

Paper CC-017

PivotTable – A Powerful Tool for Data Exploration in the Pharmaceutical Industry

Wei (Lisa) Lin & June Cheng, Merck & Co, Inc., Upper Gwynedd, PA

ABSTRACT

PivotTable is a powerful Microsoft® Excel tool used to create interactive tables that automatically extract, organize, and summarize large amounts of data. Additionally, PivotTable reports can be used to analyze data, make comparisons, detect patterns and relationships, and discover trends. As statistical programmers working in a large pharmaceutical company, we have found PivotTable to be a very useful tool for communicating with our clients because it is fast, easy and interactive. We mainly use these PivotTable reports for exploratory/ad-hoc analysis before formal reporting tables are produced. This paper will show how to use patients' Adverse Event (AE) information to create pivot tables from a SAS® dataset.

INTRODUCTION

PivotTable is one of the most powerful and useful features of Microsoft® Excel. PivotTable allows for the pivoting of data via drag-and-drop techniques, and can be used to create a table that produces statistical information. The biggest advantage of PivotTable is that meaningful information can be generated and extracted from a large table of information within a matter of minutes, or even seconds. In this paper, AE data will be used as an example to demonstrate how to create a pivot table for clinical review.

STEPS TO CREATE A PIVOT TABLE:

Step 1: Generate AE dataset in SAS.

In the pharmaceutical industry, AE information is critical for evaluating the safety of a drug. Imagine we have a SAS AE dataset containing AE information such as Subject Identifier (subjid), Report Term of Adverse Events (aeterm), Body System or Organ Class of AE (aebodsys), Serious Events (aeser), Causality of AE (aerel), Severity of AE (aesev), AE start date (aestdt), AE end date (aeendt), etc. The AE dataset will be merged with the DM (Demographic) dataset to contain all the information needed. Below is an example of a resulting AE dataset:

