Visualizing Clinical Trial Data: Small Data, Big Insights

Michael Drutar and Elliot Inman, SAS Institute Inc., Cary, NC

ABSTRACT

Data visualization is synonymous with big data, for which billions of records and millions of variables are analyzed simultaneously. But that does not mean that data scientists analyzing clinical trial data that include only thousands of records and hundreds of variables cannot take advantage of data visualization methodologies. This paper presents a comprehensive process for loading standard clinical trial data into SAS® Visual Analytics, an interactive analytic solution. The process implements template reporting for a wide variety of point-and-click visualizations.

THE DATA SCIENTIST AS SOFTWARE DEVELOPER

For almost a hundred years, the standard format for publication of data has been printed reports, typically journal articles or regulatory submissions with tables, graphs, and narrative. That format has been supplanted only by PDF, printed paper's first cousin. Genetically speaking, SAS Visual Analytics is not even in the same species.

For analysts whose only reporting experience is generating printed or hard copy reports of tables and graphs or perhaps a PowerPoint slide deck, creating reports in SAS Visual Analytics will be a completely new experience. In some ways, using SAS Visual Analytics might seem more like software development than analytic modeling. In fact, it is. But SAS Visual Analytics does not require any web app development or JAVA experience to use. All of the reporting presented here can be generated using simple drag-and-drop and point-and-click methods—making it exceptionally easy for a statistician or clinical trials programmer to create engaging, interactive web-based reports.

Some analysts might say: But I am a statistician, not a web designer. That might be technically true based on whatever formal job code/description Human Resources uses for your company. But analysts should be careful not to abdicate the power to create these reports. SAS Visual Analytics offers an opportunity for analysts to create interactive web-based reports that reflect the influence of variables of importance, known interactions, reliable relationships, and other findings from statistical models. Without analysts actively involved in the report development process, report writers might rely on variables of *convenience* instead of variables of *importance*. Reports will reflect a rote representation of standard hierarchies and relationships. The first pull-down will always be the experimental condition, followed by a demographic variable, and so on.

There is nothing fundamentally wrong with such a report. The only problem is that, in many cases, the actual treatment effect might be mediated by the interaction of experimental condition, multiple demographic variables, and other covariates. The effect might be limited to one end of the age distribution, but only for subjects who did not experience a particular severe side effect in the first month of the study. The real world is a complex place. There is no easy way to demonstrate that complexity using traditional printed reports. It is easy to do so in SAS Visual Analytics. The template reports described here demonstrate what you can do.

A BRIEF INTRODUCTION TO SAS VISUAL ANALYTICS

Comprehensive training on SAS Visual Analytics is beyond the scope of this paper, but a quick introduction to the software and terminology will be helpful. SAS Visual Analytics is server-based software that users access much in the same way they access SAS® Drug Development. A user has an ID and password. User privileges can be controlled by the System Administrator, who controls whether each user can load data, access raw data, and create or view reports.

For the reports presented here, the source data is standard SAS® data sets. Since SAS Visual Analytics is an in-memory software solution, data must be loaded into the SAS® LASR™ Server, a RAM-based analytics engine. That operation can be performed in a number of ways, as documented in the User Manual. Once loaded into the SAS LASR Server, data is available for analysis and reporting.

There are two main analysis and reporting interfaces within SAS Visual Analytics: SAS® Visual Analytics Explorer and SAS® Visual Analytics Designer. In some ways, those terms are misnomers. You can "design" a report as an exploration and you can "explore" data in a designer report. The templates presented in this paper were built using the designer. One advantage of the designer is that it enables a research group to provide access to meaningful reporting without risking the security and stability of the underlying data. Analysts concerned that less experienced users might misinterpret data or create misleading reports can retain greater control over the reporting by simply assigning other users to a report in Read-Only mode.

Once data is loaded into SAS Visual Analytics, report creators have access to a wide variety of objects to display, graph, and control the interaction of visual objects in a report. (See Figure 1.) As the figure shows, all of these visual objects can be dragged to the canvas for a report. This draft report is the tab labeled "Section 1." Variables listed under the "Data" tab can then be dragged onto the object, and those objects can then be deployed as a report.
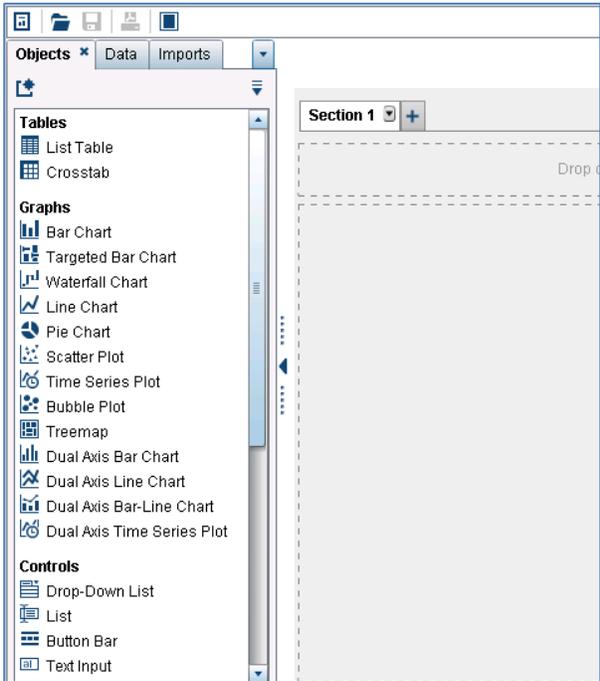

