PharmaSUG 2016 – Paper AD15

Automating Patient Narratives – The Medical Writer Loves Me!

Scott Burroughs, PAREXEL International, Durham, NC

Patient narratives have long been part of the safety review process for studies that go on to submission. For much of this time, they have been done by Medical Writers who pored through the CRFs and enter each subject's data by hand. With larger and longer studies, this can be quite the arduous task.

Newer methodologies have toyed with the use of a Mail Merge to a spreadsheet, but with the data available to us, why can't these be programmed/automated by us? They can! As long as the wording of each sentence and paragraph follows a script of sorts, it should be easy, right? Not so fast… formatting and length can be issues to tackle. This paper goes into detail on the various issues that arose doing my first patient narratives for a Medical Writer.

INTRODUCTION

Medical writers have probably feared and loathed patient narratives through the years in the pharmaceutical industry, but that could come to an end with the help of us programmers. As long as the timelines work out correctly…the data available to the programmer in time and the date the narratives are due is late enough…we can help tremendously by programming them.

The idea to possibly do patient narratives for a number of submission studies arose initially for several very long (and large) studies that lasted potentially up to 7+ years long. These were to be safety narratives for subjects who withdrew due to an AE. That reduces the number of subjects considerably, but to do them by hand would still be voluminous, considering that I was to list all adverse events that occurred over a period up to 7+ years and also list the complete medical history for subjects who are very sick.

ISSUES

1. The first issue was a simple one….can it be done programmatically? While yes, we had all of the data, but first of all, can we generate the output in the correct format that was required of them?
2. Not really a big issue, but something different for me. My previous company was late to the game for switching to CDISC standards, so this would be my first time programming using ADaM data.
3. I was given a 'script' for the wording format and order of the paragraphs/data to be presented. It may not be perfect, so there may be room to improve it or data reasons to modify it.
4. Perhaps narratives in the past didn't go longer than one page, but with the factors listed above, there's a likelihood that these studies might for some subjects. Dealing with that might not be so easy.
5. For the medical history and concomitant medications, what variables do we use….the preferred terms or the raw values? The preferred terms will not have typos, but details are often lost. Do we attempt to 'massage' the raw values to make them presentable to regulatory authorities?
6. How will data anomalies affect the output? Or any ad hoc changes to the report…

METHODS/SOLUTIONS

1. "The first issue was a simple one….can it be done programmatically? While yes, we had all of the data, but first of all, can we generate the output in the correct format that was required of them?"

As an experienced programmer, I know that pretty much any data can be programmed to be output in whatever data set I want. Whether or not that physical output can be done is the primary concern here before doing any data set manipulation.

Since the output would be going to the Medical Writer, I wanted it to be both as ‘wording’ perfect as possible (as little changing of the sentence format, wording, spelling, etc., by them) and ‘format’ perfect as possible, meaning as little format re-working as possible. Their software for the reports are in MS Word, so I wanted to be able to output the results in an RTF file.

But how do I output a table at the top (for overall subject info) , and freeform paragraphs (without frames/separators, etc.).? ODS RTF is how we output to an RTF file. For the table at top, PROC REPORT will get the job done, with a little fancy manipulation. Upon research, ODS RTF text= will allow the entire rest of the subject’s report to be done in one long text string in appended macro variables.

The table at the top was to look like this:

Protocol Id:	XYZPDQ123
Subject Number:	987123
Treatment:	Treatment A
AE(s) Leading to Discontinuation and/or Withdrawal (preferred term):	Adverse event preferred term

Everything in the first column is text and static, while everything in the second column is data driven and dynamic. However, note that these are columns in a table, so the first line across are the variable labels, and everything underneath are the data values. With the 2nd columns, I had to make the variable label=Study ID and the values from 3 different variables.

The code to get the header is such:

```
%macro wreeporte ;
proc report data=taybull nowd headline headskip
        split="~" missing
        style(report)={outputwidth=6.85 in}
        style(header)={background=white font_size=12pt font_face="Arial Narrow" frame=box
                 font_weight=medium just=left}
        ;
  column var1 var2;
  define var1     / width=30 flow "Protocol Id:"
    style(column)={background=white font_size=12pt font_face="Arial Narrow" frame=box
               cellwidth=2.5 in};
  define var2     / width=50 flow "&var2labill"
    style(column)={background=white font_size=12pt font_face="Arial Narrow" frame=box
               cellwidth=4 in};
run;
%mend wreeporte ;
```

For the body, since I don’t have SAS® 9.4, I couldn’t use the nice PROC ODSTEXT, which probably would have yielded the easiest and nice solution. I tried several options, including PROC REPORT (directly, but it goes to a new page, since the header is also a PROC REPORT). I can’t remember if I tried some form of PROC GREPLAY, with PROC GSLIDE and text or ODS LAYOUT, but that might have been a steep learning curve.

