ABSTRACT

The success of any clinical trial depends on the accuracy and integrity of the study conduct and the data produced from the trial. As in any experiment, data plays the central role and almost everybody involved in a clinical trial generates, maintains, or explains data. Hence, we can all agree: It is vital that the data is clean and, more importantly, that it is fully utilized to make everyone’s job easier and efficient.

A plethora of software, programming languages, and tools are employed by various contributors in clinical research across the industry to help make sense of clinical and operational data. The scope of and investment in such tools depends on the budget and organizational priorities. There are organizations that operate with moderate or minimalistic software resources. This paper will explore various ways in which SAS programmers (either statistical programmers or clinical programmers) can help other departments in such organizations.

The paper will list some examples of such offerings, sometimes with a sample approach, like, data listings, CRF tracking metrics, patient profiles, and site summary metrics; patient profiles with specific data points to help CRAs; patient safety summaries; and custom safety narrative templates. This paper also deals with some of the graphical representation of the data before the actual statistical programming starts for tables, listings, and figures. It also describes unique ways for summarizing clinical trial data to make it easy to spot errors in data about individual subjects or clinical sites.

INTRODUCTION

Clinical research is a collaborative process where various departments with varied skills and responsibilities work together with a common goal. This collaboration culminates in a successful, quality submission with biostatisticians and statistical programmers playing a key role in summarizing the reports reviewed by regulatory authorities. The quality of the data-driven decisions is limited by the quality of the data. If your data have errors, your decisions will be error prone as well. Detecting inadvertent errors and fraudulent data is paramount at every step. In general, CROs or sponsors often hire people to handle data entry and analysis, but how can you tell if you have the right staffing, especially with the type of reports that are being developed to see data more clearly? Can the database programmer produce these reports? Or, should we go to a statistical programmer who can utilize a gamut of capabilities in SAS? What about a useful software that isn’t available because of financial limitations? One of the best ways to ensure data integrity is to create the reports that are discussed in this paper with appropriate data management plans from your clinical data management team with the help of a statistical programmer.

The idea is to use statistical programmers in addition to the database programmer in the initial stages of trial.

Why is it important? Data integrity is what enables organizations to get a clear picture of the trial, which, in turn, makes decision making efficient.
Figure 1. Following the Data Trail and Traditional Statistical (SAS®) Programmer Role: From the book SAS Programming in pharmaceutical industry.

WHAT ARE SOME OF THE THINGS THAT STATISTICAL (SAS®) PROGRAMMERS CAN OFFER

Quality of Data - Clinical Data Management
- CRF completion/tracker
- Site summary and query rate

Patient Safety and Efficacy - Drug Safety
- Safety review report
- RECIST listings

Decision-Making Visuals by Statistical Programmer (web-based reports) - Clinical Operations
- Simple demographics
- Sites with AE/SAE count
- Complete site analysis
- Laboratory data
- Patients meeting inclusion/exclusion criteria

Consumers of these reports could be data managers, biostatisticians, medical writers, clinicians, CRAs, or other decision makers on the sponsor’s team. These reports are more than the DVS listings, edit check programs, or patient profiles/narratives that a programmer is doing on regular basis.

QUALITY OF DATA

CRF Completion
The earlier the missing CRF pages report (Display 1) is created and implemented, the sooner we will be alerted to any problems and patterns in the data collection process. Additionally, regular and thorough review of this type of report results in a more accurate trial database, a shorter time between the end of a trial and a final database and, ultimately, a higher quality trial.
<Let's Check Data Integrity>, continued

The specification from the data management team may look like this example in Table 1.

