Paper 396-2008

Oracle Clinical® for SAS® Programmers

Kevin Lee, Therakos, Exton, PA

ABSTRACT

This paper is intended for SAS programmers who are interested in understanding the difference in the database structure between Oracle Clinical and SAS. It also helps SAS programmers to use SAS/ACCESS® to extract raw data from Oracle Clinical. This paper will discuss the database structure of Oracle Clinical and its relationship with the extracted SAS data.

INTRODUCTION OF ORACLE CLINICAL

Oracle Clinical (OC) is the software package that is widely used in the pharmaceutical environment as a data entry and storage tool. It captures the data entered from CRF (Case Report Form) and stores the data in its database system. Unlike SAS, the OC builds the database differently.

INTRODUCTION OF ORACLE CLINICAL DATABASE

Building OC database requires the understanding of hierarchy structure in OC. The OC database developer generally builds up the OC database in the order of DVG (Discrete Value Group), Questions, QG (Question Groups), DCM (Data Collection Module), DCI (Data Collection Instrument) and DCI Books. However, its hierarchy structure is the opposite way.

THE HIERARCHY STRUCTURE OF ORACLE CLINICAL DATABASE

- **DCI (Data Collection Instrument) Books**: The DCI books are a collection of DCI. It provides the order of data entry in OC.

- **DCI (Data Collection Instrument)**: The DCI is a DCM or a group of DCM. One DCI usually represents each page of CRF.

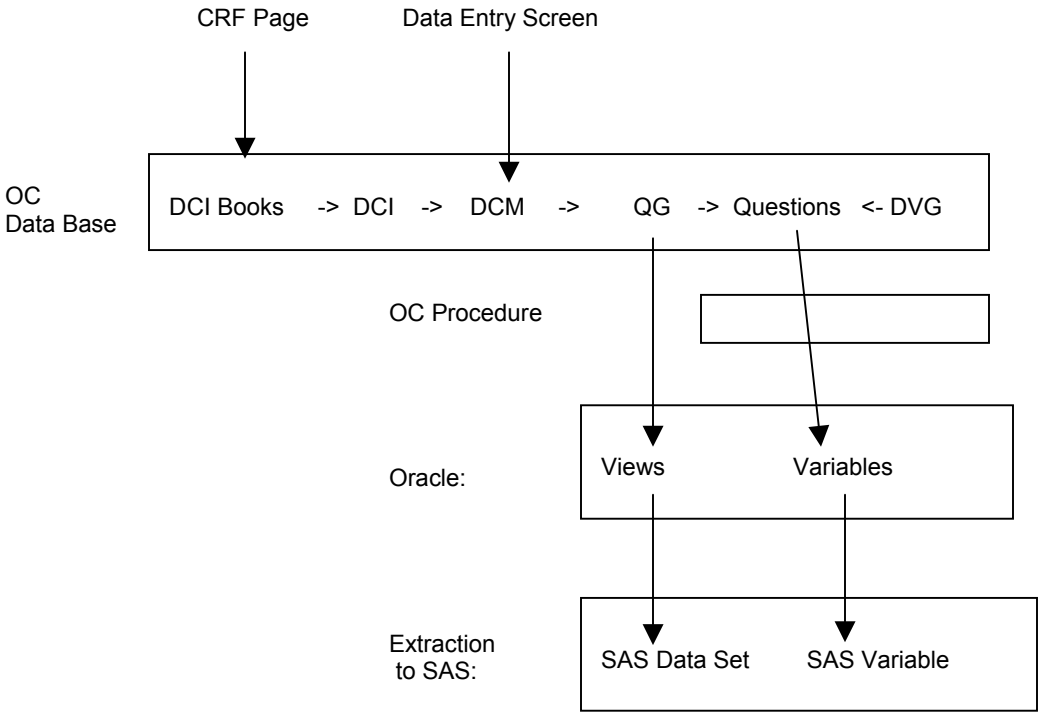
- **DCM (Data Collection Module)**: The DCM is a QG or a group of QG, in which the related data can be entered in a single clinical visit. As we will discuss later, one DCM represents each screen of the data entry.

