

## Large batch processing of TLF pagination auto-inspection, 10 secs is enough

Michael Yang, PRA Health Sciences;  
Vera Zhu, PRA Health Sciences

### ABSTRACT

During programming work, we need to check TLF pagination issue(s). Some checking tools mentioned in the historical literature usually call Word software through VBA to read RTF page number and other information in order to judge whether TFL is paged normally. However, it is precisely because of the introduction of Word software, when the total number of TFLs and pages for inspection is too large, the speed of batch processing quickly becomes very slow. In this paper, we propose a new approach that enables us to significantly improve the checking speed. It opens the RTF file by OpenTextFile instead of Word, and uses regular expression to capture the key information. It is concluded that checking hundreds of TLFs(more than 10,000 pages in total) can be completed correctly in ten seconds. Besides, this tool can also find out whether the RTF files have been manually modified or not.

### INTRODUCTION

In the daily work of generating tables and listings, we often encounter RTF pagination issue. There are mainly 5 types of common pagination issues:

1. The page number duplicates in the footnote (see Figure 1).
2. Some page numbers are missing in the footnote (see Figure 2).
3. Total page number in the footnote calculated incorrectly (see Figure 3).
4. There is extra page with info that does not belong to contents of table. (see Figure 4).
5. Page break occurs within a single logical page (see Figure 5).

Page 25 of 27

Table 14.1.10 Concomitant Medications by ATC Classification Level 1 and WHO Drug Dictionary Preferred Term  
(Safety Analysis Set)

Anatomical Therapeutic Chemical Classification (Level 1) Preferred Term [n (%)]	Treatment A (N=137)	Treatment B (N=134)
SENSORY ORGANS		
KETOROLAC	1 (0.7)	0
PREDNISOLONE ACETATE	1 (0.7)	0

Source: Dataset: ADCM, ADSL, Program: t\_cm.sas, Output: t-14-01-10-cm.rtf, Generated on: 01JAN2019 00:00

Page 25 of 26

Page 26 of 27

Table 14.1.10 Concomitant Medications by ATC Classification Level 1 and WHO Drug Dictionary Preferred Term  
(Safety Analysis Set)

Anatomical Therapeutic Chemical Classification (Level 1) Preferred Term [n (%)]	Treatment A (N=137)	Treatment B (N=134)
.Uncoded	5 (3.6)	6 (4.5)

Source: Dataset: ADCM, ADSL, Program: t\_cm.sas, Output: t-14-01-10-cm.rtf, Generated on: 01JAN2019 00:00

Page 25 of 26

Figure 1. Duplicate Page Number

Table 14.1.2.9 Demographic Characteristics  
(Modified Intent-to-treat Analysis Set)

Characteristic	TRT A (N=70)	TRT B (N=99)
Age (years)		
n	70	99
Mean (std)	62.6 (7.74)	63.0 (8.91)
Median	64.0	64.0
Q1, Q3	56.0, 68.0	57.0, 69.0
Min, Max	41, 77	36, 80

Source: Dataset: ADSL, Program: t\_demo.sas, Output: t-14-01-02-09-demo.rtf, Generated on: 01JAN2019 00:00

Page 2 of 5

Page 3 of 4

Table 14.1.2.9 Demographic Characteristics  
(Modified Intent-to-treat Analysis Set)

Characteristic	TRT A (N=70)	TRT B (N=99)
Age Group [n (%)]		
< 65 years	36 (51.4)	51 (51.5)
≥ 65 years	34 (48.6)	48 (48.5)

Source: Dataset: ADSL, Program: t\_demo.sas, Output: t-14-01-02-09-demo.rtf, Generated on: 01JAN2019 00:00

Page 4 of 5

## Figure 2. Missing Page Number in Footnote

Page 1 of 2

Listing 5 Infections  
(Safety Population)

Randomized Treatment: Treatment A		Genus and Species	Superinfection/ New Infection
Subject ID/ Age/Race/Sex	Specimen Collection Date Time (Day)	Visit	
100-0001/50/W/F	2018-01-01T00:00 (7)	Day 7/EOT	XXXXXXXXX Superinfection

Source: Dataset: ADXX, Program: l\_inf.sas, Output: l\_listing5.rtf, Generated on: 2019-01-01T00:00

Page 1 of 1

Page 2 of 2

Listing 5 Infections  
(Safety Population)

Randomized Treatment: Treatment B		Genus and Species	Superinfection/ New Infection
Subject ID/ Age/Race/Sex	Specimen Collection Date Time (Day)	Visit	
100-0002/50/B/M	2018-01-01T00:00 (7)	Day 7/EOT	XXXX New Infection