I discovered the use of ODS RTF text= and various styles that helped make it look just like something typewritten on the page.

The rather simple looking code to print the up to 6 paragraphs is such, where the text printed is all listed between the two double quotes you see at beginning and end:

```
ods rtf text="^S={leftmargin=0.5in rightmargin=0.5in font_size=12pt font_face='Times New Roman'}
^n
&p1
^2n
&p2
^2n
&p3
^2n
&p4a
^2n
&p4b
^2n
&p5
";
```

Each of the 6 paragraphs is separated by notation to add a blank line between them.

2. For this task, I would be programming using CDISC (ADaM) data for the first time, as my previous company (now my client) was late to the game for making the switch. Could I handle it? I've been programming in SAS in the pharma industry for over 20 years…..of course I can handle it. ☺ It's mostly different variable names and occasional changes in format (like horizontal vs vertical, etc.)

3. For the script that was given to me, I was able to work with the medical writer to 'massage' it some, for better readability and flow. Other modifications were for keeping the code flexible enough to take into account plural/non-plural, gender differences (like 'his vs. her'), or quantitative differences that would change wording. ('No concomitant medications were reported' vs. 'Concomitant medications included').

4. With the study lasting over 7 years and the condition studies being rather severe, the number of concomitant medications taken and/or the conditions listed in the medical history can be quite large, leading to the possibility of the text going longer than one page. That creates issues in programming and due to how the paragraphs are listed (each as one macro variable), I didn't want to deal with breaking up paragraphs (sections). So if paragraph normally would have naturally spilled onto the next page, it needed to be kept whole onto the next page. With the RTF output, I didn't think I had the luxury of using a 'lines left' type of function warning me that not enough room was left to print the entire paragraph on the current page.

Because of this, I needed to keep track of the size of each of the two portions of the second paragraph, which contained the medical history and the concomitant meds, to see if each subject's total text size will cause it to spill onto a new page.

5. For the medical history and concomitant medications, what variables do we use….the preferred terms/dictionary terms or the raw values? The preferred terms will not have typos, but details are often lost. The decision was to use the raw values, which meant a lot of typos/long form text/dates they wanted 'massaged' or abbreviated for consistency/saving space. Because of many instances of mixed case values and the massaging changing things, it seemed best to make all lower case and add build a dictionary of terms that get changed, so future narratives can has fewer changes needed each time.

For each new study, I had to send out a list of all new terms (2, one for med hx and one for conmeds) that haven't already been reviewed and changed for the medical writer to review. I had to then add each new term to the program's 'dictionary' of changes.

Here's an examples of the code I used to modify the terms (after I set them to lowercase) any time they arose. You can see because these are the verbatim terms, misspellings are included.

```
 **** strings length 7 cap 9th only****;
 do yy='vitamin','vitmain' ;
   z4=index(cmtrt,yy);
   if z4>0 then do;
     caps=1;
     substr(cmtrt,z4+8,1)=upcase(substr(cmtrt,z4+8,1));
   end;
 end;
```

From this example, I am actually capitalizing the 2nd letter after the end of the string (such as 'vitamin d').

6. A data anomaly that would affect the programming that I thought about ahead of time was the case where 2 (or more) events led to withdrawal. With the way the program was created, where there is one observation per subject (and are thus very 'wide' data sets), multiple AE's that led to W/D would affect the wording and structure of all paragraphs except the 2nd. For this situation (that I haven't yet encountered), I had programmed for it in the data manipulation part, but not the reporting part.

An ad hoc request that came up afterward was the request to add any deaths – including the event that caused the death (after study completion) - to the narratives. While not the same as a 2nd event that led to W/D, it still meant some decent-sized effect to the programming and several of the paragraphs.

PUTTING THE DATA TOGETHER

In order to put all the data together, I had to make all the sections 'horizontal' for one obs per subject. That meant a lot of PROC TRNANSPOSE, especially with the many conmeds and medical history terms these subjects had.