SAS System Viewer - [ae.sas7bdat]														
File Edit View Window Help														
	SUBJID	SITENO	AGE	RACE	SEX	DOB	AETERM	AEDECOD	AEBODSYS	AESTDT	AEENDT	AESEV	AEREL	AESER
1	32922	0026	42	BLACK	F	05NOV1964	abdominal bloat	Abdominal bloat	General disorder	15JAN2008	.	MILD	NOT RELAT	N
2	32907	0010	45	WHITE	F	04SEP1962	Abdominal pain	Abdominal pain	General disorder	17MAR2008	.	MODERATE	NOT RELAT	N
3	32935	0013	44	BLACK	M	28AUG1963	Abdominal pain	Abdominal pain	General disorder	18MAY2008	.	MILD	NOT RELAT	N
4	32864	0048	42	WHITE	M	30APR1965	abcess	Abscess	Investigations	20NOV2007	.	MODERATE	NOT RELAT	N
5	32843	0047	41	WHITE	M	16FEB1968	abscess formati	Abscess	Investigations	10MAR2008	06APR2008	MODERATE	NOT RELAT	N
6	32930	0022	47	WHITE	M	05OCT1960	Overdose	Accidental over	procedural compl	04DEC2007	.	MILD	NOT RELAT	Y
7	32916	0026	43	WHITE	F	08MAR1964	swelling tendon	Achilles thicke	Musculoskeletal	01JAN2008	17JAN2008	MILD	NOT RELAT	N
8	32815	0047	41	WHITE	M	06FEB1966	acute gastroent	Acute gastroent	Investigations	04JAN2008	06JAN2008	MODERATE	NOT RELAT	N
9	32842	0041	43	WHITE	M	20MAY1964	acute rhinitis	Acute rhinitis	Investigations	20SEP2007	21OCT2007	MILD	NOT RELAT	N
10	32877	0015	35	BLACK	F	05APR1972	acute stress di	Acute stress di	Psychiatric diso	06APR2008	.	SEVERE	RELATED	N
11	32939	0034	43	WHITE	M	02MAR1964	ALT increase	Alanine increas	Investigations	07FEB2008	.	MODERATE	RELATED	N
12	32926	0049	38	WHITE	M	18MAY1969	ALT increased	Alanine increas	Investigations	10MAR2008	15MAY2008	MILD	NOT RELAT	N
13	32890	0004	46	BLACK	M	20MAY1961	Increased Alani	Alanine increas	Investigations	01APR2008	.	MODERATE	NOT RELAT	N
14	32875	0035	40	WHITE	M	29NOV1966	increased ALT	Alanine increas	Investigations	21NOV2007	.	MODERATE	RELATED	N
15	32890	0004	46	BLACK	M	20MAY1961	Increased Alkal	Alkaline increa	Investigations	01APR2008	.	MODERATE	NOT RELAT	N
16	32914	0034	45	WHITE	M	11OCT1962	Ammonium increa	Ammonia increas	Investigations	10JAN2008	.	MILD	NOT RELAT	N
17	32828	0048	58	WHITE	M	04MAY1949	anemia	Anaemia	Blood disorders	06FEB2008	.	SEVERE	NOT RELAT	N
18	32045	0012	35	BLACK	M	15OCT1971	anal fissure	Anal fissure	General disorder	23JAN2008	29FEB2008	MILD	NOT RELAT	N
19	32827	0044	61	WHITE	M	19MAY1946	angina pectoris	Angina pectoris	Cardiac disorder	08JAN2008	.	SEVERE	NOT RELAT	N
20	32803	0046	46	WHITE	M	14SEP1960	fracture right	Ankle fracture	procedural compl	14JAN2008	21JAN2008	MODERATE	NOT RELAT	N
21	32061	0007	43	WHITE	M	12APR1964	Anxiety	Anxiety	Psychiatric diso	25JAN2008	.	MILD	NOT RELAT	N
22	32813	0029	46	WHITE	M	21MAR1961	Anxiety	Anxiety	Psychiatric diso	18SEP2007	.	MODERATE	NOT RELAT	N
23	32908	0040	46	BLACK	M	15MAR1961	anxiety	Anxiety	Psychiatric diso	01DEC2007	.	MODERATE	RELATED	N
24	32053	0036	52	WHITE	M	26JUL1955	anxious-depress	Anxiety depress	Psychiatric diso	08MAR2008	.	MODERATE	NOT RELAT	N
25	32914	0034	45	WHITE	M	11OCT1962	AST increase	Aspartate incre	Investigations	13DEC2007	.	MODERATE	RELATED	N
26	32915	0034	53	WHITE	M	09MAY1954	AST increase	Aspartate incre	Investigations	13DEC2007	.	MODERATE	RELATED	N
27	32890	0004	46	BLACK	M	20MAY1961	Increased Aspar	Aspartate incre	Investigations	01APR2008	.	MODERATE	NOT RELAT	N
28	32063	0038	40	WHITE	F	16SEP1967	asthenie	Asthenia	General disorder	12DEC2007	.	MODERATE	NOT RELAT	N
29	32905	0038	30	WHITE	F	11JAN1977	asthenie	Asthenia	General disorder	12DEC2007	.	MILD	NOT RELAT	N
30	32864	0048	42	WHITE	M	30APR1965	MRSA	Aureus infectio	Investigations	20NOV2007	.	MODERATE	NOT RELAT	N

Step 2: Export SAS dataset to Excel.

Now this SAS dataset needs to be exported to an Excel file. There are a couple of ways to do that. One is to use the export wizard in SAS: click on **"Export Data"** under **"File"**, follow the steps to enter the libname and name of the dataset, and enter the location and name of the Excel file to export.

Another method to export SAS dataset to an Excel file is by using following SAS statements:

```
proc export data = A.AE
  outfile = "C:\Documents and Settings\Linlis\My Documents\Test\lae.xls"
  dbms = EXCEL;
run;
```

After the Excel file AE.xls has been created, Excel can be opened and AE.xls can be used to create a pivot table. This step can be automated in SAS with the following statements:

```
options noxwait noxsync;
```

```
x "start excel";
```

```
data _null_ ;  
  x = sleep(1) ;  
run ;
```

```
filename sasexcel DDE 'excel|system';
```

```
data _null_ ;  
  file sasexcel;  
put '[open("C:\Documents and Settings\Linlis\My Documents\Test\lae.xls")]';  
run;
```

Following are the explanations of these statements:

- ❖ Options noxwait noxsync - sets the SAS system options which are instructions that affect the SAS session. They control the way that SAS performs operations such as SAS system initialization, hardware and software interfacing, and the input, processing, and output of jobs and SAS files. The noxwait automatically closes the prompt window after the command has completed. The noxsync instructs the system to process as soon as the command is issued.
- ❖ The x command - can be used to execute operating system commands without leaving SAS. The form is x 'operating system command'. Here the statement is used to open the Excel application.
- ❖ Data _null_ - this is used to execute a DATA step without creating a SAS data set. Using _NULL_ causes SAS to execute the DATA step as if it was creating a new data set, but no observations or variables are written to an output data set. This process can be a more efficient use of computer resources if the DATA step is being used for a function such as report writing, for which the output of the DATA step does not need to be stored as a SAS data set.
- ❖ Sleep - is used to allow the Excel window to initialize prior to attempting to open the file. The length of the wait period can be adjusted up or down as needed.
- ❖ Filename - associates a "file reference name" which will later be used in the SAS program with an external data file or an output device.
- ❖ DDE - Dynamic Data Exchange. The primary function of DDE is to allow Windows applications to share data.
- ❖ FILE statement - specifies the fileref sasexcel defined in the previous DDE filename statement.
- ❖ PUT statements - writes executable Excel macro commands to open the Excel file.

At this point, AE.xls is opened in Excel:

