
Reconciling CRF and SAS Data

Jim DePasquale
Synteract, Inc.



Background

- **Must perform audit of CRF data compared to what data is in SAS data set**
 - We do not export by plate, rather by creating data sets that may contain data from more than 1 plate
 - For example, a vital signs data set
 - Little correlation between a DF record and a SAS record
- **Use patient profiles to reconcile CRF data and SAS data**



Background

- Legacy system was SAS 6 based software
 - Required data processing staff to order variables within a dataset and the data sets within the report
 - Result was a manual and time consuming process to create patient profile report.
- Already create an annotated CRF document that maps CRF data to SAS data sets



Overview

- Patient profile generator program reads in ACRF file and creates SAS program
- SAS program created is a series of PROC REPORTS used to create 1 RTF document per patient
- Output document is in the order of the patient's CRF book allowing for easy comparison to original CRFs



Step1: Annotated CRF

- Text based document that maps CRF data to SAS data set

```

james@syncom171 /tmp
Window Edit Options Help
-----
#PLT FLD  VARIABLE  DATASET  TYPE  LEN  FORMAT  LABEL
-----
001 7    PATID    IE      Char  4    blank   Subject Identifier For Internal Purposes
001 7    SUBJID  IE      Char  8    blank   Subject Identifier for the Study
001 8    SITEID  IE      Char  4    blank   Study Site Identifier
001 11   IEYN    IE      Char  1    blank   Did Subject Meet All Eligibility Crit.
001 13   IEESTCD IE      Char  3    blank   Exception Criteria Short Name
001 14   IERESP  IE      Char  50   blank   Reason for Enrollment Exception
001 15   IESPON  IE      Char  25   blank   Individual Approving Exception
001 16   IEEXDC  IE      Char  11   blank   Date of Exception - Char
001 16   IEEXDT  IE      Num   8    YYYYMMDD10. Date of Exception - Num
-----
#PLT FLD  VARIABLE  DATASET  TYPE  LEN  FORMAT  LABEL
-----
002 7    PATID    DM      Char  4    blank   Subject Identifier for Internal Purposes
002 7    SUBJID  DM      Char  8    blank   Subject Identifier for the Study
002 7    PATID    IE      Char  4    blank   Subject Identifier for Internal Purposes
*acrf* 386 lines, 28177 characters
    
```



Step1: Annotated CRF(cont.)

- Perl script reads the ACRF file and uses the DFTips file to find the location of the field in the PS file used for study setup
- Output is each plate image with field number superimposed along with the variable attributes in a SAS PROC CONTENTS-like format



Step1: Annotated CRF (cont.)

SCREENING		
Site Number	Subject Number	Subject Initials
□□□□	□□□□	□□□□
INCLUSION/EXCLUSION CRITERIA		
Did the Subject meet all eligibility criteria? <input type="checkbox"/> Yes <input type="checkbox"/> No		
*If NO, please complete the following		
Criteria # NOT Met (Please enter number of Criteria Not Met)		
□□	Exception explanation:	_____ Date Granted: _____
□□	Exception granted by:	_____ Date: □□/□□/□□
□□	Exception explanation:	_____ Date Granted: _____
□□	Exception granted by:	_____ Date: □□/□□/□□
□□	Exception explanation:	_____ Date Granted: _____
□□	Exception granted by:	_____ Date: □□/□□/□□
□□	Exception explanation:	_____ Date Granted: _____
□□	Exception granted by:	_____ Date: □□/□□/□□
V1 09/02/2004 Confidential CRF		

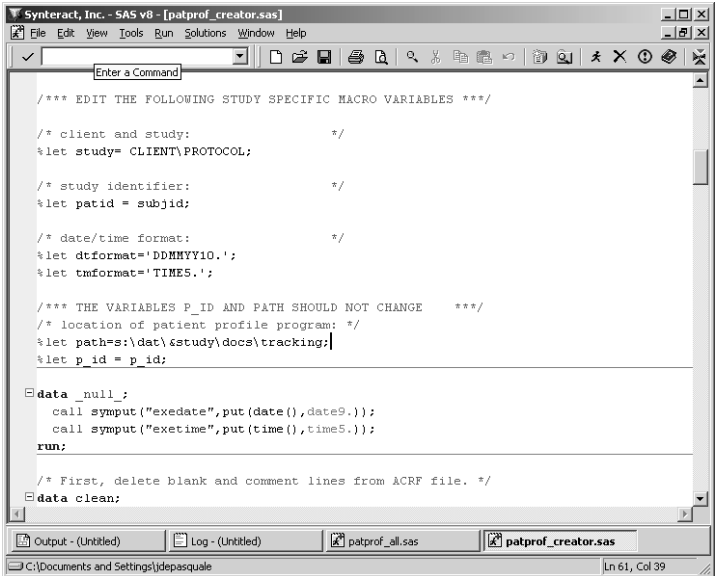
Data Set - Prepared by Department, Inc. - on October 12, 2004						
#	Variable	Default	Type	Length	Format	Label
1.	STUDID		Char	4		Subject Identifier for Internal Response
2.	STUDID		Char	4		Subject Identifier for the Study
3.	STUDID		Char	4		Study Site Identifier
4.	STUDID		Char	2		Exception Criteria Short Name
5.	STUDID		Char	50		Reason for Enrollment Exception
6.	STUDID		Char	10		Individual Approving Exception
7.	STUDID		Char	11		Date of Exception - Char
8.	STUDID		Char	8	YYMMDD10	Date of Exception - Num

