

A dynamic-automated approach to reviewing RTF files

Youying Yu, Eli Lilly and Company, Indianapolis, Indiana, USA
Hrideep Antony, Syneos Health, North Carolina, USA

A dynamic-automated approach to reviewing RTF files

- ▶ Youying Yu is a Consultant Statistician-Computation with Eli Lilly. She has Master's degree in Statistics and 20 years experience in pharmaceutical industry with focus on statistical analysis/SAS, clinical trial, and NDA application. She enjoys to apply innovation ideas to improve efficiency during NDA submission process. Youying likes music, singing, painting, and calligraphy at her spare time.
- ▶ Hrideep Antony is a Principal Statistical Programmer with Syneos Health, and has over 14 years of extensive SAS experience. Hrideep enjoys developing and providing innovative/efficacious solutions to streamline the FDA submission process. He has been given opportunities to present at multiple SAS conferences in the past. Hrideep enjoys spirituality, music, travelling, volunteering at nursing homes/rehabilitation facilities, learning and sharing knowledge.

Introduction

- ✓ Did the current data transfer significantly change any of the results?
- ✓ Did any of the p-value's change significantly in those 600 outputs since we re-ran the programs after the latest data transfer?
- ✓ Could you come up with a list of outputs that has a TEAE of dizziness out of those 600 outputs? By the way we need that list before noon's meeting.