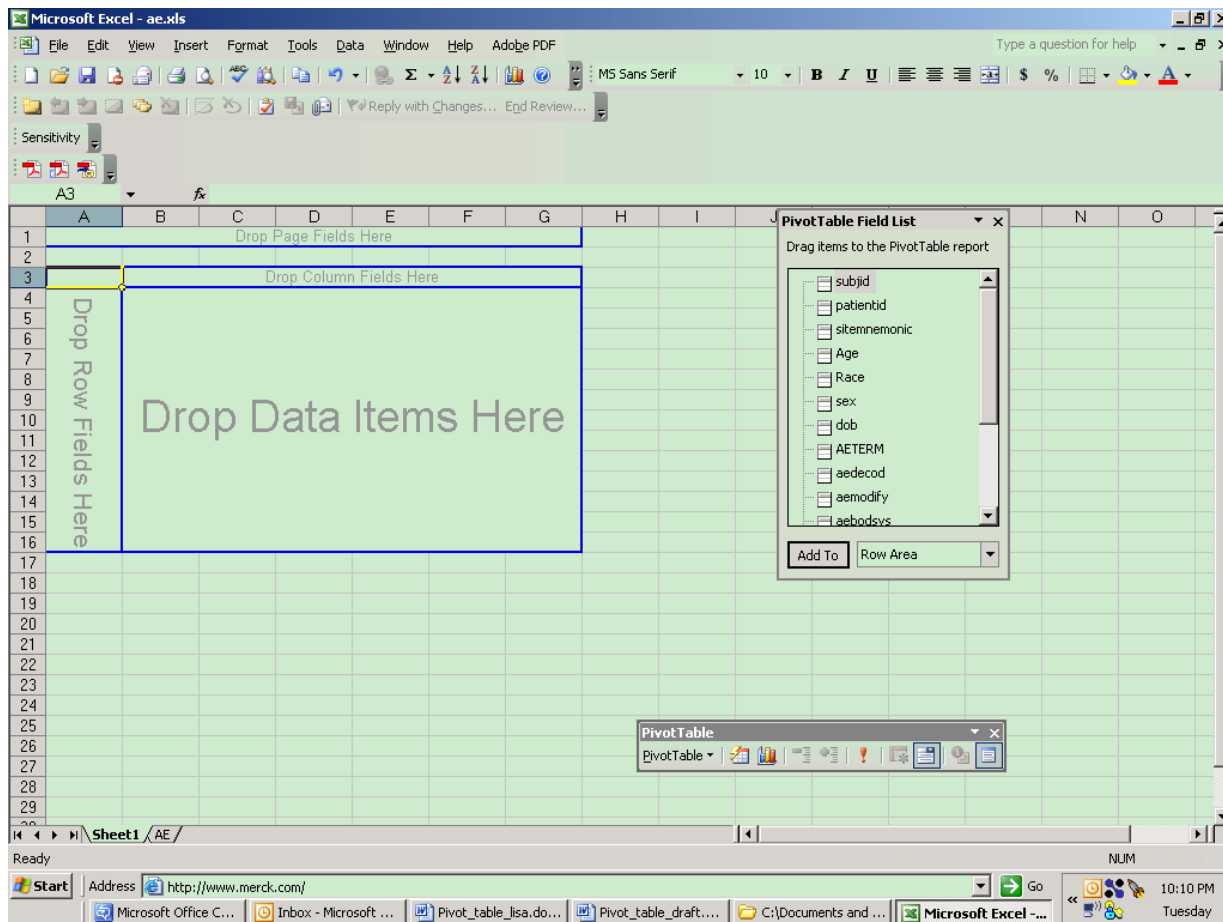
	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	SUBJID	SITENO	AGE	RACE	SEX	DOB	AETERM	AEDECOD	AEBODSYS	AESTDT	AEENDT	AESEV	AEREL	AESER
2	32828	0048	58	WHITE	M	5/4/1949	anemia	Anaemia	Blood disorders	2/6/2008		SEVERE	NOT RELATED	N
3	32871	0021	38	WHITE	F	1/1/1969	Submaxillary enl	Lymph nodes enlarged	Blood disorders	1/10/2008	4/3/2008	MILD	NOT RELATED	N
4	32827	0044	61	WHITE	M	5/19/1946	angina pectoris	Angina pectoris	Cardiac disorders	1/8/2008		SEVERE	NOT RELATED	N
5	32813	0029	46	WHITE	M	3/21/1961	Heart Flutter	Cardiac flutter	Cardiac disorders	2/7/2008	4/17/2008	MILD	NOT RELATED	N
6	32827	0044	61	WHITE	M	5/19/1946	LAD heart disea	Coronary artery disease	Cardiac disorders	1/9/2008		MILD	NOT RELATED	N
7	32827	0044	61	WHITE	M	5/19/1946	Diastolic dysfun	Diastolic dysfunction	Cardiac disorders	1/8/2008		MILD	NOT RELATED	N
8	32813	0029	46	WHITE	M	3/21/1961	Palpitations	Palpitations	Cardiac disorders	2/7/2008	4/17/2008	MILD	NOT RELATED	N
9	32922	0026	42	BLACK	F	11/5/1964	palpitations	Palpitations	Cardiac disorders	12/21/2007		MILD	NOT RELATED	N
10	32014	0010	51	WHITE	M	7/18/1955	Right Earache	Earache	Endocrine disorders	8/9/2007	8/10/2007	MILD	NOT RELATED	N
11	32890	0004	46	BLACK	M	5/20/1961	hyperthyroidism,	Graves' disease	Endocrine disorders	1/21/2008	4/11/2008	MODERATE	NOT RELATED	N
12	32890	0004	46	BLACK	M	5/20/1961	hyperthyroidism	Hyperthyroidism	Endocrine disorders	1/21/2008		MILD	NOT RELATED	N
13	32892	0007	39	WHITE	M	1/12/1968	Hypothyroidism	Hypothyroidism	Endocrine disorders	12/6/2007		MODERATE	NOT RELATED	N
14	32930	0022	47	WHITE	M	10/5/1960	Ear Drum	Tympanic perforation	Endocrine disorders	12/18/2007	12/18/2007	MILD	NOT RELATED	N
15	32851	0046	56	ASIAN	M	7/4/1951	konjunktivitis	Conjunctivitis	Eye disorders	12/25/2007	1/2/2008	MODERATE	NOT RELATED	N
16	32801	0046	66	WHITE	M	7/18/1940	cornea-right eye	Perforated corneal ulcer	Eye disorders	3/28/2008	4/10/2008	SEVERE	NOT RELATED	N
17	32922	0026	42	BLACK	F	11/5/1964	abdominal bloati	Abdominal bloating	General disorders	1/15/2008		MILD	NOT RELATED	N
18	32907	0010	45	WHITE	F	9/4/1962	Abdominal pain	Abdominal pain	General disorders	3/17/2008		MODERATE	NOT RELATED	N
19	32935	0013	44	BLACK	M	8/28/1963	Abdominal pain	Abdominal pain	General disorders	5/18/2008		MILD	NOT RELATED	N
20	32045	0012	35	BLACK	M	10/15/1971	anal fissure	Anal fissure	General disorders	1/23/2008	2/29/2008	MILD	NOT RELATED	N

Step 3: Create Pivot Table.

