

Using Boolean Functions to Excel at Validation Tracking

Noory Kim, Statistics & Data Corporation

ABSTRACT

Our statistical programming teams use Excel to keep track of the completion and validation of project outputs. Large spreadsheets can provide detailed information on individual outputs but not a quick summary of overall progress. How can we enable all team members to be more mindful of the overall progress and thus take better ownership of meeting project deadlines? This paper will show how to combine Boolean formulas in Excel to count metrics such as (1) the number of topline outputs that passed validation and (2) the number of outputs that have passed validation on or after the primary completion date. The reader can use these formulas to create a dashboard in a separate tab of the validation tracker. This method does not require VBA or SAS®, making it relatively simple to implement. By including the dashboard in the validation tracker itself, the reader can provide an easy way for all programmers working on a project to check its overall progress.

INTRODUCTION

A validation tracking spreadsheet is an essential tool for measuring the progress of a project to ensure timely delivery of that project deliverables. Yet the validation tracker itself does not necessarily come with a quick and easy way to check the progress of a project, especially a large one with many tables and listings.

We can address this by using formulas in Excel to keep ongoing counts of outputs completed and validated. Figure 1 shows an example of a status tab added to a validation tracker.

Yet while counting rows that satisfy a single criterion is relatively easy to figure out (typically requiring only a single formula), counting rows that satisfy multiple criteria may not be as apparent. In this paper, I will show how to do so using Boolean array formulas.

	A	B	C	D	E	F	G	H
1				SDTM	ADs	Tables	Listings	Figures
2								
3	All		Number of Outputs Planned	56	12	76	34	20
4								
5		Since Start of Study	Number of Outputs Ready	56	12	74	30	1
6			Number Validated (not necessarily Passed)	56	12	50	29	0
7			Number Passed	56	12	41	20	0
8								
9		Since Reference Date	Number of Outputs Ready (on or after reference date)	56	12	74	30	1
10			Number Validated after Primary Completion Date and Reference Date	56	12	38	13	0
11			Number Passed after Primary Completion Date and Reference Date	56	12	35	10	0
12								
13								
14	Topline		Number of Topline Outputs Planned			52	24	2
15								
16		Since Start of Study	Number of Topline Outputs Ready			52	24	1
17			Number of Topline Validated (not necessarily Passed)			50	24	0
18			Number of Topline Passed			41	20	0
19								
20		Since Reference Date	Number of Topline Outputs Ready (on or after reference date)			52	24	1
21			Number of Topline Validated after Primary Completion Date and Reference Date			38	11	0
22			Number of Topline Passed after Primary Completion Date and Reference Date			35	10	0
23								
24								
25	Reference Date			11/30/2021	12/31/2021	1/31/2022	1/31/2022	1/31/2022

Figure 1: An Example of a Status Tab Added to the Validation Tracker. The counts in the top half are across all outputs, while the counts in the second half are across topline outputs. Within each half, the upper portion counts outputs from the beginning of the project, while the lower portion counts outputs only on or after the reference date specified in the last row (row 25).

Having such a Status tab can enable you to quickly compile a status report for your team. For example, you could adjust the reference dates (seen in the row 25 in Figure 1) to get the figures you need. Here's an example of such a status report:

Hi Team,

Thanks for your hard work so far! Here's where things currently stand with Project XYZ.

Tables

- 52 topline tables
- 52 primary tables in place
- 38 validated since 31JAN (73%)
- 35 passed since 31JAN (60%)

Listings

- 24 topline listings
- 24 primary listings in place
- 11 validated since 31JAN (46%)
- 10 passed since 31JAN (42%)

Keep up the great work!

Note that 'since' rather than 'as of' is used here in this report, as we want to make sure validation dates are current (no earlier than the chosen reference date – 31JAN2022 in this case).

HOW TO ACCESS EXCEL FORMULAS EN MASSE

In Excel, you can toggle between (a) the normal view (in which the results of formulas are displayed) and (b) the formula view (where the formulas themselves are displayed). The keyboard shortcut is CTRL + ` (the grave accent, above the left tab key). Figure 2 shows an example of such a view, which displays the formulas used to obtain the counts in Figure 1 (columns D to G). These can be copied into a text editor, where they can be viewed and edited more easily than in Excel itself.