<table>
<thead>
<tr>
<th>Site Id</th>
<th>PI Name</th>
<th>Visit ID</th>
<th>Visit Name</th>
<th>Missing Page ID</th>
<th>Missing Page Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Dan</td>
<td>1001</td>
<td>Cycle 2 Day 1</td>
<td>90</td>
<td>Maintenance Hydration</td>
</tr>
<tr>
<td>1</td>
<td>Dan</td>
<td>1001</td>
<td>Cycle 2 Day 2</td>
<td>50</td>
<td>Maintenance Hydration</td>
</tr>
<tr>
<td>1</td>
<td>Dan</td>
<td>1001</td>
<td>Cycle 2 Day 3</td>
<td>50</td>
<td>Maintenance Hydration</td>
</tr>
<tr>
<td>2</td>
<td>Ren</td>
<td>1002</td>
<td>Cycle 2 Day 3</td>
<td>70</td>
<td>DCF-HRI</td>
</tr>
<tr>
<td>2</td>
<td>Ren</td>
<td>1002</td>
<td>Cycle 3 Day 3</td>
<td>50</td>
<td>Maintenance Hydration</td>
</tr>
<tr>
<td>2</td>
<td>David</td>
<td>1003</td>
<td>Cycle 0 Day 2</td>
<td>10</td>
<td>Patient Visit</td>
</tr>
<tr>
<td>3</td>
<td>David</td>
<td>1003</td>
<td>Cycle 0 Day 2</td>
<td>10</td>
<td>Patient Visit</td>
</tr>
</tbody>
</table>

Table 1. Example of Specification

Site Summary and Query Rate

This report will give us the metrics and the status of a subject, e.g., whether the subject has entered the study and what the current CRF status is for a site or a subject, like pages received, entered, reviewed and cleaned, queries open, queries closed, etc. See Display 2.
Display 2. Site Summary Report

We can produce this type of reports by sponsor and study. An example of this can be seen in Table 2.

<table>
<thead>
<tr>
<th>Sponsor</th>
<th>Study</th>
<th>NA</th>
<th>No Data</th>
<th>Entered</th>
<th>Completed</th>
<th>Monitored</th>
<th>Issue</th>
<th>DM Reviewed</th>
<th>Frozen</th>
<th>Signed</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sponsor1</td>
<td>Study</td>
<td>0</td>
<td>1216</td>
<td>85</td>
<td>36</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>41</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Sponsor2</td>
<td>Study</td>
<td>15</td>
<td>804</td>
<td>43</td>
<td>1584</td>
<td>63</td>
<td>0</td>
<td>133</td>
<td>2972</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Sponsor3</td>
<td>Study</td>
<td>0</td>
<td>1194</td>
<td>179</td>
<td>1904</td>
<td>3</td>
<td>0</td>
<td>49</td>
<td>2943</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Site Summary Report by Sponsor

PATIENT SAFETY AND EFFICACY

Safety Review Report

This report will help us identify all safety information in a single shot for each subject. Similar reports can be used across various studies. See Display 3.
## Let's Check Data Integrity, continued

### PATIENT

<table>
<thead>
<tr>
<th>Patient # / Initials</th>
<th>Dosing Group</th>
<th>Date ICF Signed</th>
<th>Date of Screening Visit</th>
<th>DOB</th>
<th>Sex</th>
<th>ECOG at Screening</th>
<th>ECOG at C101</th>
</tr>
</thead>
<tbody>
<tr>
<td>01001/R-V</td>
<td>SAFETY RUNIN PHASE</td>
<td>14FEB2013</td>
<td>14FEB2013</td>
<td>10DEC1940</td>
<td>MALE</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

