

Generating the Demographics and Adverse Event Tables Using Excel and VBA

David Franklin, TheProgrammersCabin.com

ABSTRACT

Microsoft Excel is not regarded as a package to do statistical analysis for a number of reasons, namely a spreadsheet is not technically advanced enough to ensure proper data/documentation management and workflows, and statistical functions are lacking or not beyond the capabilities of serious statistical computational needs. Despite this, Excel is still the leading software package used to do statistics. This paper looks at two distinct table types, the Demographics and the Adverse Event tables; and three distinct categories of data, the categorical, continuous and occurrence categories, and shows that Excel with VBA can indeed be used to produce these tables. While not used for publication or serious reporting, these could be used as a way to look at the data before 'official' reports are available. Along the way, we will also see how an Adverse Event Timeline Plot may be created.

INTRODUCTION

Why use Excel? It is not regarded as a serious package to do statistical analysis. But it is easy to use, most people have access to it, it easier to get a person to do a simple Demographics table using Excel than find a person who knows SAS or R to do the same. However, Excel has one big disadvantage, managing data – getting an updated set of data for a study and rerunning SAS programs to produce output is far quicker and easier than importing data into Excel sheets and processing results.

That being said, Excel does have an advantage of seeing results in real time, the intimacy of reviewing numbers on a screen without switching between files.

This paper shows how it is possible to use Excel to produce output using different techniques which are given in the paper.

THE DATA

Two datasets will be used in this paper, ADSL:

	A	B	C	D	E	F	G	H	I	J
1	STUDYID	USUBJID	SUBJID	SAFFL	AGE	AGEU	SEX	RACE	ETHNIC	ARMCD
2	NewSauceA	NewSauceA-101	101	Y	23	YEARS	M	OTHER	NOT HISPANIC OR LATINO	SAUCE
3	NewSauceA	NewSauceA-102	102	Y	73	YEARS	M	WHITE	NOT HISPANIC OR LATINO	PLACEBO
4	NewSauceA	NewSauceA-103	103	Y	32	YEARS	F	MULTIPLE		SAUCE
5	NewSauceA	NewSauceA-201	201	Y	31	YEARS	F	BLACK OR AFRICAN AMERICAN	NOT HISPANIC OR LATINO	PLACEBO
6	NewSauceA	NewSauceA-202	202	N	86	YEARS	M	WHITE	NOT HISPANIC OR LATINO	SCRNFAL
7	NewSauceA	NewSauceA-203	203	Y	49	YEARS	M	ASIAN	NOT HISPANIC OR LATINO	SAUCE
8	NewSauceA	NewSauceA-204	204	Y	58	YEARS	F	WHITE	NOT HISPANIC OR LATINO	PLACEBO
9	NewSauceA	NewSauceA-301	301	Y	51	YEARS	F	MULTIPLE		SAUCE
10	NewSauceA	NewSauceA-302	302	Y	36	YEARS	M	BLACK OR AFRICAN AMERICAN	NOT HISPANIC OR LATINO	PLACEBO
11	NewSauceA	NewSauceA-303	303	Y	65	YEARS	M	WHITE	NOT HISPANIC OR LATINO	SAUCE

and ADAE:

	A	B	C	D	E	F
1	USUBJID	SUBJID	ARMCD	AESEQ	AEDECOD	AEBODSYS
2	NewSauceA-101	101	SAUCE	1	Flatulence	Gastrointestinal disorders
3	NewSauceA-101	101	SAUCE	2	Constipation	Gastrointestinal disorders
4	NewSauceA-101	101	SAUCE	3	Fatigue	General disorders and administration site conditions
5	NewSauceA-102	102	PLACEBO	1	Headache	Nervous system disorders
6	NewSauceA-102	102	PLACEBO	2	Dizziness	Nervous system disorders
7	NewSauceA-102	102	PLACEBO	3	Muscle spasms	Infections and infestations
8	NewSauceA-103	103	SAUCE	1	Food poisoning	Gastrointestinal disorders
9	NewSauceA-201	201	PLACEBO	1	Nausea	Gastrointestinal disorders
10	NewSauceA-203	203	SAUCE	1	Diarrhoea	Gastrointestinal disorders
11	NewSauceA-204	204	PLACEBO	2	Diarrhoea	Gastrointestinal disorders
12	NewSauceA-301	301	SAUCE	1	Constipation	Gastrointestinal disorders
13	NewSauceA-301	301	SAUCE	2	Palpitations	Cardiac disorders
14	NewSauceA-301	301	SAUCE	3	Vomiting	Gastrointestinal disorders
15	NewSauceA-302	302	PLACEBO	1	Haemorrhoids	Gastrointestinal disorders
16	NewSauceA-303	303	SAUCE	1	Anxiety	Psychiatric disorders
17	NewSauceA-303	303	SAUCE	2	Vomiting	Gastrointestinal disorders