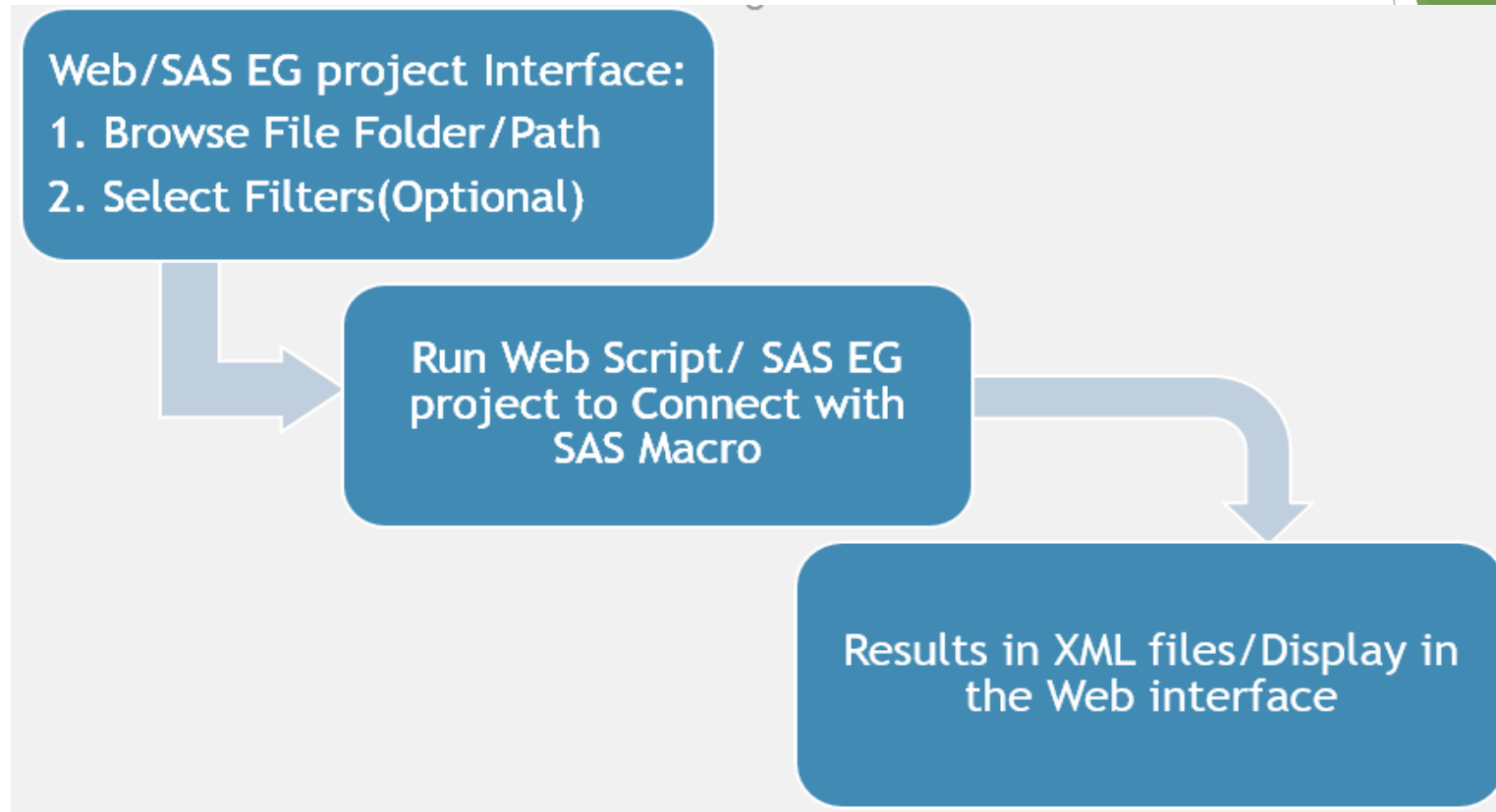


Process Flow

- The HTML and SAS Enterprise Guide interface which acts as an interface between the user and communicates with the SAS macro
- The %reviewfiles SAS utility macro that reads in the RTF files and processes the data.
- The XML output files, which provides the summary of findings along with the corresponding titles and file names for quick access

