PharmaSUG China 2024 - Paper CC-10039

A SAS Macro to Cross Check Abbreviations/scripts between TLFs Body Content and Footnotes

Ke Xu, Dizal Pharmaceutical

ABSTRACT

Footnotes play a crucial role in tables, listings, and figures (TLFs), providing essential details such as the annotation of abbreviations and additional explanations denoted by script indicators. These annotations are vital for reviewers to comprehend the TLFs and implementation of analysis. Since these abbreviations/scripts are usually data-driven, inconsistencies between the main body of the TLFs and the footnotes can frequently arise if statistical programmers forget to update footnotes. Sometimes, abbreviations/scripts are in the body of TLFs, but there are no related annotations in footnotes, or vice versa. Manual checks, which rely solely on the human eye, are time-consuming and prone to miss some abbreviations/scripts. To address this, this paper introduces a SAS macro to identify inconsistencies of abbreviations/scripts between the body of TLFs and footnotes. The data sources for this macro are RTF files for TLFs, along with the metadata of figures. This macro can generate an EXCEL report that details the inconsistencies. It has been demonstrated that this tool will enhance both the efficiency and accuracy of the review process of checking abbreviations/scripts between TLFs body content and footnotes.

INTRODUCTION

In addition to the title and body, footnotes also play a crucial role in tables, listings, and figures (TLFs). Footnotes can provide essential details such as the annotation of abbreviations and additional explanations denoted by script indicators, which is helpful for reviewers to understand the TLFs accurately. Statistical programmers are responsible for the production of TLFs with accuracy, consistency, and efficiency.

Since these abbreviations/scripts are usually data-driven, inconsistencies between the main body of the TLFs and the footnotes can frequently arise if statistical programmers forget to update footnotes. Sometimes, abbreviations/scripts are in the body of TLFs, but there are no related annotations in footnotes, or vice versa. These inconsistencies may lead to confusion for reviewers. Although manual checking which relies solely on the human eye can reduce inconsistencies to some extent, it is time-consuming and inaccurate.

This paper introduces a SAS macro to identify inconsistencies of abbreviations/scripts between the body of TLFs and footnotes. The data sources for this macro are RTF files for TLFs, along with the metadata of figures. This macro can generate an EXCEL report that details the inconsistencies.

STEP 1: RETRIEVE ALL RTF FILES IN FOLDERS AND EXPORT TO SAS DATA SETS

Firstly, it is needed to retrieve the rtf file names in related folders. A macro called READFILE is used to obtain all rtf file names automatically and return a SAS data set FILEINFO to store them.

```
%readfile(phy_path=/* (required) Path where input rtf file is placed*/
          );
```

FILEINFO

	PATH	FILENAME	DATASETNAME	LANGUAGE
1	/data/dizaldwh/ar-dev/test/test/test/reports/tables	t-dm-saf	T_DM_SAF	EN
2	/data/dizaldwh/ar-dev/test/test/test/reports/tables	t-mh-tumor-saf	T_MH_TUMOR_SAF	EN
3	/data/dizaldwh/ar-dev/test/test/test/reports/tables	t-dor-eva	T_DOR_EVA	EN

Display 1. Screen Print of FILEINFO SAS Data Set to Store rtf File Names

Then, a macro called MX_READRTF is designed to read the rtf file, including title, body, and footnotes, and save these contents into a SAS data set. The title, body, and footnotes are divided into different categories marked by the variable of TYPE.

```
    %mx_readrtf(infile_path=/*(required) Path where input rtf file    is placed*/
               ,infile_name=/*(required) Input rtf file name*/
               ,datasetname=/*(optional) output SAS Data Set name*/
               ,lib=/*(optional) Libname where output SAS Data Set is placed, default
as not assigned, work library*/
               ,language=
               ,debug=/*(optional) debug (keep temporary Data Sets?) Y/N, default is
N*/

               );
```

Lastly, based on the FILEINFO data set, call execute macro MX_READRTF to read all rtf files into SAS data sets and set them together.

```
data _null_;
    set FILEINFO;
          call execute('%mx_readrtf(language=' || strip(language)          ||
          ',infile_path=' || strip(path) || ', infile_name=' ||
          strip(filename)||', datasetname='  || strip(datasetname) || ');');
run;
```

T_DOR_EVA

	TYPE	COL1	COL2	COL3	DATASETNAME
1	1	Table 14.2.2.1 Summary of Duration of Response Evaluable for Response Set			T_DOR_EVA
2	2		25 mg (N = 100)	50 mg (N = 100)	T_DOR_EVA
3	3	Follow-up Time (months)			T_DOR_EVA
4	3	Median (95% CI) ^{super a}	14.6 (NE, NE)	17.8 (NE, NE)	T_DOR_EVA
5	4	Please refer to last page for complete footnotes.			T_DOR_EVA
6	4	Footnotes			T_DOR_EVA
7	4	Source: ADTTE			T_DOR_EVA
8	4	Data extraction: 2024-07-16			T_DOR_EVA
9	4	N: Number of participants in the analysis set for each treatment group; CI: Confidence Interval; NE: Not Estimable.			T_DOR_EVA
10	4	The percentage was calculated based on number of responders as the denominator.			T_DOR_EVA
11	4	^{super a} Estimated by Kaplan-Meier method.			T_DOR_EVA