STEP 2: Create SAS PROC REPORTS

- SAS program that parses the ACRF text file to create a SAS program that creates the patient reports
- Only need to modify study specific information: paths on network, unique patient identifier and date format
- Program determines the relationship between a plate, data set and the order of the variables on the page



STEP 2: Create SAS PROC REPORTS (cont.)



```

Synteract, Inc. - SAS v8 - [patprof_creator.sas]
File Edit View Tools Run Solutions Window Help
Enter a Command

/**** EDIT THE FOLLOWING STUDY SPECIFIC MACRO VARIABLES ****/

/* client and study: */
%let study= CLIENT\PROTOCOL;

/* study identifier: */
%let patid = subjid;

/* date/time format: */
%let dtformat='DDMMYY10.';
%let tmformat='TIMES.';


/**** THE VARIABLES P_ID AND PATH SHOULD NOT CHANGE ****/
/* location of patient profile program: */
%let path=:\dat\study\docs\tracking;
%let p_id = p_id;

data _null;
  call symput ("exedate",put (date (),date9.));
  call symput ("exetime",put (time (),time5.));
run;

/* First, delete blank and comment lines from ACRF file. */
data clean;

```

Output - (Untitled) Log - (Untitled) patprof_all.sas patprof_creator.sas
C:\Documents and Settings\jdepasquale Ln 61, Col 39



Step 3: PROC REPORTS

- The PROC REPORT statements contain the order of the variables and the length of the variable
- Before each PROC REPORT there is a DATA step
 - Appropriate subsetting of data by patient and visit, if needed
 - Patient and visit
 - Variable labels can be modified
 - Observation number is added to record

Step 3: PROC REPORTS (cont.)

- Once plate and data set relationship is mapped out, it can be repeated via macro call for each visit



Step 3: PROC REPORTS (cont.)

```
Syneract, Inc. - SAS v8 - [patprof_all.sas]
File Edit View Tools Run Solutions Window Help
.....
* plate:001 data set: IE ;
.....
%macro IE (page,num);

title6 "plate:001 dataset:IE ";
title7 "Page %page";

data IE ;
  set db.IE ;
  by subjid;
  if subjid eq &p_id and seqnum eq &nnum;

  line=_n_;

  label ietestcd = "Criteria # NOT Met"
        iereas = "Exception Explanation"
        iexdc = "Date Granted"
        line = "Obs";

run;

proc report data=IE headline headskip missing split='@' nowindow
  columns line formname studyid domain FATID SEQNUM
         SUBJID
         SITEID
         IEYN
         IETESTCD
         IEREAS
         IESPON
         IEXDC
         line;
  define line / display width=4 center flow;
  define formname / display width=40 center flow;
  define domain / display width=12 center flow;

```



Step 4: The Report

- Program utilizes the SAS Output Delivery System (ODS) to create a RTF document suitable for viewing and printing in MS Word
- Output document is easy to read and follows the order of the patient's CRF book
- Footer fields contain QC sign off area for documentation



Step 4: The Report (cont.)

A screenshot of an Adobe Acrobat window displaying a clinical trial report form. The window title is 'Adobe Acrobat - [original.pdf]'. The form includes a header with a barcode and fields for 'Site Number' (01011), 'Subject Number' (2010), and 'Subject Initials' (redacted). Below this is the 'INCLUSION/EXCLUSION CRITERIA' section. A question asks 'Did the Subject meet all eligibility criteria?' with 'Yes' and 'No' options. Two exceptions are listed: '102' and '103', both with explanations 'IGFOP-3 CRITERIA INCREASED TO 10%' and 'OF 5D OF -2'. Each exception includes a date granted and the name of the person granting the exception. The bottom of the form has a 'Date Granted' field and a 'Exception granted by' field with a signature line.

Step 4: The Report (cont.)

page 1 of 46


plate#01 dataset:IE
Page 1

Obs	Form Name	Study Identifier	Domain Abbreviation	Subject Identifier for Internal Purposes	DataSet Sequence Number	Subject Identifier for the Study	Study Site Identifier
95	INCLUSION CRITERIA	DFUG05	IE	5010	1	00015010	0001
96	INCLUSION CRITERIA	DFUG05	IE	5010	1	00015010	0001

Did Subject Meet All Eligibility Crt.	Criteria # NOT Met	Biocroton Description	Biocroton Individual Approving Biocroton	Date Created	Cha
N	102	10% CRITERIA INCREASED TO 10% OF ID OF 2	MD	13AUG2004	95
N	103	100% CRITERIA INCREASED TO 10% OF ID OF 2	MD	13AUG2004	96

a:\data\test\protocol\dsort\adtag\0015010_28OCT2004_allref_28OCT2004_1454

Reviewed by: _____
Date reviewed: _____



Advantages

- Program works well across multiple studies
- Minimal manipulation of SAS data sets
 - most of the time, none at all
- Report is easy to compare with original CRF data



What's Next?

- Incorporate visit map and page map information in order to further automate report creation
- Interface that allows data management staff to change certain variable attributes such as column width and labels without overwriting existing data



Questions?

Jim DePasquale
Application Developer II
Synteract

