Clinical Trials Terminology for SAS Programmers

Sy Truong, Meta-Xceed, Inc, Fremont, CA

ABSTRACT

The drug development process is a clinical process that has its own language. It is not required that SAS programmers function as a MD or a regulatory expert, but working knowledge of the terminology is important to be effective. This paper will walk through the drug development process from discovery to Phase IV. It will explain a wide range of acronyms such as IND, NDA, GCP and MedDRA. It will also describe some of the terminologies used within the process of clinical trials as a drug is developed and submitted to the FDA. This will give SAS programmers a larger perspective and context to their work during the analysis and reporting of clinical trials data.

INTRODUCTION

This paper will tell a fictitious story about a college graduate named James who is starting a new position at a pharmaceutical company. Each new term James encounters is presented in ***bold*** and ***italicized*** for emphasis. As he enters a new professional world, he meets many people and learns new processes that are filled with unfamiliar vocabulary and acronyms. As James settles into his new job as a SAS programmer, he learns the meaning of these terminologies and becomes more productive in his work.

GETTING THE JOB

After an enjoyable summer of R&R following his graduation from the University of California, James browses through the want ads to confront the adult world of employment. James has a vague notion of what a ***Pharmaceutical*** company does in that it performs research and development of drugs. He sees ads for ***Biotechnology*** companies which is a general term used to explain a technique of using living organisms within biological systems to develop micro-organisms for a particular purpose. The end products from ***Biotech*** and ***Pharmaceutical*** companies are usually drugs. These companies are sometimes referred to as the ***Biopharmaceutical*** industry.

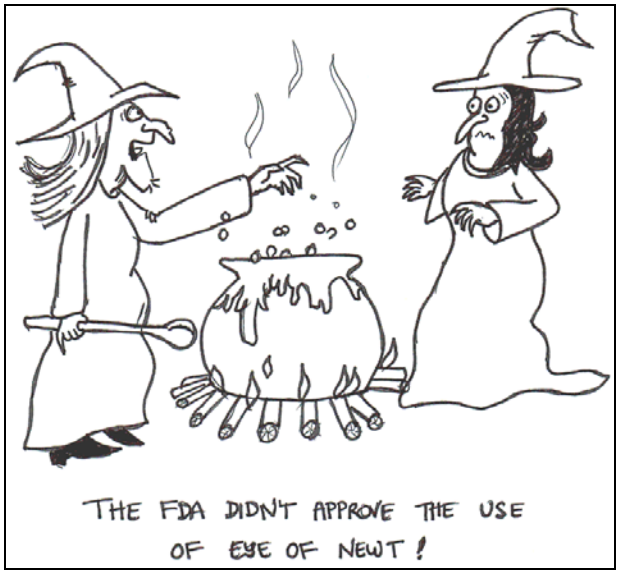
James was successful at acquiring a job as a ***Statistical Programmer*** which requires him to program using the SAS language to analyze clinical data and produce reports for the ***FDA***. He was familiar with the ***Food and Drug Administration*** but learned that they set many of the regulations that affect his job. During his search, he also saw other job titles including: ***Bioanalyst***, ***Clinical Data Analyst***, ***Statistical Programmer Analyst*** and ***SAS Programmer***. It turns out that different companies have different names for the same job.

STARTING THE JOB

James started his first day in a small cubicle at Genenco, and his only interaction was with Barbara, a ***Biostatistician,*** who was his boss. James' degree included many statistical courses but a PhD in statistics or ***biostatistics***, such as the one which Barbara had, is required to manage the statistics department at Genenco. After setting James up with a computer account and a desktop that was faster than anything he had ever laid his hands on, Barbara delivered a big binder which contained a ***Protocol*** for his first ***clinical study***. It was a monster document that must have been at least a hundred pages. The ***protocol*** outlined all the procedures and contained detailed plans of the study. It had the ***study design***, statistical methodology, and acted as a road map for those involved in conducting the study. James' first task was to read and understand what the study was all about. He was new to clinical trials and was just learning about the

concept of a ***controlled experiment.*** The protocol noted that the clinical trial had patients grouped into different groups such as those in the ***placebo controlled group*** which had no active drug. This is how comparisons are made within the ***controlled clinical trial.***