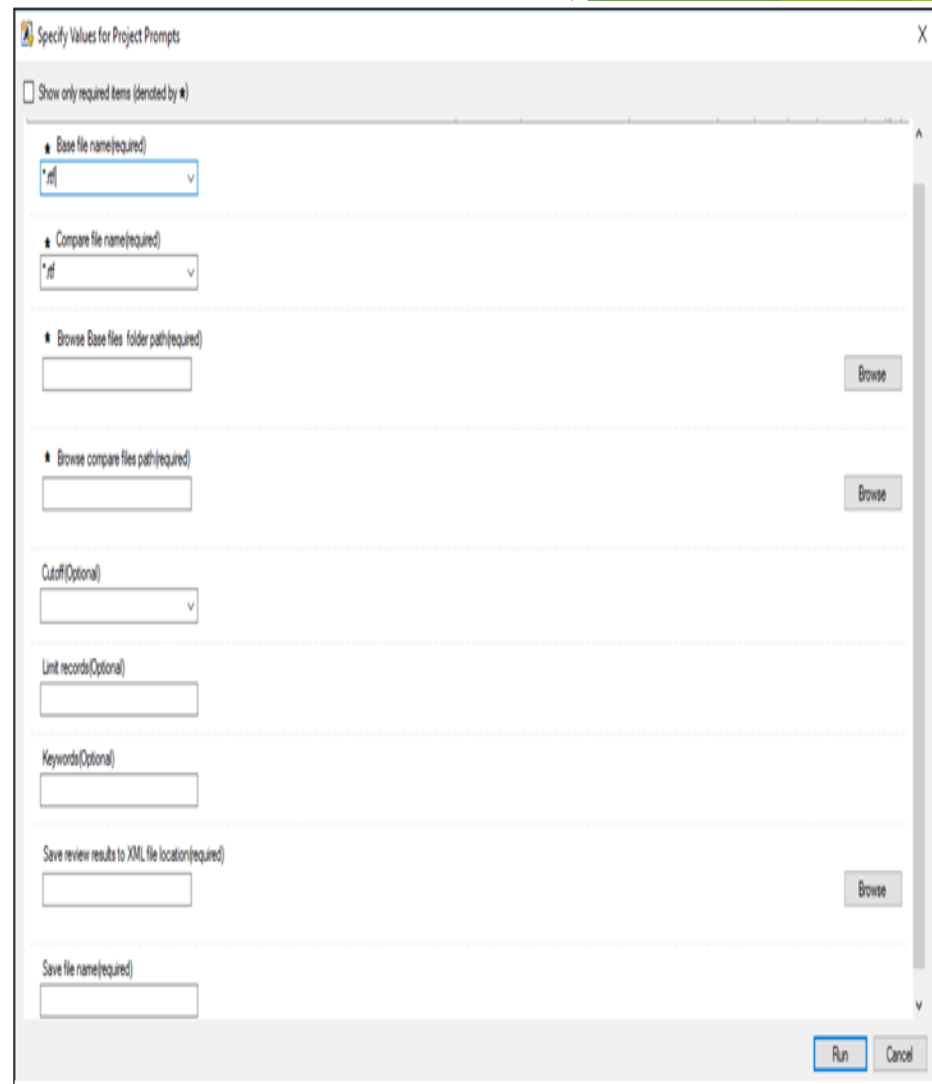


SAS Enterprise Guide Interface



Input parameters

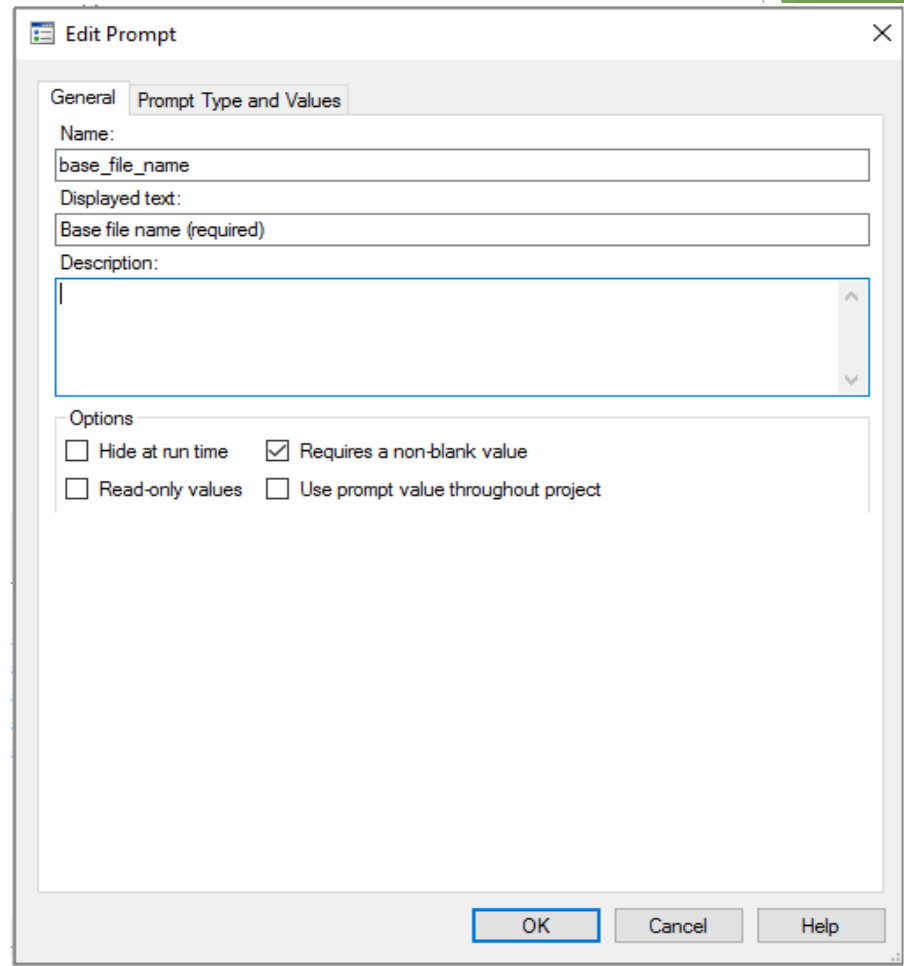
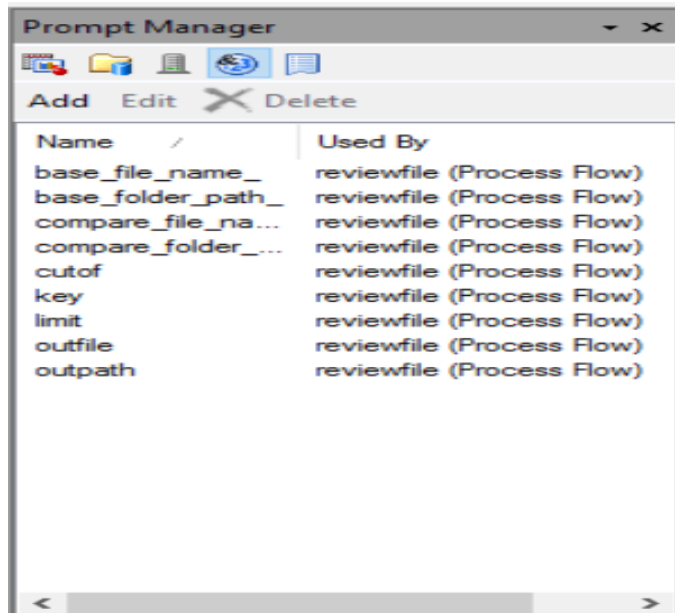
- **Base file name and Compare filename parameters:** used to specify the RTF file name to be reviewed at the location specified in the **folder path parameters**
- **Keywords Parameter** which is optional, allows the user to search the files for a specific keyword such as P-value.
- **Cutoffs parameter(optional):** To fetch the differences beyond a specific cutoff value
- **Limit records parameter(optional):** To limit the number of records of output that are displayed, which is limited to 6 observations in the current example
- **Save review results to folder parameter:** To save the comparison results in an XML file format with the file name specified using the script below.
- **Run button** will initiate the review process



