PharmaSUG 2013 - Paper IB10

SAS® PROGRAMMER TO CLINICAL SAS PROGRAMMER

Gayatri Karkera, inVentiv Health Care, Mumbai, India
Neha Mohan, inVentiv Health Care, Mumbai, India

ABSTRACT

In the clinical programming industry, as hiring managers, we tend to select candidates with varied academic backgrounds spanning from life sciences to abstract subjects like mathematics, statistics and computer sciences. Many a times our selection bias for candidates with mathematics, statistics and computer sciences, as compared to candidates from other academic backgrounds, is emanating purely from their training of programming skills. Anyone who has been in the clinical programming industry for long would agree that SAS® programming in this industry requires lot more than just programming. In this note, we attempt to bring out some dimensions of journey from pure SAS programmer to clinical SAS programmer. The trajectory of this exciting journey is full of learning. We share some of our learning and experiences basis our industry observations.

INTRODUCTION

For an experienced pure SAS programmer or for a beginner who has not been ever exposed to clinical programming, one of the most common challenge that managers face is to how effectively the programmer be introduced to the subject and where to begin his journey. A programmer not trained in the clinical field may typically be used to linear programming style or object oriented programming style or both. This training assumes the availability of the algorithm or the solution of the problem at hand while the programmer is expected to manage the constructs and the grammar of his/her computer language. And this is one reason, why in most cases, programmers find it difficult, at least in their initial days, understanding dimensions of clinical programming. Here the programmer has to think beyond his language constructs and grammar, beyond the algorithms and solutions provided, at many a times, beyond the thinking of a biostatisticians.

We found that starting with the questions based on some basic concepts, it is possible to find a starting point in bigger and complex picture of clinical programming, and identify major milestones where clinical programmers have significant roles to play. In this paper we will restrict our discussions around the following concepts only.

- Objectives of clinical trials
- Phases in a clinical trial
- Cross - functional interaction - role of the sponsor, investigator or clinician, statistician, data managers and medical writers.
- Patient enrollment, concepts of center/site, signing of consent, inclusion/exclusion criteria and randomization.
- Interdependency of the study documents and data
- Concepts from programming prospective
- Regulatory submissions

OBJECTIVES OF CLINICAL TRIALS

Unlike most applications that are programmed by the pure SAS programmers, clinical programming requires a programmer to understand the objective of the clinical trial study, its phases, the study design etc.

Clinical trials are intended to find answers to a research question by means of generating data for proving or disproving a hypothesis. The first step in initiating a clinical trial is writing the protocol. This is either done within the sponsor or the contract research organization (CRO).

In a clinical trial once the phase of the trial is decided the respective documents (Protocol, Case Report Form, Statistical analysis plan etc.) are then created accordingly. In every trial, the protocol is prepared with the help of the clinician and the biostatistician. After an investigator receives the protocol, he/she will determine the feasibility of conducting the clinical trial. Some of the very basic questions that are considered typically are:

- Does the investigator have the necessary patient population?
- Is the protocol well designed and clearly understood by the study team at the trial site(s)?
- Are the study timelines reasonable?

UNDERSTANDING PHASES IN A CLINICAL TRIAL

Any clinical programmer needs to understand all the phases involved in clinical trial, their relative importance, , characteristics that differentiate them, types of inferences derived in each of the phases and corresponding statistical hypothesis established for the same purpose, etc.

Phase I trials:

Phase I trials are usually trials with smaller sample sizes ranging from 10 to 30. Typically, healthy volunteers are recruited to do the safety testing of a new drug unless the new drug is for life-threatening diseases such as cancer, HIV etc. where actual patients are put on to trials. Phase I trials are important because they are the first step, after successful animal trials, in finding a potentially new treatments for the disease under consideration.
The aim is, primarily, to

- Determine the safe dose range
- Discover the side effects
- Estimate tolerability
- Characterize pharmacokinetic and pharmacodynamics activities

In Phase I trials the statistical programmer would generally perform programming of PK/PD Analysis and safety parameters using descriptive statistics, binary data analysis etc., and in some cases survival data analysis as well. Since this phase only involves safety profiling, largely the programming efforts are not that intense as they would be for in other phases.