By lunch time, James was able to read through parts of the protocol but there were many parts which he did not understand. James recalled that during his day long interview, he met Cindy and Ralph. Cindy was a ***Clinical Research Associate*** (***CRA***) who had a strong clinical background since she was a ***Registered Nurse*** (***RN***). After several emails and phone calls, James realized that Cindy's job required her to travel a lot. She was currently visiting a ***CRO*** **(*Contract Research Organization)*** which Genenco contracts with to handle all ***data management*** aspects of several studies. The ***CRO*** was installing a new ***Electronic Data Capture*** (***EDC)*** system which was intended to give Cindy faster access to the clinical information. This was also sometimes referred to as a ***computer assisted data collection*** system.



Since James was unsuccessful at contacting Cindy, he got in touch with Ralph who works in ***Regulatory***. Ralph interfaces with the **FDA** and performs internal audits at Genenco to ensure that everyone is doing their job according to ***CFR Part 11*** which is the ***Code of Federal Regulations*** set by the FDA to regulate food, drug, ***biologics*** and device industries. The ***part 11*** specifically deals with the creation and maintenance of electronic records.

James was able to set up a meeting with Ralph later that week to help explain some of the terms within the protocol. The protocol was authored by Irving who was the ***Investigator*** on the study. James had never met Irving so did not work up the courage to contact him. By reviewing the protocol, it described that Irving was a MD, PhD and the author of the ***crucial treatment plan***. Irving collaborated with Paul who was the ***PI*** or ***Principal Investigator*** for this trial. Paul managed the entire team of investigators including Irving.

REGULATORY WOLRD

James realized that he was lucky enough to catch some of Ralph's time. At the meeting, Ralph started with the basics by explaining how the information is collected on patients or ***human subjects*** during the conduct of the study. This information is written down on a ***CRF*** or ***Case Report Form***. These forms collect information such as ***demographic*** and ***adverse events***. The ***demographic*** information is sometimes referred to as ***DEMOG***. The **case report form** contains characteristics of the subject including things such as sex, age and ***medical history***. The ***medical history*** information is collected on its own form separate from the demographic form. The ***Adverse Event CRF***, also known as ***AE***, records ***Side Effects*** or ***Adverse Effects*** from the drug or other treatments. All the information collected is known as ***Source Data***, which include important documents because they contain the ***core*** information required to reconstruct the essential intellectual capital of the study. Ralph continued to explain that Genenco is the ***sponsor*** company who is responsible for the management, financing and conduct of the entire trial.

STUDY DESIGN

James learned that in the current study, the ***subjects*** are ***randomized*** into distinct groups. This means that they are randomly assigned to groups so that each subject has an equal chance to be assigned to the ***placebo control*** or ***active treatment groups***. At the point when they are ***randomized*****,** they are assigned to their drug which is also referred to as a **baseline.** This is important because there are other analyses that measure the **change from baseline** to draw statistical conclusions. The different treatment groups will later be compared to verify for ***statistical significance***. The group that is assigned to the ***placebo control group*** gets treated with an inactive drug. The ***placebo*** is also sometimes referred as the ***sugar pill***, which is an inactive substance designed to look like the drug being tested. This is intended to avoid any psychological affects that a subject may have when taking the drug. In this case, the ***control groups*** are ***blinded*** in the sense that they do not know if the drug that they are taking contains the active ingredient. If the study had only the ***control groups blinded***, it would be classified as a ***single blinded*** study. However, in this case, neither Irving the ***investigator***, nor the subjects knew which group has the ***active treatment***. This study is therefore a ***double blinded*** study. The acronym for ***double blinded*** is ***DB*** which

confused James since he used this to describe databases. The secrecy of a **double blinded** study was a surprise to James since he thought that everyone would know what they are taking, including the people administering the drugs. If all was out in the open, this would be referred to as an ***open-label study***.