	A	B	C	D	E	F	G	H
1				SDTM	ADs	Tables	Listings	Figures
3	All		Number of Outputs Planned	=SUM(IF(ISBLA	=SUM(IF(ISBLA	=SUM(IF(ISBLA	=SUM(IF(ISBLA	=SUM(IF(ISBLA
5		Since Start of Study	Number of Outputs Ready	=SUM(IF(ISBLA	=SUM(IF(ISBLA	=SUM(IF(ISBLA	=SUM(IF(ISBLA	=SUM(IF(ISBLA
6			Number Validated (not necessarily Passed)	=SUM(IF(ISBLA	=SUM(IF(ISBLA	=SUM(IF(ISBLA	=SUM(IF(ISBLA	=SUM(IF(ISBLA
7			Number Passed	=SUM(IF(ISBLA	=SUM(IF(ISBLA	=SUM(IF(ISBLA	=SUM(IF(ISBLA	=SUM(IF(ISBLA
9		Since Reference Date	Number of Outputs Ready (on or after reference date)	=SUM(IF(ISBLA	=SUM(IF(ISBLA	=SUM(IF(ISBLA	=SUM(IF(ISBLA	=SUM(IF(ISBLA
10			Number Validated after Primary Completion Date and Reference Date	=SUM(IF(ISBLA	=SUM(IF(ISBLA	=SUM(IF(ISBLA	=SUM(IF(ISBLA	=SUM(IF(ISBLA
11			Number Passed after Primary Completion Date and Reference Date	=SUM(IF(ISBLA	=SUM(IF(ISBLA	=SUM(IF(ISBLA	=SUM(IF(ISBLA	=SUM(IF(ISBLA
14	Topline		Number of Topline Outputs Planned			=SUM(IF(Table	=SUM(IF(Listir	=SUM(IF(Figur
15		Since Start of Study	Number of Topline Outputs Ready			=SUM(IF(Table	=SUM(IF(Listir	=SUM(IF(Figur
17			Number of Topline Validated (not necessarily Passed)			=SUM(IF(Table	=SUM(IF(Listir	=SUM(IF(Figur
18			Number of Topline Passed			=SUM(IF(Table	=SUM(IF(Listir	=SUM(IF(Figur
20		Since Reference Date	Number of Topline Outputs Ready (on or after reference date)			=SUM(IF(Table	=SUM(IF(Listir	=SUM(IF(Figur
21			Number of Topline Validated after Primary Completion Date and Reference Date			=SUM(IF(Table	=SUM(IF(Listir	=SUM(IF(Figur
22			Number of Topline Passed after Primary Completion Date and Reference Date			=SUM(IF(Table	=SUM(IF(Listir	=SUM(IF(Figur
25	Reference Date			44530	44561	44592	44592	44592

Figure 2: Viewing Formulas in Excel. The formula view can be accessed by the keyboard shortcut CTRL + ` (grave accent, above the left tab key). Here we can glimpse the formulas used to generate the counts seen in Figure 1.

FORMULAS USED IN THE STATUS TAB

The general form of the formulas used in this paper follow this pattern:

$$= \text{SUM}(\text{BAF}_1 * \text{BAF}_2 * \dots * \text{BAF}_n)$$

where each $\text{BAF}_\#$ is a Boolean array formula (BAF) that returns an array of Boolean values, one for each value in an input array.

ASSUMPTIONS IN THE FOLLOWING EXAMPLES

In the examples which follow below, we will assume that the Tables tab in the validation tracker includes the following columns:

Column Letter	A	B	G	H	M	N
Column Title	Table Number	Topline	Primary Completion Date	Final Run Date (of Primary Program)	Validation Date	Pass/Issues
Sample Value	14.1.1.1	Y	01-Jan-2022	28-Feb-2022	01-Mar-2022	Pass

Table 1: Columns in the Tables tab referred to by the examples in this paper.

We will also assume that there are no more than 776 tables (with none listed after row 777). In addition, cell F25 in the Status tab holds a reference date for comparing dates (as seen in Figure 1 and Figure 2.)

BOOLEAN ARRAY FORMULAS

Here are five individual Boolean formulas used to generate counts in the Status tab shown in Figure 1.

BAF1: Check if a cell is not blank.

= IF(ISBLANK(*input array*), 0, 1)

Return 1 for each cell not blank, and 0 for each blank cell. Note that by ordering 0 then 1 within the IF clause, we are effectively introducing a NOT Boolean operator (which is applied to the ISBLANK function).

Example: The following looks at cells in the A column of the 'Tables' tab. It counts the number of non-empty cells (up to the 777th row).

= IF(ISBLANK(Tables!\$A\$2:\$A\$777), 0, 1)

BAF2: Check if a cell has the substring "Pass" (case insensitive).

= ISNUMBER(SEARCH("Pass", *input array*))

Return 1 for each cell containing "Pass", and 0 for each cell not containing "Pass". The search function returns a number only for cells that contain "Pass" (case insensitive), and a #VALUE! error for cells that do not contain it.

