

CLINICAL TRIALS INTERVIEW QUESTIONS

1. Describe the phases of clinical trials?

Ans:- These are the following four phases of the clinical trials:

Phase 1: Test a new drug or treatment to a small group of people (20-80) to evaluate its safety.

Phase 2: The experimental drug or treatment is given to a large group of people (100-300) to see that the drug is effective or not for that treatment.

Phase 3: The experimental drug or treatment is given to a large group of people (1000-3000) to see its effectiveness, monitor side effects and compare it to commonly used treatments.

Phase 4: The 4 phase study includes the post marketing studies including the drug's risk, benefits etc.

2. Describe the validation procedure? How would you perform the validation for TLG as well as analysis data set?

Ans:- Validation procedure is used to check the output of the SAS program generated by the source programmer. In this process validator write the program and generate the output. If this output is same as the output generated by the SAS programmer's output then the program is considered to be valid. We can perform this validation for TLG by checking the output manually and for analysis data set it can be done using PROC COMPARE.

3. How would you perform the validation for the listing, which has 400 pages?

Ans:- It is not possible to perform the validation for the listing having 400 pages manually. To do this, we convert the listing in data sets by using PROC RTF and then after that we can compare it by using PROC COMPARE.

4. Can you use PROC COMPARE to validate listings? Why?

Ans:- Yes, we can use PROC COMPARE to validate the listing because if there are many entries (pages) in the listings then it is not possible to check them manually. So in this condition we use PROC COMPARE to validate the listings.

5. How would you generate tables, listings and graphs?

Ans:- We can generate the listings by using the PROC REPORT. Similarly we can create the tables by using PROC FREQ, PROC MEANS, and PROC TRANSPOSE and PROC REPORT. We would generate graph, using proc Gplot etc.

6. How many tables can you create in a day?

Ans:- Actually it depends on the complexity of the tables if there are same type of tables then, we can create 1-2-3 tables in a day.

7. What are all the PROCS have you used in your experience?

Ans:- I have used many procedures like proc report, proc sort, proc format etc. I have used proc report to generate the list report, in this procedure I have used subjid as order variable and trt_grp, sbd, dbd as display variables.

8. Describe the data sets you have come across in your life?

Ans:- I have worked with demographic, adverse event , laboratory, analysis and other data sets.

9. How would you submit the docs to FDA? Who will submit the docs?

Ans:- We can submit the docs to FDA by e-submission. Docs can be submitted to FDA using

Define.pdf or define.Xml formats. In this doc we have the documentation about macros and program and E-records also. Statistician or project manager will submit this doc to FDA.

10. What are the docs do you submit to FDA?

Ans:- We submit ISS and ISE documents to FDA.

11. Can u share your CDISC experience? What version of CDISC SDTM have you used?

Ans: I have used version 3.1.1 of the CDISC SDTM.

12. Tell me the importance of the SAP?

Ans:- This document contains detailed information regarding study objectives and statistical methods to aid in the production of the Clinical Study Report (CSR) including summary tables, figures, and subject data listings for Protocol. This document also contains documentation of the program variables and algorithms that will be used to generate summary statistics and statistical analysis.

13. Tell me about your project group? To whom you would report/contact?

My project group consisting of six members, a project manager, two statisticians, lead programmer and two programmers.

I usually report to the lead programmer. If I have any problem regarding the programming I would contact the lead programmer.

If I have any doubt in values of variables in raw dataset I would contact the statistician. For example the dataset related to the menopause symptoms in women, if the variable sex having the values like F, M. I would consider it as wrong; in that type of situations I would contact the statistician.

14. Explain SAS documentation.

SAS documentation includes programmer header, comments, titles, footnotes etc. Whatever we type in the program for making the program easily readable, easily understandable are in called as SAS documentation.

15. How would you know whether the program has been modified or not?

I would know the program has been modified or not by seeing the modification history in the program header.

16. Project status meeting?

It is a planetary meeting of all the project managers to discuss about the present Status of the project in hand and discuss new ideas and options in improving the Way it is presently being performed.

17. Describe clin-trial data base and oracle clinical

Clintrial, the market's leading Clinical Data Management System (CDMS). Oracle Clinical or OC is a database management system designed by Oracle to provide data management, data entry and data validation functionalities to Clinical Trials process. 18. Tell me about MEDRA and what version of MEDRA did you use in your project? Medical dictionary of regulatory activities. Version 10

19. Describe SDTM?

CDISC's Study Data Tabulation Model (SDTM) has been developed to standardize what is submitted to the FDA.

20. What is CRT?

Case Report Tabulation, Whenever a pharmaceutical company is submitting an NDA, company has to send the CRT's to the FDA.

21. What is annotated CRF?

Annotated CRF is a CRF(Case report form) in which variable names are written next the spaces provided to the investigator. Annotated CRF serves as a link between the raw data and the questions on the CRF. It is a valuable toll for the programmers and statisticians.

22. What do you know about 21CRF PART 11?

Title 21 CFR Part 11 of the Code of Federal Regulations deals with the FDA guidelines on electronic records and electronic signatures in the United States. Part 11, as it is commonly called, defines the criteria under which electronic records and electronic signatures are considered to be trustworthy, reliable and equivalent to paper records.

23 What are the contents of AE dataset? What is its purpose?

What are the variables in adverse event datasets?The adverse event data set contains the SUBJID, body system of the event, the preferred term for the event, event severity. The purpose of the AE dataset is to give a summary of the adverse event for all the patients in the treatment arms to aid in the inferential safety analysis of the drug.

24 What are the contents of lab data? What is the purpose of data set?

The lab data set contains the SUBJID, week number, and category of lab test, standard units, low normal and high range of the values. The purpose of the lab data set is to obtain the difference in the values of key variables after the administration of drug.

25.How did you do data cleaning? How do you change the values in the data on your own?

I used proc freq and proc univariate to find the discrepancies in the data, which I reported to my manager.

26.Have you created CRT's, if you have, tell me what have you done in that?

Yes I have created patient profile tabulations as the request of my manager and and the statistician. I have used PROC CONTENTS and PROC SQL to create simple patient listing which had all information of a particular patient including age, sex, race etc.

27. Have you created transport files?

Yes, I have created SAS Xport transport files using Proc Copy and data step for the FDA submissions. These are version 5 files. we use the libname engine and the Proc Copy procedure, One dataset in each xport transport format file. For version 5: labels no longer than 40 bytes, variable names 8 bytes, character variables width to 200 bytes. If we violate these constraints

your copy procedure may terminate with constraints, because SAS xport format is in compliance with SAS 5 datasets.

```
Libname sdtm "c:\sdtm_data";Libname dm xport "c:\dm.xpt";  
Proc copy;  
In = sdtm;  
Out = dm;  
Select dm;  
Run;
```

28. How did you do data cleaning? How do you change the values in the data on your own?

I used proc freq and proc univariate to find the discrepancies in the data, which I reported to my manager.

29. Definitions?

CDISC- Clinical data interchange standards consortium.They have different data models, which define clinical data standards for pharmaceutical industry.

SDTM – It defines the data tabulation datasets that are to be sent to the FDA for regulatory submissions.

ADaM – (Analysis data Model)Defines data set definition guidance for creating analysis data sets.

ODM – XML – based data model for allows transfer of XML based data .

Define.xml – for data definition file (define.pdf) which is machine readable.

ICH E3: Guideline, Structure and Content of Clinical Study Reports

ICH E6: Guideline, Good Clinical Practice

ICH E9: Guideline, Statistical Principles for Clinical Trials

Title 21 Part 312.32: Investigational New Drug Application

30. Have you ever done any Edit check programs in your project, if you have, tell me what do you know about edit check programs?

Yes I have done edit check programs .Edit check programs – Data validation.

1.Data Validation – proc means, proc univariate, proc freq.Data Cleaning – finding errors.

2. Checking for invalid character values. Proc freq data = patients; Tables gender dx ae / nocum nopercnt; Run; Which gives frequency counts of unique character values.
3. Proc print with where statement to list invalid data values. [systolic blood pressure - 80 to 100][diastolic blood pressure - 60 to 120]
4. Proc means, univariate and tabulate to look for outliers. Proc means - min, max, n and mean. Proc univariate - five highest and lowest values [stem leaf plots and box plots]
5. PROC FORMAT - range checking
6. Data Analysis - set, merge, update, keep, drop in data step.
7. Create datasets - PROC IMPORT and data step from flat files.
8. Extract data - LIBNAME. 9. SAS/STAT - PROC ANOVA, PROC REG.
10. Duplicate Data - PROC SORT Nodupkey or NoduplicateNodupkey - only checks for duplicates in BYNoduplicate - checks entire observation (matches all variables) For getting duplicate observations first sort BY nodupkey and merge it back to the original dataset and keep only records in original and sorted.
11. For creating analysis datasets from the raw data sets I used the PROC FORMAT, and rename and length statements to make changes and finally make a analysis data set.

31. What is Verification?

The purpose of the verification is to ensure the accuracy of the final tables and the quality of SAS programs that generated the final tables. According to the instructions SOP and the SAP I selected the subset of the final summary tables for verification.

E.g Adverse event table, baseline and demographic characteristics table. The verification results were verified against with the original final tables and all discrepancies if existed were documented.

32. What is Program Validation?

Its same as macro validation except here we have to validate the programs i.e according to the SOP I had to first determine what the program is supposed to do, see if they work as they are supposed to work and create a validation document mentioning if the program works properly and set the status as pass or fail. Pass the input parameters to the program and check the log for errors.

33. What do you know about ISS and ISE, have you ever produced these reports?

ISS (Integrated summary of safety):Integrates safety information from all sources (animal, clinical pharmacology, controlled and uncontrolled studies, epidemiologic data). "ISS is, in part, simply a summation of data from individual studies and, in part, a new analysis that goes beyond what can be done with individual studies."ISE (Integrated Summary of efficacy)ISS & ISE are critical components of the safety and effectiveness submission and expected to be submitted in the application in accordance with regulation. FDA's guidance Format and Content of Clinical and Statistical Sections of Application gives advice on how to construct these summaries. Note that, despite the name, these are integrated analyses of all relevant data, not summaries.

34. Explain the process and how to do Data Validation?

I have done data validation and data cleaning to check if the data values are correct or if they conform to the standard set of rules.A very simple approach to identifying invalid character values in this file is to use PROC FREQ to list all the unique values of these variables. This gives us the total number of invalid observations. After identifying the invalid data ...we have to locate the observation so that we can report to the manager the particular patient number.Invalid data can be located using the data `_null_` programming.

Following is e.g

```
DATA _NULL_;
```

```
INFILE "C:PATIENTS,TXT" PAD;FILE PRINT; ***SEND OUTPUT TO THE  
OUTPUT WINDOW;
```

```
TITLE "LISTING OF INVALID DATA";
```

```
***NOTE: WE WILL ONLY INPUT THOSEVARIABLES OF INTEREST;INPUT  
@1 PATNO $3.@4 GENDER $1.@24 DX $3.@27 AE $1.;
```

```
***CHECK GENDER;IF GENDER NOT IN ('F','M',' ') THEN PUT PATNO=  
GENDER=;
```

```
***CHECK DX;
```

```
IF VERIFY(DX,' 0123456789') NE 0  
THEN PUT PATNO= DX=;
```

```
***CHECK AE;
```

```
IF AE NOT IN ('0','1',' ') THEN PUT PATNO= AE=;  
RUN;
```

For data validation of numeric values like out of range or missing values I used proc print with a where statement.

```
PROC PRINT DATA=CLEAN.PATIENTS;  
WHERE HR NOT BETWEEN 40 AND 100 AND  
HR IS NOT MISSING OR  
SBP NOT BETWEEN 80 AND 200 AND  
SBP IS NOT MISSING OR  
DBP NOT BETWEEN 60 AND 120 AND  
DBP IS NOT MISSING;TITLE "OUT-OF-RANGE VALUES FOR  
NUMERICVARIABLES";  
ID PATNO;  
VAR HR SBP DBP;  
RUN;
```

If we have a range of numeric values '001' – '999' then we can first use user defined format and then use proc freq to determine the invalid values.

```
PROC FORMAT;  
VALUE $GENDER 'F','M' = 'VALID' ' ' = 'MISSING'OTHER = 'MISCODED';  
VALUE $DX '001' - '999' = 'VALID' ' ' = 'MISSING'OTHER = 'MISCODED';  
VALUE $AE '0','1' = 'VALID' ' ' = 'MISSING'OTHER = 'MISCODED';  
RUN;
```

One of the simplest ways to check for invalid numeric values is to run either PROC MEANS or PROC UNIVARIATE. We can use the N and NMISS options in the Proc Means to check for missing and invalid data. Default (n nmiss mean min max stddev). The main advantage of using PROC UNIVARIATE (default n mean std skewness kurtosis) is that we get the extreme values i.e lowest and highest 5 values which we can see for data errors. If u want to see the patid for these particular observationsstate and ID patno statement in the univariate procedure.

35. Roles and responsibilities?

Programmer:

Develop programming for report formats (ISS & ISE shell) required by the regulatory authorities. Update ISS/ISE shell, when required.

Clinical Study Team:

Provide information on safety and efficacy findings, when required. Provide updates on safety and efficacy findings for periodic reporting.

Study Statistician

Draft ISS and ISE shell. Update shell, when appropriate. Analyze and report data in approved format, to meet periodic reporting requirements.

36. Explain Types of Clinical trials study you come across?

Single Blind Study

When the patients are not aware of which treatment they receive.

Double Blind Study

When the patients and the investigator are unaware of the treatment group

assigned.

Triple Blind Study

Triple blind study is when patients, investigator, and the project team are unaware of the treatments administered.

37. What are the domains/datasets you have used in your studies?

Demog

Adverse Events

Vitals

ECG

Labs

Medical History

PhysicalExam etc

38. Can you list the variables in all the domains?

Demog: Usubjid, Patient Id, Age, Sex, Race, Screening Weight, Screening Height, BMI etc

Adverse Events: Protocol no, Investigator no, Patient Id, Preferred Term, Investigator Term, (Abdominal dis, Freq urination, headache, dizziness, hand-food syndrome, rash, Leukopenia, Neutropenia) Severity, Seriousness (y/n), Seriousness Type (death, life threatening, permanently disabling), Visit number, Start time, Stop time, Related to study drug?

Vitals: Subject number, Study date, Procedure time, Sitting blood pressure, Sitting Cardiac Rate, Visit number, Change from baseline, Dose of treatment at time of vital sign, Abnormal (yes/no), BMI, Systolic blood pressure, Diastolic blood pressure.

ECG: Subject no, Study Date, Study Time, Visit no, PR interval (msec), QRS duration (msec), QT interval (msec), QTc interval (msec), Ventricular Rate (bpm), Change from baseline, Abnormal.

Labs: Subject no, Study day, Lab parameter (Lparm), lab units, ULN (upper limit of normal), LLN (lower limit of normal), visit number, change from baseline, Greater than ULN (yes/no), lab related serious adverse event (yes/no). Medical History: Medical Condition, Date of Diagnosis (yes/no), Years of onset or occurrence, Past condition (yes/no), Current condition (yes/no).

PhysicalExam: Subject no, Exam date, Exam time, Visit number, Reason for exam, Body system, Abnormal (yes/no), Findings, Change from baseline (improvement, worsening, no change), Comments

39. Give me the example of edit checks you made in your programs? Examples of Edit Checks

Demog: Weight is outside expected range Body mass index is below expected
(check weight and height)
Age is not within expected range.
DOB is greater than the Visit date or not..
Gender value is a valid one or invalid. etc
Adverse Event
Stop is before the start or visit Start is before birthdate Study medicine
discontinued due to adverse event but completion indicated (COMPLETE =1)

Labs

Result is within the normal range but abnormal is not blank or 'N'
Result is outside the normal range but abnormal is blank

Vitals

Diastolic BP > Systolic BP

Medical History

Visit date prior to Screen date Physical Physical exam is normal but comment
included

40. What are the advantages of using SAS in clinical data management? Why should not we use other software products in managing clinical data?

ADVANTAGES OF USING A SAS®-BASED SYSTEM

Less hardware is required.

A Typical SAS®-based system can utilize a standard file server to store its databases and does not require one or more dedicated servers to handle the application load. PC SAS® can easily be used to handle processing, while data access is left to the file server. Additionally, as presented later in this paper, it is possible to use the SAS® product SAS®/Share to provide a dedicated server to handle data transactions.

Fewer personnel are required.

Systems that use complicated database software often require the hiring of one or more DBA's (Database Administrators) who make sure the database software is running, make changes to the structure of the database, etc.

These individuals often require special training or background experience in the particular database application being used, typically Oracle.

Additionally, consultants are often required to set up the system and/or studies since dedicated servers and specific expertise requirements often complicate the process. Users with even casual SAS® experience can set up studies. Novice programmers can build the structure of the database and design screens. Organizations that are involved in data management almost always have at least one SAS® programmer already on staff. SAS® programmers will have an understanding of how the system actually works which would allow them to extend the functionality of the system by directly accessing SAS® data from outside of the system. Speed of setup is dramatically reduced. By keeping studies on a local file server and making the database and screen design processes extremely simple and intuitive, setup time is reduced from weeks to days. All phases of the data

management process become homogeneous. From entry to analysis, data reside in SAS® data sets, often the end goal of every data management group. Additionally, SAS® users are involved in each step, instead of having specialists from different areas hand off pieces of studies during the project life cycle. No data conversion is required. Since the data reside in SAS® data sets natively, no conversion programs need to be written. Data review can happen during the data entry process, on the master database. As long as records are marked as being double-keyed, data review personnel can run edit check programs and build queries on some patients while others are still being entered. Tables and listings can be generated on live data. This helps speed up the development of table and listing programs and allows programmers to avoid having to make continual copies or extracts of the data during testing.

43. Have you ever had to follow SOPs or programming guidelines? SOP describes the process to assure that standard coding activities, which produce tables, listings and graphs, functions and/or edit checks, are conducted in accordance with industry standards are appropriately documented. It is normally used whenever new programs are required or existing programs required some modification during the set-up, conduct, and/or reporting clinical trial data.

44. Describe the types of SAS programming tasks that you performed: Tables? Listings? Graphics? Ad hoc reports? Other? Prepared programs required for the ISS and ISE analysis reports. Developed and validated programs for preparing ad-hoc statistical reports for the preparation of clinical study report. Wrote analysis programs in line with the specifications defined by the study statistician. Base SAS (MEANS, FREQ, SUMMARY, TABULATE, REPORT etc) and SAS/STAT procedures (REG, GLM, ANOVA, and UNIVARIATE etc.) were used for summarization, Cross-Tabulations and statistical analysis purposes. Created Statistical reports using Proc Report, Data _null_ and SAS Macro. Created, derived and merged and pooled datasets, listings and summary tables for Phase-I and Phase-II of clinical trials.

45. Have you been involved in editing the data or writing data queries? If your interviewer asks this question, the u should ask him what he means by editing the data... and data queries...

41. Are you involved in writing the inferential analysis plan? Table's specifications?

42. What do you feel about hardcoding?

Programmers sometime hardcode when they need to produce report in urgent. But it is always better to avoid hardcoding, as it overrides the database controls in clinical data management. Data often change in a trial over time, and the hardcode that is written today may not be valid in the future. Unfortunately, a hardcode may be forgotten and left in the SAS program, and that can lead to an incorrect database change.

43. How do you write a test plan?

Before writing "Test plan" you have to look into on "Functional specifications". Functional specifications itself depends on "Requirements",

so one should have clear understanding of requirements and functional specifications to write a test plan.

44. What is the difference between verification and validation?

Although the verification and validation are close in meaning, "verification" has more of a sense of testing the truth or accuracy of a statement by examining evidence or conducting experiments, while "validate" has more of a sense of declaring a statement to be true and marking it with an indication of official sanction.

45. What other SAS features do you use for error trapping and data validation?

Conditional statements, if then else.

Put statement

Debug option.

46. What is PROC CDISC?

It is new SAS procedure that is available as a hotfix for SAS 8.2 version and comes as a part with SAS 9.1.3 version.

PROC CDISC is a procedure that allows us to import (and export XML files that are compliant with the CDISC ODM version 1.2 schema.

For more details refer SAS programming in the Pharmaceutical Industry text book.

47) What is LOCF?

Pharmaceutical companies conduct longitudinal studies on human subjects that often span several months. It is unrealistic to expect patients to keep every scheduled visit over such a long period of time. Despite every effort, patient data are not collected for some time points. Eventually, these become missing values in a SAS data set later. For reporting purposes, the most recent previously available value is substituted for each missing visit. This is called the Last Observation Carried Forward (LOCF). LOCF doesn't mean last SAS dataset observation carried forward. It means last non-missing value carried forward. It is the values of individual measures that are the "observations" in this case. And if you have multiple variables containing these values then they will be carried forward independently.

48) ETL process:

Extract, transform and Load:

Extract:

The 1st part of an ETL process is to extract the data from the source systems. Most data warehousing projects consolidate data from different source systems.

Each separate system may also use a different data organization / format. Common data source formats are relational databases and flat files, but may

include non-relational database structures such as IMS or other data structures such as VSAM or ISAM.

Extraction converts the data into a format for transformation processing. An intrinsic part of the extraction is the parsing of extracted data, resulting in a check if the data meets an expected pattern

Transform: The transform stage applies a series of rules or functions to the extracted data from the source to derive the data to be loaded to the end target. Some data sources will require very little or even no manipulation of data. In other cases, one or more of the following transformations types to meet the business and technical needs of the end target may be required:

Selecting only certain columns to load (or selecting null columns not to load) · Translating coded values (e.g., if the source system stores 1 for male and 2 for female, but the warehouse stores M for male and F for female), this is called automated data cleansing; no manual cleansing occurs during ETL · Encoding free-form values (e.g., mapping "Male" to "1" and "Mr" to M) ·

Joining together data from multiple sources (e.g., lookup, merge, etc.) · Generating surrogate key values · Transposing or pivoting (turning multiple columns into multiple rows or vice versa) · Splitting a column into multiple columns (e.g., putting a comma-separated list specified as a string in one column as individual values in different columns) ·

Applying any form of simple or complex data validation; if failed, a full, partial or no rejection of the data, and thus no, partial or all the data is handed over to the next step, depending on the rule design and exception handling. Most of the above transformations itself might result in an exception, e.g. when a code-translation parses an unknown code in the extracted data. Load: The load phase loads the data into the end target, usually being the data warehouse (DW).

Depending on the requirements of the organization, this process ranges widely. Some data warehouses might weekly overwrite existing information with cumulative, updated data, while other DW (or even other parts of the same DW) might add new data in a historized form, e.g. hourly. The timing and scope to replace or append are strategic design choices dependent on the time available and the business needs. More complex systems can maintain a history and audit trail of all changes to the data loaded in the DW.

As the load phase interacts with a database, the constraints defined in the database schema as well as in triggers activated upon data load apply (e.g. uniqueness, referential integrity, mandatory fields), which also contribute to the overall data quality performance of the ETL process.

CDISC SDTM INTERVIEW QUESTIONS

1) What do you know about CDISC and its standards?

CDISC stands for Clinical Data Interchange Standards Consortium and it is developed keeping in mind to bring great deal of efficiency in the entire drug development process. CDISC brings efficiency to the entire drug development process by improving the data quality and speed-up the whole drug development process and to do that CDISC developed a series of standards, which include Operation data Model (ODM), Study data Tabulation Model (SDTM) and the Analysis Data Model ADaM).

2) Why people these days are more talking about CDSIC and what advantages it brings to the Pharmaceutical Industry?

A) Generally speaking, Only about 30% of programming time is used to generate statistical results with SAS®, and the rest of programming time is used to familiarize data structure, check data accuracy, and tabulate/list raw data and statistical results into certain formats. This non-statistical programming time will be significantly reduced after implementing the CDISC standards.

3) What are the challenges as SAS programmer you think you will face when you first implement CDISC standards in you company?

A) With the new requirements of electronic submission, CRT datasets need to conform to a set of standards for facilitating reviewing process. They no longer are created solely for programmers convenient. SDS will be treated as specifications of datasets to be submitted, potentially as reference of CRF design. Therefore, statistical programming may need to start from this common ground. All existing programs/macros may also need to be remapped based on CDISC so one can take advantage to validate submission information by using tools which reviewer may use for reviewing and to accelerate reviewing process without providing unnecessary data, tables and listings. With the new requirements from updating electronic submission and CDISC implementation, understanding only SAS® may not be good enough to fulfill for final deliverables. It is a time to expand and enhance the job skills from various aspects under new change so that SAS® programmers can take a competitive advantage, and continue to play a main role in both statistical analysis and reporting for drug development.