**Figure 1. Drag-and-Drop Graphs and Controls in SAS Visual Analytics**

But the true value of SAS Visual Analytics is the interactivity between these controls and graphs. In traditional scientific reporting, a graph is a fixed-format, static representation of a relationship. The X variable is on the horizontal axis and the Y variable is on the vertical axis (See Figure 2.) This graph was built using standard print-ready proportions. On paper, this graph would probably appear with a table of the actual values. This basic graph, surfaced as html on a web page, could have additional functionality, enabling users to scroll over bars to see the actual counts.
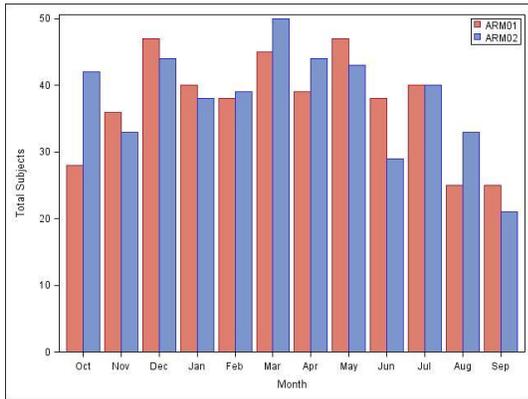
Figure 2. Traditional Fixed-Format Graph

In SAS Visual Analytics, that same graph is a dynamic object. (See Figure 3.) The format of this paper does not do justice to the dynamic nature of the graph, but the following graph is not just an image, but a control. Users can interact with the graph in many ways, including:

- clicking on a bar to select subjects in a particular experimental condition at that time and thus filter other graphs and tables in that report

- clicking on the legend to highlight that experimental condition only

- selecting multiple adjacent or non-adjacent bars to capture data from a particular set of subjects, such as the first or last enrolled patients, to explore other differences
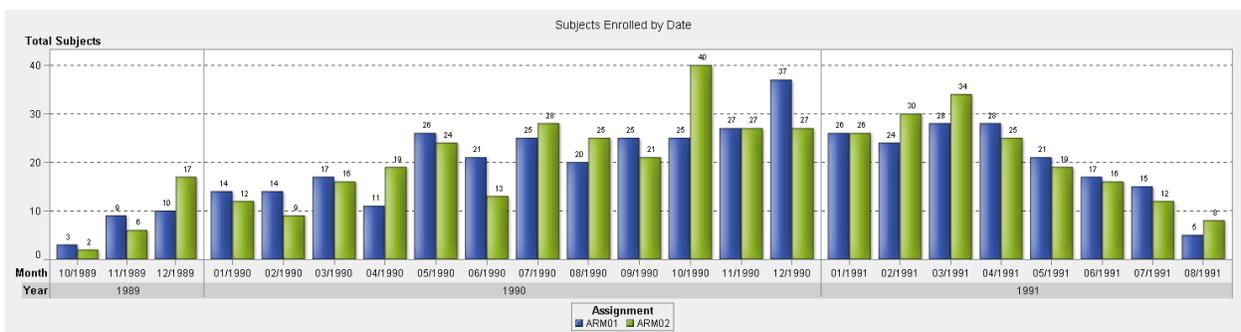
Figure 3. Dynamic Bar Chart in Visual Analytics

The same interactivity exists for line graphs, pie charts, treemaps, and a wide variety of other objects. There are many different ways to represent the data and create opportunities for interaction, which is why the role of the data scientist remains so critical to the development of these types of reports. With so many reporting options available, it is always important to ensure that interactivity reflects the nature of the underlying data.

OVERVIEW OF THE TEMPLATE REPORTS

The following sections of this paper describe the implementation of six template reports. The table below provides a quick overview of each report. This overview is followed by the Report Template Gallery, which shows each report with a description of the data, objects used, and possible interactivity.

Report	Use Case
Trial Assignment Analysis	This report presents demographic data on age, race, gender, and some initial biometric measures (height, weight, blood pressure). The focus of the report is a comparison of the two arms of a single study: one represented in green and the other represented in blue. Users can select and filter by various demographic variables to explore the degree to which the assignment to the two arms of the study was balanced.
Enrollment History	This report presents the demographic data with an emphasis on when subjects were enrolled in the study. The same blue/green color scheme in the Trial Assignment Analysis report is used for continuity. But the X axis of the bar chart is the date of enrollment in the study, and the Y axis is the count of participants. Users can select data by age group, race, and even investigator. The purpose of the report is to enable users to review and explore the timing of fulfillment of the study protocol.
Lab Test Comparison	This report includes typical filtering options by sex and race, but the main purpose of the report is to enable users to compare one or more labs to display the low, high, and normal range results across labs. A pie chart at the bottom shows the totals labs administered and the detail table provides more specific results from each lab.
Patient Lab Record	This report enables users to search for one particular subject in this study. The report displays the data for the patient in tabular form (assignment, age, height, weight) and in dynamic format with a bar chart and detail table showing the results of every lab test for that patient using the low, normal, and high categories.
Adverse Event Duration	This report enables users to select one particular view of the adverse event data, focusing on the first event, first serious event, last event, or last serious event, using a radial check box. The report displays the results in a bubble plot, distinguishing each experimental condition of the study with data on specific body systems. Other filters for age, race, and sex allow for additional interactions with the data.
Adverse Events over Time	This report enables users to explore adverse event data by the investigator and/or specific subjects to see the incidence of mild, moderate, and severe adverse events, and to see the duration of events over time. The duration graph itself can be used as a filter to drive the day-of-study graph in the report.

Table 1. Overview of Six Template SAS Visual Analytics Reports

The Report Template Gallery identifies which data sets were used and any necessary transformations of the data that were required. It does not detail every edit to the standard clinical trial data. For example, "M" and "F" were replaced with "Male" and "Female" before the data was loaded into SAS Visual Analytics. Variable values were formatted for readability using mixed-case fonts and abbreviations were used for certain very long strings. MMDDYY dates were converted to monthly values. But all of the variables included in these template reports should be readily available from any clinical study.