Example:

= ISNUMBER(SEARCH("Pass", Tables!\$N\$2:\$N\$777))

BAF3: Check if a date is on or after a reference date.

= IF(*input array* >= *cell with reference date*, 1, 0)

Return 1 for dates in input array that are on or after the reference date, and 0 for dates before the reference date.

Example: The following formulas compares date values in column G in the Tables tab with the reference date in F25 of the Status tab.

= IF(Tables!\$G\$2:\$G\$777>=F\$25, 1, 0)

BAF4: Check if the validation date is on or after the primary date.

= IF(ISERROR(DATEDIF(*input array 1*, *input array 2*, "d")), 0, 1)

where *input array 1* has dates of primary programs or outputs and *input array 2* has validation dates. Return 1 if for rows where the validation date is on or after the (non-missing) primary date, and 0 for other rows.

(The DATEDIF function returns a numeric value if the validation date is on or after the primary date. Otherwise, it returns an error rather than returning a negative number.)

Example:

= IF(ISERROR(DATEDIF(Tables!\$G\$2:\$G\$777, Tables!\$M\$2:\$M\$777, "d")), 0, 1)

BAF5: Check if a primary output is indicated as a topline output.

= IF(*input array*, 1, 0)

In our case, topline outputs are indicated with a 'Y' in column B in the Tables tab.

= IF(Tables!\$B\$2:\$B\$777="Y", 1, 0)

COMBINING BOOLEAN ARRAY FORMULAS

By multiplying individual Boolean array formulas (BAFs), we are checking to see if *all* conditions corresponding to those formulas hold true. This returns 1 if all conditions hold true, and 0 otherwise. The SUM function adds up the number of rows that satisfy all conditions specified.

Let's walk through a couple of examples:

Example 1: Counting the number of topline tables that have passed validation

To count the number of topline tables, we need to count the number of rows satisfying the following criteria:

Criteria 1: Is the table indicated as a topline output?

We use BAF5 to check which rows have 'Y' in column B.

= IF(Tables!\$B\$2:\$B\$777="Y", 1, 0)

This appears with the following (non-Excel) notation in Table 2 below:

= BAF5(B)

where the letter inside the parentheses refers to the column with the input array.

Criteria 2: Has the table passed validation?

We use BAF2 to check which rows have the substring 'Pass' in column N.

= ISNUMBER(SEARCH("Pass", Tables!\$N\$2:\$N\$777))

This appears in Table 2 as

= BAF2(N)

We then multiply these criteria to see which rows satisfy all of them. This will return an output array of length 776, one for each row.

= IF(Tables!\$B\$2:\$B\$777="Y", 1, 0) * ISNUMBER(SEARCH("Pass", Tables!\$N\$2:\$N\$777))

This appears in Table 2 as

= BAF5(B) * BAF2(N)

Finally, we sum up the values of the output array to get our count.

= SUM(IF(Tables!\$B\$2:\$B\$777="Y", 1, 0)
* ISNUMBER(SEARCH("Pass", Tables!\$N\$2:\$N\$777)))

In Table 2 this appears as

= SUM(BAF5(B) * BAF2(N))

Example 2: Counting the number of topline tables that have passed validation on or after the primary completion date

Criteria 1: Is the table indicated as a topline output?

As we saw in Example 1, this appears in Table 2 as

= BAF5(B)

Criteria 2: Has the table has passed validation?

As we saw in Example 1, this appears in Table 2 as

= BAF2(N)

Criteria 3: Is the date of validation on or after the date of primary completion?

We use BAF4 to check which rows have a validation date (in column M) that is on or after the primary completion date (in column J).

= IF(ISERROR(DATEDIF(Tables!\$G\$2:\$G\$777,Tables!\$M\$2:\$M\$777,"d")), 0, 1)

This appears in Table 2 as

= BAF4(G,M)

where the first letter in the parentheses refers to the column with input array 1, and the second letter refers to the column with input array 2.

We multiply these three criteria to see which rows satisfy all of them.

= IF(Tables!\$B\$2:\$B\$777="Y", 1, 0)
* ISNUMBER(SEARCH("Pass", Tables!\$N\$2:\$N\$777))
* IF(ISERROR(DATEDIF(Tables!\$G\$2:\$G\$777,Tables!\$M\$2:\$M\$777,"d")), 0, 1)

Using the (non-Excel) notation of Table 2, we can write this as

= BAF5(B) * BAF2(N) * BAF4(G,M)

Finally, we sum up the values of the output array to get our count.