- **QG (Question Group)**: The QG is, like its wording explain, the group of questions. It collects the related data by grouping related questions.

- **Questions**: The question captures and stores the data entered. In SAS, it can be viewed as a variable.

- **DVG (Discrete Value Group)**: The DVG is the discrete value for a question. It is also known as code list. In SAS, it can be regarded as a format.

The diagram for OC Database structure and its relationship with SAS



THE ORACLE CLINICAL PROCEDURE

There are 2 types of OC Procedures – Validation and Derivation.

The Validation Procedures are usually used for validation of the data entry – SAS programmers can do this validation part by the edit check programs.

The Derivation Procedures create the derived variables in OC database. For example, the variable of age can be derived from both consent date and birth date that are entered from Case Report Form (CRF).

THE REAL EXAMPLE

The paper will show the example of CRF and Data Entry Screen in OC. It will introduce the corresponding hierarchical structure of Oracle Clinical database. It will also introduce the source codes in OC and the SAS codes that can extract OC database into SAS environment.

THE SAMPLE PAGE OF ANNOTATED CRF

Screening

Protocol Study000	Site No. 000	Subject No. ☐☐☐☐	Subject Initials ☐☐☐

Informed Consent	
Date Informed Consent signed: (must be prior to all study procedures)	CONSDT ☐☐/☐☐☐/200☐ dd mmm yyyy
Time Informed Consent signed:	CONSTM ☐☐:☐☐ (24 hour clock)

Demographics			
Date of Birth: ☐☐ / ☐☐☐ / ☐☐☐☐ dd mmm yyyy BIRTHDT		Gender: SEX	☐Male (1) ☐Female (2)
Ethnicity (Check one box only):	RACE ☐ Caucasian (1) ☐ African American (2) ☐ Asian (3)	☐ Hispanic (4) ☐ Native American (5) ☐ Other (99) specify: _____ RACESP	

Please review inclusion criteria below and check (√) appropriate boxes. IEINA

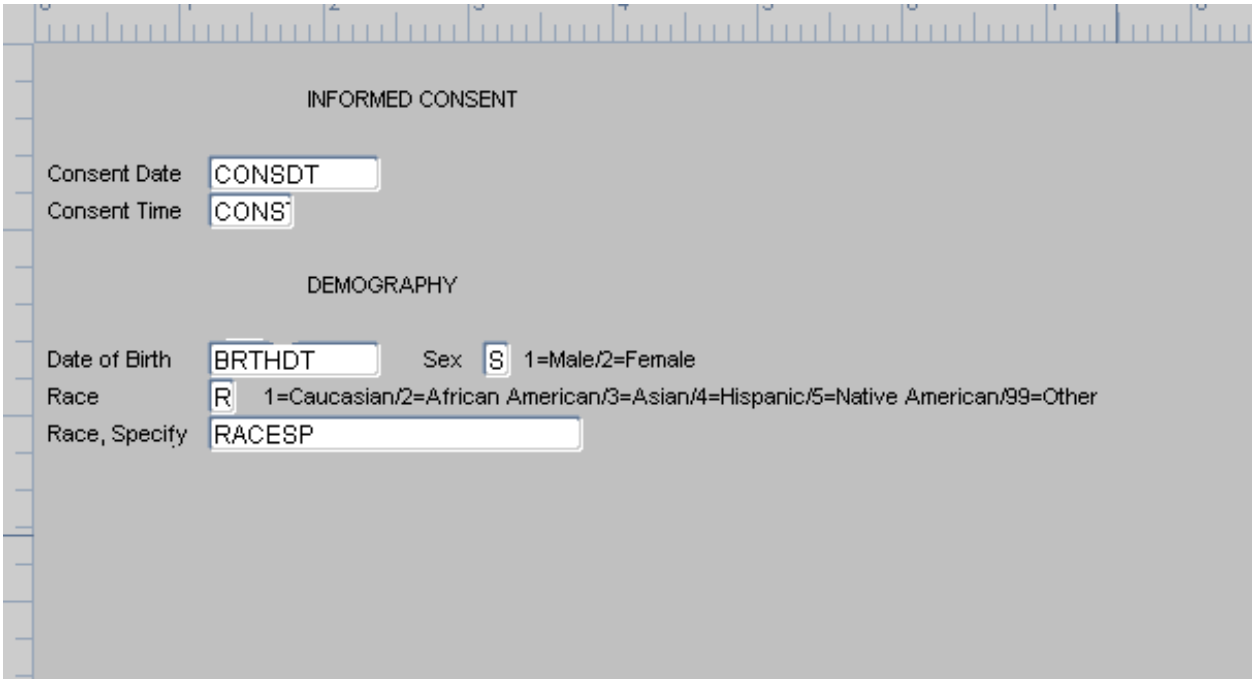
Inclusion Criteria IEINEX		Yes (1)	No (0)	N/A (2)
1.	Subject is a healthy man or woman.	☐	☐	
2.	If male, subject must be surgically sterile or agree to use an appropriate method of contraception.	☐	☐	☐ (Females Only)
3.	If female and of childbearing potential, subject must be surgically sterile.	☐	☐	☐ (Males Only)
4.	Subject is at least 18 years old or older.	☐	☐	
5.	Subject is able to understand the study procedures, agree to participate in the study program, and voluntarily provide written informed consent.	☐	☐	