References:

Pharmasug/2007/fc/fc05

pharmasug/2003/fda compliance/fda055

1) What do you understand about SDTM and its importance?

SDTM stands for Standard data Tabulation Model, which defines a standard structure for study data tabulations that are to be submitted as part of a product application to a regulatory authority such as the United States Food and Drug Administration (FDA) 2.

In July 2004 the Clinical Data Interchange Standards Consortium (CDISC)

published standards on the design and content of clinical trial tabulation data sets, known as the Study Data Tabulation Model (SDTM). According to the CDISC standard, there are four ways to represent a subject in a clinical study: tabulations, data listings, analysis datasets, and subject profiles⁶.

Before SDTM:

There are different names for each domain and domains don't have a standard structure. There is no standard variables list for each and every domain.

Because of this FDA reviewers always had to take so much pain in understanding themselves with different data, domain names and name of the variable in each analysis dataset. Reviewers will have spent most of the valuable time in cleaning up the data into a standard format rather than reviewing the data for the accuracy. This process will delay the drug development process as such.

After SDTM:

There will be standard domain names and standard structure for each domain. There will be a list of standard variables and names for each and every dataset. Because of this, it will become easy to find and understand the data and reviewers will need less time to review the data than the data without SDTM standards. This process will improve the consistency in reviewing the data and it can be time efficient.

The purpose of creating SDTM domain data sets is to provide Case Report Tabulation (CRT) data FDA, in a standardized format. If we follow these standards it can greatly reduce the effort necessary for data mapping. Improper use of CDISC standards, such as using a valid domain or variable name incorrectly, can slow the metadata mapping process and should be avoided⁴.

2) PROC CDISC for SDTM 3.1 Format 2?

Syntax The PROC CDISC syntax for CDISC SDTM is presented below. The DATA= parameter specifies the location of your SDTM conforming data source. PROC CDISC MODEL=SDTM;SDTM SDTMVersion = "3.1";DOMAINDATA DATA = results. AE DOMAIN = AE CATEGORY = EVENT;RUN;

3) What are the capabilities of PROC CDISC 2?

PROC CDISC performs the following checks on domain content of the source:

Verifies that all required variables are present in the data set

Reports as an error any variables in the data set that are not defined in the domain

Reports a warning for any expected domain variables that are not in the

data set

Notes any permitted domain variables that are not in the data set

Verifies that all domain variables are of the expected data type and proper length

Detects any domain variables that are assigned a controlled terminology specification by the domain and do not have a format assigned to them.

The procedure also performs the following checks on domain data content of the source on a per observation basis:

Verifies that all required variable fields do not contain missing values

Detects occurrences of expected variable fields that contain missing values

Detects the conformance of all ISO-8601 specification assigned values; including date, time, date time, duration, and interval types

Notes correctness of yes/no and yes/no/null responses,

4) What are the different approaches for creating the SDTM 3?

There are 3 general approaches to create the SDTM datasets:

- a) Build the SDTM entirely in the CDMS,
- b) Build the SDTM entirely on the “back-end” in SAS,
- c) or take a hybrid approach and build the SDTM partially in the CDMS and partially in SAS.

BUILD THE SDTM ENTIRELY IN THE CDMS

It is possible to build the SDTM entirely within the CDMS. If the CDMS allows for broad structural control of the underlying database, then you could build your eCRF or CRF based clinical database to SDTM standards.

Advantages:

- Your “raw” database is equivalent to your SDTM which provides the most elegant solution.
- Your clinical data management staff will be able to converse with end-users/sponsors about the data easily since your clinical data manager and the end-user/sponsor will both be looking at SDTM datasets.
- As soon as the CDMS database is built, the SDTM datasets are available.

Disadvantages:

- This approach may be cost prohibitive. Forcing the CDMS to create the SDTM structures may simply be too cumbersome to do efficiently.
- Forcing the CDMS to adapt to the SDTM may cause problems with the operation of the CDMS which could reduce data quality.

BUILD THE SDTM ENTIRELY ON THE “BACK-END” IN SAS

Assuming that SAS is not your CDMS solution, another approach is to take the clinical data from your CDMS and manipulate it into the SDTM with SAS programming.

Advantages:

- The great flexibility of SAS will let you transform any proprietary CDMS structure into the SDTM. You do not have to work around the rigid constraints of the CDMS.
- Changes could be made to the SDTM conversion without disturbing clinical data management processes.
- The CDMS is allowed to do what it does best which is to enter, manage, and clean data.

Disadvantages: • There would be additional cost to transform the data from your typical CDMS structure into the SDTM.

Specifications, programming, and validation of the SAS programming transformation would be required.

- Once the CDMS database is up, there would then be a subsequent delay while the SDTM is created in SAS.

This delay would slow down the production of analysis datasets and reporting. This assumes that you follow the linear progression of CDMS -> SDTM -> analysis datasets (ADaM).

- Since the SDTM is a derivation of the “raw” data, there could be errors in translation from the “raw” CDMS data to the SDTM.
- Your clinical data management staff may be at a disadvantage when speaking with end-users/sponsors about the data since the data manager will likely be looking at the CDMS data and the sponsor will see SDTM data.

BUILD THE SDTM USING A HYBRID APPROACH

Again, assuming that SAS is not your CDMS solution, you could build some of the SDTM within the confines of the CDMS and do the rest of the work in SAS. There are things that could be done easily in the CDMS such as naming data tables the same as SDTM domains, using SDTM variable names in the CDMS, and performing simple derivations (such as age) in the CDMS. More complex SDTM derivations and manipulations can then be performed in SAS.

Advantages:

- The changes to the CDMS are easy to implement.
- The SDTM conversions to be done in SAS are manageable and much can be automated.

Disadvantages:

- There would still be some additional cost needed to transform the data from the SDTM-like CDMS structure into the SDTM. Specifications, programming, and validation of the transformation would be required.
- There would be some delay while the SDTM-like CDMS data is converted to the SDTM.
- Your clinical data management staff may still have a slight disadvantage when speaking with endusers/ sponsors about the data since the clinical data manager will be looking at the SDTM-like data and the sponsor will see the true SDTM data.

5) What do you know about SDTM domains?

A basic understanding of the SDTM domains, their structure and their interrelations is vital to determining which domains you need to create and in assessing the level to which your existing data is compliant. The SDTM consists of a set of clinical data file specifications and underlying guidelines. These different file structures are referred to as domains. Each domain is designed to contain a particular type of data associated with clinical trials, such as demographics, vital signs or adverse events.

The CDISC SDTM Implementation Guide provides specifications for 30 domains. The SDTM domains are divided into six classes.

The 21 clinical data domains are contained in three of these classes:
Interventions,
Events and
Findings.

The trial design class contains seven domains and the special-purpose class contains two domains (Demographics and Comments).

The trial design domains provide the reviewer with information on the criteria, structure and scheduled events of a clinical trial. The only required domain is demographics.

There are two other special purpose relationship data sets, the Supplemental Qualifiers (SUPPQUAL) data set and the Relate Records (RELREC) data set. SUPPQUAL is a highly normalized data set that allows you to store virtually any type of information related to one of the domain data sets. SUPPQUAL domain also accommodates variables longer than 200, the first 200 characters should be stored in the domain variable and the remaining should be stored in it5.

6) What are the general guidelines to SDTM variables?

Each of the SDTM domains has a collection of variables associated with it. There are five roles that a variable can have:
Identifier,
Topic,
Timing,
Qualifier,
and for trial design domains,

Rule. Using lab data as an example, the subject ID, domain ID and sequence (e.g. visit) are identifiers.

The name of the lab parameter is the topic, the date and time of sample collection are timing variables,

the result is a result qualifier and the variable containing the units is a variable qualifier.

Variables that are common across domains include the basic identifiers study ID (STUDYID), a two-character domain ID (DOMAIN) and unique subject ID (USUBJID).

In studies with multiple sites that are allowed to assign their own subject identifiers, the site ID and the subject ID must be combined to form USUBJID.

Prefixing a standard variable name fragment with the two-character domain ID generally forms all other variable names.

The SDTM specifications do not require all of the variables associated with a domain to be included in a submission. In regard to complying with the SDTM standards, the implementation guide specifies each variable as being included in one of three categories:

Required, Expected, and Permitted⁴.

REQUIRED – These variables are necessary for the proper functioning of standard software tools used by reviewers. They must be included in the data set structure and should not have a missing value for any observation.

EXPECTED – These variables must be included in the data set structure; however it is permissible to have missing values.

PERMISSIBLE – These variables are not a required part of the domain and they should not be included in the data set structure if the information they were designed to contain was not collected.

7) Can you tell me more About SDTM Domains⁵?

SDTM Domains are grouped by classes, which is useful for producing more meaningful relational schemas. Consider the following domain classes and their respective domains.

- Special Purpose Class – Pertains to unique domains concerning detailed information about the subjects in a study.

Demography (DM), Comments (CM)

- **Findings Class** – Collected information resulting from a planned evaluation to address specific questions about the subject, such as whether a subject is suitable to participate or continue in a study.

Electrocardiogram (EG)

Inclusion / Exclusion (IE)

Lab Results (LB)

Physical Examination (PE)

Questionnaire (QS)
Subject Characteristics (SC)
Vital Signs (VS)

- **Events Class** – Incidents independent of the study that happen to the subject during the lifetime of the study.

Adverse Events (AE)
Patient Disposition (DS)
Medical History (MH)

- **Interventions Class** – Treatments and procedures that are intentionally administered to the subject, such as treatment coincident with the study period, per protocol, or self-administered (e.g., alcohol and tobacco use).

Concomitant Medications (CM)
Exposure to Treatment Drug (EX)
Substance Usage (SU)

- **Trial Design Class** – Information about the design of the clinical trial (e.g., crossover trial, treatment arms) including information about the subjects with respect to treatment and visits.

Subject Elements (SE)
Subject Visits (SV)
Trial Arms (TA)
Trial Elements (TE)
Trial Inclusion / Exclusion Criteria (TI)

Trial Visits (TV)

7) Can you tell me how to do the Mapping for existing Domains?

First step is the comparison of metadata with the SDTM domain metadata. If the data getting from the data management is in somewhat compliance to SDTM metadata, use automated mapping as the 1st step.

If the data management metadata is not in compliance with SDTM then avoid auto mapping. So do manual mapping the datasets to SDTM datasets and the mapping each variable to appropriate domain.

The whole process of mapping include: *Read in the corporate data standards into a database table.

- Assign a CDISC domain prefix to each database module.
- Attach a combo box containing the SDTM variable for the selected domain to a new mapping variable field.
- Search each module, and within each module select the most appropriate CDISC variable.
- Then search for variables mapped to the wrong type Character not equal to Character; Numeric not equal to Numeric.

- Review the mapping to see if any conflicts are resolvable by mapping to a more appropriate variable.
- We need to verify that the mapped variable is appropriate for each role.
- Then finally we have to ensure all 'required' variables are present in the domain6.

8) What do you know about SDTM Implementation Guide, Have you used it, if you have can you tell me which version you have used so far?

SDTM Implementation guide provides documentation on metadata (data of data) for the domain datasets that includes filename, variable names, type of variables and its labels etc. I have used SDTM implementation guide versions 3.1.1/3.1.2

9) Can you identify which variables should we have to include in each domain?

A) SDTM implementation guide V 3.1.1/V 3.1.2 specifies each variable is being included in one of the 3 types.

REQUIRED –They must be included in the data set structure and should not have a missing value for any observation.

EXPECTED – These variables must be included in the data set; however it is permissible to have missing values.

PERMISSIBLE – These variables are not a required part of the domain and they should not be included in the data set structure if the information they were designed to contain was not collected.

10) Can you give some examples for MAPPING *6?

Here are some examples for SDTM mapping:

- Character variables defined as Numeric
- Numeric Variables defined as Character
- Variables collected without an obvious corresponding domain in the CDISC SDTM mapping. So must go into SUPPQUAL
- Several corporate modules that map to one corresponding domain in CDISC SDTM.
- Core SDTM is a subset of the existing corporate standards
- Vertical versus Horizontal structure, (e.g. Vitals)
- Dates – combining date and times; partial dates.
- Data collapsing issues e.g. Adverse Events and Concomitant Medications.
- Adverse Events maximum intensity
- Metadata needed to laboratory data standardization.

10) Explain the Process of SDTM Mapping?

A list of basic variable mappings is given below *4.

DIRECT: a CDM variable is copied directly to a domain variable without any changes other than assigning the CDISC standard label.

RENAME: only the variable name and label may change but the contents remain the same.

STANDARDIZE: mapping reported values to standard units or standard terminology

REFORMAT: the actual value being represented does not change, only the format in which is stored changes, such as converting a SAS date to an ISO8601 format character string.

COMBINING: directly combining two or more CDM variables to form a single SDTM variable.

SPLITTING: a CDM variable is divided into two or more SDTM variables.

DERIVATION: creating a domain variable based on a computation, algorithm, series of logic rules or decoding using one or more CDM variables.

11) What are the Common Issues in Mapping Dummy corporate standards to CDISC (SDTM) Standards?

- Character variables defined as Numeric
- Numeric Variables defined as Character
- Variables collected without an obvious corresponding domain in the CDISC SDTM mapping. So must go into SUPPQUAL
- Several corporate modules that map to one corresponding domain in CDISC SDTM.
- Dictionary codes not in SDTM parent module, so if needed must be collected in SUPPQUAL.
- Core SDTM is a subset of the existing corporate standards
- Different structure of Lab CDISC Domain e.g. baseline flag.
- Vertical versus Horizontal structure, (e.g. Vitals)
- Additional Metadata needed to describe the source in SUPPQUAL
- Dates – combining date and times; partial dates.
- Data collapsing issues e.g. Adverse Events and Concomitant Medications.
- Adverse Events maximum intensity
- Metadata needed to laboratory data standardization.

Ref: Mapping Corporate Data Standards to the CDISC Model (SAS Paper) by David Parker, AstraZeneca, Manchester, United Kingdom

The **Analysis Data Model** describes the general structure, metadata, and content typically found in Analysis Datasets and accompanying documentation. The three types of metadata associated with analysis datasets (analysis dataset metadata, analysis variable metadata, and analysis results metadata) are described and examples provided.

(source:CDISC Analysis Data Model: Version 2.0)

Analysis datasets (AD) are typically developed from the collected clinical trial data and used to create statistical summaries of efficacy and safety data.

These AD's are characterized by the creation of derived analysis variables and/or records. These derived data may represent a statistical calculation of an important outcome measure, such as change from baseline, or may represent the last observation for a subject while under therapy. As such, these datasets are one of the types of data sent to the regulatory agency such as FDA.

The CDISC **Analysis Data Model** (ADaM) defines a standard for Analysis Dataset's to be submitted to the regulatory agency. This provides a clear content, source, and quality of the datasets submitted in support of the statistical analysis performed by the sponsor.

In **ADaM**, the descriptions of the AD's build on the nomenclature of the SDTM with the addition of attributes, variables and data structures needed for statistical analyses. To achieve the principle of clear and unambiguous communication relies on clear AD documentation. This documentation provides the link between the general description of the analysis found in the protocol or statistical analysis plan and the source data.

12) Can you explain AdAM or AdAM datasets *7?

References:

- 1) <http://support.sas.com/rnd/base/xmlengine/proccdisc/cdiscsdm.html>
- 2) <http://www.fda.gov>
- 3) [pharmasug/2005/fda-compliance/fc01.pdf](http://pharmasug.com/2005/fda-compliance/fc01.pdf)
- 4) <http://www2.sas.com/proceedings/forum2008/207-2008.pdf>
- 5) http://analytics.ncsu.edu/sesug/2006/PO08_06.PDF
- 6) <http://www.lexjansen.com/phuse/2005/cd/cd11.pdf>
- 7) <http://www.pharmasug.org/2005/FC03.pdf>

Apart from those .. you may also need to prepare for these questions too...

Robert Stemplinger:

- 1) How many years experience you have working with CDISC standards?
- 2) What have you been done as per CDISC standards.

(Tell me the usual process flow or the procedure you have followed regarding implementation of CDISC standards)

- 3) For how many studies so far you have done SDTM mapping.
- 4) Have you ever been asked to create specifications for SDTM mapping.

If yes, how do you create specification document for mapping.

5) Do you have experience doing the mapping as per the sponsor standards.

6) a) Tell me few details about the databases you have worked with so far?

b) Which database do you think you had most trouble with? (Inform, Rave, Clintrial or Oracle clinical)

7) How do you validate

a) annotated CRF

b) Specification Document

c) SDTM datasets

d) Case Report Tabulations (CRT-DDS)

8) a) How do you verify all the standards has been maintained as per the SDTM implementation guide?

b) How do you perform validation checks on SDTM v 3.1.1 or 3.1.2 datasets? (WEBSDM/Open CDISC or PROC CDISC)?

9) What you will do when you find a problem as part of the validation process?

10) What kind of macros you have developed which can be useful in creating SDTM standard datasets?

11) Do you like to create a single program for each domain and then include in a batch program or

just one big program for all the domains.

12) Do you have any experience talking to the client on regular basis? If, yes... share with me your experience?

13) Do you have experience working with people in different time zone?

14) Do you have experience or knowledge about WEBSDM checks or Open CDISC?

15) Do you know PROC CDISC?

16) How do you create Define file (XML or PDF), if you already had experience creating one?

17) If you are working as a validator, how do you communicate with the main programmer?

18) How many weeks time you think you need to finish creating the SDTM datasets? (Just for programming)?

How many weeks, if you also been asked to develop specifications?

19) Is there any sample program you can write or show ... which will give us an idea about you SAS programming skills?

20) What's the challenging part regarding the whole SDTM mapping process?

21) For which domain do you think you always need to be very careful? and why?

22) If I ask you to create SDTM mapping specification document? what documents or files you need and why?

23) Do you know anything about splitting domains. (or Can you split the domains rather than creating one big domain)?

24) What is value level meta data?

25) What do you know about controlled terminology and for which domains you need controlled terminology?

26) What are RELREC and SUPPQUAL domains.

27) Can you share with me any differences you know between implementation guide v3.1.1 and v3.1.2?

28) How do you determine the time line, If the client asked you to provide one for the SDTM mapping conversion process?

29) Is there any way to apply attributes to the SDTM variables other than just manually typing all the details about (length/label/format/informat etc) in an attrib statement?

30) You have been asked to create a domain (not included in implementation guide) for CRF, what you will do or how do you create one?

Here are few more questionsexclusive to SDTM Mapping....

CDISC SDTM Questions You might be asked in an interview

- 1) Have you used - -STAT variable anytime. If yes, why and in what kind of domain you used that variable.
- 2) I see in your CV that you have experience in developing SDTM domains based on IG 3.1.1, V3.1.2 and V3.1.3. Can you share some of the differences between each version of Implementation Guide? (Difference between SDTM IG 3.1.1 vs. V3.1.2 and V3.1.2 vs. V3.1.3)
- 3) Can you give me an example of a variable which can be used to group some of the records?
- 4) Tell me your experience using - -SPEC variable.
- 5) What's the significance of - -PRESP variable and tell me what do you know about - -OCCUR variable.
- 6) Can you give me an example of a Topic Variable in:
 - a) Intervention Domains
 - b) Event Domains
 - c) Finding Domains
- 7) What's your experience creating the Related Records domain (RELREC)? Can you give me few examples of the domains you've used to create a RELREC SDTM domain?
- 8) What's your experience creating the Findings About (FA) and Clinical Events (CE) domains. What's the difference between the FA and CE domains?
- 9) Can you give me few examples of the kind of data you are going to map it to FA and CE domains.
- 10) Why can't we include Clinical Event data in AE domain?
- 11) What's your experience creating the custom domains? How do you create a custom domain?
- 12) What you do, if you have a CRF page and all of the information collected on it aren't related to any specific SDTM domain.
- 13) When do you create a SUPPQUAL or Custom domain?
- 14) If you have any experience creating a custom domain, can you share, what kind of the data that was and what's the PREFIX you have used for the domain name.
- 15) Tell me about the difficult thing you have to do or manage when you work as a SDTM standards implementer.
- 16) Have you use - -OBJ variable. If you are, in which domain? And what's the significance.
- 17) Tell me about Required/Expected or Permissible variables in SDTM domains.
- 18) Have you created any Tumor Domains? Can you give use few examples of the tumor domains you have created.

FREQUENTLY ASKED SAS WRITTEN TEST/INTERVIEW QUESTIONS

PART 1

Objective : Gearing up for a SAS interview??

This post is a prequel of my earlier posts dealing with SAS Interview question part1 and Part 2. The post lists some objective type question with answers (Don't see the answers first .. ☺) which can appear as part of initial written test screening included as a part of selection process of many companies. You can use this post to test where you stand before you jump directly into the battlefield. This post can also be used as part of SAS knowledge improvement and also as a knowledge check.

I will be putting such posts from time to time which you can use as a quick knowledge check. So best of luck folks.

QUESTIONS :

1. The following SAS program is submitted:

```
data test;
set sasuser.employees;
if 2 le years_service le 10 then
amount = 1000;
else if years_service gt 10 then
amount = 2000;
else
amount = 0;
amount_per_year = years_service / amount;
run;
```

Which one of the following values does the variable AMOUNT_PER_YEAR contain if an employee has been with the company for one year?

- A. 0
 - B. 1000
 - C. 2000
 - D. . (missing numeric value)
2. The contents of the raw data file AMOUNT are listed below:

```
-----10-----20-----30
```

```
$1,234
```

The following SAS program is submitted:

```
data test;
infile 'amount';
input @1 salary 6.;
if _error_ then description = 'Problems';
else description = 'No Problems';
```

run;

Which one of the following is the value of the DESCRIPTION variable?

- A. Problems
- B. No Problems
- C. ' ' (missing character value)
- D. The value can not be determined as the program fails to execute due to errors.

3. The contents of the raw data file NAMENUM are listed below:

-----10-----20-----30

Joe xx

The following SAS program is submitted:

```
data test;
infile 'namenum';
input name $ number;
run;
```

Which one of the following is the value of the NUMBER variable?

- A. xx
- B. Joe
- C. . (missing numeric value)
- D. The value can not be determined as the program fails to execute due to errors.

4. The contents of the raw data file AMOUNT are listed below:

-----10-----20-----30

\$1,234

The following SAS program is submitted:

```
data test;
infile 'amount';
input @1 salary 6.;
run;
```

Which one of the following is the value of the SALARY variable?

- A. 1234
- B. 1,234
- C. \$1,234
- D. . (missing numeric value)

5. Which one of the following statements is true regarding the SAS automatic `_ERROR_` variable?

- A. The `_ERROR_` variable contains the values 'ON' or 'OFF'.
- B. The `_ERROR_` variable contains the values 'TRUE' or 'FALSE'.
- C. The `_ERROR_` variable is automatically stored in the resulting SAS data set.
- D. The `_ERROR_` variable can be used in expressions or calculations in the DATA step.

6. Which one of the following is true when SAS encounters a data error in a DATA step?

- A. The DATA step stops executing at the point of the error, and no SAS data set is created.
- B. A note is written to the SAS log explaining the error, and the DATA step continues to execute.
- C. A note appears in the SAS log that the incorrect data record was

saved to a separate SAS file for further examination.

D. The DATA step stops executing at the point of the error, and the resulting DATA set contains observations up to that point.

7. The following SAS program is submitted:

```
data work.totalsales (keep = monthsales{12} );  
set work.monthlysales (keep = year product sales);  
array monthsales {12} ;  
do i=1 to 12;  
monthsales{i} = sales;  
end;  
run;
```

The data set named WORK.MONTHLYSALES has one observation per month for each of five years for a total of 60 observations.

Which one of the following is the result of the above program?

A. The program fails execution due to data errors.

B. The program fails execution due to syntax errors.

C. The program executes with warnings and creates the WORK.TOTALSALES data set.

D. The program executes without errors or warnings and creates the WORK.TOTALSALES data set.

8. The following SAS program is submitted:

```
data work.totalsales;  
set work.monthlysales(keep = year product sales);  
retain monthsales {12} ;  
array monthsales {12} ;  
do i = 1 to 12;  
monthsales{i} = sales;  
end;  
cnt + 1;  
monthsales{cnt} = sales;  
run;
```

The data set named WORK.MONTHLYSALES has one observation per month for each of five years for a total of 60 observations.