The screenshot shows a dialog box titled "Specify Values for Project Prompts". It contains several input fields and buttons:

- ☐ Show only required items (denoted by *)
- Base file name(required):** A text box containing "*.rtf" with a dropdown arrow.
- Compare file name(required):** A text box containing "*.rtf" with a dropdown arrow.
- Browse Base files (folder path(required):** A text box with a "Browse" button to its right.
- Browse compare files path(required):** A text box with a "Browse" button to its right.
- Cutoff(Optional):** A text box with a dropdown arrow.
- Limit records(Optional):** A text box.
- Keywords(Optional):** A text box.
- Save review results to XML file location(required):** A text box with a "Browse" button to its right.
- Save file name(required):** A text box.
- At the bottom right, there are "Run" and "Cancel" buttons.

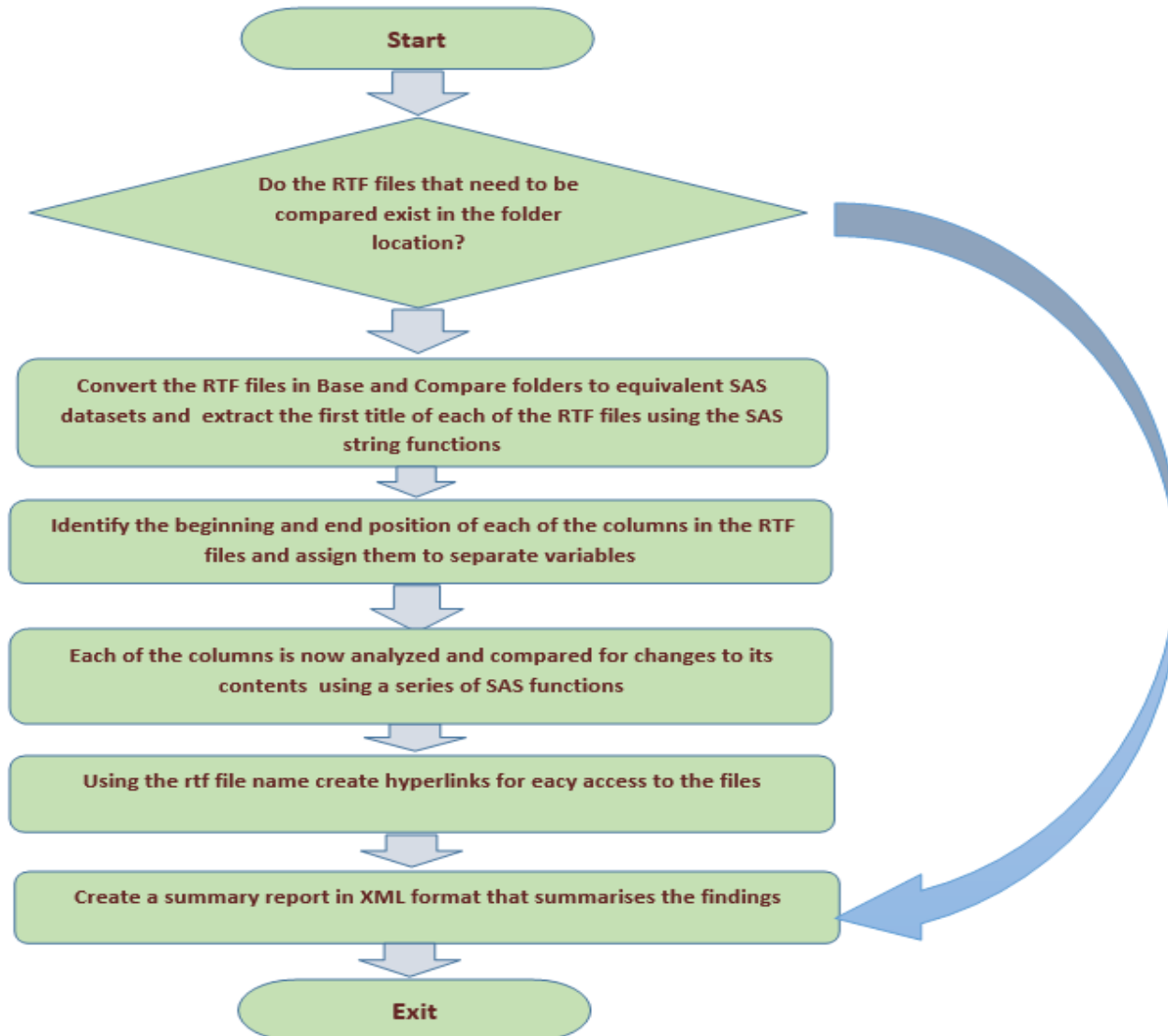
Creating Interface using SAS Enterprise Guide using Prompt Manager