REPORT TEMPLATE GALLERY

This section shows each report with a description of the data, objects used, and interactivity. A screenshot of the report and a screenshot reflecting the underlying connections between objects is included. Note that the screenshots presented here have been edited to fit within the printed page. In use, the reports are dynamic and include multiple options for resizing to change user focus and/or fill a larger screen. There are six reports:

- Trial Assignment Analysis
- Enrollment History
- Lab Test Comparison
- Patient Lab Record
- Adverse Event Duration
- Adverse Events over Time

A specification table describes the report functionality.

TRIAL ASSIGNMENT ANALYSIS

Specification	Description
Source Data	Demographic data file with the experimental assignment and the first measurement of height, weight, and blood pressure from associated data files. Data is at the one-patient-per-row level. This is a main data structure used in later reports.
Data Preparation	Most data elements are available in the format necessary for use in this report, but one modification was required. The line graph representing the Age Distribution at the top of the report required the transformation of the Age variable from a numeric to a character string. The line graph requires a character variable on the X axis. This simple format transformation was done to the SAS data set before the data was loaded into the SAS LASR Analytic Server. The Total Subjects count in the Detail Table was not in the original data loaded into the SAS LASR Analytic Server. The Frequency variable is automatically generated by SAS Visual Analytics for any data set.
Visual Analytics Objects	Line Plot (Age Distribution) Bar Chart (Males, Females) Crosstab (Detail Table)
Interactivity	This report begins top-down with selections made in the Age Distribution graph, driving changes in the Males and Females bar charts and Detail Table. Note as well that selections of race within the Males and Females graphs operate sideways, with bar chart selections on the right or left affecting the bar chart on the opposite side and the Detail Table.

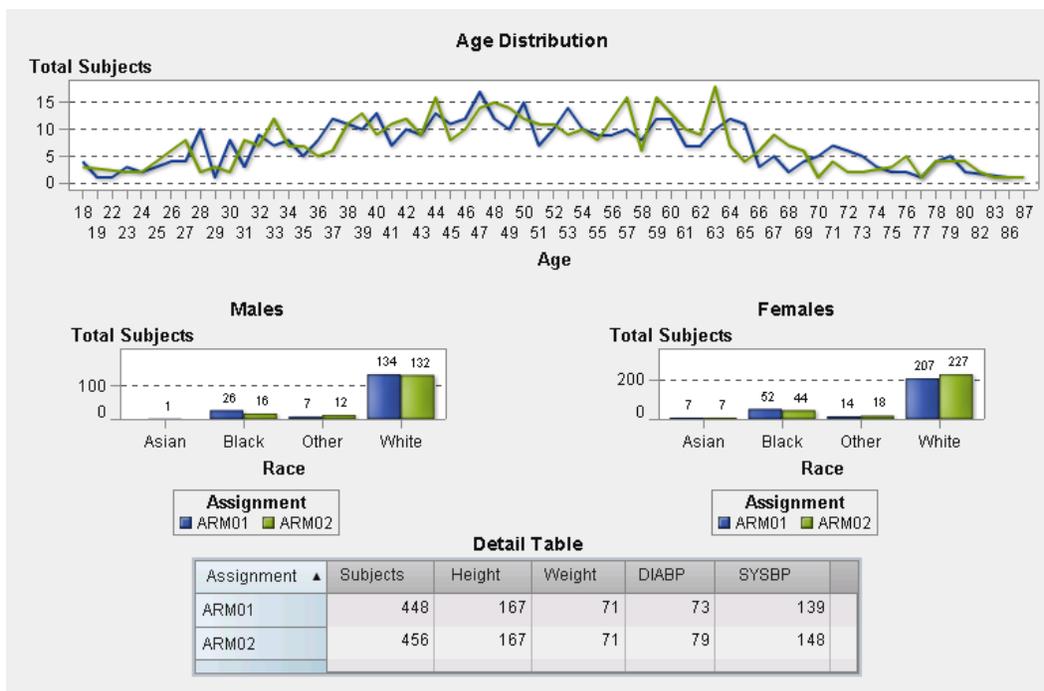**Table 2. Specifications for the Trial Assignment Analysis Report**

