Statistical Review and Validation of Patient Narratives

Indrani Sarkar, inVentiv Health Clinical, Indiana, USA
Bradford J. Danner, Sarah Cannon Research Institute, Tennessee, USA

ABSTRACT

Patient narratives are often a required component to a safety review or a clinical summary report being submitted to a regulatory agency. Events that will often require a narrative are deaths, other serious adverse events, and significant adverse events determined to be of special interest because they are clinically relevant. The manner in which they may be reported can vary, but nearly always include records of a patient's demographics, medical history, drug administration, and adverse events over the course of a clinical study. Combining the various sources of data into one cohesive presentation can be challenging, and possibly more difficult or tedious, the validation and review of each individual profile produced. Considering all the various sources of data which need to be checked and cross-referenced, time and effort can become substantial. The main purpose of this paper is to briefly describe a method, utilizing SAS®, by which we first confirm the list of subjects and events compiled by study clinicians is consistent to the study data, followed by a presentation of outputs from which comparisons may be made during statistical review and/or validation.

INTRODUCTION

Patient profiles present information collected for a subject over the course of a clinical trial that allow a medical writer or reviewer to draft a narrative to coincide with an event of interest. Possible events of interest are commonly serious and or study drug related adverse events, events that lead to withdrawal from the study, and events that result in death while on study. Instead of sifting through a variety of listing sources (demographics, medical history, adverse events, concomitant medications, etc.) for the subject of interest, the information of that one subject can be listed or graphically presented together in one place. Seems simple enough, however, production and validation of patient narratives by statistical SAS programmers can be quite challenging, considering the variety of data that are required and shifting demands and requirements from the reviewers. Discrepancies can arise between programming and the review team, and between events of interest requested from the clinical team, and those that are found in the data.

Generally speaking, we have found that patient narratives can be designed and delivered in two ways; the first alternative is that one file is produced per subject; a second alternative is that all individual narratives for a given batch are included in one large file. Producing individual files can make validation and review seem easier since they can be done on a case by case basis, however, management of the potentially large number of files can be annoying for the medical reviewers. Alternatively, dumping many profiles into one large file makes file management easy, but review of all the narratives becomes tedious when changes required for one narrative can inadvertently affect many others. The method by which our sponsor preferred to have their narratives was the multiple subjects within one or few files.

The process we generally followed in producing the patient profiles begins with discussions with the sponsor and medical writing team to determine the specifications of profiles to be used for narrative writing purposes. In addition to the document specifications, the clinicians for the studies will compile a list of subjects, commonly referred to as the narrative tracker, and the events of interest that will require narratives. This list is then used as the primary source for identifying subjects and events in the data. Concurrently we are required to check that the list of subjects and events provided by the clinicians relative to the raw data is complete and accurate and that no events that would warrant a patient narrative are omitted. The profiles are then programmed, and validated by a peer SAS programmer, and then subsequently reviewed by the study statistician(s).

As validation programmers or statistical reviewers, there are several things we need to check with final profiles going to the clinicians and medical reviewers. The most obvious of which are adherence to the document specifications, and that the information presented in the profiles is cosmetically appealing and correct. More challenging and frustrating we have found is 1) the verification that all subjects and events of interest are included in the final profile document and as reflected in the narrative tracker compiled by the study clinicians, and 2) quickly organizing diagnostic output with which to compare against the patient narrative document. Two techniques we have had reasonable success with will be briefly presented – use of SAS to detect, present, and resolve discrepancies between the clinician's narrative tracker and the CRF data; and programming which produces individual per patient, yet linked, html files to compare against the larger narrative document.

METHODS AND RESULTS

To produce the set of patient profiles required for the narratives, clinicians will send us a list of subjects and their events of interest in a spreadsheet. Typically, this tracker will contain dozens of records, therefore a manual review becomes very time consuming and increases the possibility of human error. To minimize the time and chances of making errors, SAS code is used to perform the reconciliation.

Numerous data sets (adverse events, demographics, medical history, disposition, test compound administration, for example) need to be screened for all the subjects listed with an event of interest in the sponsor's tracker. SAS is used to check all the relevant datasets for subjects with events of interest, and then compares it with those sent by the sponsor in the tracker. All the discrepancies are compiled and stored in a spreadsheet. The spreadsheet will contain a column showing the patient ID, the reason for discrepancy, and whether the subject and event are present in datasets, or present in the tracker.