Which one of the following is the result of the above program?

A. The program fails execution due to data errors.

B. The program fails execution due to syntax errors.

C. The program runs with warnings and creates the WORK.TOTALSALES data set with 60 observations.

D. The program runs without errors or warnings and creates the WORK.TOTALSALES data set with 60 observations.

9. The following SAS program is submitted:

```
data work.january;  
set work.allmonths (keep = product month num_sold cost);  
if month = 'Jan' then output work.january;  
sales = cost * num_sold;  
keep = product sales;  
run;
```

Which variables does the WORK.JANUARY data set contain?

A. PRODUCT and SALES only

- B. PRODUCT, MONTH, NUM_SOLD and COST only
 C. PRODUCT, SALES, MONTH, NUM_SOLD and COST only
 D. An incomplete output data set is created due to syntax errors.
10. The contents of the raw data file CALENDAR are listed below:
 -----10-----20-----30
 01012000
 The following SAS program is submitted:
 data test;
 infile 'calendar';
 input @1 date mmdyy10.;
 if date = '01012000'd then event = 'January 1st';
 run;
 Which one of the following is the value of the EVENT variable?
 A. 01012000
 B. January 1st
 C. . (missing numeric value)
 D. The value can not be determined as the program fails to execute due to errors.
11. A SAS program is submitted and the following SAS log is produced:
 2 data gt100;
 3 set ia.airplanes
 4 if mpg gt 100 then output;
 22 202
 ERROR: File WORK.IF.DATA does not exist.
 ERROR: File WORK.MPG.DATA does not exist.
 ERROR: File WORK.GT.DATA does not exist.
 ERROR: File WORK.THEN.DATA does not exist.
 ERROR: File WORK.OUTPUT.DATA does not exist.
 ERROR 22-322: Syntax error, expecting one of the following: a name, a quoted string, (, ;, END, KEY, KEYS, NOBS, OPEN, POINT, _DATA_, _LAST_, _NULL_.
 ERROR 202-322: The option or parameter is not recognized and will be ignored.
 5 run;
 The IA libref was previously assigned in this SAS session.
 Which one of the following corrects the errors in the LOG?
 A. Delete the word THEN on the IF statement.
 B. Add a semicolon at the end of the SET statement.
 C. Place quotes around the value on the IF statement.
 D. Add an END statement to conclude the IF statement.
12. The contents of the raw data file SIZE are listed below:
 -----10-----20-----30
 72 95
 The following SAS program is submitted:
 data test;
 infile 'size';
 input @1 height 2. @4 weight 2;
 run;

Which one of the following is the value of the variable WEIGHT in the output data set?

- A. 2
- B. 72
- C. 95
- D. . (missing numeric value)

13. A SAS PRINT procedure output of the WORK.LEVELS data set is listed below:

Obs name level

1 Frank 1

2 Joan 2

3 Sui 2

4 Jose 3

5 Burt 4

6 Kelly .

7 Juan 1

The following SAS program is submitted:

```
data work.expertise;
set work.levels;
if level = . then
expertise = 'Unknown';
else if level = 1 then
expertise = 'Low';
else if level = 2 or 3 then
expertise = 'Medium';
else
expertise = 'High';
run;
```

Which of the following values does the variable EXPERTISE contain?

- A. Low, Medium, and High only
- B. Low, Medium, and Unknown only
- C. Low, Medium, High, and Unknown only
- D. Low, Medium, High, Unknown, and ' ' (missing character value)

14. The contents of the raw data file EMPLOYEE are listed below:

-----10-----20-----30

Ruth 39 11

Jose 32 22

Sue 30 33

John 40 44

The following SAS program is submitted:

```
data test;
infile 'employee';
input employee_name $ 1-4;
if employee_name = 'Ruth' then input idnum 10-11;
else input age 7-8;
run;
```

Which one of the following values does the variable IDNUM contain when the name of the employee is "Ruth"?

- A. 11
- B. 22
- C. 32
- D. . (missing numeric value)

15. The contents of the raw data file EMPLOYEE are listed below:

-----10-----20-----30

Ruth 39 11

Jose 32 22

Sue 30 33

John 40 44

The following SAS program is submitted:

```
data test;
infile 'employee';
input employee_name $ 1-4;
if employee_name = 'Sue' then input age 7-8;
else input idnum 10-11;
run;
```

Which one of the following values does the variable AGE contain when the name of the employee is "Sue"?

- A. 30
- B. 33
- C. 40
- D. . (missing numeric value)

16. The following SAS program is submitted:

```
libname sasdata 'SAS-data-library';
data test;
set sasdata.chemists;
if jobcode = 'Chem2'
then description = 'Senior Chemist';
else description = 'Unknown';
run;
```

A value for the variable JOBCODE is listed below:

JOBCODE

chem2

Which one of the following values does the variable DESCRIPTION contain?

- A. Chem2
- B. Unknown
- C. Senior Chemist
- D. ' ' (missing character value)

17. The following SAS program is submitted:

```
libname sasdata 'SAS-data-library';
data test;
set sasdata.chemists;
if jobcode = 'chem3'
then description = 'Senior Chemist';
else description = 'Unknown';
```

run;

A value for the variable JOBCODE is listed below:

JOBCODE

CHEM3

Which one of the following values does the variable DESCRIPTION contain?

A. chem3

B. Unknown

C. Senior Chemist

D. ' ' (missing character value)

18. Which one of the following ODS statement options terminates output being written to an HTML file?

A. END

B. QUIT

C. STOP

D. CLOSE

19. The following SAS program is submitted:

```
proc means data = sasuser.shoes;
```

```
where product in ('Sandal' , 'Slipper' , 'Boot');
```

```
run;
```

Which one of the following ODS statements completes the program and sends the report to an HTML file?

A. ods html = 'sales.html';

B. ods file = 'sales.html';

C. ods file html = 'sales.html';

D. ods html file = 'sales.html';

20. The following SAS program is submitted:

```
proc format;
```

```
value score 1 - 50 = 'Fail'
```

```
51 - 100 = 'Pass';
```

```
run;
```

```
proc report data = work.courses nowd;
```

```
column exam;
```

```
define exam / display format = score.;
```

```
run;
```

The variable EXAM has a value of 50.5.

How will the EXAM variable value be displayed in the REPORT procedure output?

A. Fail

B. Pass

C. 50.5

D. . (missing numeric value)

21. The following SAS program is submitted:

```
options pageno = 1;
```

```
proc print data = sasuser.houses;
```

```
run;
```

```
proc means data = sasuser.shoes;
```

```
run;
```

The report created by the PRINT procedure step generates 5 pages of output.

What is the page number on the first page of the report generated by the MEANS procedure step?

- A. 1
 - B. 2
 - C. 5
 - D. 6
22. Which one of the following SAS system options displays the time on a report?
- A. TIME
 - B. DATE
 - C. TODAY
 - D. DATETIME
23. Which one of the following SAS system options prevents the page number from appearing on a report?
- A. NONUM
 - B. NOPAGE
 - C. NONUMBER
 - D. NOPAGENUM
24. The following SAS program is submitted:
- ```
footnote1 'Sales Report for Last Month';
footnote2 'Selected Products Only';
footnote3 'All Regions';
footnote4 'All Figures in Thousands of Dollars';
proc print data = sasuser.shoes;
footnote2 'All Products';
run;
```
- Which one of the following contains the footnote text that is displayed in the report?
- A. All Products  
Sales Report for Last Month  
All Products
  - B. All Products  
All Regions  
All Figures in Thousands of Dollars
  - C. Sales Report for Last Month  
All Products  
All Regions  
All Figures in Thousands of Dollars
  - D. Sales Report for Last Month  
All Products  
All Regions  
All Figures in Thousands of Dollars
25. The following SAS program is submitted:
- ```
proc means data = sasuser.houses std mean max;  
var sqfeet;  
run;
```
- Which one of the following is needed to display the standard deviation with only two decimal places?
- A. Add the option MAXDEC = 2 to the MEANS procedure statement.
 - B. Add the statement MAXDEC = 7.2; in the MEANS procedure step.
 - C. Add the statement FORMAT STD 7.2; in the MEANS procedure

step.

D. Add the option `FORMAT = 7.2` option to the `MEANS` procedure statement.

26. Unless specified, which variables and data values are used to calculate statistics in the `MEANS` procedure?
- A. non-missing numeric variable values only
 - B. missing numeric variable values and non-missing numeric variable values only
 - C. non-missing character variables and non-missing numeric variable values only
 - D. missing character variables, non-missing character variables, missing numeric variable values, and non-missing numeric variable values

27. The following SAS program is submitted:

```
proc sort data = sasuser.houses out = houses;
by style;
run;
proc print data = houses;
```

```
run;
```

Click on the Exhibit button to view the report produced.

```
style bedrooms baths price
CONDO 2 1.5 80050
3 2.5 79350
4 2.5 127150
2 2.0 110700
RANCH 2 1.0 64000
3 3.0 86650
3 1.0 89100
1 1.0 34550
SPLIT 1 1.0 65850
4 3.0 94450
3 1.5 73650
TWOSTORY 4 3.0 107250
2 1.0 55850
2 1.0 69250
4 2.5 102950
```

Which of the following SAS statement(s) create(s) the report?

- A. `id style;`
- B. `id style;`
`var style bedrooms baths price;`
- C. `id style;`
`by style;`
`var bedrooms baths price;`
- D. `id style;`
`by style;`
`var style bedrooms baths price;`

28. A realtor has two customers. One customer wants to view a list of homes selling for less than \$60,000. The other customer wants to view a list of homes selling for greater than \$100,000. Assuming the PRICE variable is numeric, which one of the following PRINT procedure steps will select all desired observations?
- A. `proc print data = sasuser.houses;`
`where price lt 60000;`
`where price gt 100000;`
`run;`
- B. `proc print data = sasuser.houses;`
`where price lt 60000 or price gt 100000;`
`run;`
- C. `proc print data = sasuser.houses;`
`where price lt 60000 and price gt 100000;`
`run;`
- D. `proc print data = sasuser.houses;`
`where price lt 60000 or where price gt 100000;`
`run;`
29. The value 110700 is stored in a numeric variable. Which one of the following SAS formats is used to display the value as \$110,700.00 in a report?
- A. `comma8.2`
- B. `comma11.2`
- C. `dollar8.2`
- D. `dollar11.2`
30. The SAS data set SASUSER.HOUSES contains a variable PRICE which has been assigned a permanent label of "Asking Price". Which one of the following SAS programs temporarily replaces the label "Asking Price" with the label "Sale Price" in the output?
- A. `proc print data = sasuser.houses;`
`label price = "Sale Price";`
`run;`
- B. `proc print data = sasuser.houses label;`
`label price "Sale Price";`
`run;`
- C. `proc print data = sasuser.houses label;`
`label price = "Sale Price";`
`run;`
- D. `proc print data = sasuser.houses label = "Sale Price";`
`run;`
31. The SAS data set BANKS is listed below:
- ```
BANKS
name rate
FirstCapital 0.0718
DirectBank 0.0721
VirtualDirect 0.0728
```
- The following SAS program is submitted:
- ```
data newbank;
do year = 1 to 3;
```

```
set banks;  
capital + 5000;  
end;  
run;
```

Which one of the following represents how many observations and variables will exist in the SAS data set NEWBANK?

- A. 0 observations and 0 variables
- B. 1 observations and 4 variables
- C. 3 observations and 3 variables
- D. 9 observations and 2 variables

32. The following SAS program is submitted:

```
data work.clients;  
calls = 6;  
do while (calls le 6);  
calls + 1;  
end;  
run;
```

Which one of the following is the value of the variable CALLS in the output data set?

- A. 4
- B. 5
- C. 6
- D. 7

33. The following SAS program is submitted:

```
data work.pieces;  
do while (n lt 6);  
n + 1;  
end;  
run;
```

Which one of the following is the value of the variable N in the output data set?

- A. 4
- B. 5
- C. 6
- D. 7

34. The following SAS program is submitted:

```
data work.sales;  
do year = 1 to 5;  
do month = 1 to 12;  
x + 1;  
end;  
end;  
run;
```

Which one of the following represents how many observations are written to the WORK.SALES data set?

- A. 0
- B. 1
- C. 5
- D. 60

35. A raw data record is listed below:

-----10-----20-----30

1999/10/25

The following SAS program is submitted:

```
data projectduration;  
infile 'file-specification';  
input date $ 1 - 10;
```

run;

Which one of the following statements completes the program above and computes the duration of the project in days as of today's date?

- A. duration = today() - put(date,ddmmyy10.);
- B. duration = today() - put(date,yymmdd10.);
- C. duration = today() - input(date,ddmmyy10.);
- D. duration = today() - input(date,yymmdd10.);

36. A raw data record is listed below:

-----10-----20-----30

Printing 750

The following SAS program is submitted:

```
data bonus;  
infile 'file-specification';  
input dept $ 1 - 11 number 13 - 15;
```

run;

Which one of the following SAS statements completes the program and results in a value of 'Printing750' for the DEPARTMENT variable?

- A. department = trim(dept) number;
- B. department = dept input(number,3.);
- C. department = trim(dept) || put(number,3.);
- D. department = input(dept,11.) || input(number,3.);

37. The following SAS program is submitted:

```
data work.month;  
date = put('13mar2000'd,ddmmyy10.);  
run;
```

Which one of the following represents the type and length of the variable DATE in the output data set?

- A. numeric, 8 bytes
- B. numeric, 10 bytes
- C. character, 8 bytes
- D. character, 10 bytes

38. The following SAS program is submitted:

```
data work.products;  
Product_Number = 5461;  
Item = '1001';  
Item_Reference = Item/'Product_Number';  
run;
```

Which one of the following is the value of the variable

ITEM_REFERENCE in the output data set?

A. 1001/5461

B. 1001/ 5461

C. . (missing numeric value)

D. The value can not be determined as the program fails to execute due to errors.

39. The following SAS program is submitted:

```
data work.retail;  
cost = '20000';  
total = .10 * cost;  
run;
```

Which one of the following is the value of the variable TOTAL in the output data set?

A. 2000

B. '2000'

C. . (missing numeric value)

D. ' ' (missing character value)

40. Which one of the following SAS statements correctly computes the average of four numerical values?

A. average = mean(num1 - num4);

B. average = mean(of num1 - num4);

C. average = mean(of num1 to num4);

D. average = mean(num1 num2 num3 num4);

41. The following SAS program is submitted:

```
data work.test;  
Author = 'Agatha Christie';  
First = substr(scan(author,1,' '),1,1);  
run;
```

Which one of the following is the length of the variable FIRST in the output data set?

A. 1

B. 6

C. 15

D. 200

42. The following SAS program is submitted:

```
data work.test;  
Author = 'Christie, Agatha';  
First = substr(scan(author,2,' '),1,1);  
run;
```

Which one of the following is the value of the variable FIRST in the output data set?

A. A

B. C

C. Agatha

D. ' ' (missing character value)

43. The following SAS program is submitted:

```
data work.test;  
Title = 'A Tale of Two Cities, Charles J. Dickens';  
Word = scan(title,3,' ');
```

run;

Which one of the following is the value of the variable WORD in the output data set?

- A. T
- B. of
- C. Dickens
- D. ' ' (missing character value)

44. The following SAS program is submitted:

```
data work.test;
First = 'Ipswich, England';
City_Country = substr(First,1,7)!!', '!!'England';
run;
```

Which one of the following is the length of the variable CITY_COUNTRY in the output data set?

- A. 6
- B. 7
- C. 17
- D. 25

45. The following SAS program is submitted:

```
data work.test;
First = 'Ipswich, England';
City = substr(First,1,7);
City_Country = City!!', '!!'England';
run;
```

Which one of the following is the value of the variable CITY_COUNTRY in the output data set?

- A. Ipswich!!
- B. Ipswich, England
- C. Ipswich, 'England'
- D. Ipswich , England

46. Which one of the following is true of the RETAIN statement in a SAS DATA step program?

- A. It can be used to assign an initial value to `_N_`.
- B. It is only valid in conjunction with a SUM function.
- C. It has no effect on variables read with the SET, MERGE and UPDATE statements.
- D. It adds the value of an expression to an accumulator variable and ignores missing values.

47. A raw data file is listed below:

```
-----10-----20-----30
```

```
1901 2
1905 1
1910 6
1925 .
1941 1
```

The following SAS program is submitted and references the raw data file above:

```
data coins;
infile 'file-specification';
```

```
input year quantity;
```

```
run;
```

Which one of the following completes the program and produces a non-missing value for the variable TOTQUANTITY in the last observation of the output data set?

- A. totquantity + quantity;
- B. totquantity = sum(totquantity + quantity);
- C. totquantity 0;
sum totquantity;
- D. retain totquantity 0;
totquantity = totquantity + quantity;

48. A raw data file is listed below:

```
-----10-----20-----30
```

```
squash 1.10
```

```
apples 2.25
```

```
juice 1.69
```

The following SAS program is submitted using the raw data file above:

```
data groceries;  
infile 'file-specification';  
input item $ cost;
```

```
run;
```

Which one of the following completes the program and produces a grand total for all COST values?

- A. grandtot = sum cost;
- B. grandtot = sum(grandtot,cost);
- C. retain grandtot 0;
grandtot = sum(grandtot,cost);
- D. grandtot = sum(grandtot,cost);
output grandtot;

49. The following SAS program is submitted:

```
data work.total;  
set work.salary(keep = department wagherate);  
by department;  
if first.department then payroll = 0;  
payroll + wagherate;  
if last.department;  
run;
```

The SAS data set WORK.SALARY, currently ordered by DEPARTMENT, contains 100 observations for each of 5 departments.

Which one of the following represents how many observations the WORK.TOTAL data set contains?

- A. 5
- B. 20
- C. 100
- D. 500

50. The following SAS program is submitted:

```
data work.total;
```

```

set work.salary(keep = department wagerate);
by department;
if first.department then payroll = 0;
payroll + wagerate;
if last.department;
run;

```

The SAS data set named WORK.SALARY contains 10 observations for each department, currently ordered by DEPARTMENT.

Which one of the following is true regarding the program above?

- A. The BY statement in the DATA step causes a syntax error.
- B. FIRST.DEPARTMENT and LAST.DEPARTMENT are variables in the WORK.TOTAL data set.
- C. The values of the variable PAYROLL represent the total for each department in the WORK.SALARY data set.
- D. The values of the variable PAYROLL represent a total for all values of WAGERATE in the WORK.SALARY data set.

ANSWERS :

1: d	11: b	21: d	31: b	41: d
2: a	12: a	22: b	32: d	42: a
3: c	13: b	23: c	33: c	43: b
4: d	14: d	24: b	34: b	44: d
5: d	15: d	25: a	35: d	45: d
6: b	16: b	26: a	36: c	46: c or d
7: b	17: b	27: c	37: d	47: a
8: b	18: d	28: b	38: d	48: c
9: d	19: d	29: d	39: a	49: a
10: d	20: c	30: c	40: b	50: d or c

Conclusion : So by now you'll have checked the answers and you know where you stand. So keep reading SAS and next post you have to get all correct.

Will be back with some more magic of SAS knowledge. Till then Goodbye.

SAS Clinical Interview QUESTIONS and ANSWERS

What is the therapeutic area you worked earlier?

There are so many diff. therapeutic areas a pharmaceutical company can work on and few of them include, anti-viral (HIV), Alzheimer's, Respiratory, Oncology, Metabolic Disorders (Anti-Diabetic), Neurological, Cardiovascular. Few more of them, include...

Central nervous system

Neurology

Gastroenterology

Ophthalmology

Orthopedics and pain control

Pulmonary

Vaccines

Dermatology

Gene therapy

Immunology etc

What are your responsibilities?

Some of them include; not necessarily all of them....

- Extracting the data from various internal and external database (Oracle, DB2, Excel spreadsheets) using SAS/ACCESS, SAS/INPUT.
- Developing programs in SAS Base for converting the Oracle Data for a phase II study into SAS datasets using SQL Pass through facility and Libname facility.
- Creating and deriving the datasets, listings and summary tables for Phase-I and Phase-II of clinical trials.
- Developing the SAS programs for listings & tables for data review & presentation including ad-hoc reports, CRTs as per CDISC, patients listing mapping of safety database and safety tables.
- Involved in mapping, pooling and analysis of clinical study data for safety.
- Using the Base SAS (MEANS, FREQ, SUMMARY, TABULATE, REPORT etc) and SAS/STAT procedures (REG, GLM, ANOVA, and UNIVARIATE etc.) for summarization, Cross-Tabulations and statistical analysis purposes.
- Developing the Macros at various instances for automating listings and graphing of clinical data for analysis.
- Validating and QC of the efficacy and safety tables.
- Creating the Ad hoc reports using the SAS procedures and used ODS statements and PROC TEMPLATE to generate different output formats like HTML, PDF and excel to view them in the web browser.
- Performing data extraction from various repositories and pre-process data when applicable.
- Creating the Statistical reports using Proc Report, Data _null_ and SAS Macro.
- Analyzing the data according to the Statistical Analysis Plan (SAP).
- Generating the demographic tables, adverse events and serious adverse events reports.
-

Created with



download the free trial online at nitropdf.com/professional

download the free trial online at nitropdf.com/professional

Can you tell me something about your last project study design?

If the interviewer asked you this question, then you need to tell that your current project is on a phase-1 study (or phase-2/Phase-3). You also need to tell about the name of the drug and the therapeutic area of it. Here are some more details you need to lay down in front of him...

- a) Is it a single blinded or double-blinded study?
- b) Is it a randomized or non-randomized study?
- c) How many patients are enrolled.
- d) Safety parameters only (if it is a phase-1)
- e) Safety and efficacy parameters if the study is either Phase-2,3or 4.

To get the all these details always refer www.clinicaltrials.gov .

How many subjects were there?

Subjects are nothing but the patients involved in the clinical study.

Answer to this question depends on the type of the study you have involved in.

If the study is phase1 answer should be approx. between 30-100.

If the study is phase2 answer should be approx. between 100-1000.

If the study is phase3 answer should be approx. between 1000-5000.

How many analyzed data sets did you create?

Again it depends on the study and the safety and efficacy parameters that are need to determined from the study. Approx. 20-30 datasets is required for a study to get analyzed for the safety and efficacy parameters. Here is some ex. of the datasets.

DM (Demographics), MH (Medical History), AE (Adverse Events), PE (Physical Education), EG (ECG), VS (Vital Signs), CM (Concomitant Medication), LB (Laboratory), QS (Questionnaire), IE (Inclusion and Exclusion), DS (Disposition), DT (Death), XT, SV, SC (Subject Characteristics), CO (Comments), EX (Exposure), PC, PP, TI (Therapeutic Intervention), SUPPCM, SUPPEX, SUPPLB, SUPPMH, SUPPXT, SUPPEG, etc.

How did you create analyzed data sets?

Analysis datasets are nothing but the datasets that are used for the statistical analysis of the data. Analysis datasets contains the raw data and the variables derived from the raw data. Variables, which are derived for the raw data, are used to produce the TLG's of the clinical study. The safety as well as efficacy endpoints (parameters) dictate the type of the datasets are required by the clinical study for generating the statistical reports of the TLG's. Sometimes the analysis datasets will have the variables not necessarily required to generate the statistical reports but sometimes they may required to generate the ad-hoc reports.

Refer also <http://www2.sas.com/proceedings/forum2008/207-2008.pdf> to get the complete info about creation of datasets:

How many tables, listings and graphs?

Can be in between 30-100 (including TLG's)

Created with



download the free trial online at nitropdf.com/professional
download the free trial online at nitropdf.com/professional

What do you mean by treatment emergent and treatment emergent serious adverse events?

Treatment emergent adverse events and Treatment emergent serious adverse events are nothing but the adverse events and serious adverse events which were happened after the drug administration or getting worsen by the drug, if patients are already having those adverse events before drug administration.

Can you explain something about the datasets?

DEMOGRAPHIC analysis dataset contains all subjects' demographic data (i.e., Age, Race, and Gender), disposition data (i.e., Date patient withdrew from the study), treatment groups and key dates such as date of first dose, date of last collected Case Report Form (CRF) and duration on treatment. The dataset has the format of one observation per subject.