= SUM(IF(Tables!\$B\$2:\$B\$777="Y", 1, 0)
* ISNUMBER(SEARCH("Pass", Tables!\$N\$2:\$N\$777))
* IF(ISERROR(DATEDIF(Tables!\$G\$2:\$G\$777,Tables!\$M\$2:\$M\$777,"d")), 0, 1))

Using the (non-Excel) notation of Table 2, we can write this as

= SUM(BAF5(B) * BAF2(N) * BAF4(G,M))

In a similar fashion, we can construct the other formulas used to generate the counts seen in Figure 1. Table 2 shows these formulas for counting tables, using the following non-Excel notation:

BAF#(X)

BAF#(X,Y)

where BAF# refers to a formula listed on page 4, and X and Y each refer to a column letter in the Tables tab containing an input array for that formula.

The complete Excel formulas are shown in the Appendix.

Metric Category	Metric	Formula (in non-Excel Notation)
All tables	Number of Outputs Planned	=SUM(BAF1(A))
All tables since start of study	Number of Outputs Ready	=SUM(BAF1(A) * BAF1(G))
	Number Validated (not necessarily Passed)	=SUM(BAF1(A) * BAF1(M))
	Number Passed	=SUM(BAF1(A) * BAF2(N))
All tables since reference date	Number of Outputs Ready (on or after reference date)	=SUM(BAF1(A) * BAF3(G))
	Number Validated after Primary Completion Date and Reference Date	=SUM(BAF1(A) * BAF1(M) * BAF3(G) * BAF4(G,M))
	Number Passed after Primary Completion Date and Reference Date	=SUM(BAF1(A) * BAF2(N) * BAF3(G) * BAF4(G,M))
	Number Passed after Primary Final Run Date and Reference Date	=SUM(BAF1(A) * BAF2(N) * BAF3(H) * BAF4(H,M))
Topline tables	Number of Topline Outputs Planned	=SUM(BAF5(B))
Topline tables since start of study	Number of Topline Outputs Ready	=SUM(BAF5(B) * BAF1(G))
	Number of Topline Validated	=SUM(BAF5(B) * BAF1(M))
	Number of Topline Passed	=SUM(BAF5(B) * BAF2(N))
Topline tables since reference date	Number of Topline Outputs Ready (on or after reference date)	=SUM(BAF5(B) * BAF3(G))
	Number of Topline Validated after Primary Completion Date and Reference Date	=SUM(BAF5(B) * BAF1(M) * BAF3(G) * BAF4(G,M))
	Number of Topline Passed after Primary Completion Date and Reference Date	=SUM(BAF5(B) * BAF2(N) * BAF3(G) * BAF4(G,M))
	Number of Topline Passed after Primary Final Run Date and Reference Date	=SUM(BAF5(B) * BAF2(N) * BAF3(H) * BAF4(H,M))

Table 2: Products of Boolean array formulas (BAF) used. BAF# refers to a specific BAF listed on the previous section, while the letter following in parenthesis indicates the column(s) with the input arrays for that BAF (as seen in Table 1.). The actual Excel formulas used are listed in the Appendix.

CONCLUSION

This paper covered the following:

- Accessing Excel formulas en masse by toggling to and from the formula view in Excel
- Using a Boolean array function to count rows satisfying a single criterion
- Combining Boolean array functions to count rows satisfying multiple criteria

These formulas can be used in a 'Status' tab as part of the validation tracking spreadsheet, providing a quick way for all programmers of the project team to keep tabs on the progress of a project.

REFERENCES

Microsoft Support. "Calculate the difference between two dates." <https://support.microsoft.com/en-us/office/calculate-the-difference-between-two-dates-8235e7c9-b430-44ca-9425-46100a162f38>. Accessed 01MAR2022.

ACKNOWLEDGMENTS

Thanks to Bharath Donthi, Leslie Schneider, and Adam Hamm for reviewing drafts of the abstract. All errors remain the responsibility of the author.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Noory Kim
nkim@sdclinical.com

Any brand and product names are trademarks of their respective companies.

APPENDIX

Below are the complete Excel formulas used in the Status tab shown in Figure 1 (and referred to by the non-Excel notation in Table 2).

