SAS Programming in the Pharmaceutical Industry

WORLDCOMP’16 Poster Paper Submission

Chris Helwig (EMCC)
and
Julia Bao (U. of Phoenix)

juliabao@email.phoenix.edu

ABSTRACT

This project examines the use of SAS, which is both a statistical programming platform and a programming language, in the pharmaceutical industry for the purpose of analyzing clinical trial data used to gain approval of new drugs and medical devices. This material has been used to teach computer information students about statistical programming techniques, as well as to acquaint them with the field of pharmaceutical industry software development.

General Terms
Computer Science Education, SAS, Pharmaceutical, Biostatistics.

Keywords
SAS, Computer Science Education, Pharmaceutical, Biostatistics.

1. Clinical Research

The Part of process of discovering new drugs and registering them for FDA approval and marketing. [1]

Pharmaceutical companies and research agencies engage in drug discovery, development, clinical trials, manufacturing, and marketing approval. [1]

2. SAS

SAS is a statistical software system used in a range of industries. [2]

SAS uses data steps to read data into memory and proc steps to calculate descriptive or inferential statistics, generate summary reports, and create summary graphs and charts. [3]

Data InvestigatorName;
Input Country$ Site Investigator_Name$;
Datalines;
US 1 Name1
US 2 Name2
US 3 Name3
;
Run; [2]

SAS’s implementation of structured query language, can sort, summarize, select, subset, join and concatenate datasets, treating the dataset conceptually like a database table with rows and columns. [4]

Proc SQL;
Create table NewTable as
Select * from Demographic_Data;
Quit;

SAS Macros can be used to perform repetitive tasks quickly and efficiently. [5]

Macro programs start with %, Macro variables start with &. [5]

Use %Let statement, a global statement to name a macro and give it a value. [6]

The SAS proc tabulate procedure can arrange data from a data set in table format. [7] Consider a table with columns for each dosage level and rows showing number of patients, medicine effectiveness, adverse event or side affect.

Also included would be detailed demographic data on race, age, gender, and whether subject was taking other medication. [8]

Data Drug_Study;
input gender $ race $ drug $ @@;
/* Format is M or F, race, and whether drug is Active or Placebo */
proc tabulate data = drug;
table (gender race),
drug*(n pctn<gender race>='%');
/* Generate s a table showing columns with Active and Placebo with a count (N) and percentages broken down by race and gender */
SAS also has a graph capability to generate charts. The general form of the command is:

Proc Gchart Data=Demographic_Data;
Vbar Race;
Run; [9]
SAS also generates charts showing frequency distributions using proc freq.

3. Examples
Suppose there is some sample clinical data related to a diet pill, and an investigator wants to determine if the baseline average weight of study participants has gone up or gone down. SAS can calculate such baseline comparisons. [10]

Or consider an example of a clinical trial of a drug to remove warts. SAS On Demand for Academics was used to run the SAS code. [11]

```sas
/* Podiatry data
wart diameter in cm, .3 to 1.5 cm*/
Data Podiatry;
Input Patid Trt Visit diameter Visdate mmddyy10. @@;
Datalines;
11 1 1 1.5 01-02-2015 11 1 1.1 1.5 09-02-2015
11 1 1.2 1.4 09-04-2015 11 1 2.09-06-2015
11 3 1.3 09-08-2015 11 1 1.2 09-10-2015
11 1 5 0.7 09-12-2015 21 2 1 1.7 09-02-2015
21 2 2 1.7 07-03-2015 21 2 3 0.6 06-06-2015
21 2 4 0.5 09-08-2015 21 2 5 0.6 10-10-2015
41 1 1 1.4 09-06-2015 41 1 2 1.5 07-08-2015
41 1 3 1.0 09-10-2015
; Run;
Proc Means Data=Podiatry (where = (1<=visit<=2)) Noprint;
By patid;
Var diameter;
Output out=BL(drop=_TYPE_ _FREQ_) mean=BLDiameter;
Run;
Proc means data=Podiatry (where = (3<=visit<=5)) Noprint;
By patid;
Var diameter;
output out=DDT(drop =_TYPE_ _FREQ_) mean=TrtDiameter;
Run;
DATA Change_From_Baseline;
Merge BL DDT;
by Patid;
Mean_Chg = BLDiameter - TrtDiameter;
Pct_Chg = (BLDiameter - TrtDiameter)/BLDiameter*100;
Run;
Proc Print data=Change_From_Baseline; [10]
```

4. CDISC ADaM
CDISC, which stands for Clinical Data Interchange Standards Consortium, is a standards organization for the clinical data field. [12] Part of this standard consists of the Analysis Data Model, which includes items such as variable name, label, data type, length, and comment.

5. CDISC SDTM
SDTM is the CDISC standard for the Study Data Tabulation Model, which governs human clinical trial data tabulations submitted to the Food and Drug Administration (FDA) for approval. This model provides detailed guidelines for how various medical interventions, findings, and measurements need to be documented, presented and statistically analyzed prior to submission as part of a regulatory approval process. Interventions refer to therapeutic treatments or surgical procedures; findings refer to lab test results, ECG results, or other observational evaluations. [13]

6. Biostatistics
Biostatistics is the “application of statistics to a wide range of topics in biology. The science of biostatistics encompasses the design of biological experiments, especially in medicine” and drug studies; “the collection, summarization, and analysis of data
from those experiments; and the interpretation of, and inference from, the results. A major branch of this is medical biostatistics, which is exclusively concerned with medicine and health." [14]

One statistical test that may be performed with SAS is the Proc Ttest. This statistical test can be used to compare if there is a statistically significant difference between two compared groups. [15]

Another statistical test that can be performed is proc REG for regression analysis. This test involves a dependant variable that can be predicted based on one ore more independent variables. For example, regression analysis could be used to predict blood pressure based on dosage of a new drug, cholesterol level based on patient age, or degree of wound healing based on the surface area of burn wound. [16]

Another possible study would be to consider the effectiveness of different kinds of stents used in aortic aneurism surgery. For example, in an endovascular aneurism surgery, stents are moved through the arteries via incisions near the groin to the aorta. A test could be used in a clinical study of the effectiveness of different versions of such stents, comparing factors such as diameter of aneurism, likelihood of rupturing, or symptoms such as back pain or a pulsating feeling near the navel. [17]

Another possible clinical study might evaluate the effectiveness of different models of artificial knees used in knee replacement surgeries, investigating outcome results such as range of motion, pain, and mobility levels, both before and after the procedure, and comparing one artificial knee to another.

7. Training
Seventy training options are available to learn SAS pharmaceutical programming. Philadelphia University and SAS Institute Inc. are collaborating to provide such training. The program covers: SAS® Fundamentals, SAS® Report Writing and ODS, SAS® Programming, SAS® Macro Language and SAS® SQL Processing. It also covers an overview of the pharmaceutical industry, data issues, study design, data analysis and biostatistics. [18]

Another similar training program is offered by UC San Diego, which has online certificate programs in SAS and Biostatistics. [19]

SAS Institute also offers its own Pharmaceutical Certificate. [20]

8. References
[1] SAS Clinical Programming in 18 Easy Steps by Y. Prasad, p. 1. While the FDA regulates drugs in the U.S., several international equivalent organizations regulate drugs overseas, such as the European Medicines Agency, the Chinese State Food and Drug Administration, and the Indian Ministry of Health and Family Welfare’s Drugs Controller General of India. www.fda.gov/Drugs/ResourcesForYou/Consumers/PrescriptionDrugAdvertising/ucm071964.htm. The FDA regulates drug advertising as well as the drug approval process.
[3] Ibid. at p. 16.