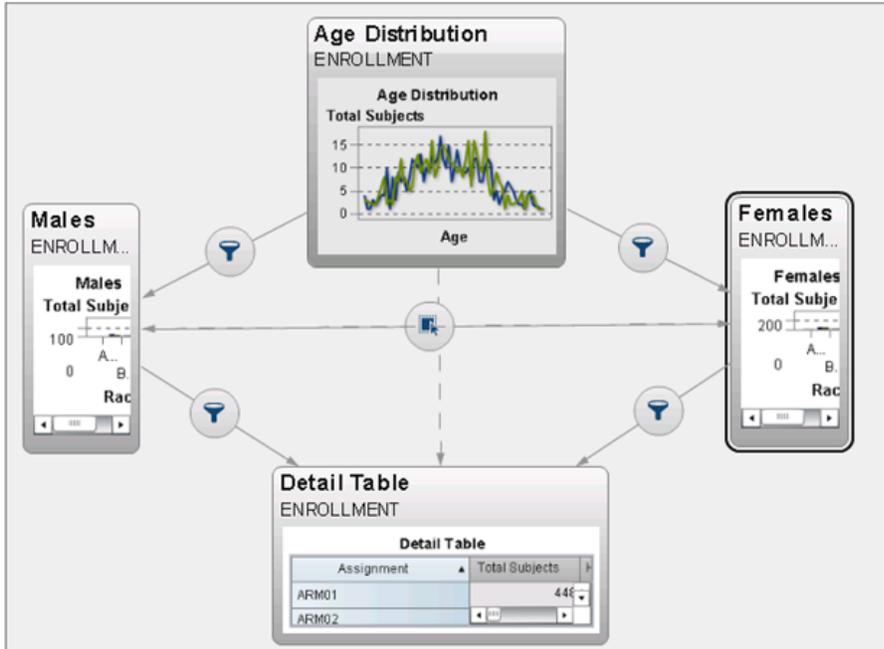

Figure 4. Trial Assignment Analysis Report

Figure 5. Trial Assignment Analysis Interaction Design

ENROLLMENT HISTORY

Specification	Description
Source Data	Same data as Trial Assignment Analysis, but this report uses other variables, such as the Investigator, culled from the same data. Data is at the one-patient-per-row level.
Data Preparation	For this report, two variables were converted into character strings, and one was calculated as follows: Created Month_Year (character variable for bar chart) Created Age Groups (character variable "10-19," "20-29") Calculated Study Duration (for Detail Table) using Last-First day over all events for each patient
Visual Analytics Objects	Bar Chart (Subjects Enrolled by Date) Bar Chart with stacked option (Subjects by Gender and Race) Line Plot (Subjects by Age Group) Crosstab (Details Table)
Interactivity	This report begins top-down, with the enrollment graph serving as a control to allow for selection of one or more groups (arm01/arm02, a range of months, and/or individual bars). That selection drives the graph, which can then be used left-to-right interactively. The Detail Table on the far right shows more data than is shown in Figure 6 and can be sorted and explored, reflecting all previous selections.

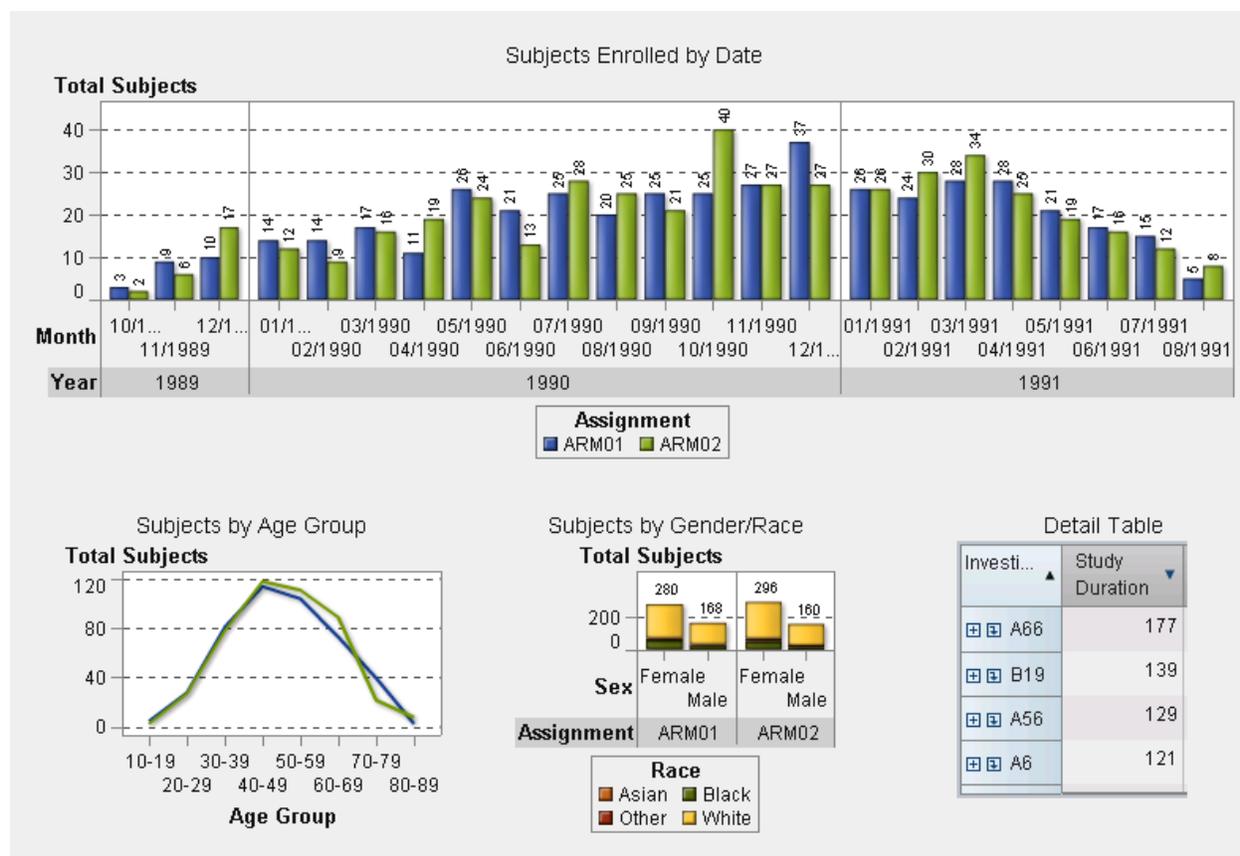**Table 3. Specifications for the Enrollment History Report**
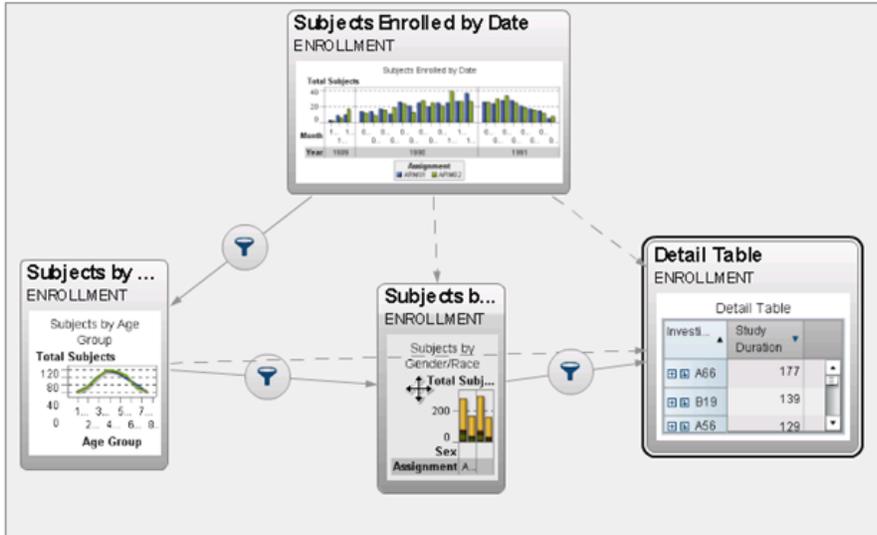