Phase II trials:

Phase II trials are required to assess the efficacy of the drug in addition to enhanced safety analysis. The sample sizes for such trials are often larger than phase I trials, and go up to 100 patients' recruitment. In a phase II trial, a new treatment is compared with another treatment already in use, or with a dummy drug (placebo). If the results of phase II trials show that a new treatment may be as good as existing treatment, or better, it then moves into phase III. In case it is neither of the case, often non-inferiority testing of the drug takes places.
These trials aim to find out

- If the new treatment is expected to work well enough to test in a larger phase III trial
- More profiling of the side effects and management of the same
- More detailed investigation about the best dose to use

Phase III trials:

Phase III trials are conducted to compare new treatments with the currently available best treatment in the market. They are generally conducted to test the effectiveness of the drug in a wider patient population to accommodate patient population variation. It therefore involves many more patients than phase I or II. These trials may compare

- A completely new treatment with the standard treatment
- Different doses or ways of giving a standard treatment

As Phase II and III trials are generally conducted to assess the effectiveness of the drug, along with safety assessment they also perform efficacy analysis. Here to analyze the efficacy endpoints the programmer would need to use more complex statistical procedure like linear mixed modeling, non-parametric testing, etc. Although the safety analysis would be similar to that in phase I, the larger sample size increases the intensity of programming efforts as subgroup analysis also gets introduced. Here subgroups could be demographic characteristics, lab parameter, prior exposure to drug/treatment, efficacy parameters etc.

Phase IV trials:

Phase IV trials are not mandatory. However, these are performed after a drug has been established to work well through first three phases of the trials and has been approved for use. These are commonly known as post-marketing trials. The aims for conducting phase IV trials are to find out

- More details about the side effects and safety of the drug
- the long term risks and benefits
- efficacy of the drug in the special populations
- publication

In addition to safety and efficacy analysis, Phase IV trials would have more of comparative analysis involved.

UNDERSTANDING CROSS-FUNCTIONAL INTERACTION

Role of the sponsor, investigator or clinician, statistician, data managers and medical writers

Clinical/ Statistical /SAS programmers implement the analysis methods on the collected data and provide the study summary tables, data listing and graphs to the statisticians, medical writers and clinicians to use in writing the clinical study report. Clinical programmers work closely with clinicians, biostatisticians and data managers and provide important interface that helps provide the linkage between CRF data and the intended analysis.

While data is being re-entered and cleaned, statistical clinical programmers may begin to write SAS programs to generate report tables, listings, and graphs according to the statistical analysis plan (SAP) and the tables shells designed by the study biostatistician. Drafts of tables, listings and graphs· are reviewed by the statisticians and more data clarification items may also be generated in this process.

Since a programmer often needs to interact with different functions, the work becomes more meaningful if they have a brief understanding of the roles performed by each of the study team member below mentioned.

Sponsor:

A sponsor is defined as “a person/entity who takes the responsibility for and initiates a clinical investigation. The sponsor may be an individual or a pharmaceutical company”.

Investigator/Clinician:

A clinical investigator involved in a clinical trial is responsible for ensuring that an investigation is conducted according to the signed investigator statement, the investigational plan, and applicable regulations; for protecting the rights, safety, and welfare of patients under the investigator's care; and for the control of drugs under investigation. The Clinical Investigator must also meet requirements set forth by the FDA (Food and Drug Administration), EMEA (European Medicines Agency) or other regulatory bodies. It is the investigator’s responsibility to see that the entire trial is conducted as required and defined in the protocol or SAP.

Biostatistician:

The role of a Biostatistician is to work with the clinician and a clinical programmer with the key responsibilities being:

- Help with the statistical design of the clinical trial study.
- Help with the formulation of hypotheses.
- Determine optimal sample size required to guarantee the level of precision.
- Work with the team preparing waiver grants.
- Ensure the methods and statistical analysis sections address each specific aim.
- Calculate the appropriate sample size for the study to address the specific aims.
- Determine the data variables to be measured and how they relate to the study objectives.
- Develop the data collection process.
- Provide methods for analysing the data.
- Interpret the results.