### VITAL SIGNS

<table>
<thead>
<tr>
<th>Visit</th>
<th>Date</th>
<th>Time</th>
<th>Sys/Dia</th>
<th>HR</th>
<th>Resp</th>
<th>Temp</th>
<th>Temp</th>
<th>Weight</th>
<th>Weight Unit</th>
<th>Height</th>
<th>Height Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Screening</td>
<td>14FEB2013</td>
<td>12:15</td>
<td>128/60</td>
<td>82</td>
<td>18</td>
<td>86</td>
<td>F</td>
<td>134</td>
<td>LB</td>
<td>66</td>
<td>N</td>
</tr>
<tr>
<td>Cycle 1 Day 1 Pre-Dose</td>
<td>16FEB2013</td>
<td>07:55</td>
<td>130/80</td>
<td>62</td>
<td>18</td>
<td>96.6</td>
<td>F</td>
<td>132.2</td>
<td>LB</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cycle 1 Day 1 Post-Dose</td>
<td>16FEB2013</td>
<td>13:25</td>
<td>118/72</td>
<td>66</td>
<td>17</td>
<td>97.8</td>
<td>F</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cycle 1 Day 1 Post-Dose</td>
<td>16FEB2013</td>
<td>14:25</td>
<td>122/72</td>
<td>63</td>
<td>16</td>
<td>97.3</td>
<td>F</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cycle 1 Day 1 Post-Dose</td>
<td>16FEB2013</td>
<td>15:25</td>
<td>128/60</td>
<td>56</td>
<td>16</td>
<td>97.7</td>
<td>F</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cycle 1 Day 1 Post-Dose</td>
<td>16FEB2013</td>
<td>16:25</td>
<td>125/70</td>
<td>87</td>
<td>17</td>
<td>97.5</td>
<td>F</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### 12-LEAD ECGS

<table>
<thead>
<tr>
<th>Visit</th>
<th>Date</th>
<th>Time</th>
<th>QTc</th>
<th>Method</th>
<th>Overall Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Screening</td>
<td>14FEB2013</td>
<td>10:51</td>
<td>378</td>
<td>MACHINE CALCULATED</td>
<td>NORMAL</td>
</tr>
</tbody>
</table>

---

**Patient # / Initials - 01001 / R-V**

### Laboratory Test Results

<table>
<thead>
<tr>
<th>Laboratory Test Short Name</th>
<th>Screening</th>
<th>C101</th>
<th>C102</th>
<th>C1015</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Result</td>
<td>Unit</td>
<td>C.S. - Y/N</td>
<td>Result</td>
</tr>
<tr>
<td>CHEMISTRY</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ALB</td>
<td>3</td>
<td>GDL</td>
<td>N</td>
<td>3</td>
</tr>
<tr>
<td>A LP</td>
<td>661</td>
<td>IUL</td>
<td>N</td>
<td>661</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COAGULATION</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INR</td>
<td>1.2</td>
<td></td>
<td>N</td>
<td>1.2</td>
</tr>
<tr>
<td>PT</td>
<td>12.2</td>
<td>SEC</td>
<td>N</td>
<td>12.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HEMATOLOGY</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BASO</td>
<td>0.2</td>
<td>OTHER X10E09UL</td>
<td>0.02</td>
<td>OTHER X10E09UL</td>
</tr>
<tr>
<td>EOS</td>
<td>0.4</td>
<td>OTHER X10E09UL</td>
<td>0.2</td>
<td>OTHER X10E09UL</td>
</tr>
</tbody>
</table>
Display 3. Safety Review Report

RECIST Listings
To verify when the response data is correct and to identify where the potential errors are, queries can be written and also used to identify areas that the CRAs might need retraining on. This can be a helpful as a real-time tool for the CRAs to use while monitoring. See Table 3.

<table>
<thead>
<tr>
<th>Site ID</th>
<th>Subject ID</th>
<th>Primary Diagnosis</th>
<th>Histologic Diagnosis</th>
<th>Assign Dose Level</th>
<th>Completion/Discontinuation</th>
<th>Visit Number</th>
<th>Visit</th>
<th>Date of Assessment</th>
<th>Lesion Type</th>
<th>Lesion Number</th>
<th>Lesion Site</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>001-01-1</td>
<td>OTHER: ESOPHAGUS</td>
<td>ADENOCARCINOMA</td>
<td>10</td>
<td>17-Jun-13</td>
<td>10</td>
<td>SCREENING</td>
<td>17-Apr-13</td>
<td>Target</td>
<td>1</td>
<td>LIVER</td>
</tr>
<tr>
<td>1</td>
<td>001-01-1</td>
<td>OTHER: ESOPHAGUS</td>
<td>ADENOCARCINOMA</td>
<td>10</td>
<td>17-Jun-13</td>
<td>10</td>
<td>SCREENING</td>
<td>17-Apr-13</td>
<td>Target</td>
<td>2</td>
<td>LIVER</td>
</tr>
<tr>
<td>1</td>
<td>001-01-1</td>
<td>OTHER: ESOPHAGUS</td>
<td>ADENOCARCINOMA</td>
<td>10</td>
<td>17-Jun-13</td>
<td>10</td>
<td>CYCLE 2</td>
<td>17-Apr-13</td>
<td>Target</td>
<td>1</td>
<td>LIVER</td>
</tr>
</tbody>
</table>

Columns Continue.

