth sheet of the downloaded file providing error information check wise.

To analyse the level of error the count of the facilities is divided into the colour brackets which gives a brief summary of number of inconsistent and probable reporting error encountered. The colour bracket is divided into five categories, i. e

1. <5%: Highlighted by dark green colour to showcase the number of facilities falling into this category which is referred to as result that is very good.
2. 5% -10%: Highlighted by light green colour to showcase the number of facilities falling into this category which is referred to as result that is good.
3. 10% -25%: Highlighted by light yellow colour to showcase the number of facilities falling into this category which is referred to as result that is moderate.
4. 25% -50%: Highlighted by orange colour to showcase the number of facilities falling into this category which is referred to as result that is poor.
5. >=50%: Highlighted by orange colour to showcase the number of facilities falling into this category which is referred to as result that is very poor.

Calculation to calculate %:

To segregate the number of consistent and probable reporting error for particular facility into percentage the following formula is used:

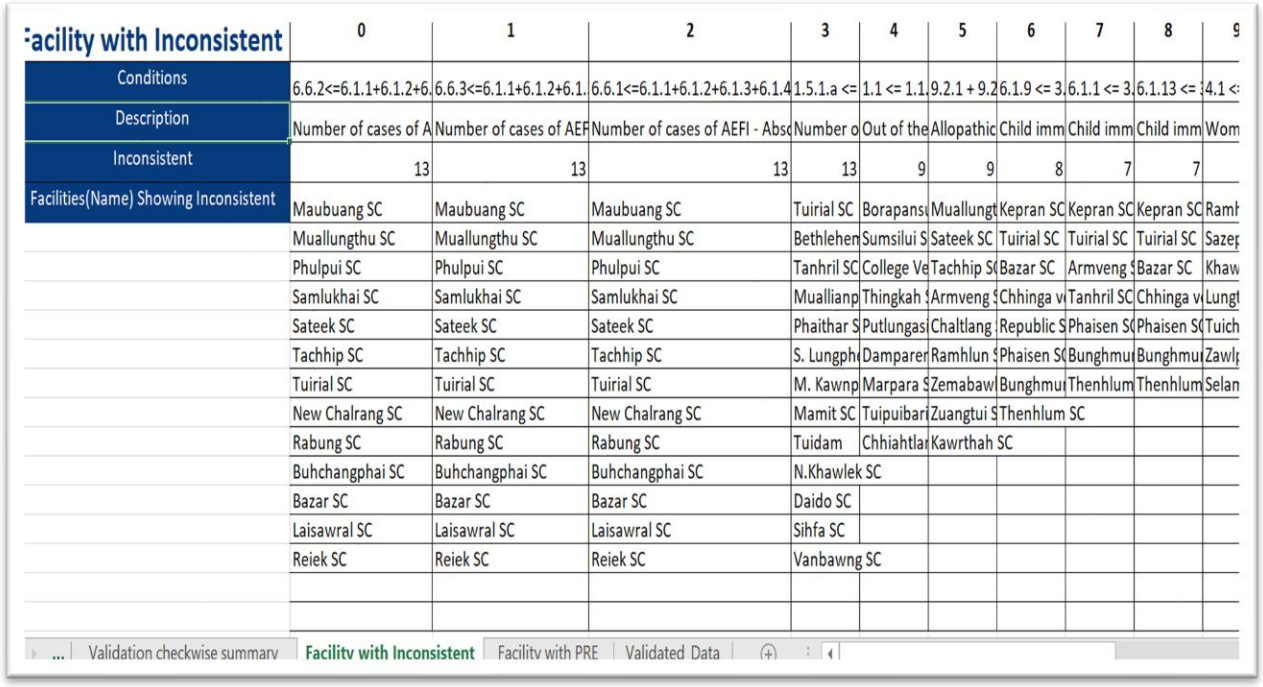
$$\%\ of\ Inconsistent = \frac{\text{Number of Inconsistent in particular facility}}{\text{Total Number of facilities}}\ X\ 100$$

$$\%\ of\ PRE = \frac{\text{Number of PRE in particular facility}}{\text{Total Number of facility}}\ X\ 100$$

After calculation of all the percentages individually the facilities percentages into the colour brackets to give a brief summary of errors in the file is done.

8.6 Facility with Inconsistencies (Sheet – 6)

This sheet provides the list of the facilities giving Inconsistencies within the dataset with the count and the condition and the description of the condition.