James had noticed that there was another study similar to the one he was currently assigned to which had the subject taking the drugs three times a day, also referred to as ***TID.*** This is Latin for "***ter in die***" which means three times a day**.** The ***Pharmacokinetics (PK)*** analysis of that study showed that with that dosing level, there were high levels of toxicity in the subject. This was an analysis of how the body processes the drug as it enters and exits the subject. The current study changed the ***standard treatment*** so subjects now take the drug ***BID***, or twice a day. Part of the reason why subjects were having so many ***serious adverse reactions*** was due to ***adverse drug reactions (ADR)*** in relation to ***concomitant drugs.*** This included other ***OTC*** or ***over the counter drugs*** that they were taking. Another change that was different between James' current study compared to a previous one was how subjects were included into the study in the first place. The change took place on the first ***Case Report Forms*** that a subject filled out, also known as the ***Inclusion and Exclusion Criteria*** form. These contain a list of criteria to evaluate if the patient was suitable for the study. For example, pregnant women were not allowed into the study due to the potential risk to the fetus. Early on during the recruitment of patients, each patient had to fill out an ***informed consent*** form which described all the potential benefits and risks involved. Ralph informed James that Genenco was required to do this due to the many federal and state laws. This concluded their conversation and James thanked Ralph for such an enlightening discussion.

TABLES LISTINGS AND GRAPHS

The topic pertaining to dosing was intriguing to James so he started to work on the ***TLGs (Tables, Listings and Graphs)*** related to ***concomitant drugs.*** The analysis of concomitant drugs was to find out if there were any ***drug interaction*** between the ***active treatment*** and other drugs that the patients were already taking. An exploratory analysis was performed to compare similarities between these drugs to show the ***bioequivalance***. James started by developing SAS programs for the ***CONMED listings,*** which listed the data chronologically by the subject identification number. This was relatively easier compared to the more sophisticated statistical reports involved in generating ***summary tables*** and ***graphs***. One of the challenging aspects of generating these listings involved the translation of drug names from source data into a preferred drug name. The drug name that is collected from the patient and recorded into the source data is also known as the ***trade name***. This is the commercial name for the drug. However, the corresponding ***generic name*** usually refers to its chemical compound. For example, if the patient took Tylenol or Anacin-3, this report will list the corresponding generic name, Acetaminophen. This is an example where drug trade names with the same active ingredient are reported with their ***preferred term*** in order to make sense when they are compared. James had to learn to use a dictionary named ***WHO-DRUG*** which listed all the drug names and how they matched to the generic drug names. This dictionary is managed by the ***World Health Organization***.

James later noticed that other reports on ***adverse events*** had a similar construct. There were multiple ***verbatim*** adverse event terms such as "head ache" and "pain in the head" which mapped to the same ***preferred term***. In this case, he was not using the ***WHO-DRUG*** dictionary, but rather ***Costart***, which was short for ***Coding Symbols for Thesaurus of Adverse Reaction Terms.*** This helped to organize adverse event listings and summary reports. All James had to do was to merge his data with ***Costart*** to acquire the associated preferred terms. It even helped him group the ***adverse event*** terms by ***body systems.*** The ***body system*** is a classification which separates adverse events into distinct groups such as those dealing with the ***cardiovascular system*** and those dealing with the ***nervous system***.

The ***data management*** group was currently going through a migration of all their work from ***Costart*** to a new dictionary named ***MedDRA.*** This is short for ***Med*** (Medical), ***D*** (Dictionary), ***R*** (Regulatory), and ***A*** (Activities). It is one of the more comprehensive databases containing terms collected in other dictionaries including:

Costart (fifth edition), ***Who-ART*** (98:3), ***J-ART*** (1996) and ***HARTS*** (Release 2.2). The dictionary is also constantly being updated with new terms, so it is one of the most comprehensive dictionaries available. There are also more sophisticated levels of classification that go beyond ***body systems*** in ***MedDRA***. Once the transition was complete, all the mapping of adverse event terms would be managed within the data management group. In the meantime, however, James worked on this mapping process and learned more about the ***adverse event coding***.