Figure 6. Enrollment History Report

Figure 7. Enrollment History Interaction Design

LAB TEST COMPARISON

Specification	Description
Source Data	This data includes the same data structure used for the Enrollment History report with data from labs added to the data structure, resulting in a data set at the one-lab-per-row level, with patients and labs repeated to reflect all labs administered to all patients.
Data Preparation	The stacked RED/GREEN/RED bars were achieved by using a Display Rule and assigning red to both Low and High, and green to Normal. Note that the Y axis is total labs administered for that lab for that experimental condition. This graph could not have been implemented using the actual value of the lab result as the labs being compared are expressed in units on very different scales. The Result in the Detail Table is the mean value for those labs. This value could be replaced with the median or another measure.
Visual Analytics Objects	Pulldown Control (Sex and Race) Slider Control (Age) List (Select a Lab Test) Bar Chart with stacked option and lattice column Pie Chart (Total Labs) List Table (Detail Table)
Interactivity	This report takes a more traditional reporting approach, allowing users to select pull-down choices for sex and race. A slider allows for selecting a range of ages. A lab test list allows the selection of one or more labs. One of the unique aspects of this interactivity is that the prompts cascade, meaning that the first selection filters the next and the next. All choices are then reflected in the bar chart, which can also be used to control the data used in the pie chart and Detail Table. The full scope of all interactivity is best shown in Figure 9, which details all of the interconnected links.

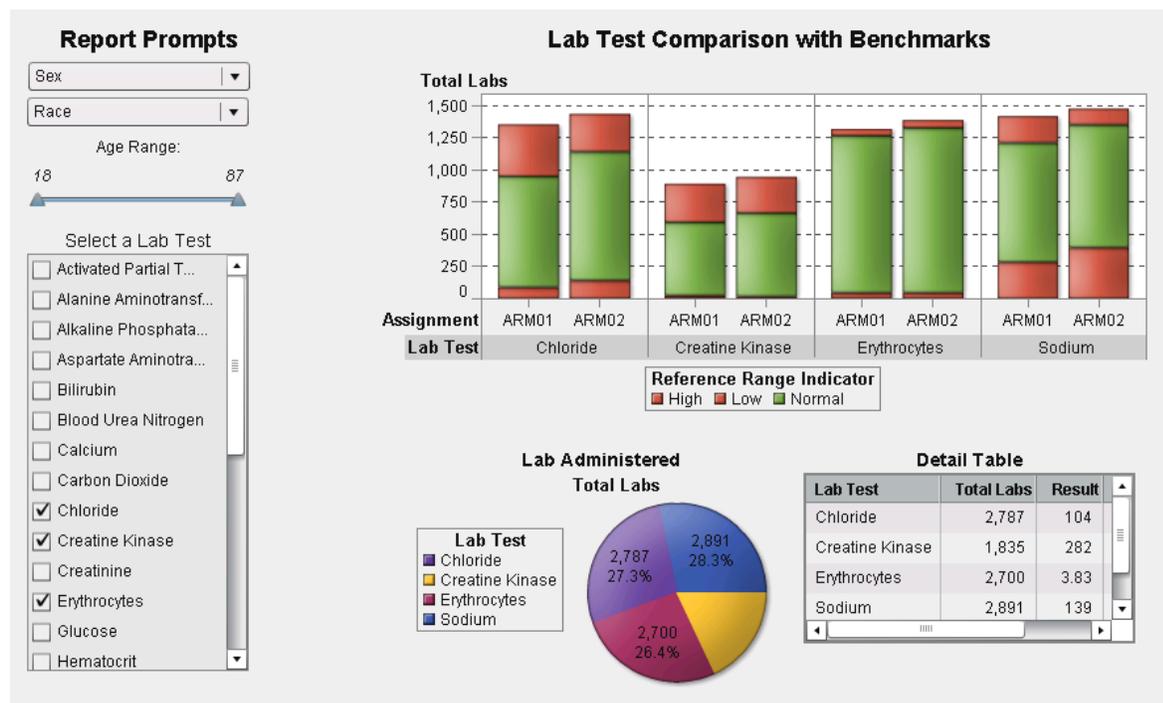**Table 4. Specifications for the Lab Test Comparison Report**
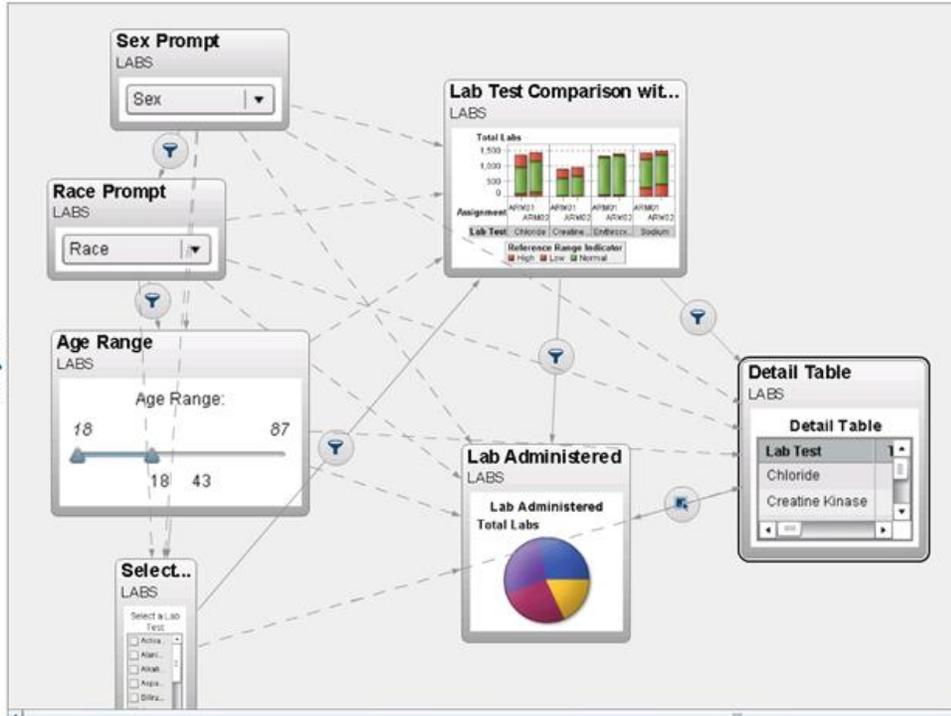