Source: Dataset: ADXX, Program: l\_inf.sas, Output: l\_listing5.rtf, Generated on: 2019-01-01T00:00

Page 1 of 1

## Figure 3. Wrong Total Page Number in Footnote

---

Listing 16.2.7.4 Treatment-Emergent Adverse Events of Interest - Nausea and Vomiting  
(Safety Population)

---

Treatment A		AE caused study discont ?
Subject ID/ Age/Race/Sex	MedDRA SOC/ Preferred Term/ Verbatim Term	
100-0001/50/W/F	Gastrointestinal disorders/ Nausea/ NAUSEA	N

---

Source: Dataset: ADAE, Program: l\_ae.sas, Output: l\_16\_2\_7\_4\_ae.rtf, Generated on: 2019-01-01T00:00

Page 15 of 15

Page 16 of 16

Directory		
Libref	Member	File
Engine	Type	SizeLast Modified
Physical Name E:-TD4280_NA1SASP RDGRID1		
Filename E:-TD4280_NA1SASP RDGRID1		
#Name	Member	File
1ADAE	DATA	52428804/08/2019 08:20:32
2FINAL	DATA	28672004/08/2019 08:20:32
3FINAL_1	DATA	28672004/08/2019 08:20:32

**Figure 4. Extra Page with wrong info in Outputs**

Page 1 of 5

Table 14.1.2.1 Demographic Characteristics  
(Full Analysis Set)

Characteristic	Treatment A (N=137)	Treatment B (N=134)
Sex [n (%)]		
Female	53 (38.7)	35 (26.1)
Male	84 (61.3)	99 (73.9)
Race [n (%)]		
White	129 (94.2)	133 (99.3)
Black or African American	1 (0.7)	0
Asian	3 (2.2)	1 (0.7)
American Indian or Alaska Native	1 (0.7)	0
Native Hawaiian or other Pacific Islander	0	0
Other	3 (2.2)	0

Page 2 of 5

Table 14.1.2.1 Demographic Characteristics  
(Full Analysis Set)

Characteristic	Treatment A (N=137)	Treatment B (N=134)
Age (years)		
n	137	134
Mean (std)	63.7 (8.32)	62.6 (8.94)
Median	65.0	65.0
Q1, Q3	60.0, 69.0	56.0, 68.0
Min, Max	36, 78	40, 80

Source: Dataset: ADSL, Program: t\_demo.sas, Output: t-14-01-02-01-demo.rtf, Generated on: 01JAN2019 00:00

Page 1 of 4

**Figure 5. Page break occurs within a single logical page**

In addition, tables and listings are preferably direct results from SAS running. In other words, manual modification of the generated tables and listings is not recommended. Because whether manually correcting the pagination issues or modification on data will bring huge potential quality hazards to the whole project.

Regular pagination issue checking tools generally do not involve checking for 1 to 4 types of pagination issues and manual modification. This useful tool introduced in this paper can solve such problems. Even if these issues are checked in large quantities, it usually takes only a few seconds to complete.

## OPERATING PRINCIPAL OF THE TOOL

### DETECT TYPE 1 TO 4 OF THE PAGINATION ISSUES

When using PROCE REPORT to generate tables and listings, programmers typically use a page variable to control page breaking. SAS will generate a new "\sectd" (means "section default") keyword in the RTF document based on the variable value. Therefore, the total number of pages (physical page number) of RTF can be obtained by counting the total number of "\sectd" in RTF documents.

Because the footnote often contain the total number of pages calculated by SAS code, the page number of tables or listings (logical page number) can be obtained from the "Y" value of "page X of Y" in the footnote, by regular expression.

In general, the type 1 to 4 of pagination issues will result in unequal total physical page number and logical page number. By detecting the difference of the two page numbers, we can say that the RTF document has a pagination issue. The logical page number is usually greater than the physical page number in RTF with type 4 pagination issue, and less for the other types.

For the pagination issue type 1, the page number where issue firstly occurs can be obtained by detecting the first equality of two adjacent page numbers (the "X" value in "page X of Y" duplicates here).

### CHECK FOR MANUAL MODIFICATION ISSUE

The original RTF document uses the keyword "\fldinst {NUMPAGES}" to represent the total number of pages on the right side of the header, which will be dynamically parsed into the total page number when opened by software such as Microsoft Word. When the RTF file is modified and saved with Word, the keyword will be replaced by another total page number keyword "\{nofpagesY\}" (Y=1, 2, 3...) generated by Word (see Figure 6).