Display 2. Screen Print of the SAS Data Set which Is Exported by Macro MX_READRTF

STEP 2: RETRIEVE ABBREVIATIONS/SCRIPTS IN TLFS

An abbreviation is defined as two or more consecutive uppercase letters. A macro GET_ABB is developed to extract abbreviations from the SAS data set which contains the context of title, body, and footnotes.

```
    %get_abb(data=/*(required) input Data Set name*/
            ,out=/*(required) output Data Set name*/
            ,pattern= %nrbquote(/\b[A-Z]{2,}\b/)=/*(required) pattern*/
            ,var=/*(required) the variable name which needs to be extract*/
            ,language=
            ,source=/*(required) where are abbreviations from*/
            );
```

DATASETNAME	ABB	SOURCE
T_DOR_EVA	CI	Abbreviations in body
T_DOR_EVA	N	Abbreviations in body
T_DOR_EVA	NE	Abbreviations in body

DATASETNAME	ABB	SOURCE
T_DOR_EVA	CI	Abbreviations in footer1
T_DOR_EVA	N	Abbreviations in footer1
T_DOR_EVA	NE	Abbreviations in footer1

Display 3. Screen Print of the Abbreviations from the Body/title and Footnotes

The scripts will appear inside the brace, so the macro GET_SCRIPT is used to get contents in the brace from the SAS data set that contains the context of TLFs. As contents in the brace are not always scripts, just need to focus on those that start with 'Super'.

```
%get_script(data=/*(required) input Data Set name*/
            ,out=/*(required) output Data Set name*/
            ,var=/*(required) the variable name which needs to be extract*/
             ,source=/*(required) where are scripts from*/
           );
```

DATASETNAME	ABB	SOURCE
T_DOR_EVA	{Super a}	Scripts in body
T_DOR_EVA	{Super b}	Scripts in body
T_DOR_EVA	{Super c}	Scripts in body

DATASETNAME	ABB	SOURCE
T_DOR_EVA	{Super a}	Scripts in footnotex
T_DOR_EVA	{Super b}	Scripts in footnotex
T_DOR_EVA	{Super c}	Scripts in footnotex

Display 4. Screen Print of the Scripts from the Body/title and Footnotes

STEP 3: CROSS CHECK ABBREVIATIONS/SCRIPTS BETWEEN TLFS BODY CONTENT AND FOOTNOTES

After extracting the abbreviations/scripts from the body/title and footnotes respectively through above steps, it will be easy to conduct the cross check of inconsistencies. Finally, a new macro U_M_GETABB_TABLE which encapsulates the above macros is designed to generate an EXCEL report that details the inconsistencies. When the EXCEL reports show up with some abbreviations/scripts that users don't want to check, users can fill 'Y' in the E column. These abbreviations/scripts filled with 'Y' will not be checked again for the specific TLFs in future checking. The parameters of NOTCHECK_TABLE and NOTCHECK_FOOTER can be used to control whether to check some abbreviations/scripts at all TLFs level.

```
    %u_m_getabb_table(notcheck_table=/*(optional) the abbreviations no need to check
in table, Eg. "COVID" "ID" "MUGA" "LUGANO" "PET" "CT"*/
                      ,notcheck_footer=/*(optional) the abbreviations no need to check
in footer, Eg. "Min" "Max" "MedDRA"*/
                      ,language=&language.
                       ,debug=/*(optional) debug (keep temperary Data Sets?) Y/N,
default is N*/
                     );
```

	A	B	C	D	E
	RTF Files Name	Source	Abbreviations_Scripts	Message	Not Check
	T_CM_PACT_SAF	Abbreviations in body	ADC	Abbreviations/Scripts not in footer: ADC	
	T_EX_INT_SAF	Abbreviations in body	SD	Abbreviations/Scripts not in footer: SD	

Display 5. Screen Print of EXCEL Report which Details the Inconsistencies between TLFs Body and Footnotes

CONCLUSION

This paper introduces a SAS macro to identify inconsistencies of abbreviations/scripts between the body of TLFs and footnotes. The data sources for this macro are RTF files for TLFs, along with the metadata of figures. This macro can generate an EXCEL report that details the inconsistencies. Because the abbreviation is defined as two or more consecutive uppercase letters, there will be false positives (Capitalized word, E.g., ORAL; a mixed case abbreviation, E.g., QTcF). These false positives could be avoided by macro parameters and Not Check column of EXCEL report. It has been demonstrated that this tool will enhance both the efficiency and accuracy of the review process of checking abbreviations/scripts between TLFs body content and footnotes.

ACKNOWLEDGMENTS

The author would like to thank Weibin Cai for his valuable input into this paper and Jundong Ma for his helpful macro MX_READRTF.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Ke Xu

Dizal Pharmaceutical

Ken.Xu@dizalpharma.com