EFFICIENT TLF REVIEW WITHIN STATISTICAL PROGRAMMING USING R AND SAS

PHUSE US Connect 2025 | March 17, 2025

Siddharth Kumar

Website LinkedIn

1. TRADITIONAL QC PROCESS
2. CHALLENGES & GAPS
3. RTF REVIEW & SELECTIVE QC
4. SAS PROGRAMMING USE CASE
5. EXPLORING R PACKAGES
6. R PROGRAMMING SCENARIO
7. R SHINY
8. SUMMARY

VALIDATION OF TLF RTF OUTPUT

QC PROCESS

- Traditionally, all outputs undergo independent QC through double programming.
- Involves double programming of the final SAS dataset.
- Manual review of RTF outputs for any Title, footnotes, page break, and formatting issues.
- There are studies where full independent qc is not required but a targeted QC approach is followed by performing a quality check by reviewing the outputs directly.

GAPS

- While effective for data verification, PROC COMPARE fails to catch formatting issues like misaligned text, incorrect font styles, or missing footnotes.
- These issues in RTF outputs would not be detected by PROC COMPARE, leading to quality concerns from reviewers.
- Constrained by resource availability.
- Manual review of over 300 RTF outputs is time-consuming and error-prone.

EFFICIENT TFL REVIEW ENHANCING QUALITY

- This presentation proposes an efficient solution using R and SAS to streamline the review process, ensuring accuracy and consistency across Tables, Listings, and Figures (TLF) package
- The programmatic workflow includes: RTF File Processing:-
- Extracting and analyzing text from RTF files
- Generating reports to highlight inconsistencies
- By optimizing these tasks within Statistical Programming Team, we significantly reduce review time, enhance accuracy, and ensure the TLF package is robust and ready to share with Sponsors and cross-functional Team

CHALLENGES

TIME-CONSUMING PROCESS ERROR PRONE VERIFICATION

- Manually reviewing hundreds of RTF outputs requires significant time and resources, creating bottlenecks in tight submission timelines
- Reviewers must check population counts, row consistency, decimal precision, alignment, spelling, titles, footnotes, and page breaks across numerous files

SELECTIVE QC

- Population counts – Big N
- Primary and Secondary endpoints counts
- Adverse Events – Overall AE, Serious AE, Related AE
- Death
- Decimal precision and alignment
- Page Breaks
- Title and Footnotes
- Spelling Check
- Formatting

EFFICIENT TFL REVIEW ENHANCING QUALITY

- The entire review process need to be streamlined using R and SAS
- Both SAS and R have capabilities to read RTF files
- Parsing and processing these datasets creates a robust programmatic review tool
- This tool can be used as a post-PROC COMPARE stage to perform a sanity check
- Results from the programmatic review would be output to an Excel file or a Smartsheet dashboard

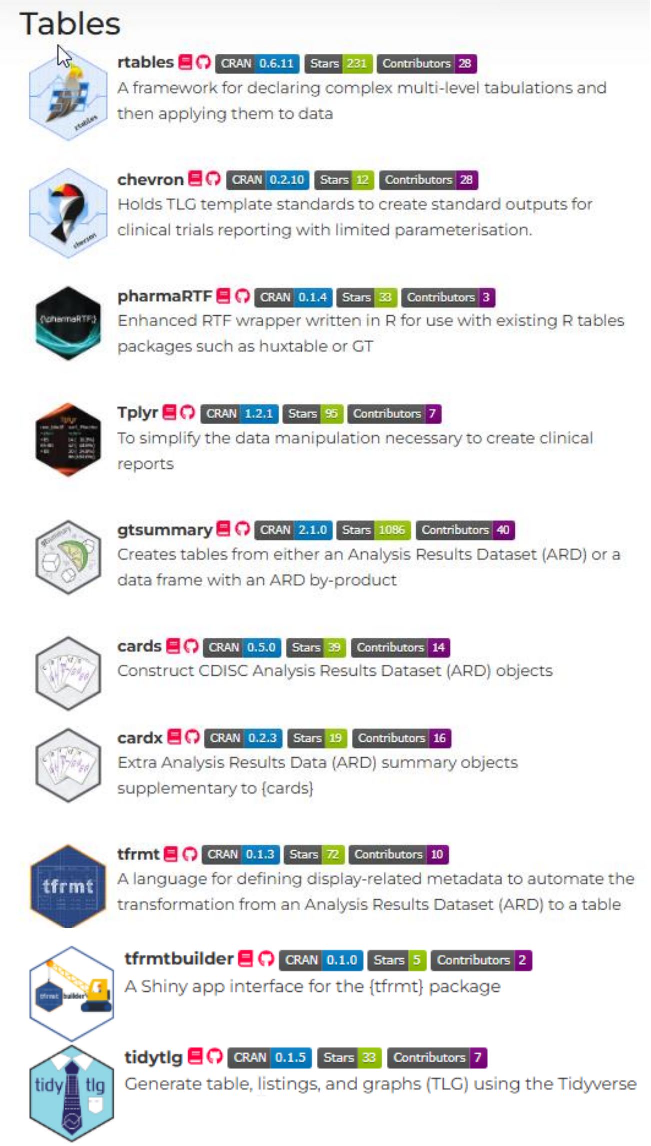
- Population Count (Big N)
- Primary & Secondary endpoints
- Adverse Events
- Serious Adverse Event
- Related Adverse Event
- Decimal Precision & alignment
- Titles and Footnotes
- Formatting, Spell Checks
- Demographic and Baseline Characteristics
- Unresolved macro variables