With the script I was given and discussion with the medical writer, we settled on 6 paragraphs. It makes sense to build each paragraph separately, and then build each paragraph by sentence, unless there is nothing dynamic about consecutive sentences.

Here's an example of the code used to put together paragraph 4, where different situations that may arise include 1) no concurrent events that started before the AE W/D and continued during, 2) One event that started before and continued during, and 3) multiple events:

```
data p4a(keep=subjid origdecod1 preconae multpreconae p4a);
 length p4a $700;
 set preconae4(where = (subjid="&num"));
 if preconae='' then p4a='No other adverse events started before and continued after the start date of
'||strip(origdecod1)||'.';
 else if multpreconae=. then
   p4a='An additional adverse event that started before and continued after the start date of '||
   strip(origdecod1)||' was reported as '||strip(preconae)||'.';
 else if multpreconae=1 then
   p4a='Additional adverse events that started before and continued after the start date of '||
   strip(origdecod1)||' were reported as '||strip(preconae)||'.';
 call symput("p4a",strip(p4a));
run;
```

The entire process per patient was in a macro that gets looped at the end over all the applicable subjects.

Here is the code used to do that:

```
options orientation=portrait;

goptions reset=goptions device=sasemf target=sasemf xmax=7in ymax=9.5in;
%MACRO loop_;

ods noptitle;
ods RTF file="&g_outfile..rtf" startpage=yes bookmark="Listing";* COMPRESS=9;

   %DO i = 1 %TO &numsubj;
      %pat(num=&&subj&i,star=*,star2=);
   %END;

   ods rtf close;
%MEND loop_;

%loop_;

%mend narrate;

%narrate;
```

An additional note is I originally built in some code to output the data to a QC data set to hopefully make it easier to QC, but my Qc person didn't use it. I think he did more spot-checking.

THE RESULTS

The following page contains an example of one page of a patient narrative of subjects who withdrew due to an AE:

Protocol Id:	XYZPDQ123
Subject Number:	654321
Treatment:	Placebo
AE(s) Leading to Discontinuation and/or Withdrawal (preferred term):	Hangnail

This 25 year-old, male subject 654321 was enrolled in a double-blind continuation study of Ear Wax Be Gone for removal of ear wax. The subject was previously enrolled in a blinded parent study (XYZPDQ122) of Ear Wax Be Gone and was randomized to 10 ml Ear Wax Be Gone. In that study, the subject received his first dose of Ear Wax Be Gone on 19NOV2007. In this double-blind continuation study, the subject received the first dose of Placebo 28 days on 17NOV2008. Treatment with Placebo was discontinued due to this event. The subject was withdrawn from the study on 13DEC2010 and the reason for withdrawal was reported as adverse event. The last dose was administered on 15NOV2010.

The subject's past medical history included helicopter parentitis, hypochondria, and addiction to taking selfies. Concomitant medications included testosterone, cortisone, cannabis 420, and Aleve.

On 27JUL2010 (981 days after the start of Placebo and 1 day after the most recent dose), the subject experienced hangnail (verbatim text: Ouchie on Selfie-Taking Finger), considered to be life-threatening in severity. During the AE, the subject received Xanax and numerous eye rolls by the investigator. The event was considered recovered/resolved on 09OCT2010.

No additional adverse events started before and continued after the start date of Hangnail.

An additional adverse event that started during or within 28 days after the Hangnail was reported as Bruised Ego (27JUL2010-09OCT2010).

The investigator reported that the causality of Hangnail was not related to Placebo.

CONCLUSION

After I finished these the first time, the Medical Writer gushed about how much of a time saver this was AND how much money it saved. The previous method for doing these was to contract out this task at $100 per subject and wait much longer for them to be produced. Apparently they were done by using the CRFs subject by subject and typed in one at a time. Some of these studies had close to 100 subjects where some subjects had many conmeds and medical history terms (over 100 for each!).

REFERENCES

CONTACT INFORMATION

BIOGRAPHY

Scott was a statistician for GlaxoSmithKline for almost 12 years until switching to a full-time programmer role 10 years ago. He has worked at PAREXEL International since March, 2015 and has worked in the Research Triangle Park, NC area since 1994. He has programmed in SAS extensively since 1992 while at a previous pharmaceutical company. He has a B.S. and an M.S. in Statistics from Virginia Tech.

CONTACT

Scott Burroughs
PAREXEL International
2520 Meridian Pkwy #200
Durham, NC 27713
scott.burroughs@parexel.com