Frequently, the differences are based on wrong adverse event terms or onset date of an event. The following generalized SAS code illustrates how a difference of this nature may be detected:

```
data aller_ds;
    length reason $200 in_crf $3 in_track $3;
    merge  anaphyx(in=a keep= usubjid ptname startdt stopdt)
           aller_trac(in=b keep= usubjid ptname startdt);
    by usubjid ptname startdt;
    if not b;
    reason= 'Event present in data but not in tracker';
    in_track = 'No';
    in_crf= 'Yes';
run;

data aller_tracker;
    length reason $200 in_crf $3 in_track $3;
    merge  anaphyx(in=a keep= usubjid ptname startdt stopdt)
           aller_trac(in=b keep= usubjid ptname startdt);
    by usubjid ptname startdt;
    if not a;
    reason= 'Event present in tracker but not in dataset';
    in_track = 'Yes';
    in_crf= 'No';
run;
```

The dataset 'anaphyx' contains all the subjects with an event of interest reported in the CRF datasets, and the dataset 'aller_trac' contains all the subjects obtained from the tracker with the same condition. The dataset 'aller_ds' contains all the subjects which are present in the raw dataset but not in the tracker. Similarly, 'aller_tracker' shows all the subjects which are present in the tracker but not in the raw dataset. The two resulting datasets are combined and used as the basis for the patients identified (PID) spreadsheet (Display 1).

	Reason for PID	Subject	AE tracker term	Start Date	In CRF data	In tracker
1	**Reason for PID**	**Subject**	**AE tracker term**	**Start Date**	**In CRF data**	**In tracker**
2	Event present in tracker but not in dataset	10071002	ASTHMA	03JAN2012	No	Yes
3	Event present in dataset but not in tracker	10071010	MISCARRIAGE OF PREGNANCY	04OCT2011	Yes	No

PID list

Ready | 100%

Display 1. Example of differences found and contained in the PID spreadsheet.

Display 1 shows discrepancies between the tracker and CRF datasets, found by the PID program. The resulting spreadsheet is then sent back to the sponsor for review. In this example (Display 1), the clinicians will revise the tracker to include the missing subject and event or by updating the adverse event term or start date, such that it is consistent with the CRF data. After multiple drafts of both the narratives and the clinician's tracking documentation, a final version of tracker is generated. The final tracker should be consistent with the CRF data and the spreadsheet generated by the PID program should be empty, demonstrating the consistency.

The second part of the review process is to check all the information presented in the narrative document. A typical subject narrative includes demographic information, medical history, test compound administration, and all adverse

events. SAS is used to read in the tracker containing all the subjects and events of interest and then collates all the relevant information from CRF datasets. Comparative outputs are organized as 1) an all-inclusive list of subjects and their events of interest, which is then subsequently linked to 2) individual per subject profiles, printed to an html file that can be viewed with a web browser, and 3) a single PDF repository which contains all the information in one file.

The following generalized SAS code illustrates the steps employed to quickly generate the individual comparative patient profiles, have them automatically linked to the all-inclusive list of subjects and events, and create the single PDF repository.

Step 1. From a list of subjects (dataset sublist) capture number of subjects and the subject numbers.

```
proc sql noprint;
select count(distinct(usubjid)) into :nsubs from sublist;
select distinct(usubjid) into :subj1-:subj%trim(%left(&nsubs)) from sublist;
quit;
```

Step 2. Set path for outputs and include print of relevant datasets.

```
%let destination=&path.profiles\;
%macro s_print(subject=);
*** simple PRINTS of comparative datasets ***;
ods proclabel "&subject. <dataset label>";
title1 "&subject. <dataset label>";
proc print data=<dataset name> noobs;
    where usubjid="&subject";
    var <vars of interest>;
run;
*** include multiple PRINTS for additional datasets  ***;
%mend s_print;
```

Step 3. Print the individual html files.

```
%macro i_print;
%do i=1 %to &nsubs.;

ods proclabel "&&subj&i..";
ods html style=normal file="&destination.narrative_&&subj&i...html";
  %s_print(subject=&&subj&i.);
ods html close;

%end;
%mend i_print;
%i_print;
```