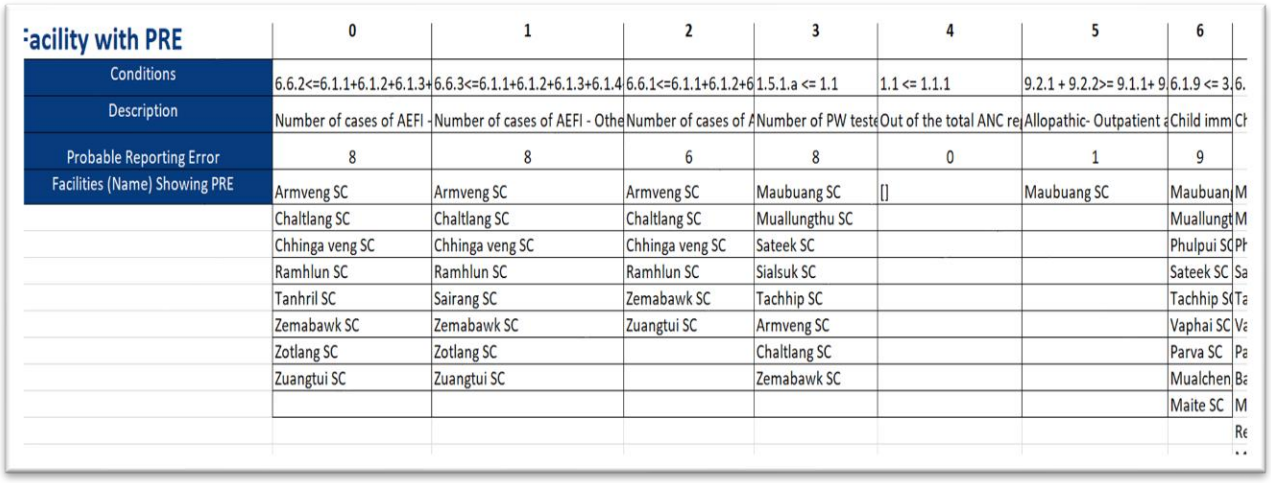
Facility with Inconsistent	0	1	2	3	4	5	6	7	8	9
Conditions	6.6.2<=6.1.1+6.1.2+6.	6.6.3<=6.1.1+6.1.2+6.1.	6.6.1<=6.1.1+6.1.2+6.1.3+6.1.4	1.5.1.a <=	1.1 <= 1.1.	9.2.1 + 9.2	6.1.9 <= 3.	6.1.1 <= 3.	6.1.13 <=	4.1 <=
Description	Number of cases of A	Number of cases of AEF	Number of cases of AEFI - Abs	Number o	Out of the	Allopathic	Child imm	Child imm	Child imm	Wom
Inconsistent	13	13	13	13	9	9	8	7	7	
Facilities(Name) Showing Inconsistent	Maubuang SC	Maubuang SC	Maubuang SC	Tuirial SC	Borapans	Muallungt	Kepran SC	Kepran SC	Kepran SC	Ramh
	Muallungthu SC	Muallungthu SC	Muallungthu SC	Bethlehen	Sumsilui S	Sateek SC	Tuirial SC	Tuirial SC	Tuirial SC	Sazep
	Phulpui SC	Phulpui SC	Phulpui SC	Tanhril SC	College Ve	Tachhip S	Bazar SC	Armveng S	Bazar SC	Khaw
	Samlukhai SC	Samlukhai SC	Samlukhai SC	Muallianp	Thingkah S	Armveng S	Chhinga v	Tanhril SC	Chhinga v	Lungt
	Sateek SC	Sateek SC	Sateek SC	Phaithar S	Putlungasi	Chaltlang	Republic S	Phaisen SC	Phaisen SC	Tuich
	Tachhip SC	Tachhip SC	Tachhip SC	S. Lungph	Damparer	Ramhlun S	Phaisen SC	Bunghmu	Bunghmu	Zawlp
	Tuirial SC	Tuirial SC	Tuirial SC	M. Kawnp	Marpara S	Zemabaw	Bunghmu	Thenhlum	Thenhlum	Selan
	New Chalrang SC	New Chalrang SC	New Chalrang SC	Mamit SC	Tuipuibari	Zuangtui S	Thenhlum SC			
	Rabung SC	Rabung SC	Rabung SC	Tuidam	Chhiahtlan	Kawrthah SC				
	Buhchangphai SC	Buhchangphai SC	Buhchangphai SC	N.Khawlek SC						
	Bazar SC	Bazar SC	Bazar SC	Daido SC						
	Laisawral SC	Laisawral SC	Laisawral SC	Sihfa SC						
	Reiek SC	Reiek SC	Reiek SC	Vanbawng SC						

... | Validation checkwise summary | Facility with Inconsistent | Facility with PRE | Validated Data

Fig 18: The sixth and the hyperlinked sheet of the downloaded file providing inconsistent detailed information check wise

8.7 Facility with Probable Reporting Error (Sheet -7)

This sheet provides the list of the facilities giving probable reporting error within the dataset with the count and the condition and the description of the condition.