CLINICAL REPORTING R PACKAGES

- R packages that are predominantly used in Clinical Reporting as cataloged on pharmaverse
- This list from pharmaverse.org encapsulates all the open source R packages(and code bases from other open source languages) that fall under the **scope of usage for end-to-end clinical reporting in pharma.**
- Pharmaverse is now a PHUSE Working Group

Source: https://pharmaverse.org/

R PACKAGES FOR GENERATION OF TFL



Tables

rtables CRAN 0.6.11 Stars 231 Contributors 28
A framework for declaring complex multi-level tabulations and then applying them to data

chevron CRAN 0.2.10 Stars 12 Contributors 28
Holds TLG template standards to create standard outputs for clinical trials reporting with limited parameterisation.

pharmaRTF CRAN 0.1.4 Stars 33 Contributors 3
Enhanced RTF wrapper written in R for use with existing R tables packages such as huxtable or GT

Tplyr CRAN 1.2.1 Stars 95 Contributors 7
To simplify the data manipulation necessary to create clinical reports

gtsummary CRAN 2.1.0 Stars 1086 Contributors 40
Creates tables from either an Analysis Results Dataset (ARD) or a data frame with an ARD by-product

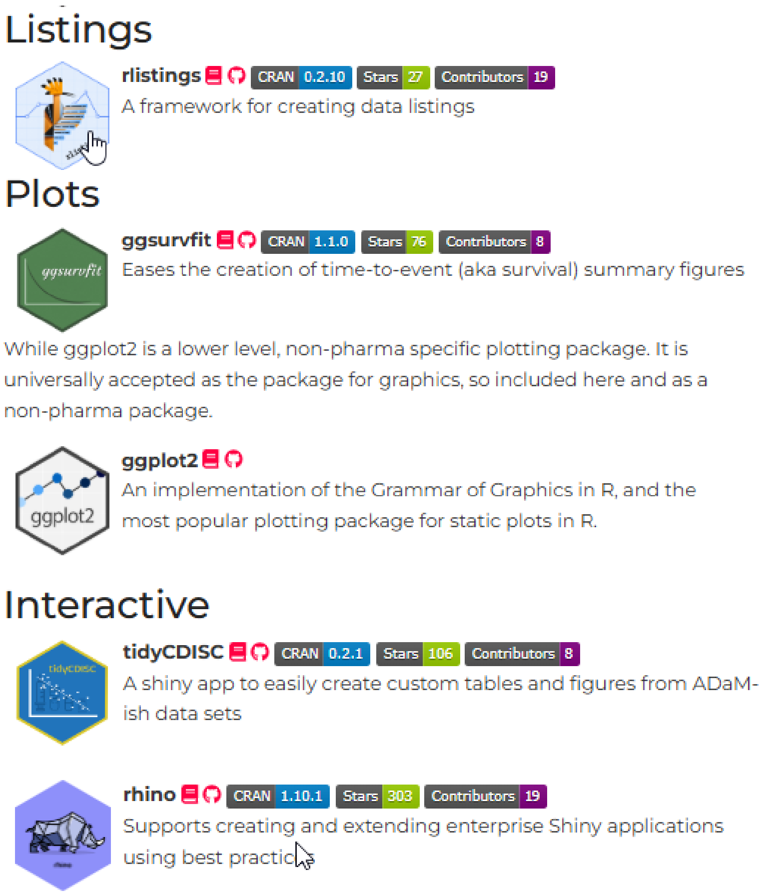
cards CRAN 0.5.0 Stars 39 Contributors 14
Construct CDISC Analysis Results Dataset (ARD) objects

cardx CRAN 0.2.3 Stars 19 Contributors 16
Extra Analysis Results Data (ARD) summary objects supplementary to {cards}

tfrmt CRAN 0.1.3 Stars 72 Contributors 10
A language for defining display-related metadata to automate the transformation from an Analysis Results Dataset (ARD) to a table

tfrmtbuilder CRAN 0.1.0 Stars 5 Contributors 2
A Shiny app interface for the {tfrmt} package

tidytlg CRAN 0.1.5 Stars 33 Contributors 7
Generate table, listings, and graphs (TLG) using the Tidyverse



