Paper SM07

Automate Validation of Statistical Reports

Neel Patel, Novartis, Prague, Czech Republic
Deepak Kumar, Novartis, Hyderabad, India

ABSTRACT

Validation of clinical reports are sometimes tedious job considering multiple factors such as titles, footnotes, counts, percentages, decimal positions etc. This becomes more cumbersome when we have page long reports with different categories and sub-groups. Comparing values and percentages manually for all categories may go for a toss in this situation. How our daily used tools and technology can help us potentially to reduce efforts and time while we can gain better quality.

In this paper, we will demonstrate how we could use SAS effectively for our daily routine comparison. How a validation programmer could use different SAS functionality and process to read final reports (plain text .rtf format files) and remove unwanted special characters from the same. We would demonstrate how we could use these features of SAS along with multiple different functions to validate reports flawless and effective way.

INTRODUCTION

Reporting outputs (tables, listings, figures) for a mid-size clinical trial can be in range of 200-350. Statistical programmers and statistician generally produce two type of outputs, safety and efficacy. Various regulators require that we assure the integrity of those reports. To achieve optimum quality there are various processes to validate those outputs.

Most common validation method is to produce the same output independently by another programmer based on the same specification and then comparing it with original outputs. If they match, the said output is then marked as validated. Though such reports are in native RTF reports we cannot use most common PROC COMPARE procedure to validate the counts, layout, footnotes, headers and various other information.

Safety outputs for most of the studies are identical and we can use automation techniques to validate those outputs. We developed a SAS program that can perform comparison automatically and can produce the same result with the same completeness. It works in a three-step method.

1. Reading RTF outputs files (i.e. original output) and converting them to a SAS dataset.
2. Automatically generating independent outputs (i.e. validation output) based on provided parameters.
3. Comparison of original and validation outputs and generating comparison report.

READING RTF OUTPUTS FILES

RTF is a plain (rich) text format and if you go through the technical details, data saved in RTF document is not readable to normal human eyes like you and me. Below is an example of how the word “PhUSE” is saved in RTF document, it is like an original word supported by many keywords, symbols, parentheses and all those are arranged in a specific pattern.

```
{\rtf1\ansi\deff0{\fonttbl {\f1 Arial;}}PhUSE}
```

Obviously this is not human readable, however RTF follows general set of syntax and by leveraging those rules and syntax we can extract the useful information from the document (and from clinical reports as well!) We identified those keywords (including but not limited to \sect, \paperw, \pard, \colortbl, \page, \par) and then used those to read the actual RTF reports and mainly safety tables. Results were encouraging.

ORIGINAL OUTPUT

Primary system organ class/ Preferred term	XXXX N=102 n (%)	YYYY N=102 n (%)
Blood and lymphatic system disorders		
Total	2 (2.0)	3 (2.9)
Anaemia	1 (1.0)	0
Antiphospholipid syndrome	1 (1.0)	0
Iron deficiency anaemia	0	3 (2.9)

EXTRACTED DATA

After some post-processing, making some changes and removing unimportant data (e.g. footnotes, white spaces) we get the following data for each record in RTF document. Next step is to find appropriate delimiter and separate those counts in two separate columns, representing each arm (we used two-arm study for in this case).

Hypothyroidism	14(13.7)9(8.8)
Goitre	3(2.9)1(1.0)
Autoimmune thyroiditis	2(2.0)2(2.0)
Hyperthyroidism	1(1.0)1(1.0)
Thyroid mass	1(1.0)1(1.0)
Oestrogen deficiency	1(1.0)0

COMPARISION WITH INDEPENDENT OUTPUT

Once we have original output ready, the next step is to generate QC output and compare the original output with it. However based on study design we might have multiple arms and that is one of the parameter in our macro, so once you pass required parameters (sorting variables, table design like SOC by PT etc.), macro will automatically generate comparison outputs.

MAKING SINGLE PACKAGE

At last we included all programs in a single macro, so all we need to select is output number, structure, sorting order, subject level dataset name and underlying dataset (AE, CM, MH) name. After finalizing all parameters, the macro will take care of all steps like reading the RTF file, preparing QC outputs and then comparing it with the original RTF output.

CONCLUSION

Our current macro and setup is a part of the validation process but cannot replace the manually or visual validation step. However, this can help in reduction of time required for the validation. This is not a replacement to the traditional validation process. Programmers and Statistician still need to check the rest of the parameters like titles, footnotes, alignments and log of the original program.

On a final note, I would like to say that the scope and future possibilities of this approach are endless. We can always update it and make it more productive and customizable based on the internal environment of an organization.

ACKNOWLEDGMENTS

We thank Anik Chatterjee (Team Lead, Novartis) for original ideation and assistance. His constant guidance helped through ideation to final paper and presentation.

RECOMMENDED READING

Book “RTF Pocket Guide” (ISBN: 9781449302047) by Sean M. Burke is worth reading, this small book explains the syntax of RTF with examples in easy language.

CONTACT INFORMATION

Your comments and questions are valued and highly encouraged. We can send SAS code upon request with valid reason. Contact details of author are available as follows.

Author Name:	Neel Patel	Co-author:	Deepak Kumar
Company:	Novartis	Company:	Novartis
Work Phone:	+420 774 681 718	Phone:	+91 9811252977
Email:	neel.patel@yahoo.com	Email:	deepak.kumar@novartis.com
Web:	www.omcool.com		