LABORATORY analysis dataset contains all subjects' laboratory data, in the format of one observation per subject per test code per visit per accession number. Here, we derive the study visits according to the study window defined in the SAP, as well as re-grade the laboratory toxicity per protocol. For a crossover study, both the visit related to the initial period and as it is related to the beginning of the new study period will be derived. If the laboratory data are collected from multiple local lab centers, this analysis dataset will also centralize the laboratory data and standardize measurement units by using conversion factors.

EFFICACY analysis dataset contains derived primary and secondary endpoint variables as defined in the SAP. In addition, this dataset can contain other efficacy parameters of interest, such as censor variables pertaining to the time to an efficacy event. This dataset has the format of one record per subject per analysis period.

SAFETY can be categorized into four analysis datasets:

VITAL SIGN analysis dataset captures all subjects' vital signs collected during the trial. This dataset has the format of one observation per subject per vital sign per visit, similar to the structure for the laboratory analysis dataset.

ADVERSE EVENT analysis dataset contains all adverse events (AEs) reported including serious adverse events (SAEs) for all subjects. A treatment emergent flag, as well as a flag to indicate if an event is reported within 30 days after the subject permanently discontinued from the study, will be calculated. This dataset has a format of one record per subject per adverse event per start date. Partial dates and missing AEs start and/or stop dates will be imputed using logic defined in the SAP.

MEDICATION analysis dataset contains the subjects' medication records including concomitant medications and other medications taken either prior to the beginning of study or during the

Created with

study. This dataset has a format of one record per subject per medication taken per start date. Incomplete and missing medication start or stop dates will be imputed using instructions defined in the SAP.

SAFETY analysis dataset contains other safety variables, whether they are defined in the SAP or not. The Safety analysis dataset, similar to Efficacy analysis dataset in structure, consists of data with one record per subject per analysis period to capture safety parameters for all subjects.

It is crucial to generate analysis datasets in a specific order, as some variables derived from one particular analysis dataset may be used as the inputs to generate other variables in other analysis datasets. For example, the time to event variables in the efficacy and safety analysis datasets are calculated based on the date of the first dose derived in the demographic analysis dataset.

Analysis datasets are generated in sequence

Demographic _____ Laboratory _____ Efficacy

Vital Sign Safety

Adverse Event

Medications

Source: www.thotwave.com/Document/.../GlobalArch/SUGI117-30_GlobalArchitecture.pdf

What is your involvement while using CDISC standards? What is mean by CDISC where do you use it?

CDISC is nothing but an organization (Clinical Data Interchange Standards Consortium), which implements industrial standards for the pharmaceutical industries to submit the clinical data to FDA.

There are so many advantages of using CDISC standards: Reduced time for regulatory submissions, more efficient regulatory reviews of submission, savings in time and money on data transfers among business.

CDISC standards is used in following activities:

Developing CRTs for submitting them to FDA to get an NDA.

Mapping, pooling and analysis of clinical study data for safety.

Creating the annotated case report form (eCRF) using CDISC-SDTM mapping.

Creating the Analysis Datasets in CDISC and non-CDISC Standards for further SAS Programming.

What do you mean when you say you created tables, listings and graphs for ISS and ISE?

<http://studysas.blogspot.com/2008/09/what-you-should-know-about-issise-isr.html>

How do you do data cleaning?

It is always important to check the data we are using- especially for the variables what we are

Created with



using. Data cleaning is critical for the data we are using and preparing.
I use Proc Freq, Proc SQL, MEANS, UNIVARIATE etc to clean the data.

I will use Proc Print with WHERE statement to get the invalid date values.

Source: http://books.google.com/books?id=dyzAV8Miv5cC&dq=data+cleaning+techniques+in+SAS&printsec=frontcover&source=bl&ots=nDNyuK3tdi&sig=hWCujfLk53KAA7no8V_c4eu_6l&hl=en&sa=X&oi=book_result&resnum=9&ct=result#PPA32,M1

Can you tell me CRT's??

Creating Case Report Tabulations (CRTs) for an NDA Electronic Submission to the FDA
<http://www2.sas.com/proceedings/sugi25/25/ad/25p031.pdf>

ABSTRACT:The Food and Drug Administration (FDA) now strongly encourages all new drug applications (NDAs) be submitted electronically. Electronic submissions could help FDA application reviewers scan documents efficiently and check analyses by manipulating the very datasets and code used to generate them. The potential saving in reviewer time and cost is enormous while improving the quality of oversight. In January 1999, the FDA released the Guidance for Industry: Providing Regulatory Submissions in Electronic Format – NDAs. As described, one important part of the application package is the case report tabulations (CRTs), now serving as the instrument for submitting datasets. CRTs are made up of two parts: first, datasets in SAS® transport file format and second, the accompanying documentation for the datasets. Herein, we briefly review the content and conversion of datasets to SAS transport file format, and then elaborate on the code that makes easy work of the accompanying dataset documentation (in the form of data definition tables) using the SAS Output Delivery System (ODS). The intended audience is SAS programmers with an intermediate knowledge of the BASE product used under any operating system and who are involved in the biotechnology industries.

Where do you use MEdDra and WHO? Can you write a code? How do you use it?

What is MedDRA?

The Medical Dictionary for Regulatory Activities (MedDRA) has been developed as a pragmatic, clinically validated medical terminology with an emphasis on ease-of-use data entry, retrieval, analysis, and display, with a suitable balance between sensitivity and specificity, within the regulatory environment. MedDRA is applicable to all phases of drug development and the health effects of devices. By providing one source of medical terminology, MedDRA improves the effectiveness and transparency of medical product regulation worldwide.

MedDRA is used to report adverse event data from clinical trials, as well as post-marketing and pharmacovigilance.

What are the structural elements of the terminology in MedDRA?

The structural elements of the MedDRA terminology are as follows:

SOC - Highest level of the terminology, and distinguished by anatomical or physiological system, etiology, or purpose

Created with



HLGT – Subordinate to SOC, superordinate descriptor for one or more HLTs
HLT – Subordinate to HLGT, superordinate descriptor for one or more PTs
PT – Represents a single medical concept
LLT – Lowest level of the terminology, related to a single PT as a synonym, lexical variant, or quasi-synonym (Note: All PTs have an identical LLT).

In what format is MedDRA distributed?

MedDRA is distributed in sets of flat ASCII delimited files. There is a different set of files for each available language. The Czech translation is distributed in UTF-8 format. For detail information as to file names, data record scheme, and record layout, sees the MedDRA ASCII and Consecutive Files Documentation document, which can be downloaded from the MedDRA MSSO Web site. MedDRA is delivered in text file format. As of MedDRA Version 11.1, the total size of all ASCII files for the English version is 12,459KB.

Source: http://www.meddramsso.com/MSSOweb/faq/meddra.htm#What_is_MedDRA

THE WHODRUG DICTIONARY:

The WHODRUG dictionary was started in 1968. The dictionary contains information on both single and multiple ingredient medications. Drugs are classified according to the type of drug name being entered, (i.e. proprietary/trade name, nonproprietary name, chemical name, etc.). At present, 52 countries submit medication data to the WHO Collaborating Center, which is responsible for the maintenance and distribution of the drug dictionary. Updates to the dictionary are offered four times per year.

Source: <http://ssc.utexas.edu/docs/sashelp/sugj/24/Coders/p108-24.pdf>

What do you mean by used Macro facility to produce weekly and monthly reports?

The SAS macro facility can do lot of things and especially it is used to...

- reduce code repetition
- increase control over program execution
- minimize manual intervention
- create modular code.

Source: http://www.gasug.org/papers/DemystifyingMacro_Fecht.pdf

And also Visit http://instruct1.cit.cornell.edu/courses/econ321/public_html/sasinfo/macro.html to get more info about macro facility.

How did you validate table's, listings and what are the other things you validated?

First, the output from the listing needs to be read into a SAS data set. Next, the validation results need to be calculated (you need to do this anyway) and then turned into a SAS data set with the same layout and properties as the one created from the original output. Last, SAS compares the original versus validation data sets by using PROC COMPARE. The results are concise, quick, accurate and 100% complete. We have to use the same procedure to validate the Tables.

Source: http://www.lexjansen.com/pharmasug/2001/proceed/techtech/tt02_quttadauro.pdf

We will also validate graphs made in SAS... but to do that we need to use SAS/GRAPH Network Visualization Workshop and using it we can validate graphs made with SAS

Created with



automatically as well as manually.

Did you see anywhere that. Patient is randomized to one drug and the patient is given another drug? if you get in which population would you put that patient into?

Although, this situation is almost impossible, but if happens I will consider that patient in the group of the drug that he was given.

What would you do if you had to pool the data related to one parallel study and one cross over study?

Say If you have a same subject in two groups taking two different drugs.. and If you had to pool these two groups how would you do it?

This situation arises when the study is a cross over design study. I would consider the same patient as two different patients of each treatment group.

What are the phases you are good at?

Phase-I,II &III.

How would you transpose dataset using data step?

```
Proc transpose data=old out=new  
prefix=DATE;  
var date;  
by name;  
run;
```

The prefix= option controls the names for the transposed variables (DATE1, DATE2, etc.) Without it, the names of the new variables would be COL1, COL2, etc.

Actually, proc transpose creates an extra variable, `_NAME_`, indicating the name of the transposed variable. `_NAME_` has a value of DATE on both observations. To eliminate the extra variable, modify a portion of the proc statement:

```
out=new (drop=_name_)
```

The equivalent data step code using arrays could be:

```
data new (keep=name date1-date3);  
set old;  
by name;  
array dates {3} date1-date3;  
retain date1-date3;  
if first.name then i=1;
```

Created with

```
else i + 1;  
dates{i} = date;  
if last.name;  
run;
```

This program assumes that each name has exactly three observations. If a name had more, the program would generate an error message when hitting the fourth observation for that name. When $i=4$, this statement encounters an array subscript out of range:

```
dates{i} = date;  
source: http://sugme.org/papers/paper.rtf
```

If some patient misses one lab how would you assign values for that missing values?? Can you write the code?

Same answer as the below question....

How do you deal with missing values?

Whenever SAS encounters an invalid or blank value in the file being read, the value is defined as missing. In all subsequent processes and output, the value is represented as a period (if the variable is numeric-valued) or is left blank (if the variable is character-valued).

In DATA step programming, use a period to refer to missing numeric values.

For example, to recode missing values in the variable A to the value 99, use the following statement:

```
IF a=. THEN a=99;
```

Use the MISSING statement to define certain characters to represent special missing values for all numeric variables. The special missing values can be any of the 26 letters of the alphabet, or an underscore. In the example below, the values 'a' and 'b' will be interpreted as special missing values for every numeric variable.

```
MISSING a b ;
```

Source; <http://ssc.utexas.edu/consulting/answers/sas/sas33.html>

Did you ever create efficacy tables?

Yes, I have created Efficacy tables. Efficacy tables are developed to get an the information about primary objectives/parameters of the study.

What is the primary and secondary end point in your last project?

Anyone can download the protocol as well as trial SAP from my website (www.sas-india.blogspot.com) or else go to www.clinicaltrials.gov , and then type any Pharmaceutical company, u remember, it will give u the list of clinical trials conducted by that company, if you just click on any one study, you will be able to see the primary and secondary objectives and all other details.

Created with

What are the stat procedures you used?

ANOVA, CATMOD, FREQ, GLM, LIFEREG, LIFETEST, LOGISTIC, NPAR1WAY, REG, TTEST, UNIVARIATE, MEANS, SUMMARY etc

Tell me something about proc mixed? (Sometimes they may ask you to write the syntax)

Syntax: <http://ftp.sas.com/samples/A55235>

PROC MIXED is a generalization of the GLM procedure in the sense that PROC GLM fits standard linear models, and PROC MIXED fits the wider class of mixed linear models. Both procedures have similar CLASS, MODEL, CONTRAST, ESTIMATE, and LSMEANS statements, but their RANDOM and REPEATED statements differ (see the following paragraphs). Both procedures use the nonfull-rank model parameterization, although the sorting of classification levels can differ between the two. PROC MIXED computes only Type I -Type III tests of fixed effects, while PROC GLM offers Types I - IV. The RANDOM statement in PROC MIXED incorporates random effects constituting the vector in the mixed model. However, in PROC GLM, effects specified in the RANDOM statement are still treated as fixed as far as the model fit is concerned, and they serve only to produce corresponding expected mean squares.

Source: <http://www.otago.ac.nz/sas/stat/chap41/sect4.htm>

What would you do, if you have to use data step functions in macro definition? Can you use all the functions in data step in macro definition?

Yes.

If I have a dataset with different subjid's and each subjid has many records? How can I obtain last but one record for each patient?

Syntax:

```
Proc sort data=old;
```

```
By subjid;
```

```
Run;
```

```
Data new;
```

```
Set old;
```

```
By subjid;
```

```
If first.subjid;
```

```
Run;
```

Or

```
proc sort data=old out=new nodupkey;
```

```
by subjid;
```

```
run;
```

Can you get some value of a data step variable to be used in any other program you do later in the same SAS session? How do you do that?

Use a macro... with a %PUT statement.

What would you do if you have to access previous records values in current record?

Created with

Using ampersand sign.... &var.

What is a p value? Why should u calculate that? What are the procedures you can use for that?

If the p-value were greater than 0.05, you would say that the group of independent variables does not show a statistically significant relationship with the dependent variable, or that the group of independent variables does not reliably predict the dependent variable. Note that this is an overall significance test assessing whether the group of independent variables, when used together reliably predicts the dependent variable, and does not address the ability of any of the particular independent variables to predict the dependent variable. Using the Proc Freq, proc anova, proc GLM & Proc Ttest we cal calculate the p-value.

What do you usually do with proc life test?

Proc Lifetest is used to obtain Kaplan-Meier and life table survival estimates (and plots). Using a strata statement in Proc Lifetest, which compare survival estimates for different groups.

Can you get survival estimates with any other procedures?

PROC LIFEREG and PROC PHREG can be used to get the survival estimates.

Can you write a code to get the survival estimates?

```
proc lifetest data=data method=km outsurv=newdata;  
time survival*status(0);  
strata study;  
run;
```

What is the difference between stratum and by statement in Proc Lifetest?

You can specify a BY statement with PROC LIFETEST to obtain separate analyses on observations in groups defined by the BY variables.

The BY statement is more efficient than the STRATA statement for defining strata in large data sets. However, if you use the BY statement to define strata, PROC LIFETEST does not pool over strata for testing the association of survival time with covariates nor does it test for homogeneity across the BY groups.

The STRATA statement indicates which variables determine strata levels for the computations. The strata are formed according to the non-missing values of the designated strata variables. The MISSING option can be used to allow missing values as a valid stratum level.

Source: <http://www.math.wpi.edu/saspdf/stat/chap37.pdf>

Which procedure do you usually use to create reports?

Proc Report, proc Tabulate and Data _null_.

Created with

What do you do, if you had to get the column names and some title in every page of your report when you create it using data_null_?

Give your data_null_ titles the "proc print" and "proc report" feel

The more you can make your "data_null_" behave like "proc print" or "proc report", when it comes to titles, the better. If the "byline" option is set then put out a dashed "byline". If not, then don't. Does your "by" variable have a label? If so, then your dashed byline should have the text of your variable label in it on the left of the equals sign. If the variable has no label then it should just be the variable name. If that's the way "proc report" or "proc print" does it then do it that way with your "data_null_". Get it to interface with #byval and #byvar entries if they exist. Give people the feel that "data_null_" reporting is no different to using "proc print" or "proc report" and you will have less opposition to your "data_null_" reports. How you do this is already in those two pages. You are going to find yourself in a situation whereby you really must do the report using data_null_ but other people are not comfortable with it because they feel it is "too different" than using "proc report". The more you can give it the same feel, the more easily you can dip into "data_null_" when you have to without people worrying.

Source: <http://www.datasavantconsulting.com/roland/nulltech.html>

How do you use the macro which is created by some other people and which is in some other folder other than SAS?

With SAS Autocall library using the SAS Autos system.

Can you tell me something regarding the macro libraries?

Macro libraries are the libraries, which stores all the macros required for developing TLG's of the clinical trial. These are very necessary in controlling and managing the macros. With the help of a %INCLUDE statement; the stored macros in the macro library can be automatically called.

Can you show me how the efficacy table looks like?

<http://studysas.blogspot.com/2008/08/tif-samples.html>

Can you show me how the safety table looks like?

<http://studysas.blogspot.com/2008/08/tif-samples.html>

Did you use ODS?

Yes, I have used the ODS(Output Delivery System), which normally used to make the output from the Tables, Listings and graphs looks pretty. ODS creates the outputs in html, pdf and rtf formats.

General syntax:

Start the output with:

Ods output---format ;

SAS statements.....

.....

Ods output-format close;

Created with



download the free trial online at nitropdf.com/professional

Your resume says you created HTML, RTF, PDF? Why you had to create three?? Can you tell me in specific why each form is used?

There are several ways of format to create the SAS output.

To publish or to place the output on the Internet we need to create the output in HTML format, by converting the output into HTML files.

We generally create the SAS output in RTF, because the RTF can be opened in Word or other word processors.

If we need to send the printable reports through email, we need to create the output in PDF.

PDF output is also needed when we send documents required to file an NDA to FDA.

What are the graphs you created?

Survival estimate graphs.

What are the procedures you used to create them?

PROC LIFETEST, PROC GCHART, PROC GPLOT, PROC GREPLAY etc.

Can you generate statistics using Proc SQI?

Yes, we can generate the statistics like N, Mean, Median, Max, Min, STD & SUM using PROC SQL. But SQL procedure cannot calculate all the above statistics by default, as it is the case with PROC MEANS.

When do you prefer Proc SQI? Give me some situation?

The SQL procedure supports almost all the functions available in the DATA step for the creation of data as well as the manipulation of the data.

When we compare the same result, obtained from SQL and with the Data step, PROC SQL requires less code and, more importantly it requires less time to execute the code.

How do you delete a macro variable?

If the macro variable is stored in the library then it is a easy to delete it. Multiple variables may be deleted by placing the variable names in the DELETE statement:

Why do you have to use proc import and proc export wizards? Give me the situation?

These two help us to transfer the files/data between SAS and external data sources.

1) What do you know about CDISC and its standards?

CDISC stands for Clinical Data Interchange Standards Consortium and it is developed keeping in mind to bring great deal of efficiency in the entire drug development process.

CDISC brings efficiency to the entire drug development process by improving the data quality and speed-up the whole drug development process and to do that CDISC

developed a series of standards, which include Operation data Model (ODM), Study data Tabulation Model (SDTM) and the Analysis Data Model ADaM).

Created with

2) Why people these days are more talking about CDSIC and what advantages it brings to the Pharmaceutical Industry?

A) Generally speaking, Only about 30% of programming time is used to generate statistical results with SAS®, and the rest of programming time is used to familiarize data structure, check data accuracy, and tabulate/list raw data and statistical results into certain formats. This non-statistical programming time will be significantly reduced after implementing the CDISC standards.

3) What are the challenges as SAS programmer you think you will face when you first implement CDISC standards in you company?

A) With the new requirements of electronic submission, CRT datasets need to conform to a set of standards for facilitating reviewing process. They no longer are created solely for programmers convenient. SDS will be treated as specifications of datasets to be submitted, potentially as reference of CRF design. Therefore, statistical programming may need to start from this common ground. All existing programs/macros may also need to be remapped based on CDISC so one can take advantage to validate submission information by using tools which reviewer may use for reviewing and to accelerate reviewing process without providing unnecessary data, tables and listings. With the new requirements from updating electronic submission and CDISC implementation, understanding only SAS® may not be good enough to fulfill for final deliverables. It is a time to expand and enhance the job skills from various aspects under new change so that SAS® programmers can take a competitive advantage, and continue to play a main role in both statistical analysis and reporting for drug development.

References:

Pharmasug/2007/fc/fc05

pharmasug/2003/fda compliance/fda055

1) What do you understand about SDTM and its importance?

SDTM stands for Standard data Tabulation Model, which defines a standard structure for study data tabulations that are to be submitted as part of a product application to a regulatory authority such as the United States Food and Drug Administration (FDA) 2. In July 2004 the Clinical Data Interchange Standards Consortium (CDISC) published standards on the design and content of clinical trial tabulation data sets, known as the Study Data Tabulation Model (SDTM). According to the CDISC standard, there are four ways to represent a subject in a clinical study: tabulations, data listings, analysis datasets, and subject profiles6.

Before SDTM:

There are different names for each domain and domains don't have a standard

Created with



nitro PDF

professional

download the free trial online at nitropdf.com/professional

download the free trial online at nitropdf.com/professional

structure. There is no standard variables list for each and every domain.

Because of this FDA reviewers always had to take so much pain in understanding themselves with different data, domain names and name of the variable in each analysis dataset. Reviewers will have spent most of the valuable time in cleaning up the data into a standard format rather than reviewing the data for the accuracy. This process will delay the drug development process as such.

After SDTM:

There will be standard domain names and standard structure for each domain. There will be a list of standard variables and names for each and every dataset. Because of this, it will become easy to find and understand the data and reviewers will need less time to review the data than the data without SDTM standards. This process will improve the consistency in reviewing the data and it can be time efficient.

The purpose of creating SDTM domain data sets is to provide Case Report Tabulation (CRT) data FDA, in a standardized format. If we follow these standards it can greatly reduce the effort necessary for data mapping. Improper use of CDISC standards, such as using a valid domain or variable name incorrectly, can slow the metadata mapping process and should be avoided⁴.

2) PROC CDISC for SDTM 3.1 Format 2?

Syntax

The PROC CDISC syntax for CDISC SDTM is presented below. The DATA= parameter specifies the location of your SDTM conforming data source.

```
PROC CDISC  
MODEL=SDTM;SDTM SDTMVersion = "3.1";DOMAINDATA DATA = results. AE  
DOMAIN = AE CATEGORY = EVENT;RUN;
```

3) What are the capabilities of PROC CDISC 2?

PROC CDISC performs the following checks on domain content of the source:

- Verifies that all required variables are present in the data set
- Reports as an error any variables in the data set that are not defined in the domain
- Reports a warning for any expected domain variables that are not in the data set
- Notes any permitted domain variables that are not in the data set
- Verifies that all domain variables are of the expected data type and proper length
- Detects any domain variables that are assigned a controlled terminology specification by the domain and do not have a format assigned to them.

The procedure also performs the following checks on domain data content of the source on a per observation basis:

Verifies that all required variable fields do not contain missing values

Created with

Detects occurrences of expected variable fields that contain missing values
Detects the conformance of all ISO-8601 specification assigned values; including date, time, date time, duration, and interval types
Notes correctness of yes/no and yes/no/null responses,

4) What are the different approaches for creating the SDTM 3?

There are 3 general approaches to create the SDTM datasets:

- a) Build the SDTM entirely in the CDMS,
- b) Build the SDTM entirely on the “back-end” in SAS,
- c) or take a hybrid approach and build the SDTM partially in the CDMS and partially in SAS.

BUILD THE SDTM ENTIRELY IN THE CDMS

It is possible to build the SDTM entirely within the CDMS. If the CDMS allows for broad structural control of the underlying database, then you could build your eCRF or CRF based clinical database to SDTM standards.

Advantages:

- Your “raw” database is equivalent to your SDTM which provides the most elegant solution.
- Your clinical data management staff will be able to converse with end-users/sponsors about the data easily since your clinical data manager and the end-user/sponsor will both be looking at SDTM datasets.
- As soon as the CDMS database is built, the SDTM datasets are available.

Disadvantages:

- This approach may be cost prohibitive. Forcing the CDMS to create the SDTM structures may simply be too cumbersome to do efficiently.
- Forcing the CDMS to adapt to the SDTM may cause problems with the operation of the CDMS which could reduce data quality.

BUILD THE SDTM ENTIRELY ON THE “BACK-END” IN SAS

Assuming that SAS is not your CDMS solution, another approach is to take the clinical data from your CDMS and manipulate it into the SDTM with SAS programming.

Advantages:

- The great flexibility of SAS will let you transform any proprietary CDMS structure into the SDTM. You do not have to work around the rigid constraints of the CDMS.
- Changes could be made to the SDTM conversion without disturbing clinical data management processes.
- The CDMS is allowed to do what it does best which is to enter, manage, and clean data.