Listings

rlistings CRAN 0.2.10 Stars 27 Contributors 19
A framework for creating data listings

Plots

ggsurvfit CRAN 1.1.0 Stars 76 Contributors 8
Eases the creation of time-to-event (aka survival) summary figures

While ggplot2 is a lower level, non-pharma specific plotting package. It is universally accepted as the package for graphics, so included here and as a non-pharma package.

ggplot2
An implementation of the Grammar of Graphics in R, and the most popular plotting package for static plots in R.

Interactive

tidyCDISC CRAN 0.2.1 Stars 106 Contributors 8
A shiny app to easily create custom tables and figures from ADaM-ish data sets

rhino CRAN 1.10.1 Stars 303 Contributors 19
Supports creating and extending enterprise Shiny applications using best practic

PHARMAVERSE

- R packages commonly used for **generation** of Tables, Listings, and Figures, as outlined on Pharmaverse

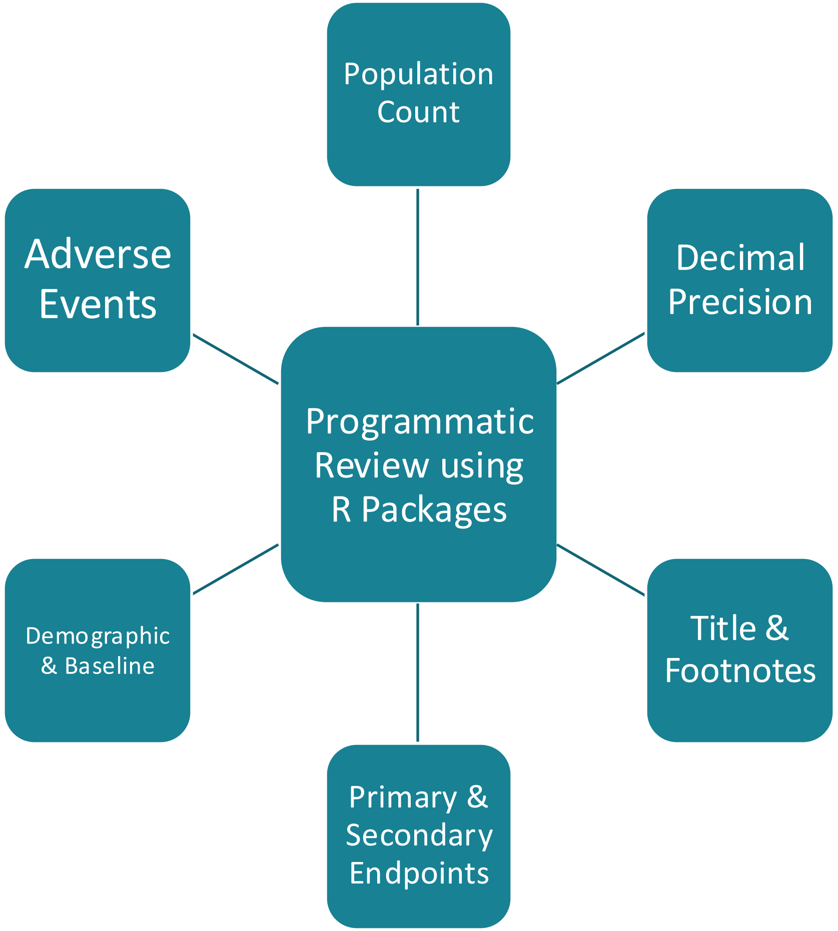
Source: https://pharmaverse.org/e2eclinical/tlg/

