Paper CT11

Fancy Data Visualisations Without Additional Toolkits

Kirsty Parker-Hodds, Veramed, Twickenham, UK

ABSTRACT

Clinical trials often involve collecting data of varying quality from sites around the world. Clinicians and other study team members may not have immediate access to the study databases which can make visualising data difficult during the blinded phase of a study when looking for data issues. This paper presents a tool created using SAS that exports data into Excel in order to display individual patient profile plots of blood pressure data over time. This Excel tool can be used by any member of the study team and has drop-down menus to display graphs of different measurements for subjects. This tool has been extended to produce static PDF outputs of each patient profile for ease of viewing two study endpoints side by side. Simple adaptations of the code allows these tools to easily be adapted and replicated over multiple studies for different purposes of displaying data.

INTRODUCTION

This paper discusses the creation of two tools that enable study teams to visualise data during the blinded phase of a clinical trial. The paper highlights the importance of creating tools that are used by everyone in the study team without the need for fancy software. Instead, packages that are widely accessible and familiar to all team members can be used to display data to a high standard. The software used for viewing the tools discussed in this paper are Microsoft Excel and Adobe PDF Reader, neither of which are currently used to their full potential to display data in the pharmaceutical industry.

THE PROBLEM

Clinical trials often have varying quality of data. Data issues and queries can arise which require discussion amongst the study team. Data visualisation is very important when discussing data issues with other study team members and ensures that everyone fully understands the problem. Although programmers and statisticians have instant access to study data, this is usually not the case for the rest of the clinical team. During a recent study that collected longitudinal blood pressure data, it became clear that it is difficult to identify data issues when looking at single data points. Given the preceding and following values, a value might be valid for one subject but an outlier for another. One method for finding any potential data issues or outliers for data collected over time is to plot the data for each subject ensuring that all data points are roughly following an expected trend.

THE TOOL

A tool is required to plot data by subject making it easy for the whole study team to visualise the data. This ensures that data issues are found more efficiently and can then be discussed amongst all team members. For the tool to be successful it requires a package that is widely accessible and familiar to everyone in the team, so that it can be used to its full potential. The tool also requires good graphics, so that when data is plotted it is clear to see and well displayed.

THE DATA

This paper focuses on using tools to analyse subjects with abnormal data points in longitudinal data measuring blood pressure. Systolic and diastolic blood pressure are the study endpoints being considered in this paper. The tools can be applied to other studies analysing different end points.

FIRST TOOL - EXCEL

The first tool is created in Microsoft Excel. This software is part of Microsoft Office, it is widely available and is familiar to most people. The tool creates multiple tabs in Excel so that data can be displayed in numerous ways. The tabs include demographic variables and the type of data (either systolic or diastolic blood pressure) which can be filtered on and in turn fed into a cover page. The cover page displays data plotted by subject and type of blood pressure, chosen by using drop down menus. The cover page displays the chosen subject's demographics and a listing of the numeric data used to create the plot.

When presenting this tool to the study team many advantages can be found: it displays data in a clear and simple manner, the use of drop down menus allows a reviewer to choose which subject and type of data to be viewed in detail, it is simple to code, full listings of data can be viewed on the different tabs and it is easy to adapt for use in another study. However, although data is displayed in an informative manner it is very time consuming to click between each subject and the specific type of blood pressure that needs to be reviewed. Therefore, finding subjects which need investigating due to abnormalities in their data is time consuming.

EXAMPLE

Below is an example of the cover page which the Excel tool creates. The diastolic blood pressure at day 1144 (visit 14) clearly does not fit the trend for this subject. However, the reading of 98mmHg is perfectly valid for another subject. Therefore, this data point may not have necessarily stood out as a potential data issue when reviewing the total dataset. The screenshot below displays an example of the cover page of the Excel tool, with the subject's demographics displayed in the top left-hand corner and the type of blood pressure being investigated next to it. The example shows how the tool is used in a study examining blood pressure, where it is important to look at the data for both systolic and diastolic blood pressure. Therefore, a drop-down menu is created for both subject and type. The data used to create the plot is also displayed as a listing on the right-hand side of the page, so that it is easy to view the numerical data. The minimum, maximum, range and treatment start and end date are displayed at the bottom of the page.