Created with

Disadvantages:

- There would be additional cost to transform the data from your typical CDMS structure into the SDTM.
Specifications, programming, and validation of the SAS programming transformation would be required.
- Once the CDMS database is up, there would then be a subsequent delay while the SDTM is created in SAS.

This delay would slow down the production of analysis datasets and reporting. This assumes that you follow the linear progression of CDMS -> SDTM -> analysis datasets (ADaM).

- Since the SDTM is a derivation of the “raw” data, there could be errors in translation from the “raw” CDMS data to the SDTM.
- Your clinical data management staff may be at a disadvantage when speaking with end-users/sponsors about the data since the data manager will likely be looking at the CDMS data and the sponsor will see SDTM data.

BUILD THE SDTM USING A HYBRID APPROACH

Again, assuming that SAS is not your CDMS solution, you could build some of the SDTM within the confines of the CDMS and do the rest of the work in SAS. There are things that could be done easily in the CDMS such as naming data tables the same as SDTM domains, using SDTM variable names in the CDMS, and performing simple derivations (such as age) in the CDMS. More complex SDTM derivations and manipulations can then be performed in SAS.

Advantages:

- The changes to the CDMS are easy to implement.
- The SDTM conversions to be done in SAS are manageable and much can be automated.

Disadvantages:

- There would still be some additional cost needed to transform the data from the SDTM-like CDMS structure into the SDTM. Specifications, programming, and validation of the transformation would be required.
- There would be some delay while the SDTM-like CDMS data is converted to the SDTM.
- Your clinical data management staff may still have a slight disadvantage when speaking with endusers/ sponsors about the data since the clinical data manager will be looking at the SDTM-like data and the sponsor will see the true SDTM data.

5) What do you know about SDTM domains?

Created with

A basic understanding of the SDTM domains, their structure and their interrelations is vital to determining which domains you need to create and in assessing the level to which your existing data is compliant. The SDTM consists of a set of clinical data file specifications and underlying guidelines. These different file structures are referred to as domains. Each domain is designed to contain a particular type of data associated with clinical trials, such as demographics, vital signs or adverse events.

The CDISC SDTM Implementation Guide provides specifications for 30 domains. The SDTM domains are divided into six classes.

The 21 clinical data domains are contained in three of these classes:

Interventions,
Events and
Findings.

The trial design class contains seven domains and the special-purpose class contains two domains (Demographics and Comments).

The trial design domains provide the reviewer with information on the criteria, structure and scheduled events of a clinical trial. The only required domain is demographics.

There are two other special purpose relationship data sets, the Supplemental Qualifiers (SUPPQUAL) data set and the Relate Records (RELREC) data set. SUPPQUAL is a highly normalized data set that allows you to store virtually any type of information related to one of the domain data sets. SUPPQUAL domain also accommodates variables longer than 200, the first 200 characters should be stored in the domain variable and the remaining should be stored in it5.

6) What are the general guidelines to SDTM variables?

Each of the SDTM domains has a collection of variables associated with it.

There are five roles that a variable can have:

Identifier,
Topic,
Timing,
Qualifier,
and for trial design domains,

Rule. Using lab data as an example, the subject ID, domain ID and sequence (e.g. visit) are identifiers.

The name of the lab parameter is the topic,
the date and time of sample collection are timing variables,
the result is a result qualifier and the variable containing the units is a variable qualifier.

Created with

Variables that are common across domains include the basic identifiers study ID (STUDYID), a two-character domain ID (DOMAIN) and unique subject ID (USUBJID).

In studies with multiple sites that are allowed to assign their own subject identifiers, the site ID and the subject ID must be combined to form USUBJID.

Prefixing a standard variable name fragment with the two-character domain ID generally forms all other variable names.

The SDTM specifications do not require all of the variables associated with a domain to be included in a submission. In regard to complying with the SDTM standards, the implementation guide specifies each variable as being included in one of three categories:

Required, Expected, and Permitted4.

REQUIRED – These variables are necessary for the proper functioning of standard software tools used by reviewers. They must be included in the data set structure and should not have a missing value for any observation.

EXPECTED – These variables must be included in the data set structure; however it is permissible to have missing values.

PERMISSIBLE – These variables are not a required part of the domain and they should not be included in the data set structure if the information they were designed to contain was not collected.

7) Can you tell me more About SDTM Domains5?

SDTM Domains are grouped by classes, which is useful for producing more meaningful relational schemas. Consider the following domain classes and their respective domains.

- Special Purpose Class – Pertains to unique domains concerning detailed information about the subjects in a study.

Demography (DM), Comments (CM)

- **Findings Class** – Collected information resulting from a planned evaluation to address specific questions about the subject, such as whether a subject is suitable to participate or continue in a study.

Electrocardiogram (EG)

Inclusion / Exclusion (IE)

Lab Results (LB)

Physical Examination (PE)

Questionnaire (QS)

Created with

Subject Characteristics (SC)
Vital Signs (VS)

• **Events Class** – Incidents independent of the study that happen to the subject during the lifetime of the study.

Adverse Events (AE)
Patient Disposition (DS)
Medical History (MH)

• **Interventions Class** – Treatments and procedures that are intentionally administered to the subject, such as treatment coincident with the study period, per protocol, or self-administered (e.g., alcohol and tobacco use).

Concomitant Medications (CM)
Exposure to Treatment Drug (EX)
Substance Usage (SU)

• **Trial Design Class** – Information about the design of the clinical trial (e.g., crossover trial, treatment arms) including information about the subjects with respect to treatment and visits.

Subject Elements (SE)
Subject Visits (SV)
Trial Arms (TA)
Trial Elements (TE)
Trial Inclusion / Exclusion Criteria (TI)

Trial Visits (TV)

7) Can you tell me how to do the Mapping for existing Domains?

First step is the comparison of metadata with the SDTM domain metadata. If the data getting from the data management is in somewhat compliance to SDTM metadata, use automated mapping as the 1st step.

If the data management metadata is not in compliance with SDTM then avoid auto mapping. So do manual mapping the datasets to SDTM datasets and the mapping each variable to appropriate domain.

The whole process of mapping include:

- *Read in the corporate data standards into a database table.
- Assign a CDISC domain prefix to each database module.
- Attach a combo box containing the SDTM variable for the selected domain to a new mapping variable field.

Created with

- Search each module, and within each module select the most appropriate CDISC variable.
- Then search for variables mapped to the wrong type Character not equal to Character; Numeric not equal to Numeric.
- Review the mapping to see if any conflicts are resolvable by mapping to a more appropriate variable.
- We need to verify that the mapped variable is appropriate for each role.
- Then finally we have to ensure all 'required' variables are present in the domain6.

8) What do you know about SDTM Implementation Guide, Have you used it, if you have can you tell me which version you have used so far?

SDTM Implementation guide provides documentation on metadata (data of data) for the domain datasets that includes filename, variable names, type of variables and its labels etc. I have used SDTM implementation guide version 3.1.1.

9) Can you identify which variables should we have to include in each domain?

A) SDTM implementation guide V 3.1.1 specifies each variable is being included in one of the 3 types.

REQUIRED – They must be included in the data set structure and should not have a missing value for any observation.

EXPECTED – These variables must be included in the data set; however it is permissible to have missing values.

PERMISSIBLE – These variables are not a required part of the domain and they should not be included in the data set structure if the information they were designed to contain was not collected.

10) Can you give some examples for MAPPING 6?

Here are some examples for SDTM mapping:

- Character variables defined as Numeric
- Numeric Variables defined as Character
- Variables collected without an obvious corresponding domain in the CDISC SDTM mapping. So must go into SUPPQUAL
- Several corporate modules that map to one corresponding domain in CDISC SDTM.
- Core SDTM is a subset of the existing corporate standards
- Vertical versus Horizontal structure, (e.g. Vitals)
- Dates – combining date and times; partial dates.
- Data collapsing issues e.g. Adverse Events and Concomitant Medications.
- Adverse Events maximum intensity
- Metadata needed to laboratory data standardization.

10) Explain the Process of SDTM Mapping?

A list of basic variable mappings is given below4.

DIRECT: a CDM variable is copied directly to a domain variable without any changes other than assigning the CDISC standard label.

RENAME: only the variable name and label may change but the contents remain the same.

STANDARDIZE: mapping reported values to standard units or standard terminology

REFORMAT: the actual value being represented does not change, only the format in which is stored changes, such as converting a SAS date to an ISO8601 format character string.

COMBINING: directly combining two or more CDM variables to form a single SDTM variable.

SPLITTING: a CDM variable is divided into two or more SDTM variables.

DERIVATION: creating a domain variable based on a computation, algorithm, series of logic rules or decoding using one or more CDM variables.

11) Can you explain AdAM or AdAM datasets??

The **Analysis Data Model** describes the general structure, metadata, and content typically found in Analysis Datasets and accompanying documentation. The three types of metadata associated with analysis datasets (analysis dataset metadata, analysis variable metadata, and analysis results metadata) are described and examples provided. (source:CDISC Analysis Data Model: Version 2.0)

Analysis datasets (AD) are typically developed from the collected clinical trial data and used to create statistical summaries of efficacy and safety data. These AD's are characterized by the creation of derived analysis variables and/or records. These derived data may represent a statistical calculation of an important outcome measure, such as change from baseline, or may represent the last observation for a subject while under therapy. As such, these datasets are one of the types of data sent to the regulatory agency such as FDA.

The CDISC **Analysis Data Model** (ADaM) defines a standard for Analysis Dataset's to be submitted to the regulatory agency. This provides a clear content, source, and quality of the datasets submitted in support of the statistical analysis performed by the sponsor.

In **ADaM**, the descriptions of the AD's build on the nomenclature of the SDTM with the addition of attributes, variables and data structures needed for statistical analyses. To achieve the principle of clear and unambiguous communication relies on clear AD

Created with



download the free trial online at nitropdf.com/professional

download the free trial online at nitropdf.com/professional

documentation. This documentation provides the link between the general description of the analysis found in the protocol or statistical analysis plan and the source data.

1. Have you used macros? For what purpose you have used?

Yes I have, I used macros in creating datasets and tables where it is necessary to make a small change through out the program where it is necessary to use the code and again.

2. How would you invoke a macro?

After I have defined a macro I can invoke it by adding the percent sign prefix to its name like this: % macro name a semicolon is not required when invoking a macro, though adding one generally does no harm.

3. How we can call macros with in data step?

We can call the macro with CALLSYMPUT

4. How do u identify a macro variable?

Ampersand (&)

5. How do you define the end of a macro?

The end of the macro is defined by %Mend Statement

6. For what purposes have you used SAS macros?

If we want use a program step for executing to execute the same Proc step on multiple data sets. We can accomplish repetitive tasks quickly and efficiently. A macro program can be reused many times. Parameters passed to the macro program customize the results without having to change the code within the macro program. Macros in SAS make a small change in the program and have SAS echo that change thought that program.

7. What is the difference between %LOCAL and % Global?

% Local is a macro variable defined inside a macro.%Global is a macro variable defined in open code (outside the macro or can use anywhere).

8. How long can a macro variable be? A token?

A component of SAS known as the word scanner breaks the program text into fundamental units called tokens. Tokens are passed on demand to the compiler. The compiler then requests token until it receives a semicolon. Then the compiler performs the syntax check on the statement.

9. If you use a SYMPUT in a DATA step, when and where can you use the macro variable?

Macro variable is used inside the Call Symput statement and is enclosed in quotes.

10. What do you code to create a macro? End one?

Created with

%MACRO and %MEND

11. What is the difference between %PUT and SYMBOLGEN?

%PUT is used to display user defined messages on log window after execution of a program where as % SYMBOLGEN is used to print the value of a macro variable resolved, on log window.

12. How do you add a number to a macro variable?

Using %eval function

13. Can you execute a macro within a macro? Describe.

Yes, Such macros are called nested macros. They can be obtained by using symget and call symput macros.

14. If you need the value of a variable rather than the variable itself what would you use to load the value to a macro variable?

If we need a value of a macro variable then we must define it in such terms so that we can call them everywhere in the program. Define it as Global. There are different ways of assigning a global variable. Simplest method is %LET.

Ex:A, is macro variable. Use following statement to assign the value of a rather than the variable itselfe.g.

```
%Let A=xyzx("&A");
```

This will assign "xyz" to x, not the variable xyz to x.

15. Can you execute macro within another macro? If so, how would SAS know where the current macro ended and the new one began?

Yes, I can execute macro within a macro, what we call it as nesting of macros, which is allowed. Every macro's beginning is identified the keyword %macro and end with %mend.

16. How are parameters passed to a macro?

A macro variable defined in parentheses in a %MACRO statement is a macro parameter. Macro parameters allow you to pass information into a macro.

Here is a simple example:

```
%macro plot(yvar= ,xvar= );  
proc plot;  
plot &yvar*&xvar;  
run;  
%mend plot;
```

17. How would you code a macro statement to produce information on the SAS log?

This statement can be coded anywhere?

OPTIONS, MPRINT MLOGIC MERROR SYMBOLGEN;

18. How we can call macros with in data step?

We can call the macro with

CALLSYMPUT,

Proc SQL and

%LET statement.

19. Tell me about call symput?

CALL SYMPUT takes a value from a data step and assigns it to a macro variable. I can then use this macro variable in later steps. To assign a value to a single macro variable,

I use CALL SYMPUT with this general form:

CALL SYMPUT ("macro-variable-name", value);

Where macro-variable-name, enclosed in quotation marks, is the name of a macro variable, either new or old, and value is the value I want to assign to that macro variable. Value can be the name of a variable whose value SAS will use, or it can be a constant value enclosed quotation marks.

CALL SYMPUT is often used in if-then statements such as this:

If age>=18 then call symput ("status","adult");

Else call symput ("status","minor");

These statements create a macro variable named &status and assign to it a value of either adult or minor depending on the variable age. Caution: We cannot create a macro variable with CALL SYMPUT and use it in the same data step because SAS does not assign a value to the macro variable until the data step executes. Data steps executes when SAS encounters a step boundary such as a subsequent data, proc, or run statement.

20. Tell me about % include and % eval?

The %include statement, despite its percent sign, is not a macro statement and is always executed in SAS, though it can be conditionally executed in a macro. It can be used to setting up a macro library. But this is a least approach.

The use of %include does not actually set up a library. The %include statement points to a file and when it executed the indicated file (be it a full program, macro definition, or a statement fragment) is inserted into the calling program at the location of the call. When using the %include building a macro library, the included file will usually contain one or more macro definitions. %EVAL is a widely used yet frequently misunderstood SAS(r) macro language function due to its seemingly simple form.

However, when its actual argument is a complex macro expression interlaced with special

Created with

characters, mixed arithmetic and logical operators, or macro quotation functions, its usage and result become elusive and problematic. %IF condition in macro is evaluated by %eval, to reduce it to true or false.

21. Describe the ways in which you can create macro variables?

There are the 5 ways to create macro variables:

%Let

%Global

Call Symput

Proc SQL

Parameters.

22. Tell me more about the parameters in macro?

Parameters are macro variables whose value you set when you invoke a macro. To add the parameters to a macro, you simply name the macro vars names parenthesis in the %macro statement.

Syntax:

```
%MACRO macro-name (parameter-1= , parameter-2= , .....parameter-n = );
```

```
macro-text%;
```

```
MEND macro-name;
```

23. What is the maximum length of the macro variable?

32 characters long.

24. Automatic variables for macro?

Every time we invoke SAS, the macro processor automatically creates certain macro var. eg: &sysdate &sysday.

25. What system options would you use to help debug a macro?

Debugging a Macro with SAS System Options. The SAS System offers users a number of useful system options to help debug macro issues and problems. The results associated with using macro options are automatically displayed on the SAS Log.

Specific options related to macro debugging appear in alphabetical order in the table below. SAS Option Description:

MEMRPT Specifies that memory usage statistics be displayed on the SAS Log.

MERROR: SAS will issue warning if we invoke a macro that SAS didn't find. Presents Warning Messages when there are misspellings or when an undefined macro is called.

SERROR: SAS will issue warning if we use a macro variable that SAS can't find.

MLOGIC: SAS prints details about the execution of the macros in the log.

MPRINT: Displays SAS statements generated by macro execution are traced on the SAS Log for debugging purposes.

SYMBOLGEN: SAS prints the value of macro variables in log and also displays text from expanding macro variables to the SAS Log.

Created with

26. If you need the value of a variable rather than the variable itself what would you use to load the value to a macro variable?

If we need a value of a macro variable then we must define it in such terms so that we can call them everywhere in the program. Define it as Global.

There are different ways of assigning a global variable.

Simplest method is %LET.

Ex:A, is macro variable. Use following statement to assign the value of a rather than the variable itselfe.g.

```
%let A=xyzx("&A");
```

This will assign "xyz" to x, not the variable xyz to x.

27. Can you execute macro within another macro? If so, how would SAS know where the current macro ended and the new one began?

Yes, I can execute macro within a macro, what we call it as nesting of macros, which is allowed. Every macro's beginning is identified the keyword %macro and end with %mend.

28. How are parameters passed to a macro?

A macro variable defined in parentheses in a %MACRO statement is a macro parameter. Macro parameters allow you to pass information into a macro. Here is a simple example:

```
%macro plot(yvar= ,xvar= );
```

```
proc plot;
```

```
plot &yvar*&xvar;
```

```
run;
```

```
%mend plot;
```

29. How would you code a macro statement to produce information on the SAS log?

This statement can be coded anywhere?

```
OPTIONS MPRINT MLOGIC MERROR SYMBOLGEN;
```

30. How we can call macros with in data step?

We can call the macro with CALLSYMPUT, Proc SQL and %LET statement.

31. What are SYMGET and SYMPUT?

SYMPUT puts the value from a dataset into a macro variable where as

SYMGET gets the value from the macro variable to the dataset.

32. What are the macros you have used in your programs?

Used macros for various puposes, few of them are..

1) Macros written to determine the list of variables in a dataset:

```
%macro varlist (dsn);
```

```
proc contents data = &dsn out = cont noprint;
```

```
run;
```

Created with

```
proc sql noprint;
select distinct name into:varname1-:varname22from cont;
quit;
```

```
%do i =1 %to &sqlobs;
%put &i &&varname&i;
%end;
%mend varlist;
%varlist(adverse)
```

2) Distribution or Missing / Non-Missing Values%macro missrep(dsn, vars=_numeric_);

```
proc freq data=&dsn.;
tables &vars. / missing;
format _character_ $missf. _numeric_ missf.;
title1 'Distribution or Missing / Non-Missing Values';
run;
%mend missrep;
%missrep(study.demog, vars=age gender bdate);
```

3) Written macros for sorting common variables in various datasets%macro sortit (datasetname, pid, investigator, timevisit)
PROC SORT DATA = &DATASETNAME;
BY &PID &INVESTIGATOR;
%mend sortit;

4) Macros written to split the number of observations in a dataset
%macro split (dsnorig, dsnsplit1, dsnsplit2, obs1);
data &dsnsplit1;
set &dsnorig (obs = &obs1);
run;data &dsnsplit2;
set &dsnorig (firstobs = %eval(&obs1 + 1));
run;%mend split;
%split(sasuser.admit,admit4,admit5,2)

33. What is auto call macro and how to create a auto call macro? What is the use of it? How to use it in SAS with macros?

SAS Enables the user to call macros that have been stored as SAS programs.

The auto call macro facility allows users to access the same macro code from multiple SAS programs. Rather than having the same macro code for in each program where the code is required, with an autocall macro, the code is in one location. This permits faster updates and

Created with

better consistency across all the programs. Macro set-up: The first step is to set-up a program that contains a macro, desired to be used in multiple programs. Although the program may contain other macros and/or open code, it is advised to include only one macro.

Set MAUTOSOURCE and SASAUTOS:

Before one can use the autocall macro within a SAS program, The MAUTOSOURCE option must be set open and the SASAUTOS option should be assigned. The MAUTOSOURCE option indicates to SAS that the autocall facility is to be activated. The SASAUTOS option tells SAS where to look for the macros. For ex: sasauto='g:\busmeas\internal\macro\';
What %put do? It displays the macro variable value when we specify %put (my first macro variable... is &.....)%
Put _automatic_ option displays all the SAS system macro variables including &SYSDATE AND &SYSTIME.

Created with



1. Describe the phases of clinical trials?

Ans:- These are the following four phases of the clinical trials:

Phase 1: Test a new drug or treatment to a small group of people (20-80) to evaluate its safety.

Phase 2: The experimental drug or treatment is given to a large group of people (100-300) to see that the drug is effective or not for that treatment.

Phase 3: The experimental drug or treatment is given to a large group of people (1000-3000) to see its effectiveness, monitor side effects and compare it to commonly used treatments.

Phase 4: The 4 phase study includes the post marketing studies including the drug's risk, benefits etc.

2. Describe the validation procedure? How would you perform the validation for TLG as well as analysis data set?

Ans:- Validation procedure is used to check the output of the SAS program, generated by the source programmer. In this process validator write the program and generate the output. If this output is same as the output generated by the SAS programmer's output then the program is considered to be valid. We can perform this validation for TLG by checking the output manually and for analysis data set it can be done using PROC COMPARE.

3. How would you perform the validation for the listing, which has 400 pages?

Ans:- It is not possible to perform the validation for the listing having 400 pages manually. To do this, we convert the listing in data sets by using PROC RTF and then after that we can compare it by using PROC COMPARE.

4. Can you use PROC COMPARE to validate listings? Why?

Ans:- Yes, we can use PROC COMPARE to validate the listing because if there are many entries (pages) in the listings then it is not possible to check them manually. So in this condition we use PROC COMPARE to validate the listings.

5. How would you generate tables, listings and graphs?

Ans:- We can generate the listings by using the PROC REPORT. Similarly we can create the tables by using PROC FREQ, PROC MEANS, and PROC TRANSPOSE and PROC REPORT. We would generate

graph, using proc Gplot etc.

6. How many tables can you create in a day?

Ans:- Actually it depends on the complexity of the tables if there are same type of tables then, we can create 1-2-3 tables in a day.

7. What are all the PROCs have you used in your experience?

Ans:- I have used many procedures like proc report, proc sort, proc format etc. I have used proc report to generate the list report, in this procedure I have used subjid as order variable and trt_grp, sbd, dbd as display variables.

8. Describe the data sets you have come across in your life?

Ans:- I have worked with demographic, adverse event , laboratory, analysis and other data sets.

9. How would you submit the docs to FDA? Who will submit the docs?

Ans:- We can submit the docs to FDA by e-submission. Docs can be submitted to FDA using

Define.pdf or define.Xml formats. In this doc we have the documentation about macros and program and E-records also. Statistician or project manager will submit this doc to FDA.

10. What are the docs do you submit to FDA?

Ans:- We submit ISS and ISE documents to FDA.

11. Can u share your CDISC experience? What version of CDISC SDTM have you used?

Ans: I have used version 1.1 of the CDISC SDTM.

12. Tell me the importance of the SAP?

Ans:- This document contains detailed information regarding study objectives and statistical methods to

Created with

aid in the production of the Clinical Study Report (CSR) including summary tables, figures, and subject data listings for Protocol. This document also contains documentation of the program variables and algorithms that will be used to generate summary statistics and statistical analysis.

13. Tell me about your project group? To whom you would report/contact?

My project group consisting of six members, a project manager, two statisticians, lead programmer and two programmers.

I usually report to the lead programmer. If I have any problem regarding the programming I would contact the lead programmer.

If I have any doubt in values of variables in raw dataset I would contact the statistician. For example the dataset related to the menopause symptoms in women, if the variable sex having the values like F, M. I would consider it as wrong; in that type of situations I would contact the statistician.

14. Explain SAS documentation.

SAS documentation includes programmer header, comments, titles, footnotes etc. Whatever we type in the program for making the program easily readable, easily understandable are in called as SAS documentation.

15. How would you know whether the program has been modified or not?

I would know the program has been modified or not by seeing the modification history in the program header.

16. Project status meeting?

It is a planetary meeting of all the project managers to discuss about the present Status of the project in hand and discuss new ideas and options in improving the Way it is presently being performed.