The programming requirements and activities are based upon the methods and statistical analysis provided by the biostatistician. Unlike other functions biostatisticians don’t have a standardized approach. Every biostatistician can have a different perspective and analysis requirements for the same objective. Hence a change of lead biostatistician during the course of a trial often leads to change in the analysis requirement in terms of the reports required to be generated and the analysis methods to be used to assess the endpoints. This often poses a challenge to the programmers as they have to accommodate these changes in requirement without extension of timelines.

Data Management:

The clinical data manager plays a key role in the setup and conduct of a clinical trial. Each of the below mentioned CDM activities are assessed for quality at regular intervals during a trial

- Case Report Form (CRF) designing
- CRF annotation
- Database design
- Data entry
- Data validation and discrepancy management
- Medical coding
- Data extraction and database locking

It is the responsibility of the programmer to notify the data management team either directly or via the biostatistician of any data issue or data discrepancy noticed in due course of the programming activities. Programmers may also need to interact with the clinical data programmers at the time of snapshot creation or data base lock.

Medical writer:

The outputs (tables, listings and figures) created by the statistical SAS programmers are used by the medical writers to produce scientific documentation. A medical writer is a specialized scientific writer who is typically not someone who performs the research. A medical writer, working with doctors, scientists, and other subject matter experts, creates documents that effectively and clearly describe research results, product use, and other medical information. The medical writer also ensures that their documents comply with regulatory, journal, or other guidelines in terms of content, format, and structure.

While preparing the scientific documents the medical writers may approach the programmers either directly or via the biostatistician to get clarity on certain outputs or in case of any additional supporting listings are required.

UNDERSTANDING CONCEPTS OF SITE/CENTER, PATIENT ENROLMENT, SIGNING OF CONSENT, INCLUSION/EXCLUSION CRITERIA'S AND RANDOMIZATION

Patient enrollment involves the identification of prospective patients and is the most time-consuming operational aspect of the clinical trial. It is estimated that, in general, the leading cause of missed clinical trial deadlines is patient recruitment, taking up to 30 percent of the clinical timeline.

The Site or Center is the hospital location or country in which the patients are recruited for the trial. Site/Center selection involves choosing the optimal recruiting sites for study participation based on the required operational management, technical and patient recruitment potential of the site based on the past experience or clinical experience capabilities of participating site.

Once a site is identified, the interested patients are then called to the site for their consent and to enroll in the study. Informed consent is a legal procedure to ensure that a patient, client, and research participants are aware of all the potential risks and costs involved in a treatment or procedure.

After the consent is recorded from the patient, he/she is screened on the medical or social standards determining whether a person may or may not be allowed to enter a clinical trial. These criteria are based on such factors as age, gender, the type and stage of a disease, previous treatment history, and other medical conditions. 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.

Post the inclusion/exclusion criteria qualification, the recruitment is started and the patient is randomized to a particular group of treatment. Randomization is the process of assigning clinical trial participants to treatment groups. Randomization gives each participant a known (usually equal) chance of being assigned to any of the groups. Successful randomization requires that group assignment cannot be predicted in advance.

UNDERSTANDING INTERDEPENDENCY OF THE STUDY DOCUMENTS AND DATA

The CRF, Protocol and the SAP together form the backbone of a clinical trial.

The protocol is the detailed document that contains all information about conducting the trial, such as inclusion/exclusion criteria, drug dispensing, study design, protocol deviations etc.

Based on the protocol the CRF is designed to collect the required data that will be needed for the analysis.

Once patient enrollment begins, the clinical data manager ensures that data is collected, validated, complete and consistent as per the CRF. The corresponding data collected through the CRF may be in a paper or electronic form which is then converted into the required data form by the clinical data programmers. This data is then checked and cleaned by the data managers to get in the right form to the statistical programmers who then develop the respective statistical reports to prove the objective of the trial. The data manager also liaises with other data providers (e.g. a central laboratory processing blood samples collected). At the completion of the clinical trial, the clinical data manager ensures that all data expected to be captured has been accounted for and that all data management activities are complete. At this stage, the data is declared final (terminology varies but common descriptions are Database Lock and Database Freeze) and the clinical data manager transfers data for statistical analysis.