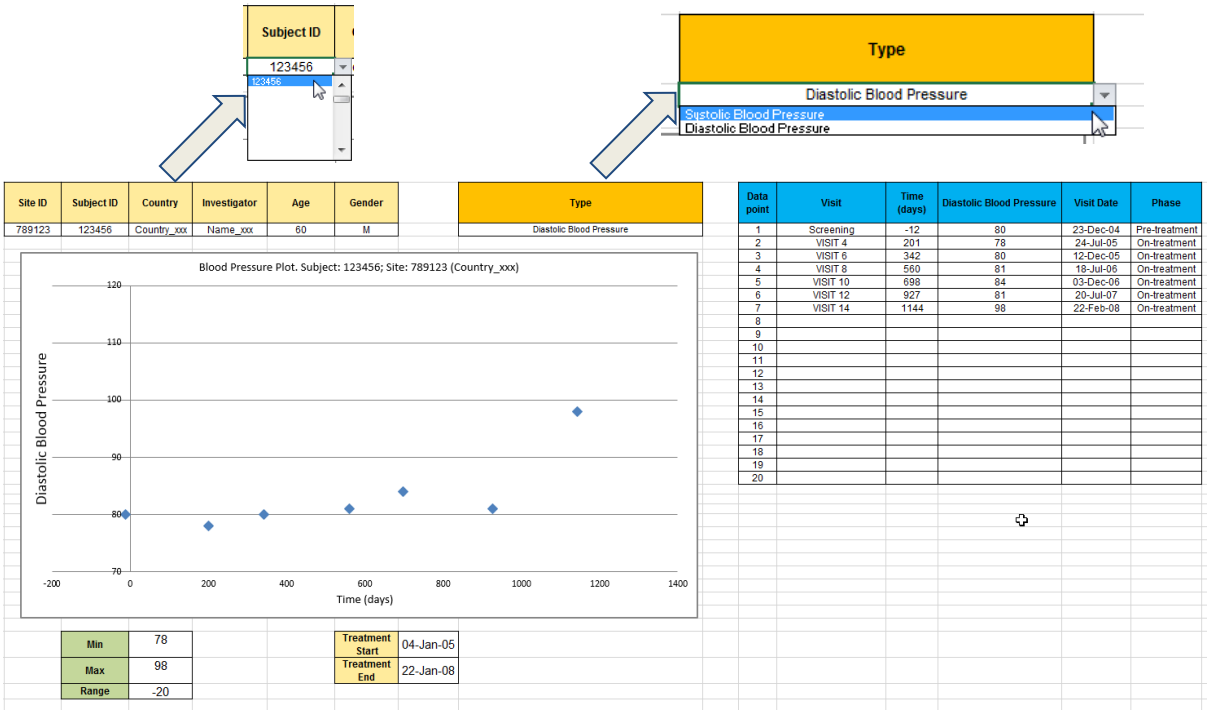

Figure 1: The Excel tool for a chosen subject and parameter

CODE

SAS is used to export the data into Excel, which is then used to program the tool. The main function used in the Excel file is VLOOKUP, this is used to find the column of data required from the tab containing the listing of all the data for subjects and different types of blood pressure. This selected column of data is then displayed on the cover page and is used to create the plot. When using VLOOKUP, error messages can occur if cells being selected are of different data types, so to avoid this an IFERROR formula can be used. This formula is IFERROR(IF(VLOOKUP(),NA(),VLOOKUP()),NA()). To create the drop-down menus for subject and type, a point and click method is used by going to data validation, changing 'Allow:' to list and then clicking 'OK'.