17. Describe clin-trial data base and oracle clinical

Clintrial, the market's leading Clinical Data Management System (CDMS). Oracle Clinical or OC is a database management system designed by Oracle to provide data management, data entry and data validation functionalities to Clinical Trials process. 18. Tell me about MEDRA and what version of MEDRA

did you use in your project? Medical dictionary of regulatory activities. Version 10

19. Describe SDTM?

CDISC's Study Data Tabulation Model (SDTM) has been developed to standardize what is submitted to the FDA.

20. What is CRT?

Case Report Tabulation, Whenever a pharmaceutical company is submitting an NDA, company has to send the CRT's to the FDA.

21. What is annotated CRF?

Case report form, it's a collection of the forms of all the patients in the trial.

22. What do you know about 21CFR PART 11?

Title 21 CFR Part 11 of the Code of Federal Regulations deals with the FDA guidelines on electronic records and electronic signatures in the United States. Part 11, as it is commonly called, defines the criteria under which electronic records and electronic signatures are considered to be trustworthy, reliable and equivalent to paper records.

23. Have you did validation in your projects?

I did validation of the fellow programmers work to ensure that the logic and intent of the program is correct and that data errors are detected. e.g. Verify error and warning messages are generated when the macro is called more than 10 times which means to add more than 10 titles.

Verify the error message when TITLENUM parameter is invalid. Verify a warning message is generated if the total length of texts specified in the input parameters LEFT, CENTER, and RIGHT is greater than 32 characters.

Also verify precedence is given to string in input parameter LEFT if the total string length is more than 32 characters. Verify there is no error/warning message generated if the macro is used within a data step and all input parameters are valid.

Created with

24. What are the contents of AE dataset? What is its purpose?

What are the variables in adverse event datasets?The adverse event data set contains the SUBJID, body system of the event, the preferred term for the event, event severity. The purpose of the AE dataset is to give a summary of the adverse event for all the patients in the treatment arms to aid in the inferential safety analysis of the drug.

25. What are the contents of lab data? What is the purpose of data set?

The lab data set contains the SUBJID, week number, and category of lab test, standard units, low normal and high range of the values. The purpose of the lab data set is to obtain the difference in the values of key variables after the administration of drug.

26.How did you do data cleaning? How do you change the values in the data on your own?

I used proc freq and proc univariate to find the discrepancies in the data, which I reported to my manager.

27.Have you created CRT's, if you have, tell me what have you done in that?

Yes I have created patient profile tabulations as the request of my manager and and the statistician. I have used PROC REPORT and Proc SQL to create simple patient listing which had all information of a particular patient including age, sex, race etc.

28. Have you created transport files?

Yes, I have created SAS Xport transport files using Proc Copy and data step for the FDA submissions. These are version 5 files. we use the libname engine and the Proc Copy procedure, One dataset in each xport transport format file. For version 5: labels no longer than 40 bytes, variable names 8 bytes, character variables width to 200 bytes. If we violate these constraints your copy procedure may terminate with constraints, because SAS xport format is in compliance with SAS 5 datasets.

```
Libname sdtm "c:\sdtm_data";Libname dm xport "c:\dm.xpt";  
Proc copy;  
In = sdtm;  
Out = dm;  
Select dm;  
Run;
```

29. How did you do data cleaning? How do you change the values in the data on your own?

I used proc freq and proc univariate to find the discrepancies in the data, which I reported to my manager.

30. Definitions?

CDISC- Clinical data interchange standards consortium. They have different data models, which define clinical data standards for pharmaceutical industry.

SDTM – It defines the data tabulation datasets that are to be sent to the FDA for regulatory submissions.

ADaM – (Analysis data Model) Defines data set definition guidance for creating analysis data sets.

ODM – XML – based data model for allows transfer of XML based data .

Define.xml – for data definition file (define.pdf) which is machine readable.

ICH E3: Guideline, Structure and Content of Clinical Study Reports

ICH E6: Guideline, Good Clinical Practice

ICH E9: Guideline, Statistical Principles for Clinical Trials

Title 21 Part 312.32: Investigational New Drug Application

31. have you ever done any Edit check programs in your project, if you have, tell me what do you know about edit check programs?

Yes I have done edit check programs .Edit check programs – Data validation.

1.Data Validation – proc means, proc univariate, proc freq.Data Cleaning – finding errors.

2.Checking for invalid character values.Proc freq data = patients; Tables gender dx ae / nocum nopercent; Run; Which gives frequency counts of unique character values.

3. Proc print with where statement to list invalid data values.[systolic blood pressure = 80 to 100|diastolic

blood pressure – 60 to 120]

4. Proc means, univariate and tabulate to look for outliers. Proc means – min, max, n and mean. Proc univariate – five highest and lowest values[stem leaf plots and box plots]

5. PROC FORMAT – range checking

6. Data Analysis – set, merge, update, keep, drop in data step.

7. Create datasets – PROC IMPORT and data step from flat files.

8. Extract data – LIBNAME.9. SAS/STAT – PROC ANOVA, PROC REG.

10. Duplicate Data – PROC SORT Nodupkey or NoduplicateNodupkey – only checks for duplicates in BYNoduplicate – checks entire observation (matches all variables)For getting duplicate observations first sort BY nodupkey and merge it back to the original dataset and keep only records in original and sorted.

11. For creating analysis datasets from the raw data sets I used the PROC FORMAT, and rename and length statements to make changes and finally make a analysis data set.

32. What is Verification?

The purpose of the verification is to ensure the accuracy of the final tables and the quality of SAS programs that generated the final tables. According to the instructions SOP and the SAP I selected the subset of the final summary tables for verification. E.g Adverse event table, baseline and demographic characteristics table. The verification results were verified against with the original final tables and all discrepancies if existed were documented.

33. What is ANNOTATED CRF?

An annotated CRF is a CRF in which the variable names are written next to the spaces provided for the investigator. It serves as a link between the database/data sets and the questions on the CRF.

34. What is Program Validation?

Its same as macro validation except here we have to validate the programs i.e according to the SOP I had to first determine what the program is supposed to do, see if they work as they are supposed to work and create a validation document mentioning if the program works properly and set the status as pass or fail. Pass the input parameters to the program and check the log for errors.

35. What do you know about ISS and ISE, have you ever produced these reports?

Created with

ISS (Integrated summary of safety):Integrates safety information from all sources (animal, clinical pharmacology, controlled and uncontrolled studies, epidemiologic data). "ISS is, in part, simply a summation of data from individual studies and, in part, a new analysis that goes beyond what can be done with individual studies."ISE (Integrated Summary of efficacy)ISS & ISE are critical components of the safety and effectiveness submission and expected to be submitted in the application in accordance with regulation. FDA's guidance Format and Content of Clinical and Statistical Sections of Application gives advice on how to construct these summaries. Note that, despite the name, these are integrated analyses of all relevant data, not summaries.

36. Explain the process and how to do Data Validation?

I have done data validation and data cleaning to check if the data values are correct or if they conform to the standard set of rules.A very simple approach to identifying invalid character values in this file is to use PROC FREQ to list all the unique values of these variables. This gives us the total number of invalid observations. After identifying the invalid data ...we have to locate the observation so that we can report to the manager the particular patient number.Invalid data can be located using the data _null_ programming.

Following is e.g

```
DATA _NULL_;
```

```
INFILE "C:PATIENTS.TXT" PAD;FILE PRINT; ***SEND OUTPUT TO THE OUTPUT WINDOW;
```

```
TITLE "LISTING OF INVALID DATA";
```

```
***NOTE: WE WILL ONLY INPUT THOSE VARIABLES OF INTEREST;INPUT @1 PATNO $3.@4  
GENDER $1.@24 DX $3.@27 AE $1.;
```

```
***CHECK GENDER;IF GENDER NOT IN ('F','M',' ') THEN PUT PATNO= GENDER=;
```

```
***CHECK DX;
```

```
IF VERIFY(DX,' 0123456789') NE 0
```

```
THEN PUT PATNO= DX=;
```

```
***CHECK AE;
```

```
IF AE NOT IN ('0','1',' ') THEN PUT PATNO= AE=;
```

```
RUN;
```

For data validation of numeric values like out of range or missing values I used proc print with a where statement.

```
PROC PRINT DATA=CLEAN.PATIENTS;  
WHERE HR NOT BETWEEN 40 AND 100 AND  
HR IS NOT MISSING OR  
SBP NOT BETWEEN 80 AND 200 AND  
SBP IS NOT MISSING OR
```

Created with

```
DBP NOT BETWEEN 60 AND 120 AND
DBP IS NOT MISSING;TITLE "OUT-OF-RANGE VALUES FOR NUMERICVARIABLES";
ID PATNO;
VAR HR SBP DBP;
RUN;
```

If we have a range of numeric values '001' – '999' then we can first use user defined format and then use proc freq to determine the invalid values.

```
PROC FORMAT;
VALUE $GENDER 'F','M' = 'VALID' ' ' = 'MISSING'OTHER = 'MISCODED';
VALUE $DX '001' - '999' = 'VALID' ' ' = 'MISSING'OTHER = 'MISCODED';
VALUE $AE '0','1' = 'VALID' ' ' = 'MISSING'OTHER = 'MISCODED';
RUN;
```

One of the simplest ways to check for invalid numeric values is to run either PROC MEANS or PROC UNIVARIATE. We can use the N and NMISS options in the Proc Means to check for missing and invalid data. Default (n nmiss mean min max stddev). The main advantage of using PROC UNIVARIATE (default n mean std skewness kurtosis) is that we get the extreme values i.e lowest and highest 5 values which we can see for data errors. If u want to see the patid for these particular observationsstate and ID patno statement in the univariate procedure.

37. Roles and responsibilities?

Programmer:

Develop programming for report formats (ISS & ISE shell) required by the regulatory authorities. Update ISS/ISE shell, when required.

Clinical Study Team:

Provide information on safety and efficacy findings, when required. Provide updates on safety and efficacy findings for periodic reporting.

Study Statistician

Draft ISS and ISE shell. Update shell, when appropriate. Analyze and report data in approved format, to meet periodic reporting requirements.

38. Explain Types of Clinical trials study you come across?

Single Blind Study

When the patients are not aware of which treatment they receive.

Double Blind Study

When the patients and the investigator are unaware of the treatment group assigned.

Triple Blind Study

Triple blind study is when patients, investigator, and the project team are unaware of the treatments administered.

39. What are the domains/datasets you have used in your studies?

Demog
Adverse Events
Vitals
ECG
Labs
Medical History
PhysicalExam etc

Created with

40. Can you list the variables in all the domains?

Demog: Usubjid, Patient Id, Age, Sex, Race, Screening Weight, Screening Height, BMI etc

Adverse Events: Protocol no, Investigator no, Patient Id, Preferred Term, Investigator Term, (Abdominal dis, Freq urination, headache, dizziness, hand-food syndrome, rash, Leukopenia, Neutropenia) Severity, Seriousness (y/n), Seriousness Type (death, life threatening, permanently disabling), Visit number, Start time, Stop time, Related to study drug?

Vitals: Subject number, Study date, Procedure time, Sitting blood pressure, Sitting Cardiac Rate, Visit number, Change from baseline, Dose of treatment at time of vital sign, Abnormal (yes/no), BMI, Systolic blood pressure, Diastolic blood pressure.

ECG: Subject no, Study Date, Study Time, Visit no, PR interval (msec), QRS duration (msec), QT interval (msec), QTc interval (msec), Ventricular Rate (bpm), Change from baseline, Abnormal.

Labs: Subject no, Study day, Lab parameter (Lparm), lab units, ULN (upper limit of normal), LLN (lower limit of normal), visit number, change from baseline, Greater than ULN (yes/no), lab related serious adverse event (yes/no). Medical History: Medical Condition, Date of Diagnosis (yes/no), Years of onset or occurrence, Past condition (yes/no), Current condition (yes/no).

PhysicalExam: Subject no, Exam date, Exam time, Visit number, Reason for exam, Body system, Abnormal (yes/no), Findings, Change from baseline (improvement, worsening, no change), Comments

41. Give me the example of edit checks you made in your programs? Examples of Edit Checks

Demog: Weight is outside expected range
Body mass index is below expected
(check weight and height)

Age is not within expected range.

DOB is greater than the Visit date or not..

Gender value is a valid one or invalid. etc

Adverse Event

Stop is before the start or visit Start is before birthdate Study medicine discontinued due to adverse event but completion indicated (COMPLETE =1)

Labs

Result is within the normal range but abnormal is not blank or 'N'
Result is outside the normal range but abnormal is blank

Vitals

Diastolic BP > Systolic BP

Medical History

Visit date prior to Screen date
Physical exam is normal but comment included

42. What are the advantages of using SAS in clinical data management? Why should not we use other software products in managing clinical data?

ADVANTAGES OF USING A SAS®-BASED SYSTEM

Less hardware is required.

Created with

A Typical SAS®-based system can utilize a standard file server to store its databases and does not require one or more dedicated servers to handle the application load. PC SAS® can easily be used to handle processing, while data access is left to the file server. Additionally, as presented later in this paper, it is possible to use the SAS® product SAS®/Share to provide a dedicated server to handle data transactions.

Fewer personnel are required.

Systems that use complicated database software often require the hiring of one or more DBA's (Database Administrators) who make sure the database software is running, make changes to the structure of the database, etc. These individuals often require special training or background experience in the particular database application being used, typically Oracle. Additionally, consultants are often required to set up the system and/or studies since dedicated servers and specific expertise requirements often complicate the process. Users with even casual SAS® experience can set up studies. Novice programmers can build the structure of the database and design screens. Organizations that are involved in data management almost always have at least one SAS® programmer already on staff. SAS® programmers will have an understanding of how the system actually works which would allow them to extend the functionality of the system by directly accessing SAS® data from outside of the system. Speed of setup is dramatically reduced. By keeping studies on a local file server and making the database and screen design processes extremely simple and intuitive, setup time is reduced from weeks to days. All phases of the data management process become homogeneous. From entry to analysis, data reside in SAS® data sets, often the end goal of every data management group. Additionally, SAS® users are involved in each step, instead of having specialists from different areas hand off pieces of studies during the project life cycle. No data conversion is required. Since the data reside in SAS® data sets natively, no conversion programs need to be written. Data review can happen during the data entry process, on the master database. As long as records are marked as being double-keyed, data review personnel can run edit check programs and build queries on some patients while others are still being entered. Tables and listings can be generated on live data. This helps speed up the development of table and listing programs and allows programmers to avoid having to make continual copies or extracts of the data during testing.

43. Have you ever had to follow SOPs or programming guidelines? SOP describes the process to assure that standard coding activities, which produce tables, listings and graphs, functions and/or edit checks, are conducted in accordance with industry standards and are appropriately documented. It is normally used whenever new programs are required or existing programs require some modification during the set-up, conduct, and/or reporting clinical trial data.

44. Describe the types of SAS programming tasks that you performed: Tables? Listings? Graphics? Ad hoc reports? Other? Prepared programs required for the ISS and ISE analysis reports. Developed and validated programs for preparing ad-hoc statistical reports for the preparation of clinical study report. Wrote analysis programs in line with the specifications defined by the study statistician. Base SAS (MEANS, FREQ, SUMMARY, TABULATE, REPORT etc) and SAS/STAT procedures (REG, GLM, ANOVA, and UNIVARIATE etc.) were used for summarization, Cross-Tabulations and statistical analysis purposes. Created Statistical reports using Proc Report, Data _null_ and SAS Macro. Created, derived and merged and pooled datasets, listings and summary tables for Phase-I and Phase-II of clinical trials.

45. Have you been involved in editing the data or writing data queries? If your interviewer asks this question, the user should ask him what he means by editing the data... and data queries...

46. Are you involved in writing the inferential analysis plan? Table's specifications?

47. What do you feel about hardcoding?

Programmers sometime hardcode when they need to produce report in urgent. But it is always better to avoid hardcoding, as it overrides the database controls in clinical data management. Data often change in a trial over time, and the hardcode that is written today may not be valid in the future. Unfortunately,

Created with

hardcode may be forgotten and left in the SAS program, and that can lead to an incorrect database change.

48. How do you write a test plan?

Before writing "Test plan" you have to look into on "Functional specifications". Functional specifications itself depends on "Requirements", so one should have clear understanding of requirements and functional specifications to write a test plan.

49. What is the difference between verification and validation?

Although the verification and validation are close in meaning, "verification" has more of a sense of testing the truth or accuracy of a statement by examining evidence or conducting experiments, while "validate" has more of a sense of declaring a statement to be true and marking it with an indication of official sanction.

50. What other SAS features do you use for error trapping and data validation?

Conditional statements, if then else.

Put statement

Debug option.

51. What is PROC CDISC?

It is new SAS procedure that is available as a hotfix for SAS 8.2 version and comes as a part with SAS 9.1.3 version.

PROC CDISC is a procedure that allows us to import (and export XML files that are compliant with the CDISC ODM version 1.2 schema.

For more details refer SAS programming in the Pharmaceutical Industry text book.

52) What is LOCF?

Pharmaceutical companies conduct longitudinal studies on human subjects that often span several months. It is unrealistic to expect patients to keep every scheduled visit over such a long period of time. Despite every effort, patient data are not collected for some time points. Eventually, these become missing values in a SAS data set later. For reporting purposes, the most recent previously available value is substituted for each missing visit. This is called the Last Observation Carried Forward (LOCF). LOCF doesn't mean last SAS dataset observation carried forward. It means last non-missing value carried forward. It is the values of individual measures that are the "observations" in this case. And if you have multiple variables containing these values then they will be carried forward independently.

53) ETL process:

Extract, transform and Load

The 1st part of an ETL process is to extract the data from the source systems. Most data warehousing projects consolidate data from different source systems.

Each separate system may also use a different data organization / format. Common data source formats are relational databases and flat files, but may include non-relational database structures such as IMS or other data structures such as VSAM or ISAM.

Extraction converts the data into a format for transformation processing. An intrinsic part of the extraction

Created with

is the parsing of extracted data, resulting in a check if the data meets an expected pattern

Transform:The transform stage applies a series of rules or functions to the extracted data from the source to derive the data to be loaded to the end target. Some data sources will require very little or even no manipulation of data. In other cases, one or more of the following transformations types to meet the business and technical needs of the end target may be required:

Selecting only certain columns to load (or selecting null columns not to load) · Translating coded values (e.g., if the source system stores 1 for male and 2 for female, but the warehouse stores M for male and F for female), this is called automated data cleansing; no manual cleansing occurs during ETL · Encoding free-form values (e.g., mapping "Male" to "1" and "Mr" to M) ·

Joining together data from multiple sources (e.g., lookup, merge, etc.) · Generating surrogate key values · Transposing or pivoting (turning multiple columns into multiple rows or vice versa) · Splitting a column into multiple columns (e.g., putting a comma-separated list specified as a string in one column as individual values in different columns) ·

Applying any form of simple or complex data validation; if failed, a full, partial or no rejection of the data, and thus no, partial or all the data is handed over to the next step, depending on the rule design and exception handling. Most of the above transformations itself might result in an exception, e.g. when a code-translation parses an unknown code in the extracted data.
Load:The load phase loads the data into the end target, usually being the data warehouse (DW).

Depending on the requirements of the organization, this process ranges widely. Some data warehouses might weekly overwrite existing information with cumulative, updated data, while other DW (or even other parts of the same DW) might add new data in a historized form, e.g. hourly. The timing and scope to replace or append are strategic design choices dependent on the time available and the business needs. More complex systems can maintain a history and audit trail of all changes to the data loaded in the DW.

As the load phase interacts with a database, the constraints defined in the database schema as well as in triggers activated upon data load apply (e.g. uniqueness, referential integrity, mandatory fields), which also contribute to the overall data quality performance of the ETL process.

Under what circumstances would you code a SELECT construct instead of IF statements?

A: I think Select statement is used when you are using one condition to compare with several conditions like.....

```
Data exam;  
Set exam;  
select (pass);  
when Physics >60;  
when math > 100;  
when English = 50;  
otherwise fail;  
run;
```

What is the one statement to set the criteria of data that can be coded in any step?

A) Options statement.

Created with

What is the effect of the OPTIONS statement ERRORS=1?

A) The -ERROR- variable has a value of 1 if there is an error in the data for that observation and 0 if it is not.

What's the difference between VAR A1 - A4 and VAR A1 -- A4?

A) There is no difference between VAR A1-A4 and VAR A1—A4. Whereas if we submit VAR A1---A4 instead of VAR A1-A4 or VAR A1—A3, you will see an error message in the log.

What do the SAS log messages "numeric values have been converted to character" mean? What are the implications?

A) It implies that automatic conversion took place to make character functions possible.

Why is a STOP statement needed for the POINT= option on a SET statement?

A) Because POINT= reads only the specified observations, SAS cannot detect an end-of-file condition as it would if the file were being read sequentially.

How do you control the number of observations and/or variables read or written?

A) FIRSTOBS and OBS option

Approximately what date is represented by the SAS date value of 730?

A) 31st December 1961

Identify statements whose placement in the DATA step is critical.

A) INPUT, DATA and RUN...

Does SAS 'Translate' (compile) or does it 'Interpret'? Explain.

A) Compile

What does the RUN statement do?

A) When SAS editor looks at Run it starts compiling the data or proc step, if you have more than one data step or proc step or if you have a proc step. Following the data step then you can avoid the usage of the run statement.

Why is SAS considered self-documenting?

A) SAS is considered self-documenting because during the compilation time it creates and stores all the information about the data set like the time and date of the data set creation, later No. of the variables, later labels, all that kind of info inside the dataset and you can look at that info using proc contents procedure.

What are some good SAS programming practices for processing very large data sets?

A) Sort them once, can use firstobs = and obs = ,

What is the difference between functions and PROCs that calculate the same simple descriptive statistics?

Created with

A) Functions can be used inside the data step and on the same data set but with proc's you can create a new data set to output the results. May be more

If you were told to create many records from one record, show how you would do this using arrays and with PROC TRANSPOSE?

A) I would use TRANSPOSE if the variables are less use arrays if the variables are more depends

What is a method for assigning first.VAR and last.VAR to the BY groupvariable on unsorted data?

A) In Unsorted data you can't use First. or Last.

How do you debug and test your SAS programs?

A) First thing is look into Log for errors or warning or NOTE in some cases or use the debugger in SAS data step.

What other SAS features do you use for error trapping and data validation?

A) Check the Log and for data validation things like Proc Freq, Proc means or some times proc print to look how the data looks like

How would you combine 3 or more tables with different structures?

A) I think sort them with common variables and use merge statement. I am not sure what you mean different structures.

Other questions:

What areas of SAS are you most interested in?

A) BASE, STAT, GRAPH, ETS Briefly

Describe 5 ways to do a "table lookup" in SAS.

A) Match Merging, Direct Access, Format Tables, Arrays, PROC SQL

What versions of SAS have you used (on which platforms)?

A) SAS 9.1.3, 9.0, 8.2 in Windows and UNIX, SAS 7 and 6.12

What are some good SAS programming practices for processing very large data sets?

A) Sampling method using OBS option or subsetting, commenting the Lines, Use Data Null

What are some problems you might encounter in processing missing values? In Data steps? Arithmetic? Comparisons? Functions? Classifying data?

A) The result of any operation with missing value will result in missing value. Most SAS statistical procedures exclude observations with any missing variable values from an analysis.

How would you create a data set with 1 observation and 30 variables from a data set with 30 observations and 1 variable?

A) Using PROC TRANSPOSE

What is the difference between functions and PROCs that calculate the same simple descriptive statistics?

A) Proc can be used with wider scope and the results can be sent to a different dataset. Functions usually

Created with

affect the existing datasets.

If you were told to create many records from one record, show how you would do this using array and with PROC TRANSPOSE?

A) Declare array for number of variables in the record and then used Do loop Proc Transpose with VAR statement

What are _numeric_ and _character_ and what do they do?

A) Will either read or writes all numeric and character variables in dataset.

How would you create multiple observations from a single observation?

A) Using double Trailing @@

For what purpose would you use the RETAIN statement?

A) The retain statement is used to hold the values of variables across iterations of the data step. Normally, all variables in the data step are set to missing at the start of each iteration of the data step. What is the order of evaluation of the comparison operators: + - * / ** ()? A) (), **, *, /, +, -

How could you generate test data with no input data?

A) Using Data Null and put statement

How do you debug and test your SAS programs?

A) Using Obs=0 and systems options to trace the program execution in log.

What can you learn from the SAS log when debugging?

A) It will display the execution of whole program and the logic. It will also display the error with line number so that you can and edit the program.