STATISTICS GEEK

While working on the ***demography summary table***, James realized that there were many statistical concepts which were new to him. He was trying to understand the details of the ***SAP,*** which was the ***Statistical Analysis Plan*** that Barbara had so carefully written out for him. It was beautifully organized with a ***TOC (Table of Contents)*** along with ***mockups*** of the tables and listing describing the layout of how they should look. It had details pertaining to the demographic listing capturing the ***baseline characteristics*** at the point of ***randomization.*** She also had text expanding on statistics, pointing out that he should apply an ***ANOVA,*** which was an ***analysis of variable***. James' statistical skills were rusty so he discussed the ***SAP*** with Barbara for clarification. She explained that she wanted the ***ANOVA*** to compare the two treatment groups within the demographic summary. This was to show the differing effects of the drugs which were to be adjusted by race, gender and other ***grouping variables***. She also wanted him to use the ***chi-squared test*** in his summary tables to verify the equality of proportions between male and female. She hoped to use this to show a 95% ***confidence interval*** in the difference between patients among the drug groups. James understood most of what she was trying to say but he made a note to look up the ***Pearson's Chi-square test*** which was beyond him at this point. Barbara also had to elaborate on the meaning of a ***confidence interval*** which gives an estimated range of values being calculated from the sample of patient data that is currently in the study.

Barbara continued to explain that the ***adverse events*** report summary tables showed a ***clinical significance*** between the different treatment groups. Many of the reports had to display ***p-values*** to signify their ***statistical significance***. The ***p-values*** were displayed for certain statistical comparisons to show the probability of accomplishing a result, if there is no difference between the ***stratified*** groups within the report. This lack of difference between the groups in the reports was also referred to as the ***null hypothesis.*** James was beginning to realize that underneath that polished appearance, Barbara was a real geek.

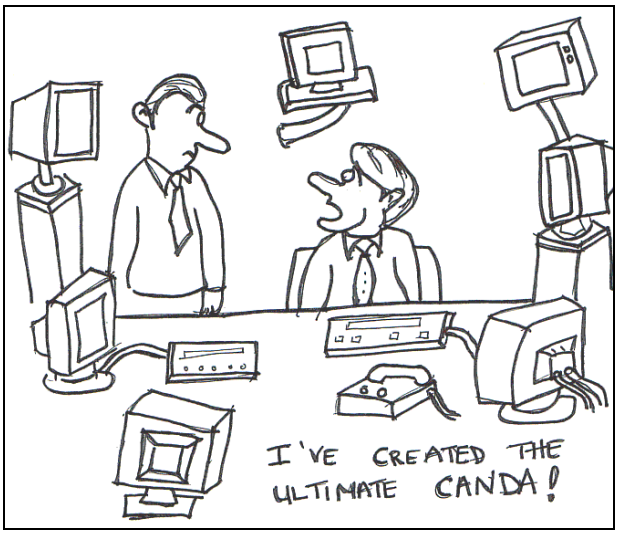
Accompanying many of the summary tables, James had to also produce ***graphs.*** In one of the ***survival analysis,*** Barbara requested a graph including a ***Kaplan-Meier*** curve showing the probability of survival. According to Barbara's request, he also created some graphs that had a ***normal distribution,*** which displayed the ***distribution*** of values in a bell shaped graph. Barbara pointed out that the curves varied in their curvature, peaking higher on some, while narrower on others. She referred to the different measurements of their curvature as ***kurtosis.*** She also referenced many of the ***univariate*** analyses**,** which dealt with one variable. For example, when they looked at the demographic characteristic for height, it was referring to just one variable. This was therefore referred to as a ***univariate analysis.*** When they looked at the analysis on the patient's overall size, it also took into consideration other variables such as weight, so this became a ***biviriate analysis***. In general, when it involves more than one variable, it is known as a ***multivariate analysis.*** Variables within the analysis were classified as either ***continuous*** or ***categorical. Continuous variables*** are things like age or weight. They are not usually limited to specific values and are numeric values. On the other hand, ***categorical variable*** are those such as race or sex. These variables usually have fixed categories and appear with check boxes in the case report form. Depending on the variables, they will affect the types of analysis and statistical models used.