NAMING RANGES

Excel has the ability to set a range of cells to a “Named Range”. In Excel 365 first highlight range then put a name in left box immediately under the ribbon, as the following example demonstrates:

	A	B	C	D	E	F	G	H	I	J
1	STUDYID	USUBJID	SUBJID	SAFFL	AGE	AGEU	SEX	RACE	ETHNIC	ARMCD
2	NewSauceA	NewSauceA-101	101	Y	23	YEARS	M	OTHER	NOT HISPANIC OR LATINO	SAUCE
3	NewSauceA	NewSauceA-102	102	Y	73	YEARS	M	WHITE	NOT HISPANIC OR LATINO	PLACEBO
4	NewSauceA	NewSauceA-103	103	Y	32	YEARS	F	MULTIPLE		SAUCE
5	NewSauceA	NewSauceA-201	201	Y	31	YEARS	F	BLACK OR AFRICAN AMERICAN	NOT HISPANIC OR LATINO	PLACEBO
6	NewSauceA	NewSauceA-202	202	N	86	YEARS	M	WHITE	NOT HISPANIC OR LATINO	SCRNFAIL
7	NewSauceA	NewSauceA-203	203	Y	49	YEARS	M	ASIAN	NOT HISPANIC OR LATINO	SAUCE
8	NewSauceA	NewSauceA-204	204	Y	58	YEARS	F	WHITE	NOT HISPANIC OR LATINO	PLACEBO
9	NewSauceA	NewSauceA-301	301	Y	51	YEARS	F	MULTIPLE		SAUCE
10	NewSauceA	NewSauceA-302	302	Y	36	YEARS	M	BLACK OR AFRICAN AMERICAN	NOT HISPANIC OR LATINO	PLACEBO
11	NewSauceA	NewSauceA-303	303	Y	65	YEARS	M	WHITE	NOT HISPANIC OR LATINO	SAUCE
12										
13										

Any named range must be unique, i.e. only one ‘AGE’, within each workbook.

SUBSETTING CAPABILITY WITH SOME FUNCTIONS

Some functions in Excel have subsetting capability, but not all. An example function call is:

```
MEDIAN ( IF ( SAFFL&ARMCD="Y" &UPPER ( G$2 ) , AGE ) )
```

where we get the MEDIAN value of named range AGE on the condition that the SAFFL="Y" and ARMCD has the value of G\$2, in this case it is either ‘Sauce’ or ‘Placebo’. The way this works is confusing at first but essentially it is the values of SAFFL and ARMCD are concatenated and compared with concatenated values of “Y” and uppercase values of cell G\$2, then uses matched rows to call function MEDIAN of values for named range AGE.

THE DEMOGRAPHICS TABLE

Below is the mock Demographics table that is going to be filled:

A	B	C	D	E
1	Yum Yum Inc:			
2	Protocol: NewSauceA			
3	Table 14.1.2 -- Subject Demographics			
4	Safety Population			
5				
6	Parameter	Sauce (N=xx)	Placebo (N=xx)	Total (N=xx)
7				
8	Age (years) [a]			
9	n	xx	xx	xx
10	Mean (SD)	xx.x (xx.xx)	xx.x (xx.xx)	xx.x (xx.xx)
11	Median	xx.x	xx.x	xx.x
12	Min	xx	xx	xx
13	Max	xx	xx	xx
14				
15	Gender			
16	n	xx	xx	xx
17	Male	xx.x (xx.x%)	xx.x (xx.x%)	xx.x (xx.x%)
18	Female	xx.x (xx.x%)	xx.x (xx.x%)	xx.x (xx.x%)
19	Missing	xx	xx	xx
20				
21	[a] Age in years at Consent.			
22	Tab: T14.1.2 (2) / Date: 2021-05-08			

As in other products it is very useful to do some calculations and other intermediate work outside the table area on the same sheet. The intermediate steps do not have to be printed in the final table as it is possible to set the print area when printing the table — it does not have to be the entire sheet.