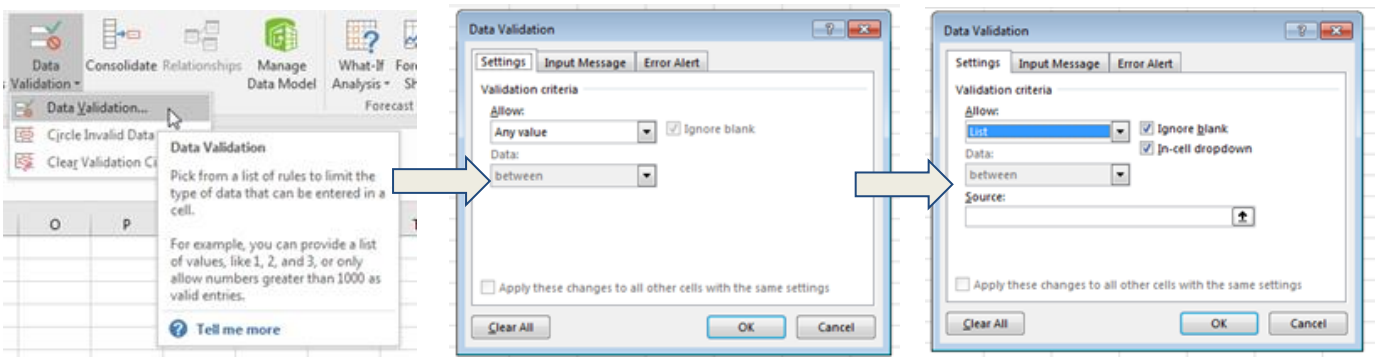

Figure 2: Creating a drop-down menu in Excel

SECOND TOOL - ADOBE

As previously discussed in this paper the Excel tool is informative and easy to use but is too time consuming to change between different subjects and types of blood pressure. Therefore, the tool is extended to PDF which can be viewed in Adobe Reader. Again, this is software most people are familiar with and feel comfortable using. By extending the tool to PDF it allowed static versions of the plots to be created. This made it quicker and more efficient for study team members to view the data for all subjects. The PDF tool enables two study endpoints to be viewed side by side for each subject, resulting in any potential data abnormalities or queries to be identified in a shorter period of time. The PDF tool included additional figure annotations such as displaying a data point in a different colour if it did not meet a certain criterion, identifying any issues with the blood pressure gauge for a subject at a visit and the addition of a vertical reference line at the screening visit and a horizontal reference line at the baseline value for each subject. It is highlighted in the plots when a new blood pressure gauge is used, this is important because different machines have slightly different calibrations. Additionally, tables are included underneath the figures to show the visit for each data point, the value at that visit, the change from baseline and the change from the previous visit.

Some tips to the PDF tool are:

Tip 1: If you want to quickly browse through the PDF pages to find subjects with remarkably different data points, then you should ensure the axes remain fixed over all pages.

Tip 2: The plots' dimensions can be adapted to perfectly fit a computer screen in full screen mode in Adobe Reader by using "Ctrl+L". This adaptation enables an ease of viewing for study team members working on small laptop screens.

Tip 3: In the PDF abnormal values are given a different colour for extra emphasis.

Tip 4: It is very simple to adapt the code to work for other studies, simply change the dataset and variables being used.

The main disadvantage of the PDF tool is the initial coding takes longer than the Excel tool.

EXAMPLE

It can be seen from the example below of the PDF tool that the type of data being measured and the subject's demographics are included in the title. The reference line for the screening and baseline visits can be seen on the figures. These reference lines enable the reader to see how much the data varies from the baseline value and whether the trend of the data for that subject stays consistent over the visits.

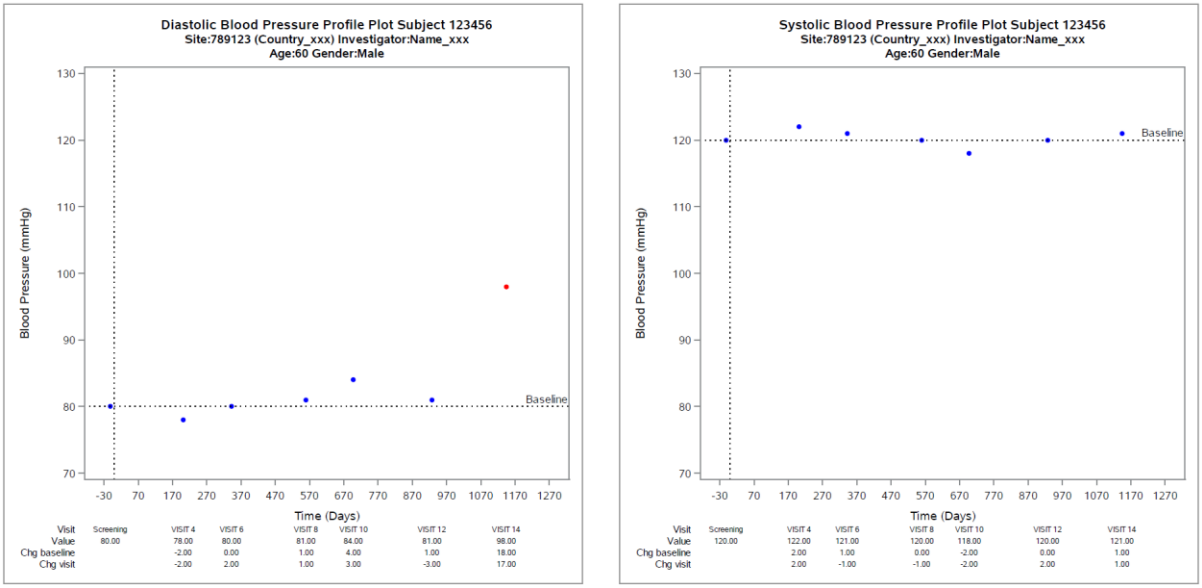

Figure 3: One PDF page, for a chosen subject displaying two study endpoints side by side