Metric Category	Metric	Excel Formula
All tables	Number of Outputs Planned	=SUM(IF(ISBLANK(Tables!\$A\$2:\$A\$777), 0, 1))
All tables since start of study	Number of Outputs Ready	=SUM(IF(ISBLANK(Tables!\$A\$2:\$A\$777), 0, 1) * IF(ISBLANK(Tables!\$G\$2:\$G\$777), 0, 1))
	Number Validated (not necessarily Passed)	=SUM(IF(ISBLANK(Tables!\$A\$2:\$A\$777), 0, 1) * IF(ISBLANK(Tables!\$M\$2:\$M\$777), 0, 1))
	Number Passed	=SUM(IF(ISBLANK(Tables!\$A\$2:\$A\$777), 0, 1) * ISNUMBER(SEARCH("Pass", Tables!\$N\$2:\$N\$777)))
All tables since reference date	Number of Outputs Ready (on or after reference date)	=SUM(IF(ISBLANK(Tables!\$A\$2:\$A\$777), 0, 1) * IF(Tables!\$G\$2:\$G\$777>=F\$25, 1, 0))
	Number Validated after Primary Completion Date and Reference Date	=SUM(IF(ISBLANK(Tables!\$A\$2:\$A\$777), 0, 1) * IF(ISBLANK(Tables!\$M\$2:\$M\$777), 0, 1) * IF(Tables!\$G\$2:\$G\$777>=F\$25, 1, 0) * IF(ISERROR(DATEDIF(Tables!\$G\$2:\$G\$777, Tables!\$M\$2:\$M\$777, "d")), 0, 1))
	Number Passed after Primary Completion Date and Reference Date	=SUM(IF(ISBLANK(Tables!\$A\$2:\$A\$777), 0, 1) * ISNUMBER(SEARCH("Pass", Tables!\$N\$2:\$N\$777)) * IF(Tables!\$G\$2:\$G\$777>=F\$25, 1, 0) * IF(ISERROR(DATEDIF(Tables!\$G\$2:\$G\$777, Tables!\$M\$2:\$M\$777, "d")), 0, 1))
	Number Passed after Primary Final Run Date and Reference Date	=SUM(IF(ISBLANK(Tables!\$A\$2:\$A\$777), 0, 1) * ISNUMBER(SEARCH("Pass", Tables!\$N\$2:\$N\$777)) * IF(Tables!\$H\$2:\$H\$777>=F\$25, 1, 0) * IF(ISERROR(DATEDIF(Tables!\$H\$2:\$H\$777, Tables!\$M\$2:\$M\$777, "d")), 0, 1))
Topline tables	Number of Topline Outputs Planned	=SUM(IF(Tables!\$B\$2:\$B\$777="Y", 1, 0))
Topline tables since start of study	Number of Topline Outputs Ready	=SUM(IF(Tables!\$B\$2:\$B\$777="Y", 1, 0) * IF(ISBLANK(Tables!\$G\$2:\$G\$777), 0, 1))
	Number of Topline Validated	=SUM(IF(Tables!\$B\$2:\$B\$777="Y", 1, 0) * IF(ISBLANK(Tables!\$M\$2:\$M\$777), 0, 1))
	Number of Topline Passed	=SUM(IF(Tables!\$B\$2:\$B\$777="Y", 1, 0) * ISNUMBER(SEARCH("Pass", Tables!\$N\$2:\$N\$777)))
Topline tables since reference date	Number of Topline Outputs Ready (on or after reference date)	=SUM(IF(Tables!\$B\$2:\$B\$777="Y", 1, 0) * IF(Tables!\$G\$2:\$G\$777>=F\$25, 1, 0))
	Number of Topline Validated after Primary Completion Date and Reference Date	=SUM(IF(Tables!\$B\$2:\$B\$777="Y", 1, 0) * IF(ISBLANK(Tables!\$M\$2:\$M\$777), 0, 1) * IF(Tables!\$G\$2:\$G\$777>=F\$25, 1, 0) * IF(ISERROR(DATEDIF(Tables!\$G\$2:\$G\$777, Tables!\$M\$2:\$M\$777, "d")), 0, 1))
	Number of Topline Passed after Primary Completion Date and Reference Date	=SUM(IF(Tables!\$B\$2:\$B\$777="Y", 1, 0) * ISNUMBER(SEARCH("Pass", Tables!\$N\$2:\$N\$777)) * IF(Tables!\$G\$2:\$G\$777>=F\$25, 1, 0) * IF(ISERROR(DATEDIF(Tables!\$G\$2:\$G\$777, Tables!\$M\$2:\$M\$777, "d")), 0, 1))
	Number of Topline Passed after Primary Final Run Date and Reference Date	=SUM(IF(Tables!\$B\$2:\$B\$777="Y", 1, 0) * ISNUMBER(SEARCH("Pass", Tables!\$N\$2:\$N\$777)) * IF(Tables!\$H\$2:\$H\$777>=F\$25, 1, 0) * IF(ISERROR(DATEDIF(Tables!\$H\$2:\$H\$777, Tables!\$M\$2:\$M\$777, "d")), 0, 1))