Step 4. Print the inclusive list of subjects and events of interest, linked to the individual patient narratives.

```
ods tagsets.Excelxp
style=normal file="&destination.check_narratives.xml"
options(frozen_headers='Yes' autofilter="All"  sheet_name="Subject List");

proc report data=sublist nowindows;
    column short usubjid comment;
    define short/display 'Narrative Category';
    define usubjid/display 'Subject ID' style={color=blue};
    define comment/display 'Event of Interest';
    compute usubjid;
          rtag="&destination.narrative_"||strip(usubjid)||".html#";
          call define(_col_,'url',rtag);
    endcomp;
run;

ods tagsets.ExcelXP close;
```
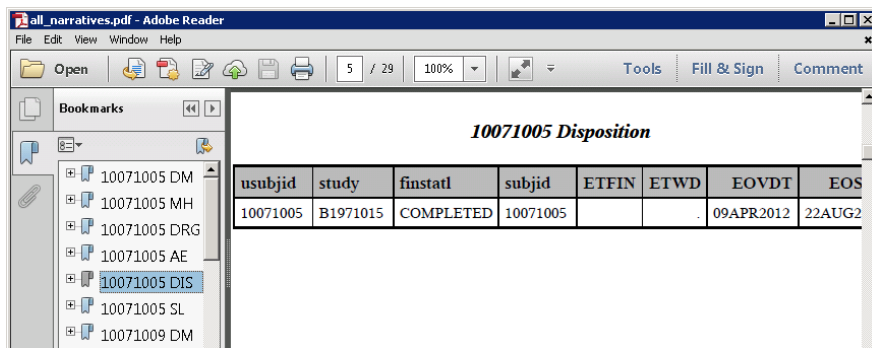
	A	B	C
1	**Narrative Category**	**Subject ID**	**Event of Interest**
2	AE of special interest	10071005	KIDNEY STONE
3	AE leading to withdrawal	10071009	ALLERGIC REACTION - NOS
4	AE leading to withdrawal	10071043	MODERATE ASTHMA EXACERBATION
5	AE leading to withdrawal	10071044	WORSENING OF EXERCISE INDUCED BRONCHOSPASM
6	AE leading to withdrawal	10071067	STAPH INFECTION
7			
8			

Subject List

Display 2. All-inclusive list of subjects and events, automatically linked to individual comparative narratives.

Display 2 illustrates all the subject numbers, their event of interest, and the reasons requiring the narrative. The hyperlinked subject numbers are associated with the corresponding html file through the use of the compute statement in the PROC REPORT, and by how the individual files were named in Step 3.

Step 5. Create one PDF repository for all information.

```
ods pdf style=printer file="&destination.all_narratives.pdf" pdftoc=1;
%macro p_print;
    %do i=1 %to &nsubs.;
        %s_print(subject=&&subj&i.);
    %end;
%mend p_print;
%p_print;
ods pdf close;
```

all_narratives.pdf - Adobe Reader

File Edit View Window Help

Open | 5 / 29 | 100% | Tools | Fill & Sign | Comment

Bookmarks

⊞ 10071005 DM
⊞ 10071005 MH
⊞ 10071005 DRG
⊞ 10071005 AE
⊞ 10071005 DIS
⊞ 10071005 SL
⊞ 10071009 DM

10071005 Disposition

usubjid	study	finstatl	subjid	ETFIN	ETWD	EOVDT	EOS
10071005	B1971015	COMPLETED	10071005		.	09APR2012	22AUG2

Display 3. All comparative narratives compiled in one compact file.

The PDF repository (Display 3) is an option for those who want to have all the information in one place instead of going through each and every html file.

CONCLUSIONS

Patient narratives are a necessary component of many regulatory submissions. Insuring the accuracy and completeness of the profiles required for patient narratives is very important. The use of SAS programming helps establish consistency between the narrative tracker compiled by study clinicians and the raw CRF datasets. We have found that the use of the PID spreadsheet to contain and review discrepancies is an efficient tool for producing a complete and accurate narrative document.

For review and validation of the narrative document, the presentation of the subject list in a spreadsheet centralizes the verification of narrative type, subjects, and event of interest to one place. And the filters automatically activated in the spreadsheet allow the user to quickly focus their review to a specific area if necessary. Embedded linkage within the spreadsheet allows us to quickly access relevant subjects, while simultaneously keeping individual narratives separate from each other, and provides an elegant and efficient method to review many narratives concurrently without struggling with clumsy or large file sizes.

Perhaps the best part of the code to produce the multiple comparative files is that it is not complex and is relatively easy to implement. Reviewers or validators can quickly compare the overall subject list to the subjects included in a large collated patient narratives document, note overall discrepancies in the spreadsheet, and send back to the programming team for revisions. Subsequently, the spreadsheet also provides a possible repository to keep track of comments and issues detected with the individual subjects in the narrative document.

ACKNOWLEDGMENTS

CONTACT INFORMATION

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

Indrani Sarkar
inVentiv Health Clinical, LLC
225 South East Street, Suite 200
Indianapolis, IN 46202
Work Phone: (812) 322-1930
E-mail: indrani.sarkar@inventivhealth.com

Bradford J Danner
Sarah Cannon Research Institute
3322 West End Avenue, Suite 900
Nashville, TN 37203
Work Phone: (615) 906-7969
E-mail: bradford.danner@scresearch.net

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.