Similarly, the SAP is also developed based on the protocol and generally contains:

- Study design and objectives
- Endpoints

- Sample size
- Randomization
- Study populations
- Statistical Analysis
    1. Demographics
    2. Primary endpoint: The main objective or endpoint of the study.
    3. Secondary endpoint(s)
    4. Safety
- Table, Figure & Listing (TFL) shells

SOME CONCEPTS FROM A PROGRAMMING PERSPECTIVE

The data that is collected through the CRF is known as the raw data. Using this raw data the required analysis datasets are created based on the study endpoints and population sets stated in SAP.

Population sets:

The most commonly used population sets are FAS (Full analysis set), ITT (Intent-to-Treat), SAF (Safety) and PPS (Per Protocol Set). These can be briefly defined as below.

FAS set: Population that contains all the treated patients.

ITT set: Patients who have at least one post baseline record.

PPS set: Patients that have at least one post baseline record and have no protocol violations.

The actual population sets for a study are defined based on the analysis objectives.

Endpoints:

Primary endpoints measure outcomes that will answer the primary (or most important) question being asked by a trial, such as whether a new treatment is better at preventing disease-related deaths than the standard therapy.

Secondary endpoints ask other relevant questions about the same study; for example, whether there is also a reduction in disease measures other than death, or whether the new treatment reduces the overall cost of treating patients.

The tables, listings and graphs are based on the endpoints and population sets that are defined in the SAP.

Being able to understand and connect data and/or tables is one of the distinguishing factors between a pure SAS programmer and a clinical SAS programmer. For instance,

- If the protocol inclusion criterion states to include only patients with age >18 years, there should be no patients with age <=18. A clinical programmer should be able to pin-point such discrepancies in the data rather than waiting for the biostatistician to notice.
- The randomized set comprises of all eligible patients, and the other population sets are derived from the randomized set. Hence, it will be quite obvious for the clinical programmer to observe on the tables that the numbers shown under these derived sets are at most the total number of patients under the randomized set.

The statistical programmer minimally knows the quintessential SAS procedure to compute the descriptive and basic inferential statistics since they recur. In addition to knowing the procedure syntax, understanding the data that will help a Statistical programmer to create meaningful outputs. For instance,

- If the requirement is to compute the number of unique patients in a treatment group, then passing data having multiple records per patient through PROC FREQ would result into incorrect output.
- A computed p-value of >1 or <0 is an indication that an incorrect procedures or data were passed.
- Passing numeric or character variables through PROC FREQ can significantly alter the results as the sort order of alphabets and numbers would differ.

We strongly recommend that clinical programmers learn as much as possible about the therapeutic area:

- What is the disease, and how are patients affected by it?
- What is the medical need? Is the goal to make a drug to cure the disease, prolong life, cause fewer side effects, or just improve the patients' quality of life?
- How does the treatment under study work? What is the mechanism of action?

REGULATORY SUBMISSIONS

One of the critical and differentiating factor for clinical programmers is they are responsible to produce outputs that involve various kinds of regulatory submissions. Below listed are some of the common ones that clinical programmer needs to acquaint oneself with.

Interim analysis:

An interim analysis is any assessment of data done during the patient enrollment or follow-up stages of a trial for the purpose of assessing center performance, the quality of the data collected, or treatment effects. Interim analysis is also called “data-dependent stopping” or “early stopping”. Interim analyses are most often used to find convincing enough evidence to say that there is a significance large treatment difference, and that the difference is convincing enough to terminate the trial at a point earlier than planned at first.
Interim analysis is also used to possibly reduce the expected number of patients and to shorten the follow-up time needed to make a conclusion.

A Data Monitoring Committee (DMC)

DMC, also called a Data and Safety Monitoring Board (DSMB), is an independent group of experts who monitor patient safety and treatment efficacy data while a clinical trial is ongoing.