PROPOSED PROGRAMMATIC QC PROCESS

Step 1

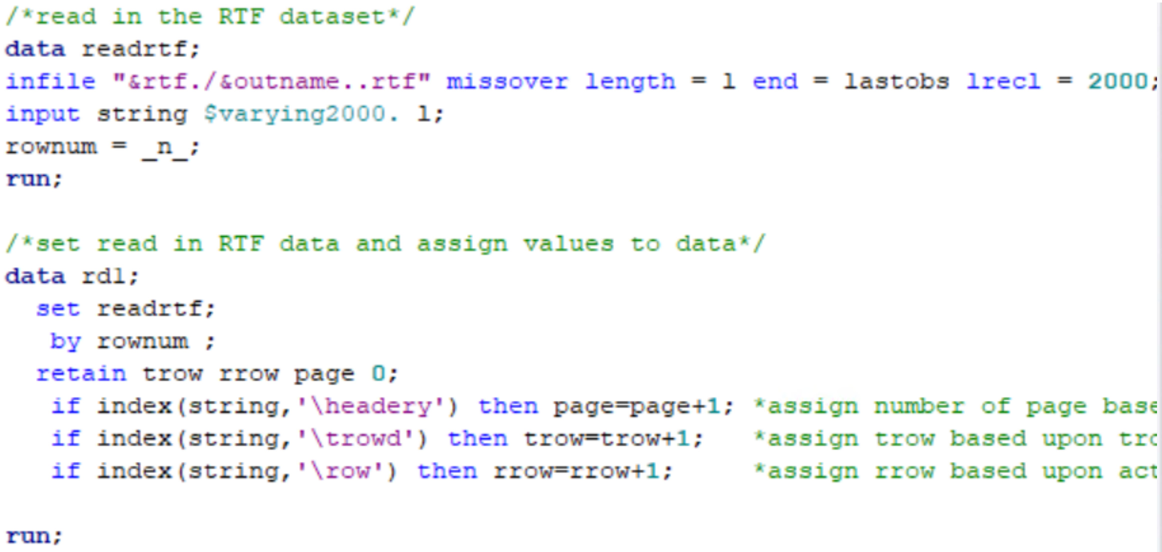
- Import RTF files into SAS
- Programmatic QC of RTF2SAS dataset
- Reduces risk and improves quality check
- Overwriting SAS Final dataset
- Eliminates need for variable naming convention

Step 2



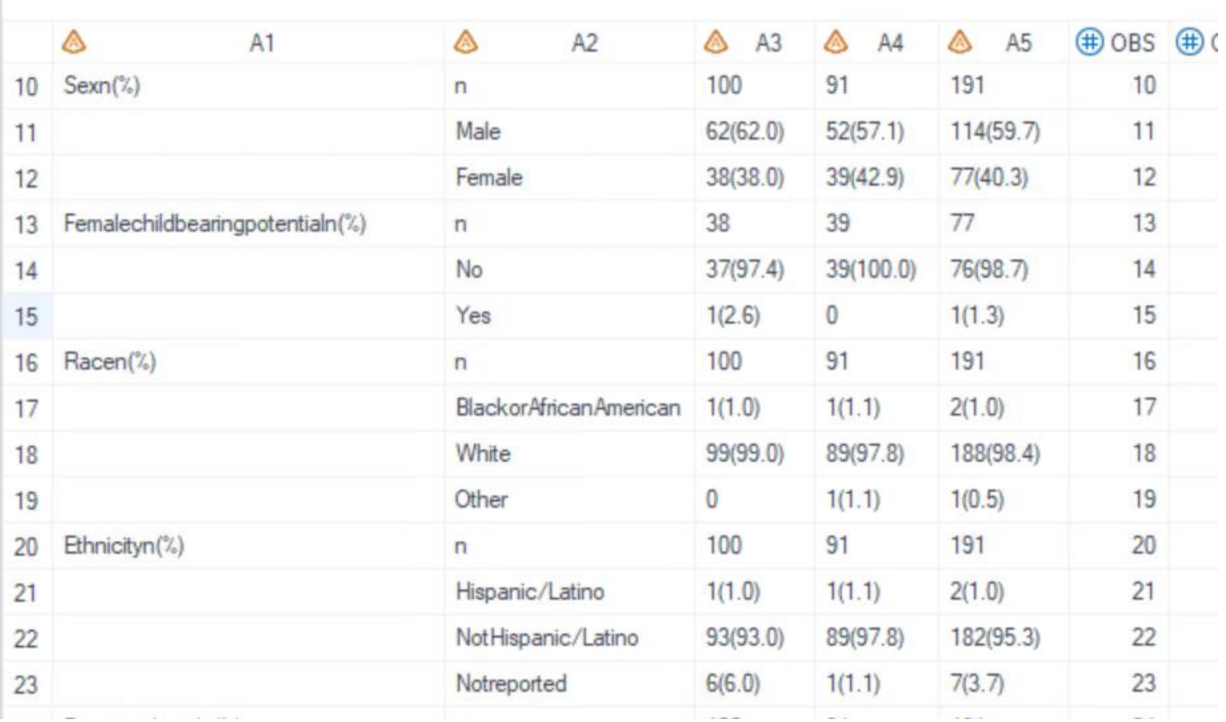
STEP 1

READ RTF FILE INTO SAS DATASET



```
/*read in the RTF dataset*/
data readrtf;
infile "&rtf./&outname..rtf" missover length = l end = lastobs lrecl = 2000;
input string $varying2000. l;
rownum = _n_;
run;

/*set read in RTF data and assign values to data*/
data rdl;
  set readrtf;
   by rownum ;
  retain trow rrow page 0;
   if index(string,'\headery') then page=page+1; *assign number of page base
   if index(string,'\trowd') then trow=trow+1;   *assign trow based upon tro
   if index(string,'\row') then rrow=rrow+1;     *assign rrow based upon act

run;
```

