

Using SAS to Automate the Production of Medical Writing Narratives

Andy York & Joe Martin

Covance CAPS

In many of the Clinical Trial studies we undertake at Covance there is a requirement for the production of Medical Narratives. These are usually produced for a defined subset of patients and they need to hold all the pertinent data for each patient together in as readable a form as possible.

The poster represented by this paper is in effect two posters in one – Part 1 (on the left hand side) covers the Six-Sigma process that was used to determine and measure “the process” – namely the time it takes to produce a Narrative. Part 2 (on the right hand side) covers the SAS process that was defined to auto-produce the header component of a Medical Writing Narrative

Part 1 - Six-Sigma Methodology

The PhUSE paper “Improving Biometrics Processes with Six-Sigma” in the “Technical Solutions” stream covers the Six-Sigma aspects of this poster in far more detail and additionally refers to several other reference projects to present an explanation of this process improvement methodology. To avoid repetition, the reader is directed to this paper in order to understand the six-sigma methodologies employed.

Table 1 below is intended to be a “teaser” to encourage you to look at the above mentioned paper. The purpose here is also to clarify the sigma score presented on the poster which shows that we obtained an improved sigma score of 5.22σ against a defect definition of narrative headers that take >8 minutes to produce. Whilst this is less than the coveted 6σ target is still represents a massive improvement over the old process score of -0.35σ where previously Narrative headers were being created manually!!

Table 1 : Table of Six-Sigma Scores vs Defects per-million Narratives

Process Capability σ_{st}	Defects per Million Narratives $_{lt}$	Long Term Yield
2	308,537	69.15%
3	66,807	93.32%
4	6,210	99.38%
5	233	99.98%
6	3.4	99.99966%

Part 2 – The SAS Solution

The Six-Sigma project determined that most of the information required for the Medical Writing Narratives is normally held in a SAS readable format or SAS datasets, so SAS was an obvious candidate for a tool to bring the necessary data together. However, while SAS is a powerful data crunching tool, it is not normally regarded as a good way to create English-like text. This poster is intended to summarise some of the challenges we faced in achieving this and also to highlight the some of the lessons learned.

The top right hand quadrant of the poster illustrates how data from some of the sources was combined to create an RTF file. Using Proc Report and macro variables we endeavoured to make the output look as much like a free format document as we could, and where there was data which had to be presented in tabular form then we tried to make this look like a normal Word table rather than a traditional SAS listing, using natural text within the cells as far as possible.

Although much of the data used was already in SAS datasets, other sources were also used. For example, lists of patients which had been chosen by Data Management were extracted from an Excel spreadsheet which had originally been created in a previous run of the process.

Because the Narratives were created automatically in this way, we were able to create prototype versions for the recipients to review and request format and content changes in a way which was not possible with the previous manual process.

In the lower right hand quadrant of the poster we have highlighted four of the challenges we faced when we first looked at this solution:

- As described already, we had to take input from a variety of sources and be able to produce refreshed reports at frequent intervals. This was very difficult with the previous system and we needed to make this work much more effortlessly.
- Creating RTF files is a standard part of SAS ODS, but unless options are setup correctly, Word is not able to do things like printing table headings out correctly at the top of pages where tables span page breaks. Unlike a listing which has a fixed number of lines, SAS cannot predict where Word will create a page break so there are some restrictions on what can be done compared with creating a Word document manually.
- We had a lot of experience in creating listings from SAS where the emphasis had typically been on accuracy. We needed to retain this but also try and think more creatively about how we could present information in a more natural way. We made some progress with this creating sentences like

“This subject is a **43-year-old Black female**, who was diagnosed with **hepatocellular carcinoma** on the **10th** of **January 2009**. “

interjecting the words in bold into the fixed text and then, in this case, changing references from “he “ to “she” later in the document. However, we did not find this “creativity” came easily to people trained as programmers and

although SAS Proc Report is an adequate tool, it is somewhat cumbersome and not particularly user-friendly.

- We took an approach of splitting the code to achieve the Narratives into two. . One set of programs was created to read the data sources and create the “data store”, and this code was specific to the study. The other set was designed to be as study independent as possible and allow formatting to be changed and enhanced without affecting the first set of programs; this second set is the part of the system where we had to find the “creativity”. As it turned out, splitting the code in this modular way worked well and saved us significant time when we subsequently created Narratives for a second study, and we expect this to also be the case on a third study on which we are currently working.

There is still plenty of room to develop the solution further; in particular making it even easier to re-use across different studies. As a CRO we have to deal with a wide variety of data structures and so having a single data store which holds all the necessary variables from the SAS (and non-SAS) sources was a key requirement for our approach.. Also this architecture makes it easier to handle change requests for the layout or content of the reports, and even create additional reports within the Narratives, because we can make such changes without touching the main body of the code.