As we do in most programming of a table the first numbers that is sought is the Population which in this case is

	G	H	I
1			
2	Sauce	Placebo	Total
3			
4	Populations		
5	5	4	9

Text on row 2 and 4 are just that, text. As for the calculations these are:

```
G5: =COUNT ( IF (SAFFL&ARMCD="Y" &UPPER (G$2) , SUBJID ) )
H5: =COUNT ( IF (SAFFL&ARMCD="Y" &UPPER (H$2) , SUBJID ) )
I5: =SUM (G5:H5)
```

For Cell I5 the SUM function has been used to produce the total rather than calling another COUNT.

To include these in the header row for the table:

```
C6: =CONCAT ("Sauce (N=", G5, ")")
D6: =CONCAT ("Placebo (N=", H5, ")")
E6: =CONCAT ("Total (N=", I5, ")")
```

and the result is:

	A	B	C	D	E
1	Yum Yum Inc:				
2	Protocol: NewSauceA				
3	Table 14.1.2 -- Subject Demographics				
4	Safety Population				
5					
6	Parameter		Sauce (N=5)	Placebo (N=4)	Total (N=9)
7					

The calculations for AGE in column C are:

```

C9:=TEXT(COUNT(IF(SAFFL&ARMCD="Y"&UPPER(G$2),AGE)), "0")
C10:=CONCAT(TEXT(AVERAGE(IF(SAFFL&ARMCD="Y"&UPPER(G$2),AGE)), "0.0"), "
(", TEXT(STDEV.S(IF(SAFFL&ARMCD="Y"&UPPER(G$2),AGE)), "0.00"), ")") <-All One Line
C11:=TEXT(MEDIAN(IF(SAFFL&ARMCD="Y"&UPPER(G$2),AGE)), "0.0")
C12:=TEXT(MIN(IF(SAFFL&ARMCD="Y"&UPPER(G$2),AGE)), "0")
C13:=TEXT(MAX(IF(SAFFL&ARMCD="Y"&UPPER(G$2),AGE)), "0")

```

Subsetting has already been described above but here two more functions are being here; first TEXT which takes a number and allows for a "format" to be applied, e.g. "0" rounds to an integer, and CONCAT which will concatenate text strings together. Note that for Mean the function AVERAGE is used, and Standard Deviation uses the STDEV.S function (this provides the Sample Standard Deviation).

For cells D9 to D13 and E9 to E13 following changes were made to the formulas above:

```

D9-13:UPPER(H$2) instead of UPPER(G$2)
E9-13:SAFFL="Y" instead of SAFFL&ARMCD="Y"&UPPER(G$2)

```

The result of all this work is:

	A	B	C	D	E
7					
8	Age (years) [a]				
9	n		5	4	9
10	Mean (SD)		44.0 (16.58)	49.5 (19.57)	46.4 (17.02)
11	Median		49.0	47.0	49.0
12	Min		23	31	23
13	Max		65	73	73
14					

Doing the calculations outside the table area is possible but example shows that it is possible to do this directly inside the table itself.

The calculations for GENDER do use an area outside the table for counts as the following demonstrates:

```

G15:=COUNT(IF(SAFFL&SEX&UPPER(ARMCD)="YM"&UPPER(G$2),SUBJID))
H15:=COUNT(IF(SAFFL&SEX&UPPER(ARMCD)="YM"&UPPER(H$2),SUBJID))
I15:=SUM(G15,H15)
G16-H16: "YF" instead of "YM"
I16:=SUM(G16,H16)

```

Now putting this inside the table requires a second set of function calls:

```

C16:=SUM(G15:G16)
C17:=CONCAT(TEXT(G15,"0")," ("&TEXT(G15*100/C$16,"0.0"),"%")
C18:=CONCAT(TEXT(G16,"0")," ("&TEXT(G16*100/C$16,"0.0"),"%")
C19:=TEXT(G5-C16,"0")
D16-19: Reference to column H instead of G
E16-19: Reference to column I instead of G

```

The end result of this is:

	A	B	C	D	E	F	G	H	I
14									
15	Gender						3	2	5
16	n		5	4	9		2	2	4
17	Male		3 (60.0%)	2 (50.0%)	5 (55.6%)				
18	Female		2 (40.0%)	2 (50.0%)	4 (44.4%)				
19	Missing		0	0	0				
20									

This is a good example of where it is easier to use some intermediate steps than to do the calculation and display in one cell.