SAS OUTPUT DATASET

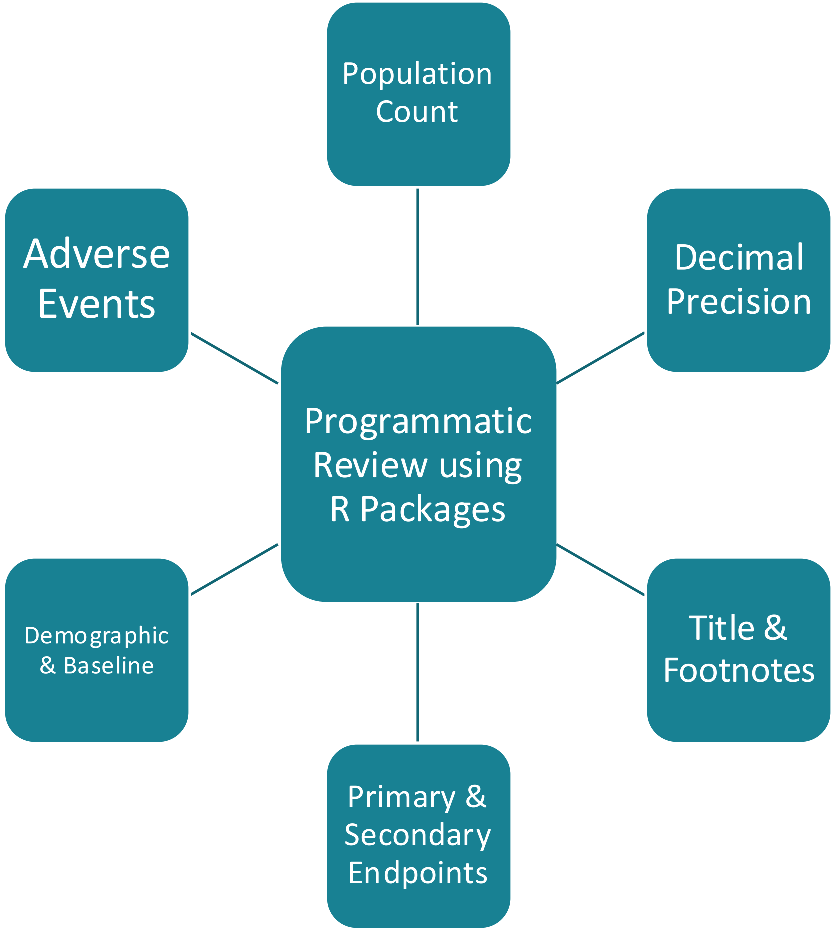


	A1	A2	A3	A4	A5	OBS
10	Sexn(%)	n	100	91	191	10
11		Male	62(62.0)	52(57.1)	114(59.7)	11
12		Female	38(38.0)	39(42.9)	77(40.3)	12
13	Femalechildbearingpotentialn(%)	n	38	39	77	13
14		No	37(97.4)	39(100.0)	76(98.7)	14
15		Yes	1(2.6)	0	1(1.3)	15
16	Racen(%)	n	100	91	191	16
17		BlackorAfricanAmerican	1(1.0)	1(1.1)	2(1.0)	17
18		White	99(99.0)	89(97.8)	188(98.4)	18
19		Other	0	1(1.1)	1(0.5)	19
20	Ethnicityn(%)	n	100	91	191	20
21		Hispanic/Latino	1(1.0)	1(1.1)	2(1.0)	21
22		NotHispanic/Latino	93(93.0)	89(97.8)	182(95.3)	22
23		Notreported	6(6.0)	1(1.1)	7(3.7)	23

- Extract data from RTF file
- Parse data
- Programmatic Check for Counts for Key demographic, Primary and Secondary endpoints