Facility with PRE	0	1	2	3	4	5	6	
Conditions	6.6.2<=6.1.1+6.1.2+6.1.3+	6.6.3<=6.1.1+6.1.2+6.1.3+6.1.4	6.6.1<=6.1.1+6.1.2+6	1.5.1.a <= 1.1	1.1 <= 1.1.1	9.2.1 + 9.2.2>= 9.1.1+ 9	6.1.9 <= 3.	6.
Description	Number of cases of AEFI -	Number of cases of AEFI - Othe	Number of cases of A	Number of PW teste	Out of the total ANC re	Allopathic- Outpatient a	Child imm	Ch
Probable Reporting Error	8	8	6	8	0	1	9	
Facilities (Name) Showing PRE	Armveng SC	Armveng SC	Armveng SC	Maubuang SC	[]	Maubuang SC	Maubuan	M
	Chaltlang SC	Chaltlang SC	Chaltlang SC	Muallungthu SC			Muallungt	M
	Chhinga veng SC	Chhinga veng SC	Chhinga veng SC	Sateek SC			Phulpui SC	PH
	Ramhlun SC	Ramhlun SC	Ramhlun SC	Sialsuk SC			Sateek SC	Sa
	Tanhril SC	Sairang SC	Zemabawk SC	Tachhip SC			Tachhip SC	Ta
	Zemabawk SC	Zemabawk SC	Zuangtui SC	Armveng SC			Vaphai SC	Va
	Zotlang SC	Zotlang SC		Chaltlang SC			Parva SC	Pa
	Zuangtui SC	Zuangtui SC		Zemabawk SC			Mualchen	Ba
							Maite SC	M
								Re

Fig 19: The seventh and the hyperlinked sheet of the downloaded file providing PRE detailed information check wise

8.8 Validated Data (Sheet -8)

This sheet is the complete raw data with the checks embedded with the sheet which gives a detailed information, row wise identification of the inconsistent and PRE with the problematic indicators because of which the user is getting errors in its dataset