Before pivoting the table, the following information should be ready:

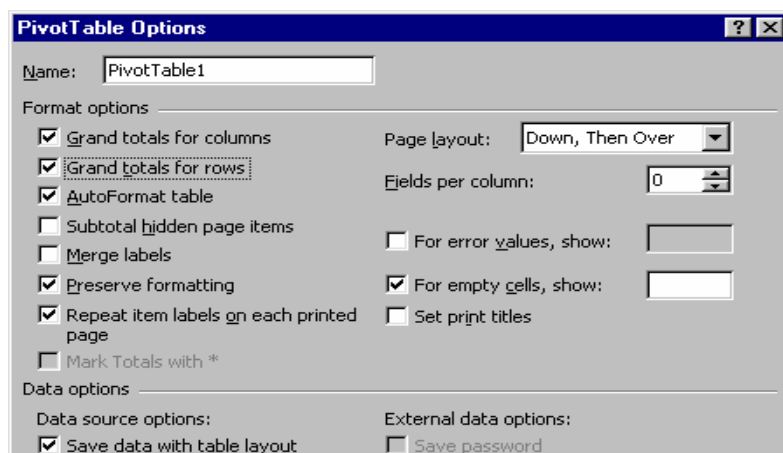
- ❖ Where is the data stored;
- ❖ Which column of data is going into which field, e.g. the optional page field, and mandatory row, column and data fields;
- ❖ Where should the pivot table go, e.g. a new worksheet or an existing one.

Now the PivotTable wizard is used to create the pivot table. From Excel's menu bar, click on **"Data"**, then click on **"PivotTable and PivotChart Report"**, and the PivotTable wizard starts up. The wizard will guide us through the steps to create a pivot table. This should generally follow the default selections until the following screen is seen after clicking on **"Finish"**.



The PivotTable Field List on the right is used to drag and drop the specific field to page, row, column and data fields. Page fields allow filtering of the entire PivotTable report to display data for a single item or all the items. Row fields and column fields are self explanatory. Data fields provide the data values to be summarized. Usually data fields contain numbers, which are combined with the Sum summary function, but data fields can also contain text, in which case the PivotTable report uses the Count summary function.

Clicking on "**Options**" instead of "**Finish**" allows more flexibility:



The base pivot table is now complete.

	A	B	C	D	E
1	SITENO	(All)			
2	SUBJID	(All)			
3	RACE	(All)			
4	SEX	(All)			
5					
6	Count of AEDECOD				
7	AEBODSYS	AEDECOD	AETERM	Total	
8	Blood disorders	Anaemia	anemia	1	
9		Anaemia Total		1	
10		Lymph nodes enlarged	Submaxillary enlarged node	1	
11		Lymph nodes enlarged Total		1	
12	Blood disorders Total			2	
13	Cardiac disorders	Angina pectoris	angina pectoris	1	
14		Angina pectoris Total		1	
15		Cardiac flutter	Heart Flutter	1	
16		Cardiac flutter Total		1	
17		Coronary artery disease	LAD heart disease	1	
18		Coronary artery disease Total		1	
19		Diastolic dysfunction	Diastolic dysfunction	1	
20		Diastolic dysfunction Total		1	
21		Palpitations	Palpitations	2	
22		Palpitations Total		2	
23	Cardiac disorders Total			6	
24	Endocrine disorders	Earache	Right Earache	1	
25		Earache Total		1	
26		Graves' disease	hyperthyroidism, Graves Disease	1	
27		Graves' disease Total		1	
28		Hyperthyroidism	hyperthyroidism	1	
29		Hyperthyroidism Total		1	

As demonstrated, the pivot table is easily constructed and the results include:

- Specific AE by Body System ;
- AE dictionary term, AE reported term;
- The count of particular AETERM (or whatever else is chosen to put at data area);
- Other information selected by the user.