What is the purpose of _error_?

A) It has only to values, which are 1 for error and 0 for no error.

How can you put a "trace" in your program?

A) By using ODS TRACE ON

How does SAS handle missing values in: assignment statements, functions, a merge, an update, sort order, formats, PROCs?

A) Missing values will be assigned as missing in Assignment statement. Sort order treats missing as second smallest followed by underscore.

How do you test for missing values?

A) Using Subset functions like IF then Else, Where and Select.

How are numeric and character missing values represented internally?

A) Character as Blank or " and Numeric as.

Which date functions advances a date time or date/time value by a given interval?

A) INTNX.

Created with

In the flow of DATA step processing, what is the first action in a typical DATA Step?

A) When you submit a DATA step, SAS processes the DATA step and then creates a new SAS data set. (creation of input buffer and PDV)

Compilation Phase

Execution Phase

What are SAS/ACCESS and SAS/CONNECT?

A) SAS/Access only process through the databases like Oracle, SQL-server, Ms-Access etc.

SAS/Connect only use Server connection.

What is the one statement to set the criteria of data that can be coded in any step?

A) OPTIONS Statement, Label statement, Keep / Drop statements.

What is the purpose of using the N=PS option?

A) The N=PS option creates a buffer in memory which is large enough to store PAGESIZE (PS) lines and enables a page to be formatted randomly prior to it being printed.

What are the scrubbing procedures in SAS?

A) Proc Sort with nodupkey option, because it will eliminate the duplicate values.

What are the new features included in the new version of SAS i.e., SAS9.1.3?

A) The main advantage of version 9 is faster execution of applications and centralized access of data and support.

There are lots of changes has been made in the version 9 when we compared with the version 8. The following are the few: SAS version 9 supports Formats longer than 8 bytes & is not possible with version 8.

Length for Numeric format allowed in version 9 is 32 where as 8 in version 8.

Length for Character names in version 9 is 31 where as in version 8 is 32.

Length for numeric informat in version 9 is 31, 8 in version 8.

Length for character names is 30, 32 in version 8.3 new informats are available in version 9 to convert various date, time and datetime forms of data into a SAS date or SAS time.

· ANYDTEW. - Converts to a SAS date value · ANYDTTMEW. - Converts to a SAS time value.

· ANYDTEW. - Converts to a SAS datetime value. CALL SYMPUTX Macro statement is added in the version 9 which creates a macro variable at execution time in the data step by ·

Trimming trailing blanks · Automatically converting numeric value to character.

New ODS option (COLUMN OPTION) is included to create a multiple columns in the output.

WHAT DIFFERENCE DID YOU FIND AMONG VERSION 6 8 AND 9 OF SAS.

The SAS 9

A) Architecture is fundamentally different from any prior version of SAS. In the SAS 9 architecture, SAS relies on a new component, the Metadata Server, to provide an information layer between the programs and the data they access. Metadata, such as security permissions for SAS libraries and where the various SAS servers are running, are maintained in a common repository.

What has been your most common programming mistake?

A) Missing semicolon and not checking log after submitting program,

Created with

Not using debugging techniques and not using Fsview option vigorously.

Name several ways to achieve efficiency in your program.

Efficiency and performance strategies can be classified into 5 different areas.

- CPU time
- Data Storage
- Elapsed time
- Input/Output
- Memory CPU Time and Elapsed Time- Base line measurements

Few Examples for efficiency violations:

Retaining unwanted datasets Not sub setting early to eliminate unwanted records.

Efficiency improving techniques:

A)

Using KEEP and DROP statements to retain necessary variables. Use macros for reducing the code.

Using IF-THEN/ELSE statements to process data programming.

Use SQL procedure to reduce number of programming steps.

Using of length statements to reduce the variable size for reducing the Data storage.

Use of Data _NULL_ steps for processing null data sets for Data storage.

What other SAS products have you used and consider yourself proficient in using?

B) A) Data _NULL_ statement, Proc Means, Proc Report, Proc tabulate, Proc freq and Proc print, Proc Univariate etc.

What is the significance of the 'OF' in X=SUM (OF a1-a4, a6, a9);

A) If don't use the OF function it might not be interpreted as we expect. For example the function above calculates the sum of a1 minus a4 plus a6 and a9 and not the whole sum of a1 to a4 & a6 and a9. It is true for mean option also.

What do the PUT and INPUT functions do?

A) INPUT function converts character data values to numeric values.

PUT function converts numeric values to character values. EX: for INPUT: INPUT (source, informat)

For PUT: PUT (source, format)

Note that INPUT function requires INFORMAT and PUT function requires FORMAT.

If we omit the INPUT or the PUT function during the data conversion, SAS will detect the mismatched variables and will try an automatic character-to-numeric or numeric-to-character conversion. But sometimes this doesn't work because \$ sign prevents such conversion. Therefore it is always advisable to include INPUT and PUT functions in your programs when conversions occur.

Which date function advances a date, time or datetime value by a given interval?

INTNX:

INTNX function advances a date, time, or datetime value by a given interval, and returns a date, time, or datetime value. Ex: INTNX(interval,start-from,number-of-increments,alignment)

INTCK: INTCK(interval,start-of-period,end-of-period) is an interval function counts the number of intervals between two give SAS dates, Time and/or datetime.

DATETIME () returns the current date and time of day.

Created with

DATDIF (sdate,edate,basis): returns the number of days between two dates.

What do the MOD and INT function do? What do the PAD and DIM functions do? MOD:

A) Modulo is a constant or numeric variable, the function returns the remainder after numeric value divided by modulo.

INT: It returns the integer portion of a numeric value truncating the decimal portion.

PAD: it pads each record with blanks so that all data lines have the same length. It is used in the INFILE statement. It is useful only when missing data occurs at the end of the record.

CATX: concatenate character strings, removes leading and trailing blanks and inserts separators.

SCAN: it returns a specified word from a character value. Scan function assigns a length of 200 to each target variable.

SUBSTR: extracts a sub string and replaces character values.Extraction of a substring:
Middleinitial=substr(middlename,1,1); Replacing character values: substr (phone,1,3)='433'; If SUBSTR function is on the left side of a statement, the function replaces the contents of the character variable.

TRIM: trims the trailing blanks from the character values.

SCAN vs. SUBSTR: SCAN extracts words within a value that is marked by delimiters. SUBSTR extracts a portion of the value by stating the specific location. It is best used when we know the exact position of the sub string to extract from a character value.

How might you use MOD and INT on numeric to mimic SUBSTR on character Strings?

A) The first argument to the MOD function is a numeric, the second is a non-zero numeric; the result is the remainder when the integer quotient of argument-1 is divided by argument-2. The INT function takes only one argument and returns the integer portion of an argument, truncating the decimal portion. Note that the argument can be an expression.

```
DATA NEW ;A = 123456 ;  
X = INT( A/1000 ) ;  
Y = MOD( A, 1000 ) ;  
Z = MOD( INT( A/100 ), 100 ) ;  
PUT A= X= Y= Z= ;  
RUN ;  
A=123456X=123Y=456Z=34
```

In ARRAY processing, what does the DIM function do?

A) DIM: It is used to return the number of elements in the array. When we use Dim function we would have to re-specify the stop value of an iterative DO statement if u change the dimension of the array.

How would you determine the number of missing or nonmissing values in computations?

A) To determine the number of missing values that are excluded in a computation, use the NMISS function.

data _null_;

Created with

```

m = . ;
y = 4 ;
z = 0 ;
N = N(m , y, z);
NMISS = NMISS (m , y, z);
run;

```

The above program results in N = 2 (Number of non missing values) and NMISS = 1 (number of missing values).

Do you need to know if there are any missing values?

A) Just use: `missing_values=MISSING(field1,field2,field3);`

This function simply returns 0 if there aren't any or 1 if there are missing values. If you need to know how many missing values you have then use `num_missing=NMISS(field1,field2,field3);`

You can also find the number of non-missing values with `non_missing=N (field1,field2,field3);`

What is the difference between: `x=a+b+c+d;` and `x=SUM (of a, b, c ,d);?`

A) Is anyone wondering why you wouldn't just use `total=field1+field2+field3;`

First, how do you want missing values handled?

The SUM function returns the sum of non-missing values. If you choose addition, you will get a missing value for the result if any of the fields are missing. Which one is appropriate depends upon your needs. However, there is an advantage to use the SUM function even if you want the results to be missing. If you have more than a couple fields, you can often use shortcuts in writing the field names. If your fields are not numbered sequentially but are stored in the program data vector together then you can use: `total=SUM(of fielda--zfield);` Just make sure you remember the "of" and the double dashes or your code will run but you won't get your intended results. Mean is another function where the function will calculate differently than the writing out the formula if you have missing values. There is a field containing a date. It needs to be displayed in the format "ddmmyy" if it's before 1975, "dd mon ccy" if it's after 1985, and as 'Disco Years' if it's between 1975 and 1985.

How would you accomplish this in data step code?

Using only PROC FORMAT.

```

data new ;
input date ddmmyy10.
;
cards;
01/05/1955
01/09/1970
01/12/1975
19/10/1979
25/10/1982
10/10/1988
27/12/1991;
run;

```

```

proc format ;
value dat low-'01jan1975'd=ddmmyy10.'01jan1975'd-'01JAN1985'd="Disco Years"

```

Created with

```
01JAN1985'd-high=date9.;  
run;
```

```
proc print;  
format date dat. ;  
run;
```

In the following DATA step, what is needed for 'fraction' to print to the log?

```
data _null_;  
x=1/3;  
if x=.3333 then put 'fraction';  
run;
```

What is the difference between calculating the 'mean' using the mean function and PROC MEANS?

A) By default Proc Means calculate the summary statistics like N, Mean, Std deviation, Minimum and maximum, Where as Mean function compute only the mean values.

What are some differences between PROC SUMMARY and PROC MEANS?

Proc means by default give you the output in the output window and you can stop this by the option NOPRINT and can take the output in the separate file by the statement OUTPUTOUT= , But, proc summary doesn't give the default output, we have to explicitly give the output statement and then print the data by giving PRINT option to see the result.

What is a problem with merging two data sets that have variables with the same name but different data?

A) Understanding the basic algorithm of MERGE will help you understand how the stepProcesses. There are still a few common scenarios whose results sometimes catch users off guard. Here are a few of the most frequent 'gotchas':

1- BY variables has different lengthsIt is possible to perform a MERGE when the lengths of the BY variables are different,But if the data set with the shorter version is listed first on the MERGE statement, theShorter length will be used for the length of the BY variable during the merge. Due to this shorter length, truncation occurs and unintended combinations could result.In Version 8, a warning is issued to point out this data integrity risk. The warning will be issued regardless of which data set is listed first:WARNING: Multiple lengths were specified for the BY variable name by input data sets.This may cause unexpected results. Truncation can be avoided by naming the data set with the longest length for the BY variable first on the MERGE statement, but the warning message is still issued. To prevent the warning, ensure the BY variables have the same length prior to combining them in the MERGE step with PROC CONTENTS. You can change the variable length with either a LENGTH statement in the merge DATA step prior to the MERGE statement, or by recreating the data sets to have identical lengths for the BY variables.Note: When doing MERGE we should not have MERGE and IF-THEN statement in one data step if the IF-THEN statement involves two variables that come from two different merging data sets. If it is not completely clear when MERGE and IF-THEN can be used in one data step and when it should not be, then it is best to simply always separate them in different data step. By following the above recommendation, it will ensure an error-free merge result.

Which data set is the controlling data set in the MERGE statement?

A) Dataset having the less number of observations control the data set in the merge statement

How do the IN= variables improve the capability of a MERGE?

A) The IN=variables What if you want to keep in the output data set of a merge only the matches (only those observations to which both input data sets contribute)? SAS will set up for you special temporary variables, called the "IN=" variables, so that you can do this and more. Here's what you have to do: signal to SAS on the MERGE statement that you need the IN= variables for the input data set(s) use the IN= variables in the data step appropriately, So to keep only the matches in the match-merge above, ask for the IN= variables and use them:

```
data three;merge one(in=x) two(in=y); /* x & y are your choices of names */by id; /* for the IN= variables for data */if x=1 and y=1; /* sets one and two respectively */run;
```

What techniques and/or PROCs do you use for tables?

A) Proc Freq, Proc univariate, Proc Tabulate & Proc Report.

Do you prefer PROC REPORT or PROC TABULATE? Why?

A) I prefer to use Proc report until I have to create cross tabulation tables, because, It gives me so many options to modify the look up of my table, (ex: Width option, by this we can change the width of each column in the table) Where as Proc tabulate unable to produce some of the things in my table. Ex: tabulate doesn't produce n (%) in the desirable format.

How experienced are you with customized reporting and use of DATA _NULL_ features?

A) I have very good experience in creating customized reports as well as with Data _NULL_ step. It's a Data step that generates a report without creating the dataset there by development time can be saved. The other advantages of Data NULL is when we submit, if there is any compilation error is there in the statement which can be detected and written to the log there by error can be detected by checking the log after submitting it. It is also used to create the macro variables in the data set.

What is the difference between nodup and nodupkey options?

A) NODUP compares all the variables in our dataset while NODUPKEY compares just the BY variables.

What is the difference between compiler and interpreter?

Give any one example (software product) that act as an interpreter?

A) Both are similar as they achieve similar purposes, but inherently different as to how they achieve that purpose. The interpreter translates instructions one at a time, and then executes those instructions immediately. Compiled code takes programs (source) written in SAS programming language, and then ultimately translates it into object code or machine language. Compiled code does the work much more efficiently, because it produces a complete machine language program, which can then be executed.

Code the table's statement for a single level frequency?

A) Proc freq data=lib.dataset;
table var;*here you can mention single variable of multiple variables seperated by space to get single frequency;
run;

What is the main difference between rename and label?

A) 1. Label is global and rename is local i.e., label statement can be used either in proc or data step where as rename should be used only in data step. 2. If we rename a variable, old name will be lost but if we label a variable its short name (old name) exists along with its descriptive name.

What is Enterprise Guide? What is the use of it?

Created with

A) It is an approach to import text files with SAS (It comes free with Base SAS version 9.0)

What other SAS features do you use for error trapping and data validation?

What are the validation tools in SAS?

A) For dataset: Data set name/debugData set: name/stmtchk

For macros: Options:mprint mlogic symbolgen.

How can you put a "trace" in your program?

A) ODS Trace ON, ODS Trace OFF the trace records.

How would you code a merge that will keep only the observations that have matches from both data sets?

A) Using "IN" variable option. Look at the following example.

```
data three;
merge one(in=x) two(in=y);
by id;
if x=1 and y=1;
run;
or
data three;
merge one(in=x) two(in=y);
by id;
if x and y;
run;
```

What are input dataset and output dataset options?

A) Input data set options are obs, firstobs, where, in output data set options compress, reuse.Both input and output dataset options include keep, drop, rename, obs, first obs.

How can u create zero observation dataset?

A) Creating a data set by using the like clause.ex: proc sql;create table latha.emp like oracle.emp;quit;In this the like clause triggers the existing table structure to be copied to the new table. using this method result in the creation of an empty table.

Have you ever-linked SAS code, If so, describe the link and any required statements used to either process the code or the step itself?

A) In the editor window we write%include 'path of the sas file';run;if it is with non-windowing environment no need to give run statement.

How can u import .CSV file in to SAS? tell Syntax?

A) To create CSV file, we have to open notepad, then, declare the variables.

```
proc import datafile='E:\age.csv'out=sarathdbms=csv replace;
getnames=yes;
proc print data=sarath;
run;
```

What is the use of Proc Sqi?

Created with

A) PROC SQL is a powerful tool in SAS, which combines the functionality of data and proc steps. PROC SQL can sort, summarize, subset, join (merge), and concatenate datasets, create new variables, and print the results or create a new dataset all in one step! PROC SQL uses fewer resources when compared to that of data and proc steps. To join files in PROC SQL it does not require to sort the data prior to merging, which is must, is data merge.

What is SAS GRAPH?

A) SAS/GRAPH software creates and delivers accurate, high-impact visuals that enable decision makers to gain a quick understanding of critical business issues.

Why is a STOP statement needed for the point=option on a SET statement?

A) When you use the POINT= option, you must include a STOP statement to stop DATA step processing, programming logic that checks for an invalid value of the POINT= variable, or Both. Because POINT= reads only those observations that are specified in the DO statement, SAS cannot read an end-of-file indicator as it would if the file were being read sequentially. Because reading an end-of-file indicator ends a DATA step automatically, failure to substitute another means of ending the DATA step when you use POINT= can cause the DATA step to go into a continuous loop.

What is the difference between nodup and nodupkey options?

A) The NODUP option checks for and eliminates duplicate observations. The NODUPKEY option checks for and eliminates duplicate observations by variable values.

Created with

CDM Fundamentals:

Q1. Being a CD manager, what is your contribution going to be, to my company?

Answer: As a CD Manager, I can assure you of accurate, complete, consistent data for reporting, to the regulatory bodies. I also communicate & coordinate with the Project Manager, Statistician, CRA, DB Manager at the clinical sites as needed to ensure the accuracy and completeness of the CT data

Q2. Who is the father of Clinical Trials

Answer: James Lind

Q3. In Health care, can you tell me the synonyms of CT

Answer: Clinical Research, Clinical Study, Medical Research

Q4. Define the CT

Answer: Clinical Trials are the comparative study of Medication against the patient's health condition.

A more comprehensive definition according to ICH is: Any investigation in human subjects intended to discover or verify the clinical, pharmacological and/or other Pharmacodynamic effects of an investigational product, and/or to identify any adverse reactions to an investigational product, and/or to study absorption, distribution, metabolism, and excretion of an investigational product with the object of ascertaining its safety and/or efficacy.

Q5. Define Unapproved Therapeutic Goods

Answer: The Drugs which did not undergo Clinical Trial are called Unapproved Therapeutic Goods

Q6. What is IND?

Answer: During the trial, the agent being tested is called an IND (Investigational New Drug)

Q7. Describe the importance of Inclusion and exclusion Criteria

Answer: Inclusion & exclusion criteria are important in that the subjects are either included in or excluded from a trial based on the inclusion and exclusion criteria

Q8. What is Meant by Masking or Blinding

Answer: Masking or blinding is the process of hiding the details whether the research subject is receiving the Investigational product or a placebo or the current standard treatment

- Single Blinding: the subject doesn't know about the treatment
- Double Blinding: Both the researcher and the patient do not know about the treatment

Q9.Emphasize the importance of masking/Blinding

Answer: Masking/ Blinding is necessary because it eliminates any bias in the treatment process being investigated

Q10.What is Placebo?

Answer: A Placebo is an inactive pill, powder, liquid which contains no active agent. The use of a Placebo helps the researcher to isolate the effect of the study treatment

Q11.What is a patient file? What information is available in it?

Answer: A Patient File (PF) contains the demographic data, Medical and treatment data about a patient or subject. It can contain paper records or can be a mixture of both paper and computer records

Q12.What are pre clinical studies?

Answer: Pre clinical studies are the animal studies that support Phase I safety and tolerance studies. They must comply with the GLP guidelines

Q13. Explain the different phases of Clinical Trials.

Answer: There are four major phases in a clinical trial.

Phase I : Human Pharmacology Trials

Phase II : Therapeutic exploratory trials

Phase III : Therapeutic Confirmatory Trials

Phase IV : Post marketing Surveillance Trials

There are different phases of CT

Pre Clinical Studies: They involve in-vitro studies and in-vivo studies on animals. Wide ranging doses are given to animals and the PK, efficacy and toxicity parameters are studied to determine the viability of further studies.

Phase 0: Human Micro Dosing Studies (normally the doses are 100 times less than the intended therapeutic doses). Single sub therapeutic doses are administered to a small number of subjects (10-15) PK and PD parameters are derived. Gives no data on safety or efficacy. To support basic go/no go decision making

Phase I:

Human Pharmacology Trials. Size - 20 to 80.

May range from several months to a year Usually to test one or more of combination of objectives.

1. Maximum tolerated dose
2. PK
3. PD
4. Early measurement of Drug activity

This phase also includes SAD, MAD and FOOD EFFECT studies.

Phase II: Therapeutic exploratory trials to determine the effective dose and the dosing regimen. May last from 1 to 2 years. Conducted after safety of the drug is confirmed in phase I.

Sample size is larger, between 20-300 Sometimes divided into Phase IIA To assess Dosing requirements. Phase IIB to study efficacy.

Phase III: Therapeutic confirmatory trials are randomized, controlled, multi-centered trials. Also called pivotal trials because they are crucial to the approval of the drug. May last from 3 to 5 years. Aimed at being definitive assessment of effectiveness of drug in comparison with the current gold standard treatment Sample size 300 – 3000

Phase IV

: Post marketing surveillance studies. Either required by the regulatory authorities or undertaken by the manufacturer for competitiveness To gather information like use of Drug in children Pregnant women, children Elderly patients Patients with renal or other failures Specific concomitant medication Also detects rare or long term adverse reactions

Q14. Describe the Scientific names for all 4 phases of trials

Answer: Phase I : Human Pharmacology Trials

Phase II : Therapeutic exploratory trials

Phase III : Therapeutic Confirmatory Trials

Phase IV : Post marketing Surveillance Trials

Q15. Distinguish Between Double Blind and Double Dummy.

Answer: Double blind is where both the subject and the researcher do not know which of the treatment the subject is receiving i.e. whether control or the study treatment.

In Double dummy, every subject is given both the control and the investigational treatment, for alternating periods.

Q16. Other name for QOL (Quality Of Life) trial.

Answer: Supportive Care Trial

Q17. What are Orphan Trials?

Answer: Orphan trials aimed at testing drugs designed to treat diseases affecting less than 200,000 people. Tested only on a small number of participants, Who are so sick that the effect of treatment, if the drug really works, is immediately apparent.

Q18. What are PK Parameters?

Answer: Pharmacokinetic parameters determine the characteristics of the drug's Absorption, Distribution, Metabolism and Excretion

(ADME).

Q19. What is ADME?

Answer: Absorption, Distribution Metabolism and Excretion

Q20. What is Pharmacodynamics?

Answer: The study of time course and the intensity of the drug effect on the organism.

Q21. What is Bioavailability?

Answer: It is the fraction of administered dose of unchanged drug that reaches the systemic circulation.

Q22. What is Bioequivalence?

Answer: Used to assess the expected in-vivo biological equivalence of two proprietary preparations of drug. If two drugs are said to be bio equivalent, then they are expected to be for all intent and purpose, same.

Q23. Is it true that Phase I trials include healthy volunteers?

Answer: Yes. But exception is made for the terminally ill patients who have no alternative therapy available.

Q24. What is the range of participants in each of the four phases of Clinical Trials?

Answer: Phase I - 20 to 80

Phase II - 200 - 300

Phase III - 300 to 3000

Phase IV - Thousands of patients who are being treated

Q25. What are the categories of Phase I Trials

Answer: SAD(Single Ascending Dose)

MAD(Multiple Ascending Dose) and Food Effect

Q26. What are the reasons for failure of different phases of trials?

Answer: Reasons for failure of Phase I trials

1. Pre-clinical animal models not equal to behavior in humans
2. Inadequate pre-clinical data
3. Change in drug formulation from pre-clinical testing to clinical testing
- [4. PK/PD](#) relationship
5. Poorly designed clinical studies
6. Drug too toxic in humans

Reasons for failure in phase II and Phase III trials

1. Infrequent ADR s
2. Drug-drug interaction
3. Drug-disease interaction

4.Genetic

5.Effectiveness insufficient

6.Economic

Q27. What are the categories of Phase II Trials

Answer: Phase IIA and Phase IIB

Q28.What is Efficacy?

Answer: The measure of the maximum strength of the drug

Q29.What is Potency?

Answer: The amount of drug required for its specific effect

Q30.What is NCE?

Answer: New Chemical Entity

Q31.What are the contents of an IND Application?

Answer:

1. The name, chemical name and structure of the NCE
2. Complete list of components of the drug
3. Quantitative composition of the drug
4. Name and address of the supplier of any new drug substance
5. Description of synthesis of any new drug substance
6. Statement of methods, facilities and controls used in manufacture and packaging of the new drug
7. Statement covering all information from pre-clinical studies and any clinical studies and experiences

With the drug

8. Copies of labels for the drug.
9. Description of scientific training and experience considered appropriate by the sponsor to qualify the investigator as a suitable expert to investigate the drug

10. Names and curriculum vitae of all the investigators
11. An outline of planned methodology to be adopted for the clinical trial

Q32. What are the contents of an NDA Application?