When there is a string such as "\{nofpagesY\}" in the RTF, it can be judged that the RTF has been artificially changed.

```
\pard\plain\intbl\sb10\sa10\qr\f3\fs20\cf1{Page {\field{\*\fldinst { PAGE }}}}{ of } \field{\*\fldinst { NUMPAGES }}\cell} \row} \trqc Before Manual Modification \cltxlrbclcverticalt\clcbpat8\clpadt10\clpadft3\clpadr10\clpadff3\cellx4260 | \yr2019\mo7\dy27\hr16\min19\version2\edmins0\{ \nofpages2\}\nofwords258 \nofchars1472\{ \nofcharws1727\}\vern107\}\{ \xmInstb1\{ \xmIns1 http://schemas.microsoft.com/office/word/2003/wordml\}\paperw15840\paperh12240\margl1440\margt1440\margb1440\gutter0\ltrsect After Manual Modification
```

Figure 6. Page Number Keyword Changed

### TECHNICAL IMPLEMENTATION OF QUICK CHECK FOR PAGINATION ISSUE AND MANUAL MODIFICATION ISSUE

This tool can automatically check above issues by opening it with OpenTextFile, retrieving relevant keywords with regular expressions, and finally transferring them into useful information. Since this process is an operation on plain text, even a large batch check takes only a few seconds.

### INSPECTION METHOD FOR PAGINATION ISSUE TYPE 5

When there is too much content in a page, pagination issue occurs in Word. At this time, since the RTF keyword "\sectd" cannot reflect whether the content of the page is too much and needs page breaking, the total number of "\sectd" in this case cannot accurately reflect the actual total number of pages.

In this case, we need to get the physical page number of RTF by calling the Word, and judge whether the RTF has page breaking issues by comparing with the logical page number in the footnote (because this method uses Word to parse the page number, the time required will increase).

Finally, the cumulative page number of each page of RTF and the cumulative number of "\sectd" of this page are compared with each other, and the first page where the 2 numbers are unequal is where the pagination issue firstly occurs.

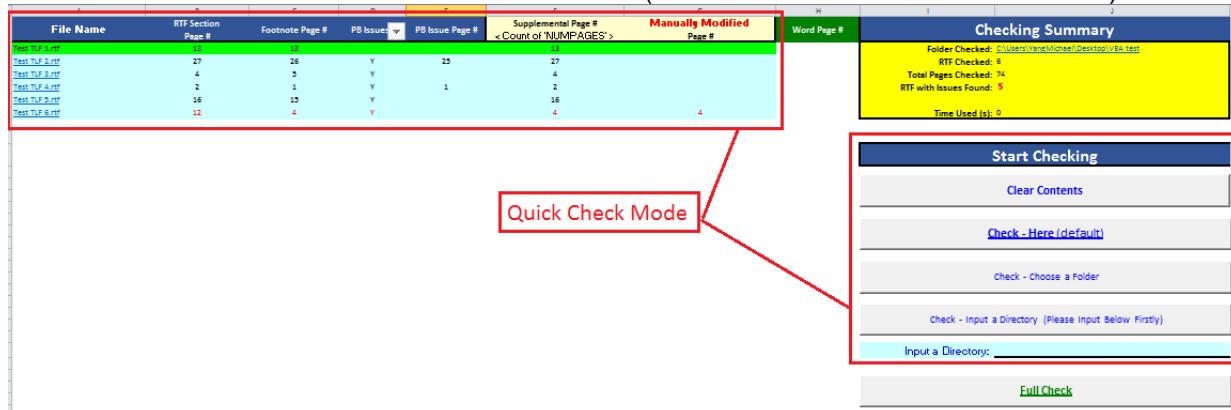
## HOW TO USE THE TOOL

There are two inspection modes:

- Quick check mode - Check the 1 to 4 types of pagination issues and manual modification issues (see Figure 7).

