A SAS Macro to Perform Consistency Check in CSR Footnote References

Jeff Xia, Merck & Co., Inc., Rahway, NJ, USA

Chandana Sudini, Merck & Co., Inc., Rahway, NJ, USA

ABSTRACT

Footnote is an important part of tables, figures, and listings (TLFs) for CSR, which might include but not limited to abundant information such as abbreviations, acronyms, additional explanations that have values to be added as a part of the TLFs. Footnote is normally provided in a sequential order such as in the format of Roman sequence, or alphabetically, etc. It is often to see inconsistency of footnote reference between the body of TLFs and the footnote section, i.e., a reference number appear in the TLF body, but there is no corresponding reference number in the footnote section, or vice versa. Catching these inconsistencies by eye-browsing is an attention demanding and error prone task to perform.

This paper introduces a SAS macro that compares the list of footnote reference number between the body of TLFs and the ones in the footnote section, and flag any discrepancies of footnote reference for each TLF. In addition, the macro generates a report to list the name of each TLF and details of the discrepancies. It is suggested to run this macro after TLFs produced as part of the dry run package, and before delivering the final TLFs to Clinical for CSR.

INTRODUCTION

TLFs for CSR consist of three main sections: title, body and footnote. Statistical programming team must produce TLFs with accuracy and consistency. Different methods can be utilized to ensure that the level of required accuracy and consistency in various parts of TLFs. , This includes but are not limited to various stakeholders' review and assorted programming tools that extract the text from TLFs, and compare them against the related datasets or CSR table shells. As a small portion of a big task, this paper focuses on how to maintain the expected consistency between the body and the footnote sections of a CSR.

Typically, the footnote comes from the mockup table shell for CSR, which aims to provide additional information about certain elements of the TLF. A footnote is marked within the body by a superscript icon, usually with a special character, number, alphabet, etc., (i.e., asterisk (*) or number (^{1}), alphabet (a)) which corresponds to the matching footnote at the bottom of the page. These footnotes are presented in a sequential order (i.e., a, b, c or I, II, III) for ease of tracking and referencing as needed. Unfortunately, there might be a disconnect between the references in the footnote section and the body of the TLF, which can result in a situation where there is a footnote with no referral in the TLF or vice versa.Inconsistencies in footnote references and the content of the TLF can cause confusion for the end user.

This paper presents a SAS macro that aims to rectify this issue and reduce the manual workload of reviewing and correcting these errors for increased efficiency and accuracy. The macro extracts the reference numbers from the CSR's body section and programmatically compares them to the reference numbers in the footnote section. Upon completion, a report is generated with details on all discrepancies between the body and the footnote so that the discrepancies can be resolved and corrected TLFs can be produced.

STEP 1: RETRIEVE LIST OF FILE NAMES IN FOLDERS

The first task is to obtain the file names in a folder including all the subfolders. A macro `listfiles` was designed to automatically retrieve the files listed in a folder, then a recursive call was used within the

macro to retrieve the files listed in the subfolders. See similar code snippet in the reference (Jeff Xia, 2022).

The next task is to read in the text contents of each rtf file, and save the contents into a SAS dataset. See below for an example of the CSR table and corresponding RTF syntax. Please note that the study number, database cutoff date, and all efficacy results are all made up numbers purely for illustration purpose.

Summary of Response Outcome in Participants with Confirmed Response Based on Central Radiology Assessment per RECIST 1.1
(Responders)

	Study Treatment (N=51)
Number of Participants with Response[a]	8
Participants Who Progressed or Died[b] (%)	**3 (37.5)**
Censored Participants (%)	**5 (62.5)**
Participants who missed 2 or more consecutive disease assessments	0 (0.0)
Participants who started new anti-cancer treatment	0 (0.0)
Participants who were lost to follow-up	1 (12.5)
Participants whose last adequate assessment was ≥ 5 months prior to data cutoff date	0 (0.0)
Ongoing response	4 (50.0)
Range of DOR (months)	12.5+ to 24.0+

[a] Includes participants with a confirmed complete response or partial response.

[b] Includes participants who progressed or died without previously missing 2 or more consecutive disease assessments.

[c] Includes participants who are alive, have not progressed, have not initiated new anti-cancer treatment, did not miss 2 or more consecutive disease assessments, are not lost to follow-up and whose last disease assessment was <5 months prior to data cutoff date.

For censored participants who met multiple criteria for censoring and do not have ongoing response, participants are included in the censoring criterion that occurred earliest.

'+' indicates there was no progressive disease by the time of last disease assessment.

Database Cutoff Date: 01JAN1999.

Source: [P001V01MK9999: adam-adsl; adtte]

Display 1. Screen Print of a Sample Table for CSR

```
\intbl\ql\sb30\sa30 \f0\fs20\fi-160\li160\ri1  {\super a} Includes participants with a confirmed complete response or partial response.
\par\fi-160\li160\ri1  {\super b} Includes participants who progressed or died without previously missing 2 or more consecutive disease assessments.
\par\fi-160\li160\ri1  {\super c} Includes participants who are alive, have not progressed, have not initiated new anti-cancer treatment, did not miss
\par\fi-160\li160\ri1  For censored participants who met multiple criteria for censoring and do not have ongoing response, participants are included ir
\par\fi-160\li160\ri1  '+' indicates there was no progressive disease by the time of last disease assessment.
\par\fi-160\li160\ri1  Database Cutoff Date: 19JAN2021.
\cell\intbl\row\pard
\trowd\trgaph36\trleft0\trqc
\cellx8640\pard
\pard
\trowd\trleft0\trqc\pard\qc \f0\fs20 Source:  [P224V03MK3475: adam-adsl; adtte]
\par }
```

Display 2. Screen Print of RTF Syntax of the Footnote Section for the Table in Display 1

The reference number in the body or the footnote section can be identified by the RTF tag "{\super". Now the task is to identify how to separate the reference numbers in the body of TLF and the ones in the footnote section. After further investigation it appears that the last occurrence of RTF Syntax of {\intbl" will mark the beginning of the footnote section.