Answer:

1. Detailed reports of pre-clinical studies
2. Detailed reports of clinical studies
3. Information on composition and manufacture of the drug and on controls and facilities used in manufacture
4. Samples of drug and its labeling
5. Full case reports of the persons who received the drug, needed only in limited circumstances
6. Patient information
7. Material previously submitted to FDA in the IND application or in periodic reports must be included by reference in the NDA

Q33. List the responses of FDA on and NDA(New Drug Application)

Answer: 1. Not Approvable

2. Approvable
3. Approval

Q34. How can one tell the significance or power of a trial?

Answer: By the size of the trial

Q35. List the tripartite countries in the ICH-GCP

Answer: USA , European Union and Japan

Q36. Describe the incidents that led to the formation of Ethical principles in Clinical Trials

Answer: The three main incidents are the Thalidomide disaster, Tuskegee syphilis study and the Nuremberg war prisoner's incidents

Q37. What does the Declaration of Helsinki say?

Answer: The Declaration of Helsinki (1964) defines rules for "research combined with clinical care" & "non-therapeutic research" they are

1. Research on human beings should be based on the results from laboratory and animal experimentations
2. Research protocols should be reviewed by an independent committee prior to initiation
3. Informed consent from research participants is necessary
4. Research should be conducted by medically scientifically qualified individuals
5. Risks should not exceed benefits

Q38.what does Nuremberg Code state?

Answer: Nuremberg code (1948) states that the voluntary consent of the human subject is absolutely essential

Q39. List the Salient feature of Belmont report

Answer: The three basic principles of Belmont report are

1. Respect for persons
Respect for persons incorporates at least two ethical convictions:
 - a. Individuals should be treated as autonomous agents
 - b. Persons with diminished autonomy are entitled to protection
 - c. Informed consent (3 elements)
 - i. Information
 - ii. Comprehension
 - iii .Voluntary ness
 - d. Subject should be given opportunity to choose what should and shall not happen to them
2. Beneficence
 - a. Human subject should not be harmed
 - b. Research should maximize the benefits and minimize the harms

- c. Risks and benefits should be assessed
- d. The nature and scope of risks and benefits must be systematically assessed

3 . Justice

- a. The benefits and risks of the research must be distributed uniformly
- b. Selection of subject: there must be a fair process and outcomes in the selection of research subjects

Q40. State the 13 core principles of ICH-GCP guidelines

Answer: 1.The clinical trials should be conducted in accordance with the ethical principles based on the declaration of Helsinki and GCP and regulatory requirements

2.Before a trial is initiated, foreseeable risks and inconveniences should be weighed against the anticipated benefit for the individual trial subject and society. A trial should be initiated and continued only if the anticipated benefits justify the risks.

3. The rights, safety, and well-being of the trial subjects are the most important considerations and should prevail over interests of science and society.

4. Available clinical and non clinical information on the product should adequate to support the proposed clinical trial

5. Trial should be scientifically sound and described in a clear, detailed protocol

6. A trial should be conducted in compliance with the protocol that has received prior institutional review board (IRB)/independent ethics committee (IEC) approval/favorable opinion.

7. The medical care given to, and medical decisions made on behalf of, subjects should always be the responsibility of a qualified physician or, when appropriate, of a qualified dentist.

8. Each individual involved in conducting a trial should be qualified by education, training, and experience to perform his or her respective task(s)

9. Freely given informed consent should be obtained from every subject prior to clinical trial participation.

10. All clinical trial information should be recorded, handled, and stored in a way that allows its accurate reporting, interpretation and verification.

11. The confidentiality of records that could identify subjects should be protected, respecting the privacy and confidentiality rules in accordance with the applicable regulatory requirement(s)

12. Investigational products should be manufactured, handled, and stored in accordance with applicable good manufacturing practice (GMP). They should be used in accordance with the approved protocol.

13. Systems with procedures that assure the quality of every aspect of the trial should be implemented

Q41. Who are the observers of ICH?

Answer: 1. The World Health Organization (WHO)

2. The European Free Trade Area (EFTA), represented at ICH by Switzerland

3. Canada, represented at ICH by Health Canada

Q42. What are the objectives of ICH?

Answers: The purpose of ICH is to make recommendations on ways to achieve greater harmonization in the interpretation and application of technical guidelines and requirements for product registration in order to reduce or obviate the need to duplicate the testing carried out during the research and development of new medicines.

• To provide a unified standard for the European Union (EU), Japan and the United States to facilitate the mutual acceptance of clinical data by the regulatory authorities in these jurisdictions

Q43. Who are the participants of Pharmacovigilance?

Answer: 1. Patients as the users of medicines

2. Doctors, pharmacists, nurses and all other healthcare professionals working with medicines regulatory authorities EMEA and those in the member states responsible for monitoring the safety of medicines

3. Pharmaceutical companies and companies importing or distributing medicines

Q44. What is the need of Pharmacovigilance?

Answer: 1. Illegal sale of medicines and drugs of abuse over the internet

2. Increased self medication practices

3. Widespread manufacture and sale of counterfeit and substandard medicines
4. Increased use of traditional medications outside the confines of traditional culture of use
5. Increased use of medications of different systems with potential for drug interactions

Q45. describe the attributes of AE

Answer: 1. Unrelated: The AE is clearly not related to the intervention

2. Unlikely: The AE is doubtfully related to the intervention
3. Possible: The AE is may be related to the intervention
4. Probable: The AE is likely related to the intervention
5. Definite: The AE is clearly related to the intervention

Q46. Schedule Y requirements according to SAE

Answer: Unsuspected adverse events are communicated from:

1. Sponsor to regulatory authorities within 14 days
2. Investigator to sponsor within 24 hours
3. Investigator to ethics committee in 7 days

Q47. Role of Principal Investigator (PI)

Answer: The Principal Investigator has the overall responsibility of the design, conduct, analysis and reporting of Clinical Trial

He has the overall responsibility for the coordination and the day-to-day management of the trial.

Q48. What is CTMS?

Answer: A CTMS describes the responsibilities of those involved in running the trial on a day-to-day basis

Q49. What is CDMS?

Answer: CDMS is the tool used to ensure that the data gathered in the course of the study is:

1. Accurate
2. Complete
3. Logical
4. Consistent

The trial data collected at the investigator site is stored in a CDMS

Q50.What is IB?

Answer: The Investigator's Brochure (IB) is a basic document which is required in a clinical trial According to the FDA regulations (Title 21 CFR 312.23), an Investigator's Brochure must contain:

1. Description of the drug substance and the formulation
2. Summary of the pharmacological and toxicological effects
3. Summary of information relating to its safety and effectiveness in humans
4. Description of possible risks and adverse reactions to be anticipated, and the precautions or special monitoring that the investigator should take.

Q51.What is Protocol Document?

Answer: A Clinical Trial Protocol is a document that describes the objective(s), design, methodology, statistical considerations, and organization of a clinical trial.

- The existence of a clinical trial protocol allows researchers at multiple locations (in a multi-center trial) to perform the study in exactly the same way, so that their data can be combined as though

They were all working together.

- The protocol also gives the study administrators (often a contract research organization) as well as the local researchers a common reference document for the researchers' duties and responsibilities

During the trial.

Q52. What is Multi-center trial (MCT)?

Answer: Multi-center trial means a clinical trial spread across various centers at different geographic locations covering varied demographic profiles.

Q53. What are the means of recruiting subjects for a clinical trial?

Answer:

1. Through volunteer database
2. Radio advertisements
3. News paper advertisements
4. TV advertisements
5. Internet recruitment
6. By posting notices at the places like to be visited by patients like clinics, pharmacies etc

Q54. What is Informed Consent?

Answer: Informed consent is the voluntary consent obtained from the research subject to participate in the research, after explaining to the person of all the risks and benefits involved in the research.

Q55. Why is randomization required in a trial?

Answer: Randomization is required in a trial to isolate the drug effect

Q56. What is CRF and what is its importance?

Answer: CRF stands for Case Report/Record Form. CRF is perhaps, the most important document after the protocol since all the clinical trial data is collected through the CRF

Q57. What is Data?

Answer: Data means Information (facts/figures) which give an accounting of the study

Q58. What is source document?

Answer: Source document means the first recording about the trial subject like original lab reports, pathology reports, surgical reports, medical records, letters from referring physicians, participant diary etc.

Q59. What are the documents required to be kept at the study site?

Answer:

Here is a list documents that need to be kept at the study site.

1. Signed FDA form 1572
 2. CVs of all investigators
 3. Signed approved protocol
 4. Informed consents / all amended informed consents
 5. Investigator's Brochure
 6. IRB approval
 7. IRB membership
 8. Assurance number
 9. Drug accountability
 10. IND safety reports
 11. Annual/interim reports
 12. All information given to the subjects
 13. CRF s on each subject (signed, dated)
 14. AE reports
 15. All source documents not kept in Medical record
 16. Meeting minutes/correspondence
 17. Signature log/equipment logs
 18. Laboratory documentation
 - a. Certification
- Abnormal range table with dates
19. Specimen handling
 - a. Instructions/labels/shipping
 20. Staff education records
 21. Financial agreements
 - a. Sponsor

b. Subject

22. study agreement grant

23. Letter of indemnification

24. Advertisements

25. End of study report

Q60.What is Common Data Elements (CDE)?

Answer: Common Data Elements mean the standardized, unique terms and phrases that delineate discreet pieces of information used to collect data on a clinical trial

Q61.What is Audit trail?

Answer: It is the data which shows that the study was conducted according to the protocol. It tells the who, when and why of the entry/changes in data. It is also defined as the "Documentation that allows reconstruction of the course of events" according to SCDM (Society for Clinical Data Management).

Q62.What is double Data Entry? What is its importance?

Answer: Double data entry is the process of entering the same data twice in pass one and pass two, by two different individuals. DDE is important because it helps in reducing the discrepancies that arise due to errors in data entry.

Q63.What are the best solutions for Clinical Data Management?

Answer: Data Analytics: AS 9 platform

EDC: Oracle clinical, phase forward, medidata solution etc

Document management Services: Documentum, Opentext, adobe solutions etc

Q64 Define Digitization

Answer: digitization is the process of converting the data into computer readable format.

Q65.What is DB closure?

Answer: When a database is closed, no further modifications are allowed on the database. The permission to further modify the data rests with a privileged few, most critical study personnel.

Q66.What is the best analytical tool?

Answer: SAS 9.0 environment

Q67.What is discrepancy resolution? Describe the process.

Answer:

As a result of batch validation, the discrepancies that arise in the database are managed by the discrepancy management team.

1. Discrepancies are resolved through the Data Clarification Form (DCF) or the Data Query Form (DQF).
2. DCF is generated and sent to the site (investigator) for clarification.
3. After the corrected values are received from the investigator as a response to the DCF, the responses are updated in the database
4. After ensuring that the data is error free, the database is locked to avoid further unauthorized Modifications

Q68. What is EDC?

Answer: The process of collection of data into a persistent form. This includes data entry (keyboard EDC, voice recognition, pen-based systems) and automated(or direct) data acquisition(bar code scanners, blood pressure cuff devices etc)

Q69.What is RDC?

Answer: Remote Data Capture. RDC involves the data entry through networked systems like internet

Q70. What is the scope of 21CFR Part11?

Answer

: a. Criteria under which electronic records and signatures are considered trustworthy, reliable, and generally equivalent to paper records and handwritten signatures.

b. part applies to records in electronic form that are created, modified, maintained, archived, retrieved, or transmitted, under any records requirements set forth in agency regulations. This part also applies to

electronic records submitted to the agency under requirements of the Federal Food, Drug, and Cosmetic Act and the Public Health Service Act, even if such records are not specifically

identified in agency regulations. However, this part

c. Does not apply to paper records that are, or have been, transmitted by electronic means.

d. Electronic records that meet the requirements of this part may be used in lieu of paper records, in accordance with § 11.2, unless paper records are specifically required.

e. Computer systems (including hardware and software), controls, and attendant documentation maintained under this part shall be

readily available for, and subject to, FDA inspection

Q71. What does Part 11 describe?

Answer: Electronic Records and Electronic signatures

Q72.What does subpart B describes?

Answer: Electronic Records (11.10 to 11.70)

11.10 Controls for closed systems.

11.30 Controls for open systems.

11.50 Signature manifestations.

11.70 Signature/record linking

Q73.What does section 11.2 describes?

Answer: Section 11.2 of 21 CFR Part 11 describes the Implementation.

Q74.What is ACT?

Answer: Act means the Federal Food Drug and Cosmetics Act ((sec. 201-903) (21U.S.C 321-393))

Q75.What is biometrics?

Answer: Biometrics means a method of verifying an individual's identity based on measurement of the individual's physical feature(s) or repeatable

action(s) where those features and/or actions are both unique to that individual and measurable.

Q76

. What is an electronic signature?

Answer: Electronic signature means a computer data compilation of any symbol or series of symbols executed, adopted, or authorized by an individual to be the legally binding equivalent of the individual's handwritten signature

Q77. Define closed systems

Answer: Closed system means an environment in which system access is controlled by persons who are responsible for the content of electronic records that are on the system

Q78. What do 11.50 of part 11 deal with?

Answer: Signature manifestation

Q79. What are the codes of ethics to be followed by the CDM professionals?

Answer: 1. Committed to following the laws and guidelines applicable to clinical research (including the

Declaration of Helsinki), to participate in the protection of the safety, dignity and well being of

patients and to maintain the confidentiality of medical records.

2. Committed to creating, maintaining and presenting quality clinical data, thus supporting accurate and timely statistical analysis, and to adhering to applicable standards of quality and truthfulness in scientific research

3. Committed to facilitating communication between clinical data management professionals and all other clinical research professionals, to maintaining competency in all areas of clinical data management, to keeping current with technological advances, and to ensuring the dissemination of information to members of the clinical research team.

4. Committed to working as an integral member of a clinical research team with honesty, integrity and respect. To making and communicating accountability for clinical data management decisions and actions within the clinical trial process.

5. Committed to maintaining and respecting proprietary knowledge at all levels, to avoiding the use of proprietary knowledge for personal gain, and to disclosing any conflict of interest. To avoiding any conduct or behavior that is unlawful, unethical or that may otherwise reflect negatively on the profession of clinical data management

6. Committed to advancing the profession of clinical data management through the development, distribution and improvement of good clinical data management practices. To aiding the professional development and advancement of colleagues within the clinical trial industry

Q80 How many chapters does the GCDMP document contain?

Answer: The GCDMP document addresses 20 chapters

Q81.What is the most recent version of GCDMP document?

Answer: Version 4

Q82. Who is GCDMP meant for?

Answer: 1.Data managers

2. Data processors

3. Statisticians

4. Site personnel

5. Clinical professionals

6. Compliance auditors

7. Regulatory affairs personnel

8. All clinical research professionals making decisions regarding or using the clinical trial data

Q83.How is privacy protection afforded to the subjects by GCDMP?

Answer: 1.Protocol review and approval by an Institutional Review Board (IRB)

2. Right to informed consent

3. Right of the subject to withdraw consent

4. Right to notice of disclosure

5. Confidential collection and submission of data

Q84.What is external data?

Answer: •Laboratory and PK/PD data

•Device data (ECG, flowmetry, vital signs, images etc)

•Electronic patient diaries

Q85.What are ASCII files?

Answer: ASCII stands for American Standard Code for Information Interchange. ASCII files are plain text files

which can be edited using any simple text editor like notepad.

Q86.What is clean data?

Answer: clean data means the data is able to support the statistical analysis, and its subsequent Presentation and interpretation, as if the data is error free.

Q87.What is the purpose of CDISC?

Answer: To help in streamlining the Clinical data lifecycle by leveraging emerging industry data standards with strong domain and technological expertise. This in turn, helps the sponsor towards production of regulatory grade products throughout all stages of the Clinical Data Lifecycle.

Q88. Why do we need CDISC?

Answer: To standardize clinical data To bring uniformity and vendor neutrality in clinical data To bring ease and cost effectiveness into data exchange Easier submission and faster evaluation of CDISC compliant data by the regulatory bodies.

Q89.List the Standard models in CDISC.

Answer:

ODM: Operational Data Model. Standards for data storage formats.

SDTM: Study Data Tabulation Model

CRT-DDS: Case Report Tabulation Data Definition Specification

Lab standards: ADaM:

Analysis Data Model

Protocol Standards:

SEND: Standards for exchange of Non-clinical data

Q90. What is the upcoming model in CDISC?

Answer:

Protocol Standards

Q91. Describe ODM standard

?

Answer: 1.ODM stands for Operational Data Model.

2. It defines standards for the storage, archival and exchange of clinical trial data

3. ODM is vendor and platform independent

4. ODM includes metadata associated with clinical data, administration data, reference data and audit trails

5. All information that needs to be shared among different Software systems during setup, operation analysis, submission and long term retention of data

6. ODM utilizes xml technology

Q92. What is the importance of lab standards?

Answer:

lab standards define the exchange of laboratory data between the lab and CRO

Q93.What is CRT DDS?

Answer: Case Report Tabulation Data Definition Specifications; also known as define .xml, is a standard for providing data definition for case report tabulation in an xml format for submission to FDA. XML is platform neutral and faster to process

Q94.What are the 4 types of data required by SDTM for FDA submission? Answer:

- Analysis data sets
- Tabulation data sets
- Patient profiles
- Listing datasets

Q95.What is metadata?

Answer:

Metadata is data about data

Q96 List the features of EDC?

Answer: Yet to answer

Q97.Differentiate between Paper based trials and Electronic trials?

Answer:

Paper based Clinical Trials are cumbersome, error prone, inflexible, extensive takes a lot of time. Archival of data is difficult. Electronic trials address all these problems to either eliminate them or to minimize them

Q98 What is the 21 CFR for good manufacturing practices (GMP)?

Answer:

210 and 211

Q99. What is the 21 CFR for good laboratory practices (GLP)?

Answer: 58

Q100.What is SDV? When is it required?

Answer:

Source Data Verification and it is required during audit trails, discrepancy management

Q101. Describe Medical Coding?

Answer:

During a multi centric trial, there tends to be a lot of variation in reporting AE. Medical coding is helpful in bringing uniformity in AE reporting.

Answer:

WHOART: WHO Adverse reporting terminology. Used in AE coding

CoSTART: Coding Symbols For

A Thesaurus Of Adverse Reaction Terms. Use in AE coding

MedDRA: Medical Dictionary For Regulatory Activities. Used in AE coding

WHODD: WHO Drug dictionary. Used in coding concomitant medication

ICD9CM: Used in Medical history coding

Q103. What is AERS? What is it's importance?

Answer:

Adverse Event Reporting System. Is used to keep track of the adverse events that may occur after a drug is marketed. It could be part of phase IV clinical trials.

Q104. Define UADR.

Answer: Unexpected Adverse Drug Reaction. Which is an ADR not documented in a protocol or IB

Q105. Define risk in Clinical Trial?

Answer:

The probable harm or discomfort caused to the trial subject

Q106. What is safety in Clinical Trial?

Answer:

Freedom from harm

Q107. What is raw data?

Answer: Records of original observations.

Q108. Who are vulnerable subjects?

Answer:

- Persons who cannot express willingness to volunteer

- Persons influenced by expectations
- Persons with incurable diseases
- Persons who are unemployed, who belong to ethnic minorities, who are homeless, minors and those who can't give consent and emergency patients.

Q109.What is meant by the well being of the subject?

Answer:

The physical and mental integrity of the subject

Q110.What is compliance?

Answer:

Adherence to all regulatory requirements

Q111What is a drug

Answer:

FDA Definition of a drug:

An active ingredient that is intended to furnish pharmacological activity or other direct effect in the diagnosis, cure, mitigation, treatment, or prevention of a disease, or to affect the structure of any function of the human body, but does not include intermediates used in the synthesis of such ingredient
More generic definition: A drug is substance which provides favorable therapeutic or prophylactic pharmaceutical benefits to the human body

Q112.What is a patent?

Answer:

A patent is the right granted by a government for any device, substance, or process that is new, inventive, and useful. The patent discloses the know-how for the invention and in return, the owner of the patent receives a 20 year period of monopoly rights to commercially exploit the invention.

Q113 What are the contents of the 21 CFR Part 58 for GLP?

Answer: Scope

Definitions

Applicability to studies performed under grants and contracts

Inspection of a testing facility

Personnel

Testing facility management

Study director

Quality assurance unit

General

Animal care facilities

Facilities for handling test and control articles

Laboratory operation areas

Specimen and data storage facilities

Equipment design

Maintenance and calibration of equipment

Standard operating procedures

Reagents and solutions

Animal care

Test and control article characterization

Test and control article handling

Mixtures of articles with carriers

Protocol

Conduct of a non-clinical laboratory study

Reporting of non-clinical laboratory study results

Storage and retrieval of records and data

Retention of records

Purpose

Grounds for disqualification

Notice of and opportunity for hearing on proposed disqualification

Final order on disqualification

Actions upon disqualification

Public disclosure of information regarding disqualification

Alternative or additional actions to disqualification

Suspension or termination of a testing facility by a sponsor

Reinstatement of a disqualified testing facility

Q114. What is the role of IRB/IEC?

Answer: IRB/IEC (Institutional Review Board/Independent Ethics Committee) acts as a third party to oversee the welfare of the trial subjects and to ensure that the trial is being conducted in accordance with the submitted protocol.

Q115. Who are the members of IRB/IEC?

Answer:

IRB/IEC may consist of clinicians, scientists, lawyers, religious leaders, and lay people to represent different view points and protect the rights of the subjects.

Q116. What are the 21 CFR documents relevant to clinical trials

Answer:

- 21 CFR Part 11
 - Electronic Records, Electronic Signatures
- 21 CFR Part 50
 - Protection of Human Subjects
- 21 CFR Part 312
 - Investigational New Drug Application

- 21 CFR Part 56
 - Institutional Review Board
 - 21 CFR Part 58
 - Good Laboratory Practices for Non - clinical Laboratory Studies
 - 21 CFR Part 202
 - Prescription Drug Advertising
 - CFR Part 210
 - Current Good Manufacturing Practice in Manufacturing, Processing, Packaging or Holding of Drugs; General
 - 21 CFR Part 211
 - Current Good Manufacturing Practice for Finished Pharmaceuticals
 - 21 CFR Part 312
 - Investigational New Drug Applications
 - 21 CFR Part 314
 - Applications for FDA Approval to Market a New Drug
 - 21 CFR Part 600
 - Biological Products: General
 - 21 CFR Part 610
 - General Biological Products Standards
 - ICH Harmonized Tripartite Guideline for Good Clinical Practice
- :

Q117 What are the contents of a Clinical Trial Protocol?

Answer: According to the ICH GCP, the following information is to be included in a protocol:

1. Protocol title

2. Name and address of Sponsor and Monitor
3. Name of authorized person
4. Name of Sponsor's medical expert
5. Name of Investigator responsible for the trial
6. Name of physician responsible for trial - related medical decisions 7.
Name of clinical laboratory and other institutions involved in the trial
8. Name and description of the clinical trial protocol
9. Summary of results from nonclinical studies
10. Potential risks and benefits to human subjects
11. Description and justification for route of administration, dosage, and treatment plan
12. Compliance to GCP
13. Description of the population to be studied
14. Reference literature and related data
15. Standard operating procedures

Q118. Who is a sponsor?

Answer:

The sponsor is the organization or individual that initiates the Ct and finances the study. The organization could be a government department, pharmaceutical company, university or

individual. It is normally a pharmaceutical company.

Q119. What is a CRO

Answer:

A CRO or Clinical Research Organization is that which is contracted by the sponsor to conduct and monitor the trial. It provides certain measure of independence to the trial and enhances the validity of trial results to be unencumbered by conflict of interest.

Q120. What are the products that are regulated by the FDA

Answer:

1. Drugs (e.g., prescriptions, OTCs, generics)
2. Biologics (e.g., vaccines, blood products)
3. Medical devices (e.g., pacemakers, contact lenses)
4. Food (e.g., nutrition, dietary supplements)
5. Animal feed and drugs (e.g., livestock, pets)
6. Cosmetics (e.g., safety, labeling)
7. Radiation emitting products (e.g., cell phones, lasers)