Figure 8. Lab Test Comparison Report

Figure 9. Lab Test Comparison Interaction Design

PATIENT LAB RECORD

Specification	Description
Source Data	Data for this report is the same as for the Lab Test Comparison report. Data is at the one-lab-per-row level with patients and labs repeated to reflect all labs administered to all patients.
Data Preparation	No additional data preparation required.
Visual Analytics Objects	Text Input Field (under Search for a Patient) List Table (Subject Details) Bar Chart data in clustered format (Lab Administration Timeline) List Table with display rules applied (Lab Results Details)
Interactivity	This report is very much a top-down, traditional report. By typing a patient ID into the text input field, a patient is selected. All data reflect that request. The bar chart with lab data is interactive in that it allows the selection of labs at a particular time or selection of labs with a particular result (low, high, normal), which are reflected in the Detail Table.

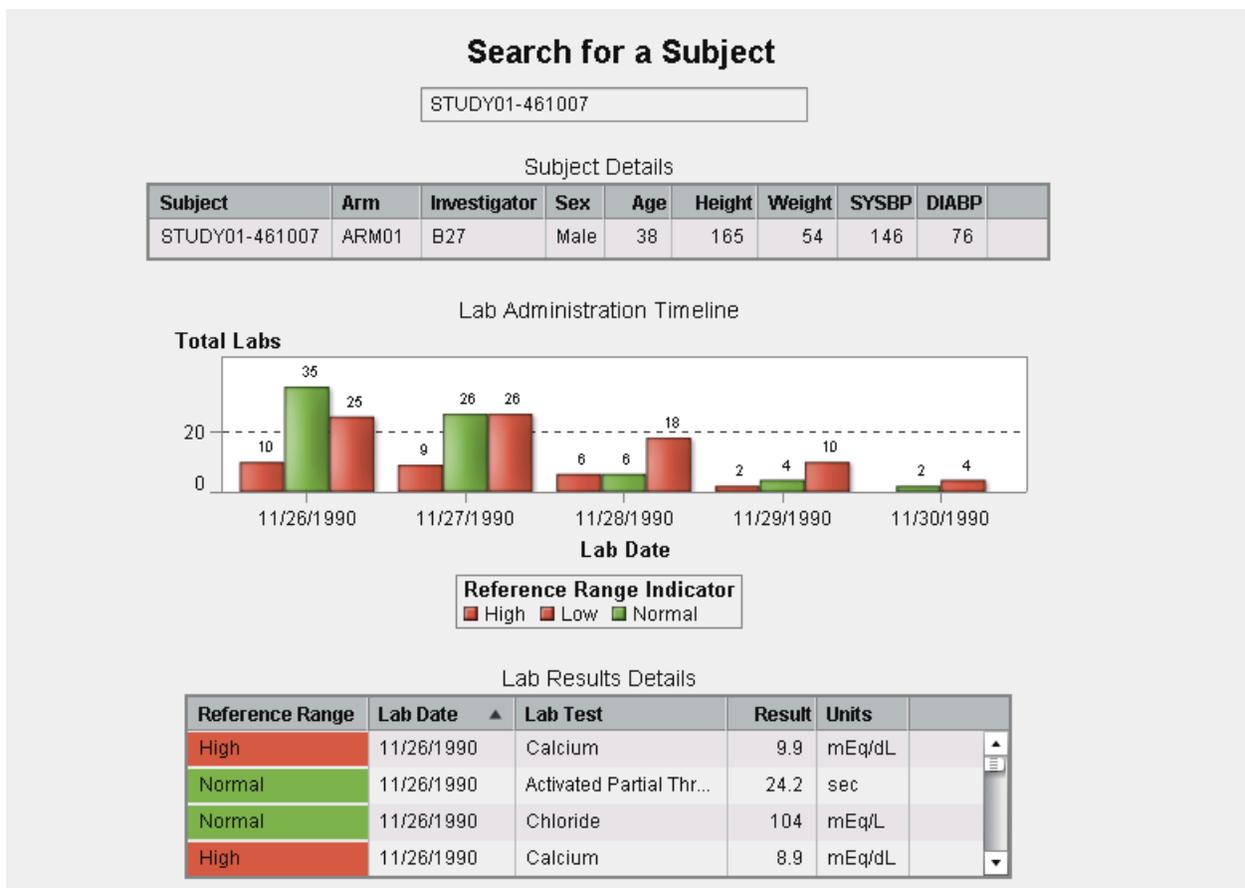**Table 5. Specifications for the Patient Lab Record Report**
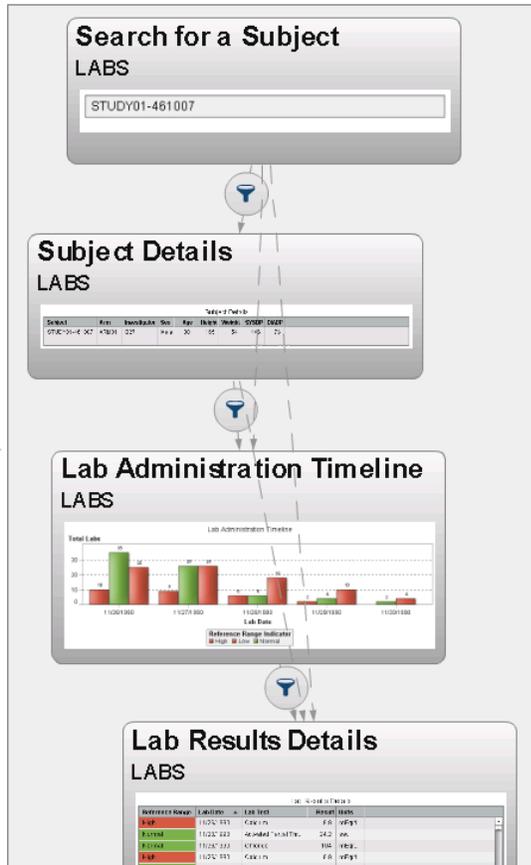