The pivot table has dynamic elements. For example, click on the drop down arrows in the page area to change selects. In the example above, "All" can be changed to the specific subjid, siteno, race, sex, etc to view the detailed AE information for the selection. It is also possible to drill down to details. Subjid, race, sex, etc. can be selected to view for the particular AETERM by double clicking on the AETERM, and a new worksheet will pop up showing all the information including subjid, site number, race, sex, date of birth, AE start date, AE intensity, AE causality, AE seriousness, etc by double clicking on the count (shown below).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	SUBJID	SITENO	AGE	RACE	SEX	DOB	AETERM	AEDECOD	AEBODSYS	AESTDT	AEENDT	AESV	AEREL	AESER
2	32828	0048	58	WHITE	M	5/4/1949	anemia	Anaemia	Blood disorders	2/6/2008		SEVERE	NOT RELATED	N
3														

Much can be done with the pivot table by dragging different variables to different areas, selecting different counts, etc. The best part about PivotTable is its ability to be manipulated via 'Trial-and-Error' and immediately show results. If the result is not what is expected, the Undo feature can be used for another trial. It must be remembered that the structure of the original table is not being changed in any way, so no harm can be done. Below is another example by adding Sex to the Column Area to see the count by sex as well as the grand total:

The screenshot shows a Microsoft Excel window with a PivotTable. The PivotTable is structured with 'AEBODSYS' and 'AEDECOD' in the Row Area, 'AETERM' in the Column Area, and 'SEX' in the Filter Area. The data is summarized by 'Grand Total'.

	A	B	C	D	E	F
1						
2	SITENO	(All)				
3	SUBJID	(All)				
4	RACE	(All)				
5						
6	Count of AEDECOD			SEX		
7	AEBODSYS	AEDECOD	AETERM	F	M	Grand Total
8	Blood disorders	Anaemia	anemia		1	1
9		Anaemia Total			1	1
10		Lymph nodes enlarged	Submaxillary enlarged node	1		1
11		Lymph nodes enlarged Total		1		1
12	Blood disorders Total			1	1	2
13	Cardiac disorders	Angina pectoris	angina pectoris		1	1
14		Angina pectoris Total			1	1
15		Cardiac flutter	Heart Flutter		1	1
16		Cardiac flutter Total			1	1
17		Coronary artery disease	LAD heart disease		1	1
18		Coronary artery disease Total			1	1

There are also some papers which discuss how to use SAS to automatically create pivot tables in Excel. The papers listed in the reference area should be consulted for further details. They basically include 3 steps:

- ❖ Prepare a single file with SAS code and the PIVOT macro calls that define the desired Excel pivot tables;
- ❖ When SAS is run, the PIVOT Macro creates a delimited text file containing exported SAS data as well as a VBScript External File containing generated Excel VBA code;
- ❖ The VBScript file creates and launches an instance of Excel that generates Excel pivot tables per your specifications.

These ideas are brilliant. The only caution is that some knowledge of VBScript may be needed to make it work with a specific Excel version. Since Microsoft® changed some of VB functions at the backend for different versions of Excel, some of the VBScript might work with one Excel version but not with another.

CONCLUSION

This paper presents PivotTable as an easy and powerful tool to perform data exploration. It also demonstrates the following main benefits: end users do not have to wait for programmers to produce formal reporting tables to perform their data exploration; end users can decide and create on the fly the additional views/summaries that occur to them while they are reviewing the data; tremendous resources are saved when programmers do not have to produce formal reporting tables for exploratory purposes. So just go ahead and give PivotTable a try!

REFERENCES

Ted Conway "Sur La Table: Creating Microsoft Excel PivotTables in a Jiffy from SAS Data" SUGI 31, March 2006

John DeMeo "Creating Microsoft Excel Pivot Tables from SAS - (Part 2)" SAS Global Forum, April 2007

CONTACT INFORMATION

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

Wei (Lisa) Lin
Merck & Co., Inc.
UG1CD-14
Po Box 1000
North Wales, PA 19454-1099
Work Phone: 267-305-6521
Email: lisa_lin@merck.com

June Cheng
Merck & Co., Inc.
UG1CD-14
Po Box 1000
North Wales, PA 19454-1099
Work Phone: 267-305-5389
Email: june_cheng@merck.com

TRADEMARKS

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are trademarks of their respective companies.