To display the final footnote with tab name and date, the following formula is used:

```
=CONCAT("Tab: ",MID(CELL("filename",A1),FIND("]",CELL("filename",A1))+1,255)," /
Date: ",TEXT(TODAY(),"yyyy-mm-dd"))
```

By default the CELL function using the FILENAME parameter will bring the complete path and tab name but the request was for the tab name only, hence the double CELL call.

The TODAY function will get the current date.

THE ADVERSE EVENT TABLE

In the last section two of the three data types, continuous and categorical, were summarized – now it is the turn of the occurrence data where there may be several observations that are only counted once.

The first step is to create three new variables (flags):

1. Sort the ADAE data by SUBJID, AEBODSYS and AEDECOD

	A	B	C	D	E	F	G	H	I
1	USUBJID	SUBJID	ARMCD	AESEQ	AEDECOD	AEBODSYS	KSUBJID	KAEBODSYS	KAEDECOD
2	NewSauceA-101	101	SAUCE	2	Constipation	Gastrointestinal disorders	1	1	1
3	NewSauceA-101	101	SAUCE	1	Flatulence	Gastrointestinal disorders	0	0	1
4	NewSauceA-101	101	SAUCE	3	Fatigue	General disorders and administration site conditions	0	1	1
5	NewSauceA-101	101	SAUCE	2	Constipation	Gastrointestinal disorders	1	1	1

2. Create labels KSUBJID, KAEBODSYS and KAEDECOD at G1, H1 and I1 respectively
3. Set 1 to G2, H2 and I2.
4. G3:=IF(B3=B2,0,1). <- if match of SUBJID with previous record set 0, else 1
5. H3:=IF(B3&F3=B2&F2,0,1) <- if match of SUBJID/AEBODSYS with previous record set 0, else 1
6. I3:=IF(B3&F3&E3=B2&F2&E2,0,1)
7. Copy G3, H3 and I3 to rest of the rows for respective columns
8. Create Names Ranges KSUBJID, KAEBODSYS and KAEDECOD for the three new variables

Once this step is done the next step is to create the template for output, as in our data:

	A	B	C	D	E
1	Yum Yum Inc:				
2	Protocol: NewSauceA				
3	Table 14.3.2 -- Adverse Events By SOC and Preferred Term				
4	Safety Population				
5					
6	Body System		Sauce	Placebo	Total
7	Preferred Term		(N=xx)	(N=xx)	(N=xx)
8	Any Adverse Event		xx.x (xx.x%)	xx.x (xx.x%)	xx.x (xx.x%)
9					
10	Cardiac disorders		xx.x (xx.x%)	xx.x (xx.x%)	xx.x (xx.x%)
11	Palpitations		xx.x (xx.x%)	xx.x (xx.x%)	xx.x (xx.x%)
12					
13	Gastrointestinal disorders		xx.x (xx.x%)	xx.x (xx.x%)	xx.x (xx.x%)
14	Constipation		xx.x (xx.x%)	xx.x (xx.x%)	xx.x (xx.x%)
15	Diarrhoea		xx.x (xx.x%)	xx.x (xx.x%)	xx.x (xx.x%)
16	Flatulence		xx.x (xx.x%)	xx.x (xx.x%)	xx.x (xx.x%)
17	Food poisoning		xx.x (xx.x%)	xx.x (xx.x%)	xx.x (xx.x%)
18	Haemorrhoids		xx.x (xx.x%)	xx.x (xx.x%)	xx.x (xx.x%)
19	Nausea		xx.x (xx.x%)	xx.x (xx.x%)	xx.x (xx.x%)
20	Vomiting		xx.x (xx.x%)	xx.x (xx.x%)	xx.x (xx.x%)
21					
22	General disorders and administration site conditions		xx.x (xx.x%)	xx.x (xx.x%)	xx.x (xx.x%)
23	Fatigue		xx.x (xx.x%)	xx.x (xx.x%)	xx.x (xx.x%)
24					
25	Infections and infestations		xx.x (xx.x%)	xx.x (xx.x%)	xx.x (xx.x%)
26	Muscle spasms		xx.x (xx.x%)	xx.x (xx.x%)	xx.x (xx.x%)
27					
28	Nervous system disorders		xx.x (xx.x%)	xx.x (xx.x%)	xx.x (xx.x%)
29	Dizziness		xx.x (xx.x%)	xx.x (xx.x%)	xx.x (xx.x%)
30	Headache		xx.x (xx.x%)	xx.x (xx.x%)	xx.x (xx.x%)
31					
32	Psychiatric disorders		xx.x (xx.x%)	xx.x (xx.x%)	xx.x (xx.x%)
33	Anxiety		xx.x (xx.x%)	xx.x (xx.x%)	xx.x (xx.x%)
34					
35	Note: Percentages are based on N				