XML output

Title	Base file	Compare file	Findings summary	Findings on key search
Summary and Analysis of Heart Rate Lowering Medication Use During Study Page 1 of 4	fqcmlowhra1	fqcmlowhra1	Acceptable differences	pvalue:exactly same or acceptable range
Summary and Analysis of Heart Rate Lowering Medication Use During Study Page 1 of 3	fqcmlowhrb1	fqcmlowhrb1	Acceptable differences	pvalue:exactly same or acceptable range
Summary of Heart Rate Lowering Medication Use During Study Page 1 of 2	fqcmlowhrc1_cr	fqcmlowhrc1_cr	Acceptable differences	pvalue:String: pvalue- not found
Summary and Analysis of Serotonergic Medication Use During Study (excluding PRN Use) Page 1 of 5	fqcmseronprna1	fqcmseronprna1	Acceptable differences	pvalue:exactly same or acceptable range
Summary and Analysis of Serotonergic Medication Use During Study (excluding PRN Use) Page 1 of 4	fqcmseronprnb1	fqcmseronprnb1	Acceptable differences	pvalue:exactly same or acceptable range
Summary of Serotonergic Medication Use During Study (excluding PRN Use) Page 1 of 4	fqcmseronprnc1_cr	fqcmseronprnc1_cr	Acceptable differences	pvalue:String: pvalue- not found
Summary of duration for TEAE of interest Page 1 of 2	fqteaedur_hl	fqteaedur_hl	Check for the differences	pvalue:String: pvalue- not found
Summary of TEAE Duration Page 1 of 1	fqteaedura	fqteaedura	Check for the differences	pvalue:String: pvalue- not found

%reviewfiles SAS macro process flow



%reviewfiles SAS macro Cont..

Step1: Check the validity of the file path

SAS Pipe option is used with the filename statement to check the legitimacy of the file path as shown in the code below.

```
%*=====;
* Using the pipe statement check the validity of the file path
%*=====;
%let proceed_flg=0;
  Filename Filename    pipe "dir &base_file_path ";
%*=====;
* using infile read in the base file folder rtf file names
%*=====;
Data _file_exist (keep=message) ;
  Infile Filename truncover;
  Input filename $200.;
  Put filename=;
  if filename="dir: cannot access" then fileok=0;
  else fileok=1;
  if fileok=0 then Message=" NOTE: Unable to open file. Please check the File path:  &ckprogram " ;
    if fileok=1 then Message=" NOTE: Success File path:  &ckprogram  is accurate" ;
  call symput ("fileok",fileok);
Run;
```

%reviewfiles SAS macro Cont..