This is the real example for CRF page. The investigator at the site gets the above information (informed consent, demographics and inclusion criteria) from the patient, fills the CRF and sends it out to Data Manager. The data manager enters the data in the data entry screen.

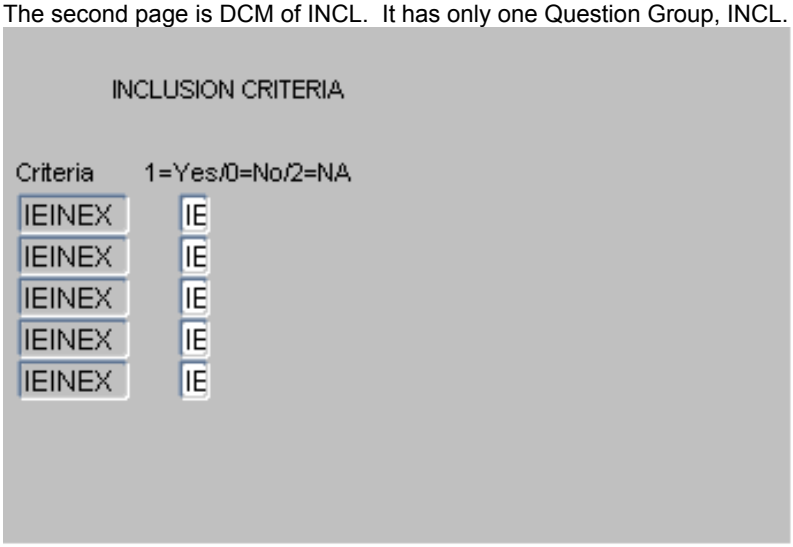
THE DATA ENTRY SCREENS IN ORACLE CLINICAL

The real database entry screen in OC looks like the following. This is the DCI of 'CONS / DEMO / INCL'. There are two screens for the sample CRF page.

The first page is the DCM of DEMO. It has two Question Groups, CONS & DM.



The second page is DCM of INCL. It has only one Question Group, INCL.



As shown above, each DCM represents each page of data entry.

THE HIERACHY STRUCTURE OF OC DATABSE FROM SAMPLE CRF PAGE

DCI Books for Study000

Name	Description
CRF_STUDY000	CRF pages for Study000
TRACKING_STUDY000	page tracking for Study000

DCI Book Pages for CRF_STUDY000

DCI Name	Display	Start Page	Clinical Event
CONS / DEMO / INCL	1	1	SCREENING

The DCI Name of 'CONS / DEMO / INCL' represents the first page of the sample CRF and its visit is 'Screening'

DCM for the DCI Name of 'CONS / DEMO / INCL'

DCM Name	Disp Seq #	Subset Name	DCM Domain
DEMO	1	DEMO	STANDARD
INCL	2	INCL	STUDY000

DCI Name has 2 DCM: DEMO and INCL. The 'standard' of DCM domain means that the DCM of DEMO is a universal DCM and 'study000' means that INCL is the study specific DCM for STUDY000. In other word, INCL is the new DCM and the modified DCM for STUDY000.