Figure 10. Patient Lab Record Report

Figure 11. Patient Lab Record Interaction Design

ADVERSE EVENT DURATION

Specification	Description
Source Data	Data on adverse events merged into the Trial Assignment/Enrollment History data structure. This data does not include lab results. Data is at the one-event-per-row level, with patients and events repeated to reflect all events experienced by all patients. Note the usual repetition of the data as described below.
Data Preparation	Unlike other data structures used in these reports, the entire data structure here is stacked, meaning that the same basic data are represented four times: once for First Adverse Event, First Severe Adverse Event, Last Adverse Event, and Last Severe Adverse Event. While the Start Day and End Day are variables on every row, the meaning of the variable is dependent on the value of the Event Timing variable on that row. The List Box Control for Event Timing forces a single selection of only one type of event, filtering the data to select only those records.
Visual Analytics Objects	List Box Controls (Age, Race, Sex) List Box Control with required single selection only (Event Timing) Bubble Plot with Hierarchy assigned to the Group variable and a lattice separating the two experimental conditions
Interactivity	Per above, all of the reporting is affected by the selection in the bottom left of the screen (the radial box that filters one of four events in the data). For this report, this is a mandatory selection. The default has been set to "First Adverse Event." With that selected, users can select age groups, race, and sex of patients, all of which are reflected in the bubble plot. Although it cannot be shown here, the bubble plot is drillable as well, with additional details of each grouping available by clicking on a bubble.

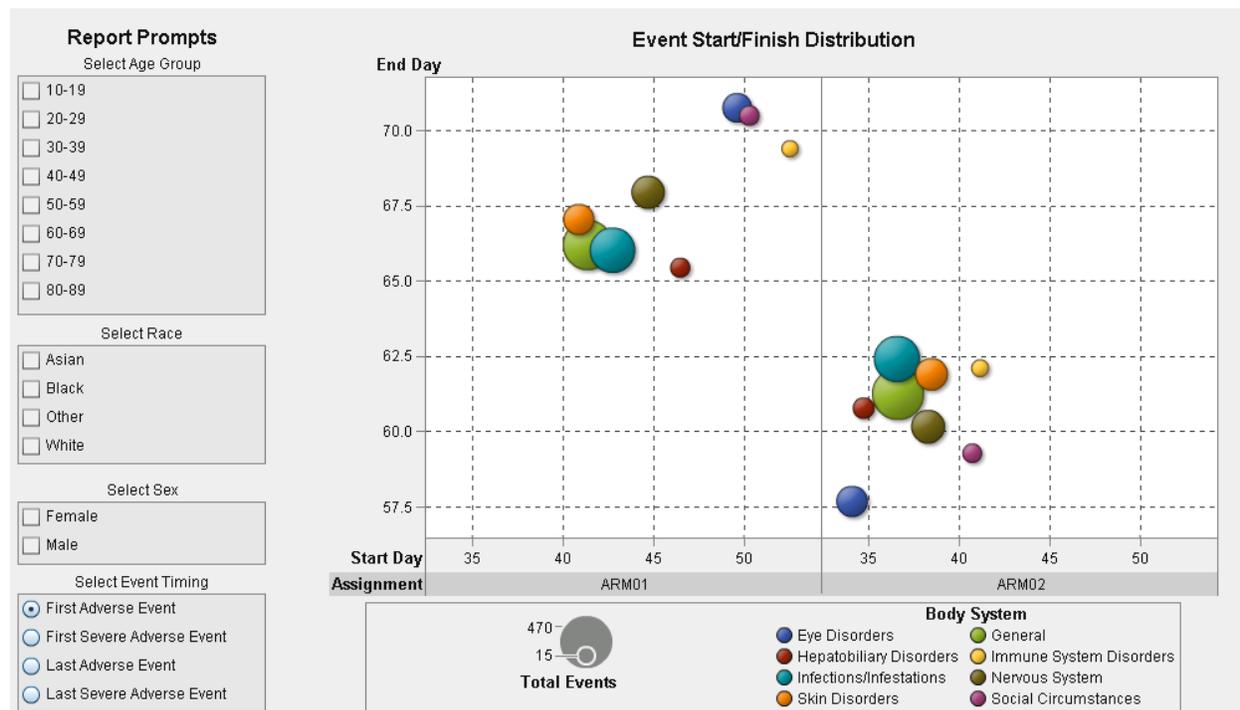**Table 6. Specifications for the Adverse Event Duration Report**

Figure 12. Adverse Event Duration Report

Figure 13. Adverse Event Duration Interaction Design

ADVERSE EVENTS OVER TIME

Specification	Description
Source Data	This is the same data as in the Adverse Event Duration report, with data at the one-event-per-row level. Patients and events are repeated to reflect all events experienced by all patients. Unlike the previous report, the data structure here is not stacked.
Data Preparation	Day of study is represented in two ways in this report. The bottom graphic showing Day of Study uses the day of the event relative to the first day of the study. But the Event Longevity graph at the top right uses adverse event data normalized so that start of the event is set to 0. This is important as the intention here is to compare the relative length of the events, not the specific day of study that the event occurred.
Visual Analytics Objects	Pull Down Control (Select Investigator) List Box Control (Subject) Vertical Bar Chart (Event Severity Duration) Horizontal Bar Chart (Event Longevity) Line Plot with Vertical Lattice using the Overlay Filled property
Interactivity	From the left, the selection of an investigator cascades down to the subjects, meaning that only those patients associated with the selected investigator are shown in the list box. Those selections are reflected in the graphs, and each graph then modifies the next, moving left-right and top-down. The longest duration events can be captured by selecting only the right tip of the bars under Event Longevity, which will then modify the graph at the bottom to reflect only the longest duration events and to allow users to see when those started for the two experimental conditions.