Step 2: Read in all the RTF files that need to be reviewed from the folder using the “Infile” and “Filename” statements

```
data base_&&dset&i ;
  infile "&&read&i" truncover;
  input base_prgm $1000.;
  retain pagefind pagefind_inc;
  length base $1000;
;*****;
' Using string functions extract the actual rtf contents from the base_prgm variable ;
;*****;
  mergevar=_n;
  x=length(base_prgm);
  rev_qa_prgm=trim(reverse(trim(base_prgm)));
  tot_pagenum=substr(compress(REV_QA_PRGM,'fo'),1,1) ;
;*****;
'in the following steps the obvious differences such as datetime stamp in the titles and file location ;
' path in the footnotes are identified;
;*****;

if index(lowercase(base_prgm), 'page') and substr(compress(REV_QA_PRGM,'fo'),1,1)+0>=0 then do; pagefind=1;pagefind_inc=pagefind; end;
if index(lowercase(base_prgm),'page')>0 and index(lowercase(base_prgm),'of')>0 then do; pagefind=1;pagefind_inc=pagefind; end;

  pagefind_inc=pagefind_inc+1;
  pagefind=0;
  if index(base_prgm,'fsl6') >0 and substr(rev_qa_prgm,1,4) ne '6lsf' then base=reverse(substr(rev_qa_prgm,1,index(rev_qa_prgm,'6lsf')-1));
run;
```

%reviewfiles SAS macro Cont..

Step 3: The converted RTF file SAS dataset equivalent data is now analyzed for the differences and the key parameter based search across all the variables and records are conducted

```
%let keyfound=;
data base_pval_&datanm(keep= base typeb org_base indexmerge index pval_full pval_location z bspval);
set base_&datanm;

retain pval_location typeb;
/*Key value search in both rows and columns to identify the values corresponding to key*/
index= index(lowercase(compress(org_base,'-')), "&key");
indexmerge= substr(base,1, index(trim(base), ' '));
if index(lowercase(compress(org_base,'-')), "&key") >0 and index(lowercase(compress(indexmerge,'-')), "&key")>0 and typeb=' ' then typeb='row';
else if index(lowercase(compress(org_base,'-')), "&key") >0 and typeb=' ' then typeb='col';
    if typeb ne ' ' then CALL SYMPUT('keyfound' , 'Y');
    else CALL SYMPUT('keyfound' , 'N');
/*Convert < amd > symbol values in pvalues to an equivalent values for analysis*/

org_base_n=TRANWRD (org_base_n, '<.001','0.0001');
org_base_n=TRANWRD (org_base_n, '>.99','0.9999');
keytyp="&key";
array arr(*)$20 pvl -pv30;
array arrn(*) pvl1 -pvb30;
cnt=0;
do j= 1 to 30;
if location>0 then do; arr(J)=scan(org_base_n,j);
arrn(J)=TRANWRD (arr(J), '_', '.') +0;
if arrn(J)>=500 then arrn(J)=.;
if arrn(J) ne . then cnt+1;
end;
end;
```

%reviewfiles SAS macro Cont..

Step 4: In this step hyperlinks for one-click access to the RTF files and XML summary output are being created using the code below. A sample output XML file is shown in figure 6.

```
base_link=cats('=HYPERLINK("", 'file:///\\xxxx', "&base_folder_path_", tflname,
'.rtf"', '&tflname', '&tflname', '&tflname')');

ods path work.templ(update) sashelp.tmplmst(read) sashelp.styles(read);
ods tagsets.ExcelXP style=sansPrinter options(embedded_titles='Yes'
        embed_titonce='no' suppress_bylines='yes'
        sheet_interval='bygroup' sheet_label=' '
        absolute_column_width="&colsp."
        autofit_height='yes'
        frozen_headers='5'
        frozen_rowheaders='yes'|
        embedded_footnote='Yes' tabcolor="blue")
        file="&outpath./&outfilenm..xml" ;
PROC REPORT DATA=allreprots2 NOWD HEADLINE HEADSKIP SPLIT="!" MISSING;
  columns Study base_rtf base_link compare_link summary_n
         %if &keyn ne %then %do; summary_key;%end;;
  define study/"Study" width=25 left flow;
  define base_rtf/"Base file" left WIDTH=100 ;
  define base_link/"Base file" left WIDTH=10 flow ; * flow;
  define compare_link/"Compare file" left WIDTH=10 flow;
  define summary_n/"Findings summary" left WIDTH=20 flow;
  %if &keyn ne %then %do;
  define summary_key/"Findings on key search" left WIDTH=20 flow;
  %end;
RUN;
```