The Population and Table Header the same as previous.

Next the counts:

	A	B	C	D	E	F	G	H	I	J	K	
1	Yum Yum Inc:											
2	Protocol: NewSauceA						Sauce	Placebo	Total			
3	Table 14.3.2 -- Adverse Events By SOC and Preferred Term											
4	Safety Population											
5							Populations					
6							5	4	9			
7	Body System		Sauce	Placebo	Total							
8	Preferred Term		(N=5)	(N=4)	(N=9)							
9	Any Adverse Event		5 (100.0%)	4 (100.0%)	9 (100.0%)		Any Adverse Event			5	4	9
10	Cardiac disorders		1 (20.0%)	0 (0.0%)	1 (11.1%)		Cardiac disorders			1	0	1
11	Palpitations		1 (20.0%)	0 (0.0%)	1 (11.1%)		Cardiac dis Palpitations			1	0	1

Note that rows in the intermediate area mirror what is in the actual table – this helps with calculations shorty. The Any Adverse Event counts are done by:

```
I7:=SUM(IF(AE_ARMCD=UPPER(G$2),KSUBJID))
J7:=SUM(IF(AE_ARMCD=UPPER(H$2),KSUBJID))
K7:=SUM(I7:J7)
```

Now bring the counts into the table and do the percentage counts:

```
C8:=CONCAT(TEXT(I7,"0")," ("&TEXT(I7*100/G$5,"0.0")&,"%")")
D8-E8:Copy of C8
```

Now for the AEBODSYS counts:

```
I9:=SUM(IF(AEBODSYS&AE_ARMCD=$A10&UPPER(G$2),KAEBODSYS))
J9:=SUM(IF(AEBODSYS&AE_ARMCD=$A10&UPPER(H$2),KAEBODSYS))
K9:=SUM(I9:J9)
```

Counts can be brought into the main table as above.

For counts with AEDECOD, in I10 and J10, replace AEBODSYS with AEDECOD, and KAEBODSYS with KAEDECOD.

To speed things up, result cells, e.g. C10:E11, can be copied from C8.