All these new statistical ***models*** were new to James and he did not wanted to destroy any of Barbara's analysis. Therefore, he requested if he could perform the same analysis upon an older ***pilot study*** for practice. The pilot study was smaller in design compared to the current

larger ***Phase III*** trial study. The current ***Phase III*** study was large in scale containing thousands of patients enrolled from ***multi-centers*** with the main objective of showing that the drug had ***efficacy.*** James was not quite ready for this so he went back and reviewed the ***Phase II*** studies designed to find the best level of dosing, while maintaining the safety of patients. He also reviewed the ***Phase I*** studies which he had difficulty getting his hands on since they were analyzed by a ***CRO*** located out of state. The ***Phase I*** studies were the first ones where the drug was tested on humans. There were only 15 patients and they were all healthy without the condition that the drug was meant to treat. It was not intended to show ***efficacy.*** Rather, it was a study to show the ***toxicity*** of the drug in the subjects. This studied the ***pharmacokinetics*** of how their bodies reacted to the drug.

ELECTRONIC SUBMISSION

As James' reports were shaping up, he had a meeting with Eric from the ***ESUB (Electronic Submission)*** group. Eric wanted to make sure that the reports which James was producing had the proper standards suggested from ***CDISC (Clinical Data Interchange Standards*** **Consortium).** Eric was very involved with ***CDISC*** and attended regular teleconferences within a subgroup named ***ADaM (Analysis Data Model Team).*** The standards dealt with ***analysis datasets*** including things such as what type of variables were to be included in a particular ***domain*** of data, such as demography. The guidelines also suggested how the report should look, including examples such as font and margin sizes. Eric described that when he first started, they were working on a ***CANDA (Computer Assisted New Drug Application).*** Those were the days before ***CDISC*** was even formed. Eric was part of a team within Genenco that would put together an ***electronic submission*** and organize it into a computer system with all the tools needed, including instructions on how the reviewer should use it. It was an exciting time since everything was new. He remembered attending ***DIA (Drug Information Association)*** conferences showing off the coolest ***SAS*** based push button system which they had developed. Even though the system was easy to use, the ***FDA*** reviewers had a different ***CANDA*** package for each ***sponsored company***. It became too difficult for them to learn a new set of tools for each company, so they later decided to standardize on using ***Adobe Acrobat Viewer*** and started to request ***PDF (Portable Document Format).*** This simplified the ***ESUB*** process since there was no longer hardware or custom software involved. However, Eric reminisced on those days when so much energy was spent on those ***CANDAs.*** Eric also recalled when Genenco was developing more than just drugs and also had ***medical devices*** as part of their ***portfolio.*** They did not have as many drugs in their ***pipeline***, so they were diversifying and developing therapies in many ***therapeutic areas.*** At that time, Eric worked with a division of the ***FDA*** named ***CBER (Center for Biologics Evaluation and Research)***. At that time, this was a separate organization from ***CDER (Center for Drug Evaluation and Research)***. When they had a ***medical device*** or ***biological product***, they worked with ***CBER.*** When they had a drug, they worked with ***CDER***. The requirements were different between the organizations. For example, for human drugs, ***CDER*** would review a ***New Drug Application (NDA)*** for approval. For a biological product or ***biologics***, they would evaluate a ***Biologics License Application (BLA),*** which is a combination of ***Product license application (PLA)*** and ***Establishment License Application (ELA).*** Things are changing now since the two organizations ***CBER*** and ***CDER*** are coming together as one organization.



CONCLUSION

It has been three months since James ventured into the world of SAS programming within the biotech company, Genenco. This was an eye opening experience for him as he prairie dogs from his small cubical, realizing that he has interacted with a lot of fascinating and knowledgeable co-workers. He feels that he is finally settling in now and getting to know the language that is spoken among this group of professionals. He is being inducted into an exclusive cult group, and it is pretty exciting.

REFERENCES

Dictionary for Clinical Trials
by Simon Day

John Wiley & Son, LTD 2002

CDER Acronym List

http://www.fda.gov/cder/handbook/acronym.htm

Thomson Center Watch Glossary

http://www.centerwatch.com/patient/glossary.html

ABOUT THE AUTHOR

Sy Truong is a Systems Developer for Meta-Xceed, Inc. They may be contacted at:

Sy Truong

48501 Warm Springs Blvd. Ste 117

Fremont, CA 94539

(510) 226-1209

sy.truong@meta-x.com