<table>
<thead>
<tr>
<th>Method of Assessment</th>
<th>Non-Target Lesions Status</th>
<th>Lesion Measurement</th>
<th>Total Dimension of Lymphoma</th>
<th>Sum of Lesion Measurements or Products</th>
<th>% Change from Baseline</th>
<th>% Change from Smallest Sum</th>
<th>Overall Responses</th>
<th>Response Description</th>
</tr>
</thead>
</table>

Table 3. RECIST Listings

CODE FOR ALL THE ABOVE REPORTS: The code for all these reports mostly uses ODS tag set or XML programming, which allows us to create multiple sheets for sites and subjects. However, the focus here should on ensuring correct patient data; the programming is an individual’s own preference. Here is one option:

```
ODS TAGSETS.EXCELXP
   OPTIONS (sheet_name="&site&nn"
   absolute_Column_Width='8,15,10,10,10,10,10,10,10,10,10,10'
   frozen_headers='3' frozen_rowheaders="3");

Proc report;
  <CODE>
  Run;

ODS TAGSETS.EXCELXP CLOSE;
ods listing;
```

CREATING DECISION-MAKING VISUALS BY STATISTICAL PROGRAMMER (SAS) (WEB-BASED REPORTS)
The consumers of these types of reports are CRAs, medical writers, clinical safety scientists, and physicians.
Let's Check Data Integrity, continued

Display 4. Complete CRF Status of a Study

Here is an example of code that can be used to examine the complete CRF status of a study:

```sas
filename odsout "C:\..\..\Desktop\Pharmasug2017";
ods html path=odsout frame="Sitel1.html"
   contents="Site_contents.html"
   body="Site_body1.html"
   nogtitle;
   title1 "SITE CRF STATUS";
   footnote j=r "SITE INFORMATION ";
   title1 "Site wise CRF status " h=9pt;
ods proclabel = "Total CRF Status";
<INDIVIDUAL PLOT STATEMENTS>
Proc SGPLOT;
<CODE>
Run;
Proc gplot;
<CODE>
```
Let's Check Data Integrity, continued

Run;

<GTDL STATEMENTS?

ods html close;
ods html;

Subject wise CRF status for siteid '01' and subject '01'

Display 5. CRF Status by Subject

Here is an example of a code that can be used to examine CRF status by subject:

```sas
proc gchart data=siteinfo;
   pie3d Site / sumvar=CRFSTATUS noheading w-outline=2
             coutline=black discrete;
run;
QUIT;
```
Multiple Site CRF STATUS

Figure 2. Multiple Site CRF and Query Status Information

The graphics in Figure 2 explain the details of the CRF status and query status for multiple sites at a glance.
<Let’s Check Data Integrity>, continued

Code for Figure 2:

```sas
proc gchart data=siteinfo;
   pie CRFSTAT / sumvar=SITE
      other=5
      otherlabel='Missing'
      group=siteid
      across=2
      clockwise
      value=none
      slice=outside
      percent=outside
      coutline=black
      noheading;
run;
quit;
```

ELIGIBILITY CRITERIA

Figure 3. Inclusion/Exclusion Criteria by Site

It is important to note that inclusion and exclusion criteria are not used to reject patients personally, but rather to identify appropriate participants and avoid chances of higher patient-selection-related risks at the site. Figure 3 shows that SITE 3 has a higher exclusion rate and waivers are granted. Here is the applicable code:

```sas
proc sgplot data=siteinfo;
   format iecatn ie.;
   vbar siteid / response=percent group=iecatn nostatlabel
      groupdisplay=stack
   xaxis label="IE Category"
   yaxis grid values=(0 to 110 by 10) label="Percentage of Criteria met";
```
Figure 4 explains that the exclusion reason labeled as EXCL30 is the reason behind all the excluded subjects from SITE 3.