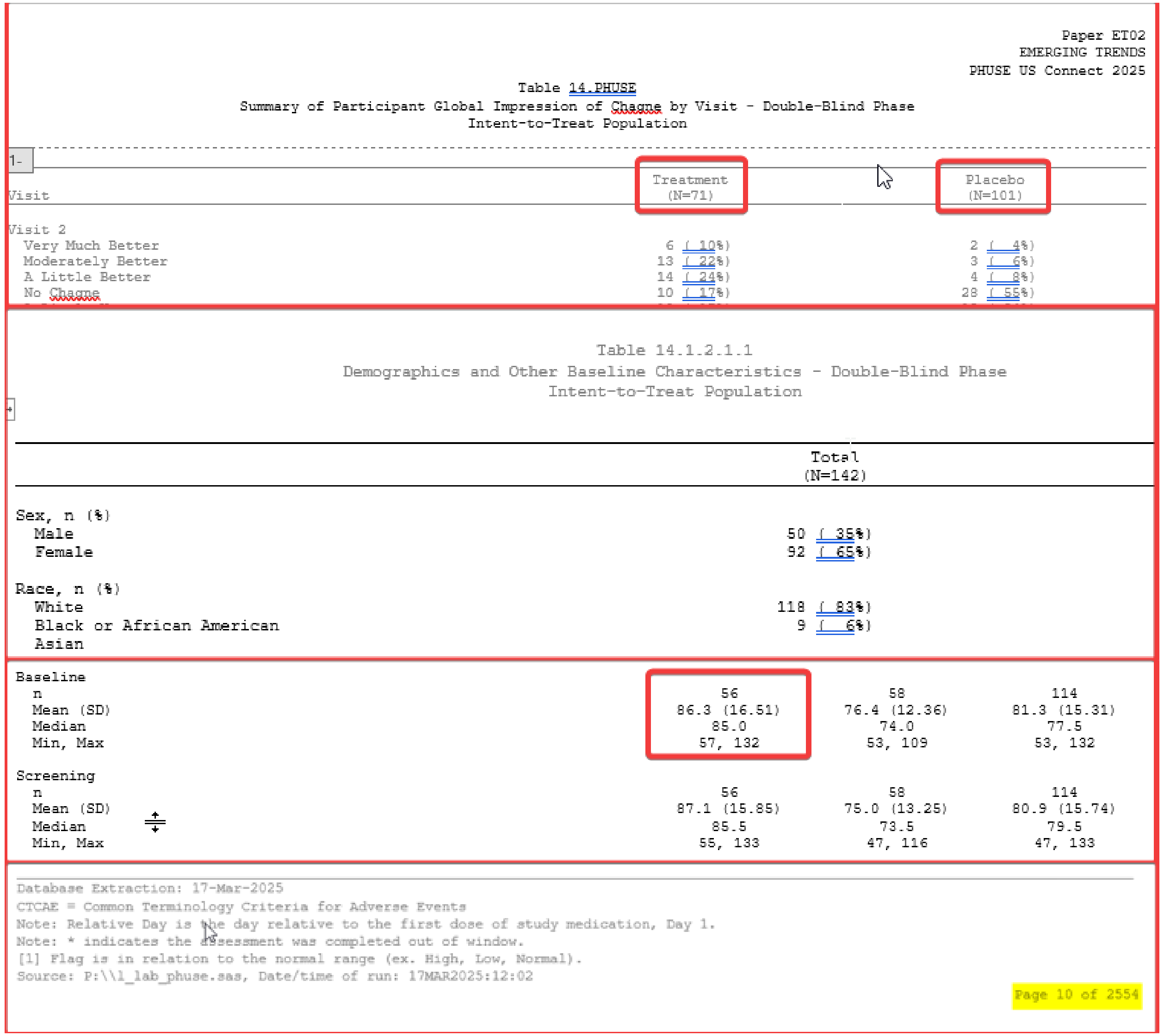
R PACKAGES USED FOR SELECTIVE QC

STEP 2



Shiny	Build interactive web applications with R
Striprtf	Extract Text from RTF File
Stringr	Consistent wrapper for common string operations
dplyr	Tool for working with data frame like objects
writexl	Export data frames to excel ‘xlsx’ format
gtsummary	Creates table summarizing datasets, regression models
Hunspell	Spell checker

REVIEW OF RTF OUTPUTS



Paper ET02
EMERGING TRENDS
PHUSE US Connect 2025

Table 14.PHUSE
Summary of Participant Global Impression of Chagne by Visit - Double-Blind Phase
Intent-to-Treat Population

Visit	Treatment (N=71)	Placebo (N=101)
Visit 2		
Very Much Better	6 (10%)	2 (4%)
Moderately Better	13 (22%)	3 (6%)
A Little Better	14 (24%)	4 (8%)
No Chagne	10 (17%)	28 (55%)

Table 14.1.2.1.1
Demographics and Other Baseline Characteristics - Double-Blind Phase
Intent-to-Treat Population

	Total (N=142)
Sex, n (%)	
Male	50 (35%)
Female	92 (65%)
Race, n (%)	
White	118 (83%)
Black or African American	9 (6%)
Asian	