How to find the last appearance of RTF syntax of "{\intbl" in the entire text file? First, the macro assigns a line number for each line of text when reading in the RTF file into SAS dataset (i.e., variable "linenumber" in the SAS code below), and then reverse the order of the text by sorting the dataset in a descending order of the variable "linenumber". Programmatically at this point it will be easier to find the first appearance of "{\intbl".

```sas
data file&i;
    filename inf  "&&file&i" ;
    infile inf truncover;;
    format  line $char2000.;
    input line 1-2000;
    linenumber = _n_ ;
run;

proc sort data = file&i;
    by descending linenumber;
run;
```

Display 3. Screen Print of SAS Code Snippet to Read in RTF text file and Reverse the Order

Since all lines of text in the dataset are in reverse order now, the following code snippet marks each line of text as footnote section from the very beginning, until it hits another {\intbl", which means the beginning of the RTF syntax belongs to the body of the RTF file.

```sas
data file&i;
    set file&i;
    length rtfname $100 location $2000 superscript $100;
    retain myfoot ;
    rtfname="&&rtf&i";
    location = "&&file&i";
    if _n_ = 1 then do;
        foot = 1;
        myfoot = foot;
    end;
    else do;
        foot = myfoot;
        if index(compress(line), "\intbl\") = 1 then do;
            myfoot = 0;
        end;
    end;
```

Display 4. Screen Print of SAS Code Snippet to Identify the Footnote Section

The footnote section always contains distinct reference numbers, there should be no duplicates at all. However, it is not always true in the body of the TLFs. It does happen that multiple references appear in a single line of text in the body. The following code snippet is designed to find all possible appearances of "{super" by using a do loop.

```
line1 = strip(line);
pos1 =  index(line1, "{\super") ;
if pos1 > 0 then do;
    pos2 = index(line1, "}");
end;

do while  (pos2 > pos1 >0) ;
    superscript = substr(line1, pos1, pos2-pos1);
    output;
    line1 = substr(line1, pos2+1);
    pos1 = 0;
    pos2 = 0;
    pos1 =  index(line1, "{\super") ;
    if pos1 > 0 then do;
        pos2 = index(line1, "}");
    end;
end;
line2 = strip(line);
```

Display 5. Screen Print of SAS Code Snippet to Identify Multiple Appearances of Reference

STEP 2: PERFORM THE COMPARSION

All lines of text from the body are saved in the dataset "readsource1". On the other hand, all lines of text from the footnote section are saved in the dataset "readsource2". Then merge the dataset "readsource1" and "readsource2" with key variables as RTF file name, location and superscript (i.e., "{\super a", "{\super b", "{\super c" etc.). The variable "message" is created to record 3 possible outcomes described below.

1. Match

2. Script in body but not in footnote

3. Script in footnote but not in body

Ideally, the outcome for all the RTF files should be "Match" for CSR package. However, if the outcome is either 2 or 3, then the reviewer should further investigate and update the TLF programs to ensure the consistency of reference numbers between the body and the footnote section.

```
proc sort data = readsource1(keep = rtfname location superscript) nodupkey;
    by rtfname location superscript;
run;

proc sort data = readsource2(keep = rtfname location superscript) nodupkey;
    by rtfname location superscript;
run;

data readsource3;
    merge readsource1(in = a)
          readsource2(in = b);
    by rtfname location superscript;
    length message $100;
    if a and not b then message = 'Script in body but not in footnote';
    else if a and not b then message = 'Script in footnote but not in body';
    else message = 'Match';
run;
```

Display 6. Screen Print of SAS Code Snippet to Comparison of References between Body and Footnote

The outcome can be generated in any format per reviewers' preference, such as RTF, PDF, or HTML. And the row with issues can be highlighted in a certain way to draw reviewers' attention. See below for a sample output of the macro:

	A	B	C	D
1	**RTF Name**	**RTF Location**	**Super Script**	**Message**
14	e0apr0t2rdor0inv.rtf	c:\SD-103\test\superscript\e0apr0t2rdor0inv.rtf	{\super a	Match
15	e0apr0t2rdor0inv.rtf	c:\SD-103\test\superscript\e0apr0t2rdor0inv.rtf	{\super b	Match
16	e0apr0t2rdor0inv.rtf	c:\SD-103\test\superscript\e0apr0t2rdor0inv.rtf	{\super c	Script in body but not in footnote

Display 7. Screen Print of Sample Output of the Presented Macro

CONCLUSION

This paper presents a macro that can be used to programmatically separate the references between the body and the footnote section in TLF for CSR, perform a comparison, and produce a report to display any discrepancies. This macro is particularly useful for the statistical programming team as they check for consistency in the TLFs for CSR.

REFERENCES

A SAS Macro to Aid Assembly Review of M5 Module in Submission. Jeff Xia, 2022. PharmaSUG proceedings.

ACKNOWLEDGMENTS

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact authors at:

Name: Jeff Xia
Enterprise: Merck & Co., Inc.
Address: 126 East Lincoln Avenue,
City, State ZIP: Rahway, NJ 07065-4607
Work Phone: 732-594-6439
E-mail: jeff.xia@merck.com
Web: www.merck.com

Name: Chandana Sudini
Enterprise: Merck & Co., Inc.
Address: 126 East Lincoln Avenue,
City, State ZIP: Rahway, NJ 07065-4607
Work Phone: 732-594-6860
E-mail: chandana.sudini@merck.com
Web: www.merck.com