The DMC is a group (typically 3 to 7 members) who are independent of the company sponsoring the trial. At least one DMC member will be a biostatistician. Clinicians knowledgeable about the disease indication should be represented. The DMC will convene at predetermined intervals (three to six months typically) and review unblinded results, i.e. results split by experimental and control arms. The DMC has the power to recommend termination of the study based on the evaluation of these results. There are typically three reasons a DMC might recommend termination of the study: safety concerns, outstanding benefit, and futility.

ISS/ISE (Integrated Summary of Safety/ Integrated Summary of Efficacy):

The ISS contains all the clinical trial data for a compound, collected from normal volunteers (from phase 1 study) and patients (all other studies). The ISS will contain details such as the extent of the exposure to study drugs by the patients, different characteristics of patients enrolled in the study, deaths –occurring during the study, counts of drop-outs from the study, potential serious adverse events (SAE), other adverse events (AE) and clinical laboratory results.

The ISS is mandatory for filing a new drug application (NDA). The safety data from different trials is pooled together and then used to identify rare AE’s.

The ISE contains a description of entire efficacy database, demographics and baseline characteristics.

These submissions are performed on trials from phase II - phase IV. Phase I studies are conducted to identify only the safety of the drug and hence do not contribute to the ISE.

CONCLUSION

The clinical domain is a very vast domain and it is difficult to have a complete expertise. We have tried to touch base on some basic topics in this paper that can be a starting point for understanding the clinical domain.

One can be a successful SAS programmer purely on the basis of the SAS expertise without any reference to clinical domain. However, it is beneficial to have basic understanding of some, if not all topics discussed in this paper to begin a career of a successful clinical programmer. A SAS programmer with the Clinical knowledge will always have an competitive edge over a purely SAS programmer as he/she will be in a position to take decisions while programming. He/she would be more independent and able to contribute more in terms of understanding and developing more meaningful reports with higher quality.

Very often as interviewers we tend to look at only the SAS skills of a programmer. However from our experience we have seen that as much as it is important for a SAS programmer to have a basic understanding of the clinical aspects for career development and growth, it is equally important for an interviewer to assess clinical knowledge in addition to SAS skills in a candidate. A person with both SAS and clinical knowledge would not only be able to do more justice to the role by generating more reliable and meaningful outputs, but this combination of knowledge would significantly eases the communication channels across functions.

REFERENCES

- Kirk Paul Lafler and Charles Edwin Shipp (2008): What's Hot, What's Not - Skills for SAS® Professionals. http://www.lexjansen.com/wuss/2008/mcp/mcp05.pdf
- Sandra Minjoe and Mario Widel (2011): Success As a Pharmaceutical Statistical Programmer http://www.lexjansen.com/pharmasug/2011/ib/pharmasug-2011-ib01.pdf
- Ming Wang: THE ROLE OF SAS PROGRAMMERS IN CLINICAL TRIAL DATA ANALYSIS http://www.lexjansen.com/nesug/nesug96/NESUG96061.pdf
- More Information http://www.wikipedia.org

ACKNOWLEDGMENTS

We would like to thank Dr. Prashant Kirkire, Country Head, India at inVentiv Health Care, for his valuable guidance and our peers and colleagues for carefully reviewing the paper with comments and suggestions.

CONTACT INFORMATION

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

Name: Gayatri Karkera
Enterprise: inVentiv Health Clinical
Address: Ground Floor, Marwah Centre, Krishanlal Marwah Marg, Andheri (E)
City, State ZIP: Mumbai-400072, Maharashtra, India
Work Phone: +91-22-4095-7364
E-mail: gayatri.karkera@inventivhealth.com ; gayatri.karkera@gmail.com

Name: Neha Mohan
Enterprise: inVentiv Health Clinical
Address: Ground Floor, Marwah Centre, Krishanlal Marwah Marg, Andheri (E)
City, State ZIP: Mumbai-400072, Maharashtra, India
Work Phone: +91-22-4095-7365
E-mail: neha.mohan@inventivhealth.com ; neha.nm@gmail.com