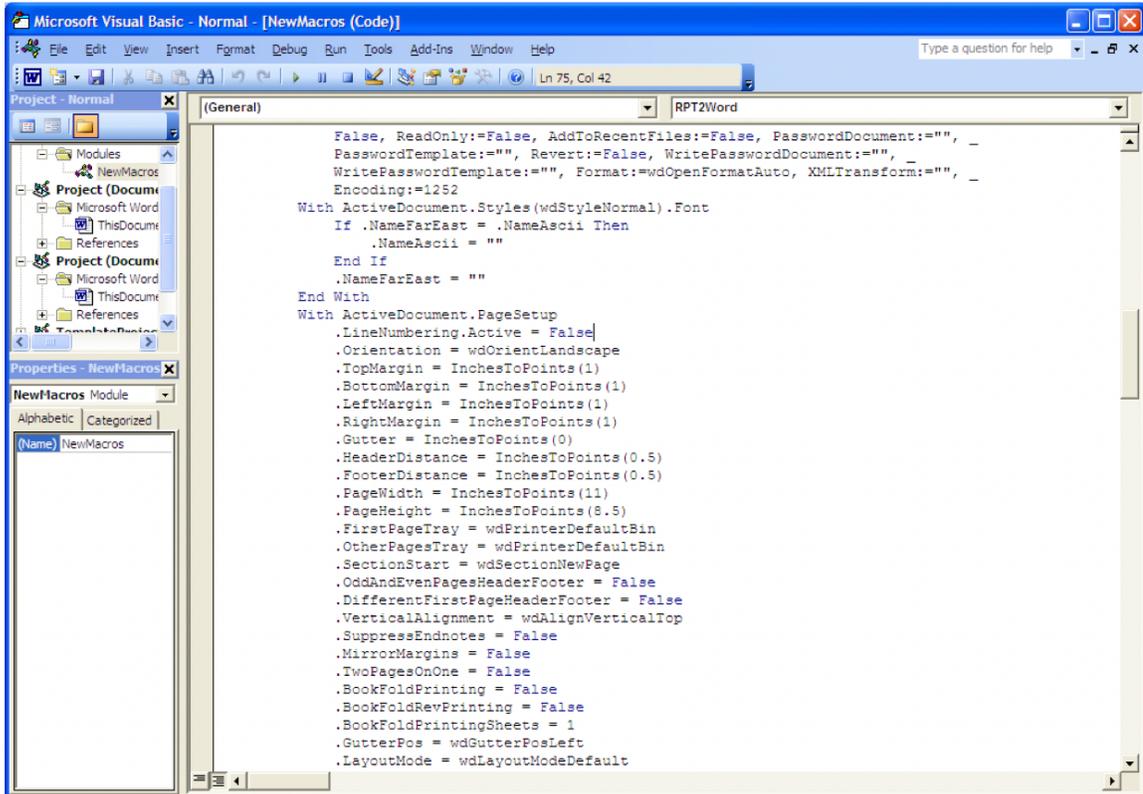
Our final Adverse Event table is

	A	B	C	D	E
1		Yum Yum Inc:			
2		Protocol: NewSauceA			
3		Table 14.3.2 -- Adverse Events By SOC and Preferred Term			
4		Safety Population			
5					
6		Body System	Sauce	Placebo	Total
7		Preferred Term	(N=5)	(N=4)	(N=9)
8		Any Adverse Event	5 (100.0%)	4 (100.0%)	9 (100.0%)
9					
10		Cardiac disorders	1 (20.0%)	0 (0.0%)	1 (11.1%)
11		Palpitations	1 (20.0%)	0 (0.0%)	1 (11.1%)
12					
13		Gastrointestinal disorders	5 (100.0%)	3 (75.0%)	8 (88.9%)
14		Constipation	2 (40.0%)	0 (0.0%)	2 (22.2%)
15		Diarrhoea	1 (20.0%)	1 (25.0%)	2 (22.2%)
16		Flatulence	1 (20.0%)	0 (0.0%)	1 (11.1%)
17		Food poisoning	1 (20.0%)	0 (0.0%)	1 (11.1%)
18		Haemorrhoids	0 (0.0%)	1 (25.0%)	1 (11.1%)
19		Nausea	0 (0.0%)	1 (25.0%)	1 (11.1%)
20		Vomiting	2 (40.0%)	0 (0.0%)	2 (22.2%)
21					
22		General disorders and administration site conditions	1 (20.0%)	0 (0.0%)	1 (11.1%)
23		Fatigue	1 (20.0%)	0 (0.0%)	1 (11.1%)
24					
25		Infections and infestations	0 (0.0%)	1 (25.0%)	1 (11.1%)
26		Muscle spasms	0 (0.0%)	1 (25.0%)	1 (11.1%)
27					
28		Nervous system disorders	0 (0.0%)	1 (25.0%)	1 (11.1%)
29		Dizziness	0 (0.0%)	1 (25.0%)	1 (11.1%)
30		Headache	0 (0.0%)	1 (25.0%)	1 (11.1%)
31					
32		Psychiatric disorders	1 (20.0%)	0 (0.0%)	1 (11.1%)
33		Anxiety	1 (20.0%)	0 (0.0%)	1 (11.1%)
34					
35		Note: Percentages are based on N			
36		Tab: T14.3.2 / Date: 2021-05-08			

VBA

The first question that comes up when talking about VBA, or its true name Visual Basic for Applications, is what is it? In its simplest form, VBA is a programming language that is inside products like Microsoft Word that allow for the building of user defined functions and automating processes, closely related to the programming language Visual Basic and its predecessors, Quick Basic and WordBasic. VBA programs can be fixed to a menu button, a keyboard shortcut, or a window icon outside the Microsoft Word environment.