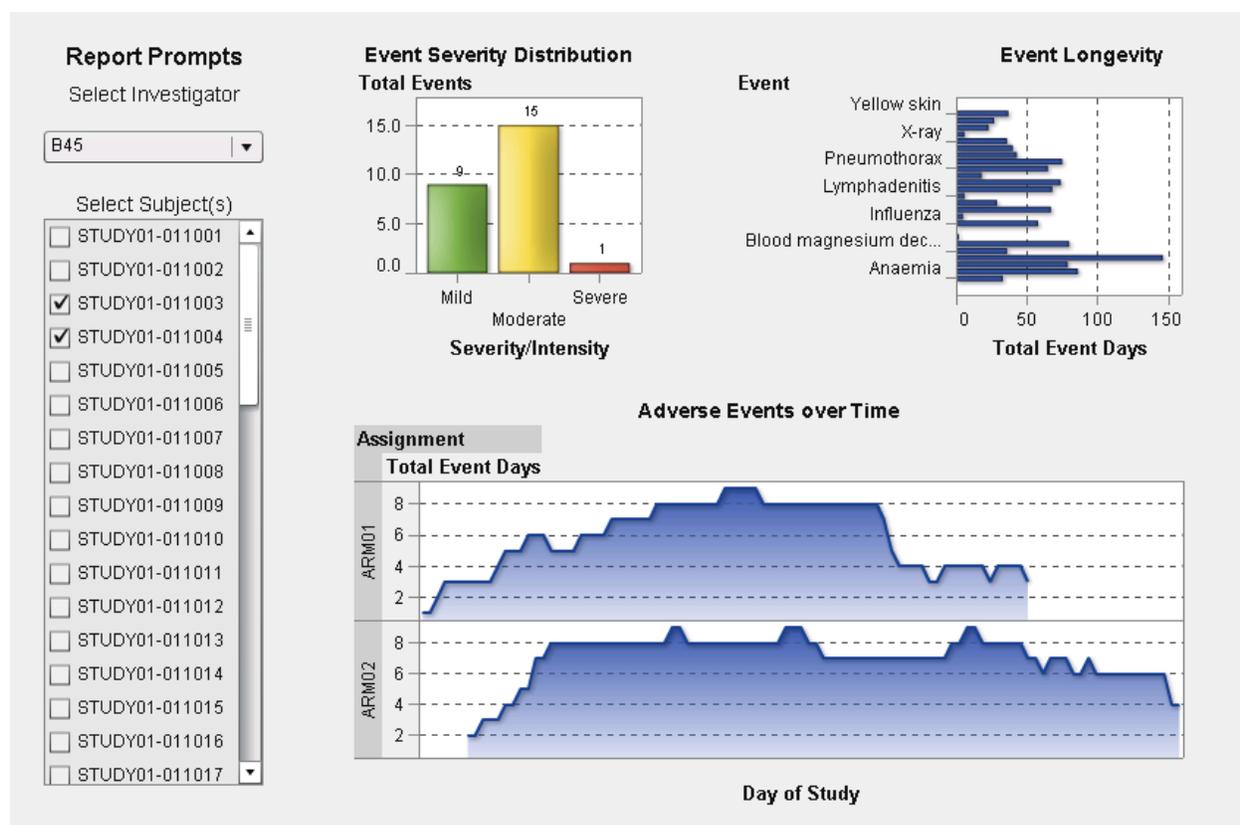**Table 7. Specifications for the Adverse Events over Time Report**

Figure 14. Adverse Events over Time Report

Figure 15. Adverse Events over Time Interaction Design

CONCLUSION

In an experiment, we define measurements, develop methods for collecting those measurements, gather data, and analyze that data, typically attempting to isolate the effects of one particular variable. To do so, we might implement experimental or statistical methods to control for the effects of other variables, factors that are not our primary concern. This is the basic method of science, especially as science is practiced in the analysis of clinical trial data where the variable of interest is a particular treatment.

That is the approach of science and no reasonable person would question the value of that approach. On the other hand, we do not live in a controlled environment in which one isolated variable has an effect. We live in a multivariate world in which a tremendous number of variables are constantly interacting with effects changing over time. The real world is a complex, dynamic place. While there have been significant improvements in the statistical models used to explain that complexity, our methods for reporting the results of such models have lagged behind.

Static reporting will always have a place in science. But newer technologies using interactive visualization of data have the potential to transform our ability to report on very complex data. To fully understand the relationship between all of the subject-state, time-dependent, and covariate relationships in data that we gather, we need a new way of reporting. SAS Visual Analytics provides data scientists with the technology to reflect these relationships. Just as advances in genetics, chemistry, and even engineering have improved the development of new compounds and therapies, advances in the technology we use to analyze clinical trial data provide the opportunity to improve our ability to learn from the data.

FURTHER READING

Full documentation of SAS Visual Analytics is available here:
http://support.sas.com/software/products/va/

CONTACT INFORMATION

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

Michael Drutar
100 SAS Campus Drive
Cary, NC 27513
SAS Institute Inc.
Michael.Drutar@sas.com
http://www.sas.com

Elliot Inman, Ph.D.
100 SAS Campus Drive
Cary, NC 27513
SAS Institute Inc.
Elliot.Inman@sas.com
http://www.sas.com