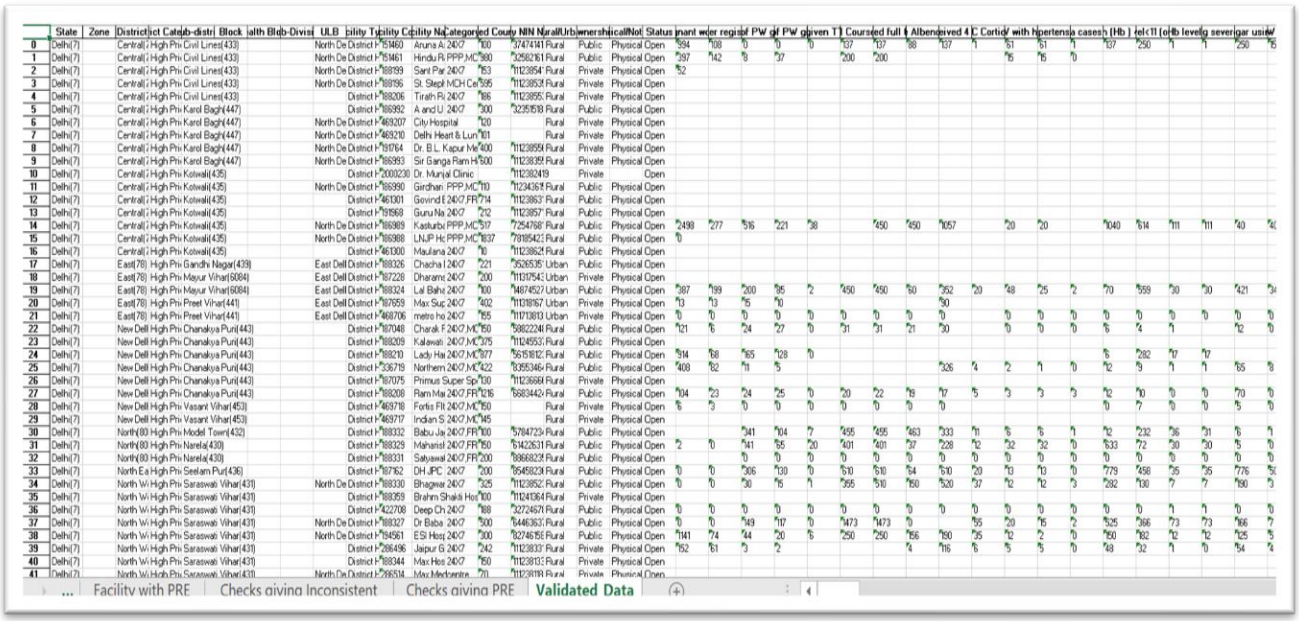


Fig 20: The eight sheet of the downloaded file providing a complete dataset uploaded by the user plus the result of the checks and the results showcased individually row wise highlighting the error present in the dataset.

9. Real Time Test Scenarios

Download report the District Hospital data from the portal https://hmis.nhp.gov.in/#!/login and open the excel sheet and save it in .xls/.xlsx format.

1) Upload the DH data into the tool *(A pop up message will show that the file is uploaded into the tool and also the link from where the file has been uploaded will be shown and also the facility type and duration will be coming).*

2) Press Validate button to validate the dataset *(A pop message will come showcasing the validation is complete for the district hospital facility).*

3) Insert four types of filters in the chorological order to filter out the dataset according to the area of interest of the user.

4) Press downloaded button to export the validated data for the tool and save the file in your desktop as per the desired location of the user *(A pop up message will be coming showing that the file has been saved at your desired location).*

5) Open the Summary sheet to check your errors in the datasets.

10. Important Terminologies & Classification of checks

1. **Consistent:** The respective validation check holds true and needs no scrutiny.
2. **Inconsistent:** The respective validation check fails and the inconsistent data item is flagged.
3. **Probable Reporting Errors**: The respective validation check may fail and it is subject to confirmation with the concerned authority. Check and verify.

10.1 Classification of checks

For related data items ("parent-child relation"):

- If the child data item has data, then the parent data item is blank, check will be interpreted as blank and needs to be highlighted.
- However, if the parent data item has a value and the child data item doesn't, then it may be highlighted as a probable reporting error. It is further subject to assessment through case studies and discussion with the state teams.

1.1.1 Out of the total ANC registered, number registered within 1st trimester (within 12 weeks)	1.1 Total number of pregnant women registered for ANC	Outcome (for 1.1.1<=1.1)
Value	Value	Consistent
Value	Value	Inconsistent (when condition fails)
Null	Value	Probable Reporting error
Value	Null	Inconsistent
Blank	Blank	Blank

For unrelated data items: When there is no relationship between the data items being compared:

1.5.1 Number of PW tested for Blood Sugar using OGTT (Oral glucose tolerance test)	1.1 Total number of pregnant women registered for ANC	Outcome (for 1.5.1<=1.1)
Value	Value	Consistent
Value	Value	Inconsistent (when condition fails)
Null	Value	Probable Reporting error
Value	Null	Inconsistent
Blank	Blank	Blank

For recurring data items: wherein service for one data items may be provided over months. It was suggested to have a permissible limit of +-50% i.e if the disparity in the two data items is more than suggested limit then it is a probable reporting error. We need to decide upon the threshold limit for these data item. These could be discussed with a public health specialist. However, the consistent and inconsistent outcome will change to probable reporting error if data items like 1.2.4 fails the threshold criteria.

1.2.4 Number of PW given 180 Iron Folic Acid (IFA) tablets	**1.1 Total number of pregnant women registered for ANC**	**Outcome (for 1.2.4<=1.1)**
Value	Value	Consistent
Value	Value	Inconsistent (when condition fails)
Null	Value	Probable Reporting error
Value	Null	Probable Reporting error
Blank	Blank	Blank

Complex Check: Involving multiple data items in either side of the check.

14.3.3 Number of Left Against Medical Advice (LAMA) cases	**14.3.1.a+14.3.1.b+14.3.2.a+14.3.2.b Inpatient (Male)-Children<18yrs+Inpatient (Male)-Adults+ Inpatient (Female)-Children<18yrs+Inpatient (Female)- Adults**	**Outcome (for 14.3.3<=14.3.1.a+14.3.1.b+14.3.2.a+14.3.2.b)**
Value	Value	Consistent
Value	Value	Inconsistent (when condition fails)
Null	Value	Probable Reporting error
Value	Null	Inconsistent
Blank	Blank	Blank