Microsoft Word also has its own integrated development environment (IDE), an example of which is shown below:



It is in the IDE that VBA source can be written, edited and compiled for use in your work.

An easy way to create a VBA macro is to record keystrokes that it can stored inside the spreadsheet. This is a way of programming for repeating analysis and can be edited in such a way that allowance for new data is possible.

A tutorial on VBA programming is beyond the scope of this paper.

ADVERSE EVENT TIMELINE PLOT

This is a very specialized plot that can display a set of Adverse Events over time, typically by patient.

The type of plot that is need is almost like a Gantt Chart but can be more easily thought of as a stacked horizontal bar chart.

Lets look at some data:

	A	B	C	D	E	F	G
1	BJID	ARMCD	AESEQ	AEDECOD	AESTDY	AEDUR	AEENDY
2	101	SAUCE	1	Flatulence	3	4	7
3	101	SAUCE	2	Constipation	4	2	6
4	101	SAUCE	3	Fatigue	7	14	21

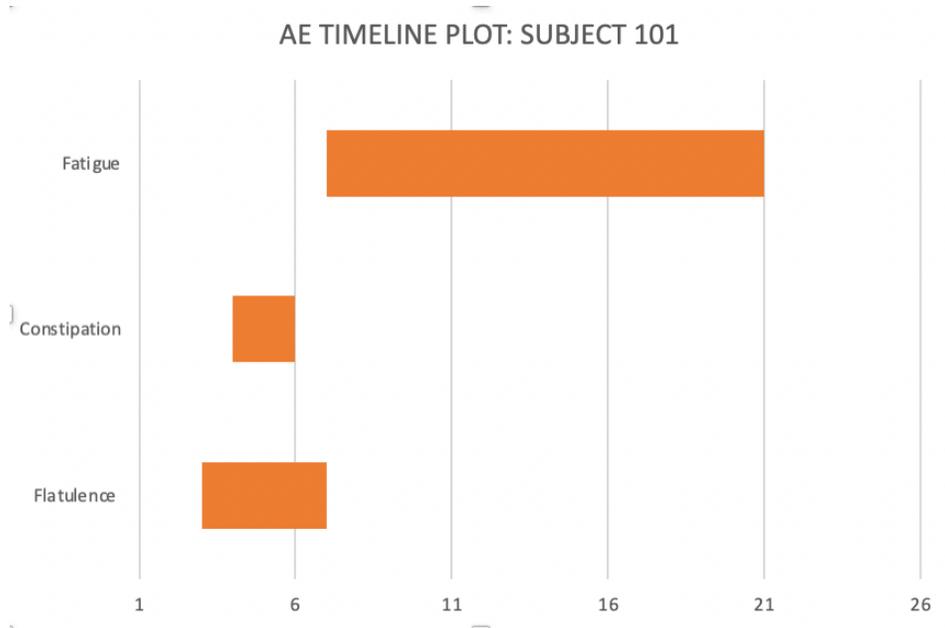
It is most important to calculate AEDUR (AE Duration) as AEENDY (AE End Day) – AESTDY (AE Start Day) which may mean a duration of 0 if the subject had the start and stop days the same.

To produce the plot for the three Adverse Events above:

1. Highlight the data for cells D2:F4

2. Select a stacked horizontal bar chart
3. Set the first series to 'No Fill' and 'No Border'
4. Change the horizontal line to a start day of 1 or greater

An example output using this method for the data selected above is shown below:



Full customization using Excel options is possible – find a customization that is useful for your display and save it. Using VBA, where initially recording the keystrokes and editing the program, is something that will make an Adverse Event Timeline easy and quick to do.

CONCLUSION

In this paper it has been demonstrated that it is possible to generate basic summary output, in the examples of a Demographics and Adverse Event summary tables. However, while possible, it does have some major challenges in that the data source is in the sheet itself and not stored in a compliant form where control is managed and secure, and there are difficulties in making the process easily expandable when updated data for a study becomes available.

CONTACT

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

David Franklin
dfranklin@TheProgrammersCabin.com

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