Output

Index tab of the XML summary output

Title	Base file	Compare file	Findings summary	Findings on key search
Summary and Analysis of NSAID Use During Study (excluding PRN Use) Page 1 of 3	fqcmnsaidnprna1	fqcmnsaidnprna1	Check for the differences	Check
Summary of duration for TEAE of interest Page 1 of 2	fqteaedur_hl	fqteaedur_hl	Exact Match	String: pvalue- not found

Detailed compare summary report

Base file	Compare file	Base file difference	compare file difference	Issue
fqteaedur_hl	fqteaedur_hl	On the >= 6th treated attack 145 (30.1) 96 (20.0) 75 (15.6) 62 (12.9) 33 (6.9) 12 (2.5) 5 (1.0) 481	On the >= 6th treated attack 146 (30.3) 96 (20.0) 75 (15.6) 62 (12.9) 33 (6.9) 12 (2.5) 5 (1.0) 481	Check
fqteaesonset_hl	fqteaesonset_hl	On the 1st treated attack 65 (18.5) 31 (9.2) 24 (7.1) 21 (6.3) 19 (5.7) 7 (2.1) 0 (0.0) 336	On the 1st treated attack 66 (18.6) 31 (9.2) 24 (7.1) 21 (6.3) 19 (5.7) 7 (2.1) 0 (0.0) 339	Check

Detailed keyword search report

Base file	Compare file	Base file difference	compare file difference	Finding
fqcmnsaidnprna1	fqcmnsaidnprna1		nabumetone 1 (0.1) [0.1] 0 (0.0) [0.0] 0.00 .083	new record
fqcmnsaidnprna1	fqcmnsaidnprna1	naproxen sodium 4 (0.3) [0.3] 11 (0.3) [0.3] 1.07 (0.34, 3.39) .906	naproxen sodium 77 (5.7) [5.8] 150 (4.4) [4.4] 0.75 (0.56, 0.99) .043	pvalue difference >.5
fqcmnsaidnprna1	fqcmnsaidnprna1		vicoprofen 0 (0.0) [0.0] 1 (0.0) [0.0] .480	new record
fqteaedur_hl	fqteaedur_hl			String: pvalue- not found

Business impact using this tool

- Better control of quality, specially in making sure that the validated version in QA is consistent with the version in PRD
- Significant amount of time is being saved by the medical reviewers to identify changes to different versions of outputs thus eliminating potential submission delays.
- The tedious process of looking for changes to the output files with large volume is made easy using the tool (eg: patient narrative reports)

Conclusion

- This tool will compare the RTF outputs and print any differences that are not caused due to timestamp differences or program mode in the file content.
- This tool can check all the files at a directory level to compare the contents of all the RTF files in the directory with the equivalent files in another location(eg: QA vs. PRD) .
- This tool can be used to identify reports by keyword and check changes to contents while controlling the acceptable range of differences (eg: P-values)
- The utility macro can be executed using the SAS EG and the folder locations can be selected using the pop-up facility that is featured in SAS EG without having to manipulate the actual program to call this macro.
- The utility macro creates an XML file output that has the summary of findings and hyperlinks to facilitate the review and easy accessibility.

Author's Contact Information

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

Youying Yu

Consultant Statistician-Computation, Global Statistical Science, Eli Lilly & Company

Work Phone: +1-317-433-1033

Email: yu_youying@lilly.com

Web: <http://www.lilly.com>

Hrideep Antony

Principal Statistical Programmer, Clinical Division, Syneos Health

Phone: +1-984-459-4785

E-mail: Hrideep.antony@syneoshealth.com

Web: <http://www.syneoshealth.com>