Note:

RTF Section Page # = Total page number by counting of "\sectd"  
 Footnote Page # = Total page number read from the footnote (Y of "page X of Y")  
 PB Issues? = Whether issues found (including pagination and modification issues)  
 PB Issue Page # = Page number where pagination issue firstly occurs  
 Supplemental Page # = Supplemental total page number by counting of 'NUMPAGES'  
 Manually Modified Page # = Total page number of modified RTF (Y of "{\nofpagesY}")  
 Test TLF 2 – The page number 25 duplicates in the footnote  
 Test TLF 3 – The page number 3 is missing in the footnote  
 Test TLF 4 – Total page number 1 in the footnote calculated incorrectly (should be 2)  
 Test TLF 5 – The last page with info that does not belong to contents of table.  
 Test TLF 6 – Modification issue found (column B to G has been marked in red)

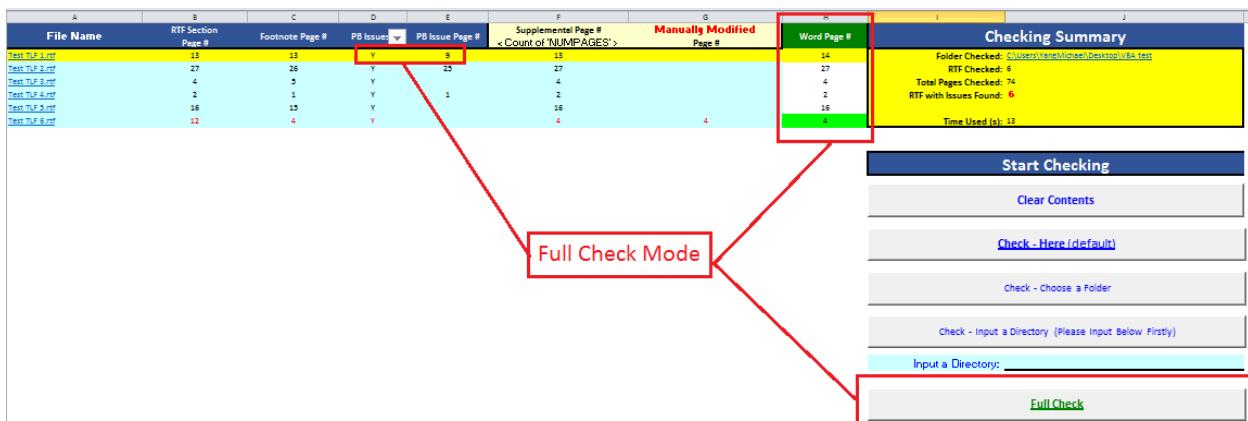


**Figure 7. Quick Check Mode**

- Full check mode - Supplemental check for type 5 pagination issue only, after quick check is completed (see Figure 8). It provides page number from Word in column H, which compared with the number in column B to find whether type 5 pagination issue exists in the RTF, and gives the page number where pagination issue firstly occurs in column E.

Note:

Word Page # = Total page number from Word  
 Test TLF 1 – Page break occurs within a single logical page



**Figure 8. Full Check Mode**

There are three ways to indicate the inspection path in the quick mode:

1. Directly check all RTF files of the current path of the tool (click on the below button);

[Check - Here \(default\)](#)

2. Manually select a path to check (click on the below button);

[Check - Choose a Folder](#)

3. Input a full path name to check (Input a directory on the underline, then click the Check button above it).

[Check - Input a Directory \(Please Input Below Firstly\)](#)

[Input a Directory:](#) \_\_\_\_\_

## CONCLUSION

In this paper an EXCEL VBA tool which can quickly batch check 5 kinds of pagination issues and manual modification issue is introduced. The tool has the following advantages:

1. It can check multiple pagination issues and provide the page number where the pagination issue firstly occur.
2. It can also check whether the contents of tables and listings have been manually modified or not.
3. The checking speed is relatively fast, it's time saving.
4. The tool is embedded in EXCEL file, which is easy to use.

In a word, this tool can simplify the checking of RTF pagination issue and manual modification issue, and make the daily work more efficient.

## REFERENCES

"VBScript Programmer's Reference", Third edition

SAS 9.2 Online product documentation. Available at support.sas.com/documentation/92/index.html

Yong Cao and Xuan Sun.2017."A convenient tool to check the pagination issue in TLF". PharmaSUG China 2017, Paper 46

## **ACKNOWLEDGMENTS**

The contents of this paper are the work of the authors and do not necessarily represent the opinions, recommendations, or practices of PRA.

## **CONTACT INFORMATION**

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

Name: Michael Yang  
Enterprise: PRA Health Sciences  
Phone: +86 139 1669 4663  
E-mail: 806922078@qq.com  
Web: [www.prahs.com](http://www.prahs.com)

Name: Vera Zhu  
Enterprise: PRA Health Sciences  
Phone: +86 18621101268  
E-mail: ZhuVera@prahs.com  
Web: [www.prahs.com](http://www.prahs.com)

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Any brand and product names are trademarks of their respective companies.