Baseline			
n	56	58	114
Mean (SD)	86.3 (16.51)	76.4 (12.36)	81.3 (15.31)
Median	85.0	74.0	77.5
Min, Max	57, 132	53, 109	53, 132
Screening			
n	56	58	114
Mean (SD)	87.1 (15.85)	75.0 (13.25)	80.9 (15.74)
Median	85.5	73.5	79.5
Min, Max	55, 133	47, 116	47, 133

Database Extraction: 17-Mar-2025
CTCAE = Common Terminology Criteria for Adverse Events
Note: Relative Day is the day relative to the first dose of study medication, Day 1.
Note: * indicates the assessment was completed out of window.
[1] Flag is in relation to the normal range (ex. High, Low, Normal).
Source: P:\\l_lab_phuse.sas, Date/time of run: 17MAR2025:12:02

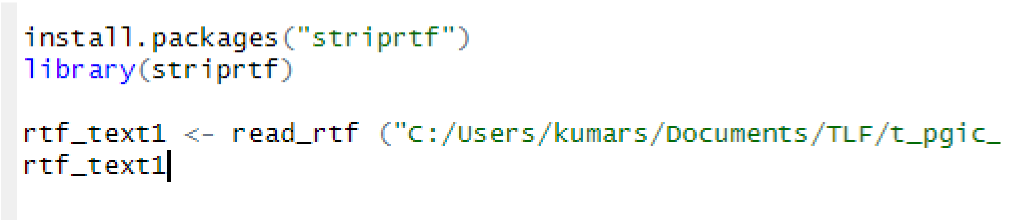
Page 10 of 2554

RTF REVIEW

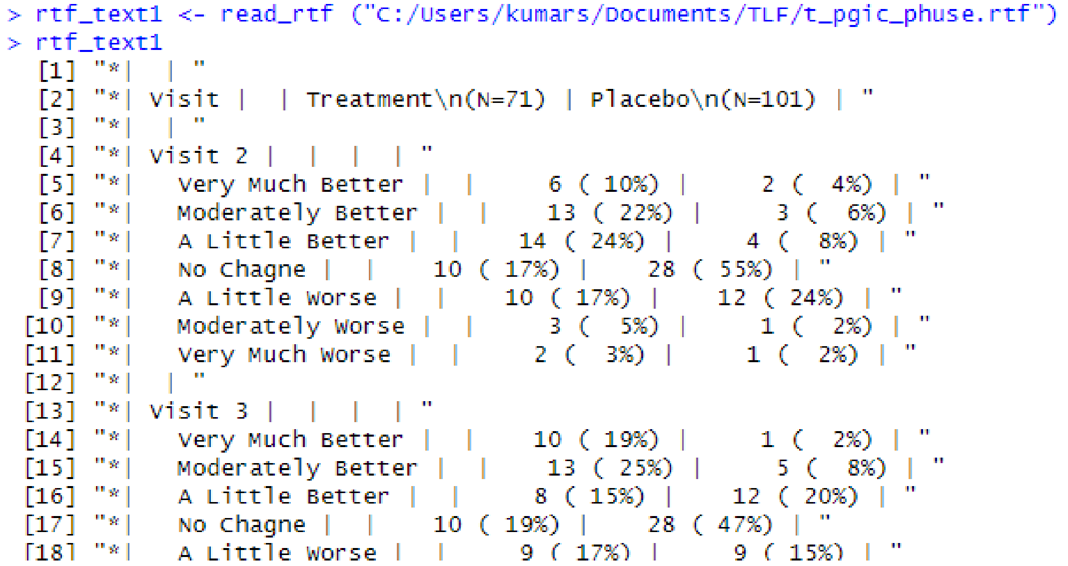
- Population Count (Big N)
- Primary & Secondary endpoints
- Adverse Events
- Serious Adverse Event
- Related Adverse Event
- Decimal Precision & alignment
- Titles and Footnotes
- Formatting, Spell Checks
- Demographic and Baseline Characteristics
- Unresolved macro variables

READ & EXTRACT USING R PACKAGES

STRIPRTF - EXTRACT TEXT FROM RTF FILE

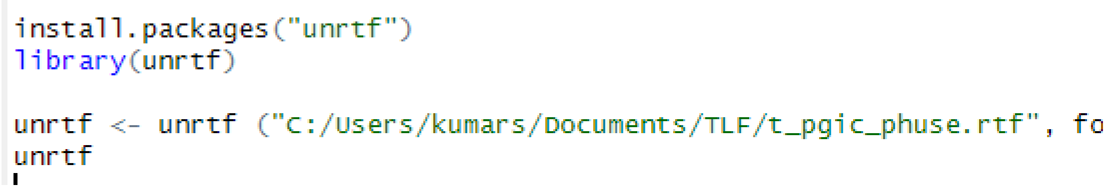


```
install.packages("striprtf")
library(striprtf)

rtf_text1 <- read_rtf ("C:/Users/kumars/Documents/TLF/t_pgic_
rtf_text1
```