![Exclusion Criteria by Site and Reason](image1)

**Figure 4. Exclusion Criteria by Site and Reason**

Figure 5 explains that site 3 has two waivers granted for EXCL30.

![Exclusion Criteria With Waivers Count](image2)

**Figure 5. Exclusion Criteria With Waivers Count**
<Let's Check Data Integrity>, continued

DEMOGRAPHICS

To see simple demographics by gender, age, race, and site:

```sas
proc sgpanel data= siteinfo;
panelby sex ethnic/
    layout=panel columns=4;
    hbox age / category=siteid;
    rowaxis display=(nolabel);
run;
```

ADVERSE EVENTS

Programmers can try multiple ways to show the AEs and SAEs, which, in turn, help medical writers, the safety department, and physicians. Some of the examples are mentioned.

Figure 6 explains which site has more AEs reported, with a reference line for 15 and 20 subjects.

Figure 7 explains which site has more SAEs reported, with a reference line for 5 and 2 subjects.
Figure 6. Adverse Event Count by Site

Figure 7. Serious Adverse Event Count by Site
CODE for Figure 6 (Same code can be used for Figure 7):

```sas
proc sgplot data=siteinfo;
  vbar siteid / response=aenum stat=sum group=siteid nostatlabel
    groupdisplay=cluster dataskin=gloss;
  refline 15 / lineattrs=(color=darkgreen) label='More………………'
    labelloc=inside labelpos=min;
  refline 25 / lineattrs=(color=darkred) label='More…..'
    labelloc=inside labelpos=min;
  xaxis display=(nolabel);
  yaxis grid label="Adverse Event Count";
run;
```

Display 7 explains how the AEs are distributed by site and the size of the bubble indicates the number of subjects in the site with AE. We can see that SITE 1 has more subjects with headaches that are moderate and severe. We can create these types of plots for a single event and check the ratios.

Display 7. Adverse Events by Site and Severity

Here is an example of code that can be used to see adverse events by sites and by severity:

```sas
proc sgplot data=aebct;
  format site. asevn sev. ;
  bubble x=site y=aeterm size=aecount / group=aesevn transparency=0.2;
  inset 'Bubble size is proportional to Number of Subjects in the Site with AE' / position=bottomright;
```
LABORATORY DATA:

Usually the labs data are collected by different labs and if we can show the difference between the ranges of results in those collections that will help in reconciliation of labs by data management.

Figure 8. Lab with Local Values

Figure 9. Lab with Central Values

Here is an example of code identifying lab values:

```plaintext
title1 "LAB VALUES BY SITE - LOCAL" h=9pt;
ODS PROCLABEL = "Albumin-Local labs BY SITE";

proc sgplot data = siteinfo;
```
CONCLUSION

There are many software programs and programming languages available to check clinical data integrity like Tableau, Spotfire, JMP, Python, and SAS visual analytics options. But these are cost-intensive and time-consuming to learn. We know every company has an SAS license and every company has an SAS programmer(s). If we can use the knowledge of an SAS programmer using SAS, we might save a lot of time and money. The latest SAS version contains many possibilities to create all type of reports. I strongly believe that sponsor-level customizations can be done in SAS but not with other software programs.

REFERENCES

Matange, Sanjay. Google search for “Sanjay Matange SAS ODS SG Graphics” will produce a list of dozens of publications, too numerous to list here.


ACKNOWLEDGMENTS

I would like to thank my manager, Michael Wisniewski, in particular, who believed me and for providing encouragement and supporting PharmaSUG participation, and my friend and well-wisher, Sridhar Patel, who has immense patience to listen my ideas and provide his valuable feedback.

RECOMMENDED READING

CONTACT INFORMATION

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

Harivardhan Jampala
Chiltern International
Address: 4000 CentreGreen Way, Suite 300, Cary, North Carolina, 27513, United States: +1 919 462 8867
Hari.Vardhan@Chiltern.com

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.