CODE

The code to produce the PDF tool is created in SAS. Some important parts of the code include:
"noautolegend" - this suppresses the automatic creation of the legend,
"dattrmap" - allows specification of the colours used depending on values in the data,
"refline" - adds a horizontal reference line for the baseline value and a vertical reference line at the screening visit.
The table at the bottom of the figures is created using "xaxistable", this requires a dataset with the variables being displayed. This method of annotating figures has many formatting options which are easy to use, some of these include "position", "location" and "labelattrs". An example of the dataset used to create the table is shown in Figure 4.

```
proc sgplot data=plots noautolegend dattrmap=myattrmap;
  where subject=&subject and type='Diastolic Blood Pressure';
  by subject type;
  title " Diastolic Blood Pressure Profile Plot Subject &subject";
  title2 "Site:&pp_siteid (&pp_country) Investigator:&pp_invnam";
  title3 "Age:&pp_age Gender:&pp_sex";
  scatter x=day y=dxa_result / group=flag markerattrs=(symbol=CircleFilled)
                               attrid=myid;
  option nobyline;
  xaxis label="Time (Days)" values=(-30 to 1270 by 50);
  yaxis label="Blood Pressure (mmHg)" values=(0.2 to 1.4 by 0.2);
  refline 0 / axis=x lineattrs=(color=black pattern=dot);
  refline &pp_baseline / axis=y lineattrs=(color=black pattern=dot)
                         label="Baseline";
  xaxistable visit_no reading_c chg_c chg_previous_visit_no_c/ position=bottom
                                                               location=outside
                                                               labelattrs=
                                                               (size=8pt);
run;
```

	REFID	Visit_No	reading_c	chg_c	chg_previous_visit_no_c
1	123456_1_Diastolic Blood Pressure	Screening	80.00		
2	123456_2_Diastolic Blood Pressure	VISIT 4	78.00	-2.00	-2.00
3	123456_3_Diastolic Blood Pressure	VISIT 6	80.00	0.00	2.00
4	123456_4_Diastolic Blood Pressure	VISIT 8	81.00	1.00	1.00
5	123456_5_Diastolic Blood Pressure	VISIT 10	84.00	4.00	3.00
6	123456_6_Diastolic Blood Pressure	VISIT 12	81.00	1.00	-3.00
7	123456_7_Diastolic Blood Pressure	VISIT 14	98.00	18.00	17.00
8	123456_1_Systolic Blood Pressure	Screening	120.00		
9	123456_2_Systolic Blood Pressure	VISIT 4	122.00	2.00	2.00
10	123456_3_Systolic Blood Pressure	VISIT 6	121.00	1.00	-1.00
11	123456_4_Systolic Blood Pressure	VISIT 8	120.00	0.00	-1.00
12	123456_5_Systolic Blood Pressure	VISIT 10	118.00	-2.00	-2.00
13	123456_6_Systolic Blood Pressure	VISIT 12	120.00	0.00	2.00
14	123456_7_Systolic Blood Pressure	VISIT 14	121.00	1.00	1.00

Figure 4: An example of the dataset needed for "xaxistable"

CONCLUSION

In conclusion, both the Excel and PDF tools are informative and easy for the rest of the study team to use. The tools enable visualisation when identifying and discussing potential data abnormalities and queries. It is initially more time consuming to code the PDF tool in comparison to the Excel tool, but now both tools exist they would be easy to replicate for different types of studies by changing the dataset going into the code and the variables used. Each tool explores a different aspect of the data and so they are best used in parallel. The PDF tool can be used to quickly compare two study endpoints across multiple subjects. This can identify any data points that need further investigation and discussion with the rest of the study team. The Excel tool can then be used to look at the desired subject in detail as the tool contains more in-depth information. This enables a thorough discussion about any data queries within the study team. This paper has shown that fancy software is not needed to create tools that visually represent data well and the rest of the study team find easy to use.

ACKNOWLEDGMENTS

I would like to thank Nick Cowans for his help and support.

CONTACT INFORMATION

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

Kirsty Parker-Hodds
Veramed Ltd
5th Floor Regal House
70 London Road
Twickenham
TW1 3QS
020 3696 7240
kirsty.parkerhodds@veramed.co.uk
www.veramed.co.uk