Question Group (QG) for the DCM of DEMO

DCM Question Group Name	Library Question Group Name	Question Group Domain	Short Name	Disp Seq #
CONS	CONS	STANDARD	CONS	1
DM	DM	STUDY000	DM	2

DEMO has 2 QG: CONS and DM. As explained for DCM domain, DM is the study specific QG. The 'Disp Seq #' represents the order of the prompt in data entry screen.

Question Group (QG) for the DCM of INCL

DCM Question Group Name	Library Question Group Name	Question Group Domain	Short Name	Disp Seq #
INCL	INCL	STANDARD	INCL	1

INCL has only one QG: INCL.

Questions for the Question Group CONS

Question Name	Question Domain	Disp Seq #	Data Type	Data Time Format	Length	DVG Name	SAS Name	SAS Label
CONSDT	STANDARD	1	DATE	DMY	8		CONSDT	Consent Date
CONSTM	STANDARD	2	TIME	HM	6		CONSTM	Consent Time

CONS has 2 Questions: CONSDT and CONSTM. The sequence number also represents the order of variable prompted in data entry screen. The data type and format are the Oracle Clinical data type and format. The length represents the maximum of data captured in data entry. The 'SAS Name' and 'SAS label' will be the name and label when extracted to SAS environment.

Questions for the Question Group DM

Question Name	Question Domain	Disp Seq #	Data Type	Data Time Format	Length	DVG Name	SAS Name	SAS Label
BIRTHDT	STANDARD	1	DATE	DMY	8		BIRTHDT	Birth Date
SEX	STANDARD	2	CHAR		1	SEX	SEX	Sex
RACE	STANDARD	3	CHAR		2	RACE	RACE	Race
RACESP	STANDARD	4	CHAR		40		RACESP	Other Race

DM has 4 Questions: BIRTHDT, SEX, RACE and RACESP. Unlike BIRTHDT and RACESP, SEX and RACE have DVG which functions as a format.

Questions for Question Group INCL

Question Name	Question Domain	Disp Seq #	Data Type	Data Time Format	Length	DVG Name	SAS Name	SAS Label
IEINEX	STANDARD	1	CHAR		6	INCLQ	INQ	Inclusion Criteria
IEINA	STANDARD	2	CHAR		2	YESNO	INA	Inclusion Response

INCL has 2 Questions: IEINEX and IEINA. Notice that the questions of IEINEX and IEINA will be renamed to INQ and INA in the Oracle view. This kind of practice is not recommended, but it can be done.

Discrete Value Group (DVG) for Questions, SEX, RACE, INQUEST and INANS

DVG Name	DVG Domain	Description	Seq #	DVG Value	DVG Long Value
SEX	STANDARD	Gender	1	M	Male
			2	F	Female
RACE	STANDARD	Race	1	1	Caucasian
			2	2	African American
			3	3	Asian
			4	4	Hispanic
			5	5	Native American
			99	99	Other
YESNO	STANDARD	Yes/No	0	NA	Not Applicable
			1	Y	Yes
			2	N	No
INCLQ	STUDY000	Inclusion Criteria	1	INCL1	Subject…
			2	INCL2	If male…
			3	INCL3	If female…
			4	INCL4	Subject is …
			5	INCL5	Subject is able…

For the question of SEX, the data entry screen will prompt the choice 1 or 2. If 1 is chosen, the screen will display the value of 'M' and OC creates 3 variables in the database : SEX(Sex), SEXN(Sex – DVN) and SEXL(Sex – DVL). SEX contains the value from 'DVG Value', SEXN the value from 'Seq #' and SEXL the value from 'DVG Long Value'.

The above Oracle Clinical structure creates the views in an Oracle Database. Since OC is the interface software for Oracle, OC database can be seen as views in Oracle, more specifically in PL/SQL. SAS programmer can relate QGs in OC to views in PL/SQL, Questions to Variables.

The views created by above OC
The views created by above OC database are CONS, DM and INCL.

The variables created in the view of CONS, DM and INCL are the following. The standard variables such as SUBJID, SITEID, SUBJINIT, VISIT and VISITNUM will be created for all the views. Usually, there will be more standard variables.

View Name	Variable Name	Variable Length	Variable Label	Variable Type
CONS	SUBJID	8	Subject ID	NUM
	SITEID	10	Site ID	CHAR
	VISIT	30	Visit	CHAR
	VISITNUM	8	Visit Number	NUM
	…..			
	CONSDT	8	Consent Date	CHAR
	CONSTM	4	Consent Time	CHAR
DM	…..			
	BIRTHDT	8	Birth Date	CHAR
	SEX	1	Sex	CHAR
	SEXN	1	Sex – DVN	NUM
	SEXL	6	Sex – DVL	CHAR
	RACE	2	Race	CHAR
	RACEN	2	Race – DVN	NUM
	RACEL	15	Race - DVL	CHAR
	RACESP	40	Other Race	CHAR
INCL	…..			
	INQ	4	Inclusion Criteria	CHAR
	INQN	1	Inclusion Criteria – DVN	NUM
	INQL	200	Inclusion Criteria - DVL	CHAR
	INA	2	Inclusion Response	CHAR
	INAN	1	Inclusion Response – DVN	NUM
	INAL	14	Inclusion Response - DVL	CHAR

The source codes for view CM
The OC is the interface of Oracle. So, whatever is created in OC is also written in Oracle. SAS programmers can view the source codes for view CM in PL/SQL Developer. Please notice that real source codes in Oracle are a lot more complicated and the following is the simplified version.

```
create or replace view study000.demo as
       select 'Study000' as study, a.subjid, a.siteid, a.visit, a.visitnum,,,,
              b.consdt, b.consttm
       from rdcm a, response_view b,,,
       ..
quit;
```

The sample codes for extracting program
This view can be extracted by SAS program. The following is the example codes for extraction SAS program.

```
libname raw "C:/Study000/Raw";
libname ocdb oracle path="" user="" password=""
       schema=" Study000";
proc sql;
       create table raw.cons as
              subjid, siteid, visitnum, visit,,,,
              consdt, constm
       from ocdb.cm
quit;
```

The above codes create the CONS data set in RAW directory. SAS programmers can create other data sets such as demo in the RAW directory in the same way.

THE ADVANTAGE OF KNOWING ORACLE CLINICAL DATABASE

The understanding of Oracle Clinical database helps the SAS programmers to link the relationship between CRF and extracted raw data. SAS programmers will be able to understand how many data sets are created for each CRF page and how many events or visits should be obtained in each data set.

It also helps SAS programmers to understand what each variable in raw data represents. The Oracle Clinical creates more data than entered in CRF. Some are standard and some derived. When merging more than one raw data set to create analysis data sets, SAS programmers will be able to find better variables for merging.

SAS programmers and OC developers can also work together to automate the process so that SAS programmers do not need to spend too much time in programming. For example, both can decide to use the consistent variable names so that SAS programmers do not need to rename the variables on each clinical trial. This process will help to save the time and efforts in both sides.

SAS programmers can have better communications and working relationships with data management teams. The understanding of OC database helps us to have a better understanding on the data management side - their issues, problems and even shortcomings.

Therefore, knowing OC database helps SAS programmers to get better in SAS programming. The job descriptions of SAS programmer go beyond programming. We interact with other team members. We need to understand how the whole study team works. The SAS programming is a part of process. Knowing more process makes us better SAS programmers.

CONCLUSION

SAS programmers need to understand that the Oracle Clinical is mainly used for the data management while SAS for statistical analysis. The OC database is built based on this idea of helping the data management. SAS programmers extract the raw data sets from OC database and generate the analysis data sets. In other words, SAS programmers convert the idea of data management to that of statistical analysis. The understanding of OC database helps to narrow the gaps between OC and SAS environment, furthermore data management and statistical analysis.

CONTACT INFORMATION

Your comments and questions are valued and welcomed. Please contact the author at

Kevin Lee
Therakos
Exton, PA
(610) 280 - 1119
Email:Kevin.kyosun.lee@gmail.com

TRADEMARKS