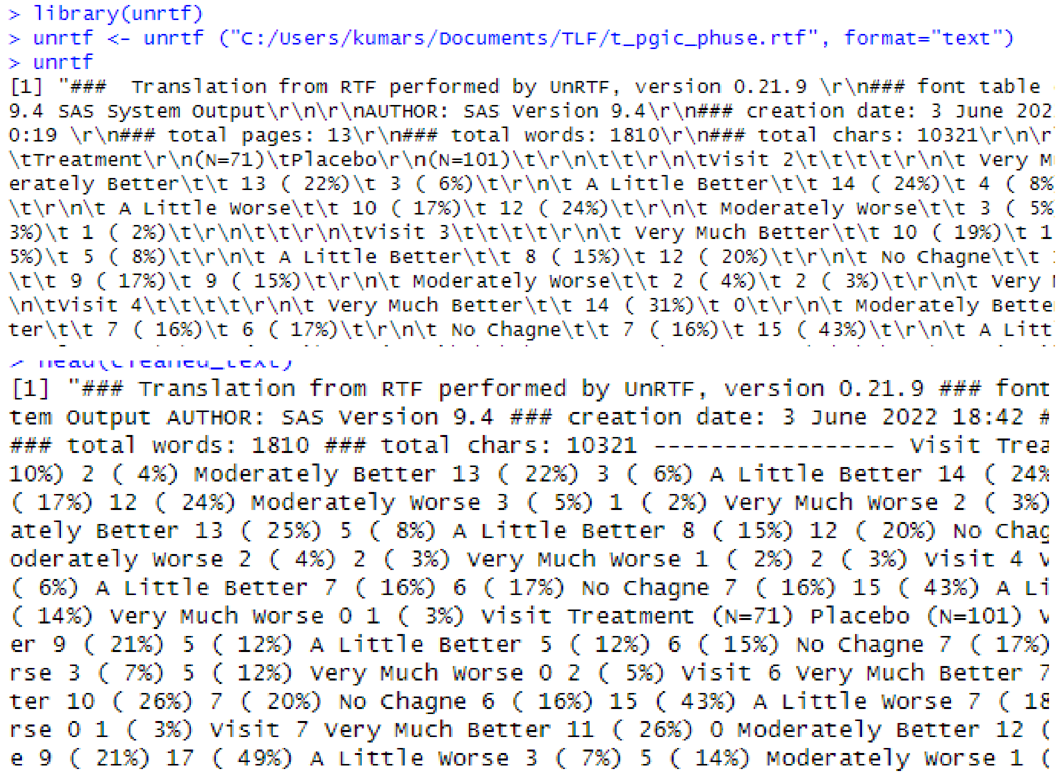


```
> rtf_text1 <- read_rtf ("C:/Users/kumars/Documents/TLF/t_pgic_phuse.rtf")
> rtf_text1
 [1] "*|  | "
 [2] "*| Visit |  | Treatment\n(N=71) | Placebo\n(N=101) | "
 [3] "*|  | "
 [4] "*| Visit 2 |  |  |  | "
 [5] "*|   Very Much Better |  |      6 ( 10%) |      2 (  4%) | "
 [6] "*|   Moderately Better |  |     13 ( 22%) |      3 (  6%) | "
 [7] "*|   A Little Better |  |     14 ( 24%) |      4 (  8%) | "
 [8] "*|   No Chagne |  |     10 ( 17%) |     28 ( 55%) | "
 [9] "*|   A Little Worse |  |     10 ( 17%) |     12 ( 24%) | "
[10] "*|   Moderately Worse |  |      3 (  5%) |      1 (  2%) | "
[11] "*|   Very Much Worse |  |      2 (  3%) |      1 (  2%) | "
[12] "*|  | "
[13] "*| Visit 3 |  |  |  | "
[14] "*|   Very Much Better |  |     10 ( 19%) |      1 (  2%) | "
[15] "*|   Moderately Better |  |     13 ( 25%) |      5 (  8%) | "
[16] "*|   A Little Better |  |      8 ( 15%) |     12 ( 20%) | "
[17] "*|   No Chagne |  |     10 ( 19%) |     28 ( 47%) | "
[18] "*|   A Little Worse |  |      9 ( 17%) |      9 ( 15%) | "
```

UNRTF - CONVERTS AN RTF DOCUMENT TO HTML, TEXT



```
install.packages("unrtf")
library(unrtf)

unrtf <- unrtf ("C:/Users/kumars/Documents/TLF/t_pgic_phuse.rtf", fo
unrtf
```



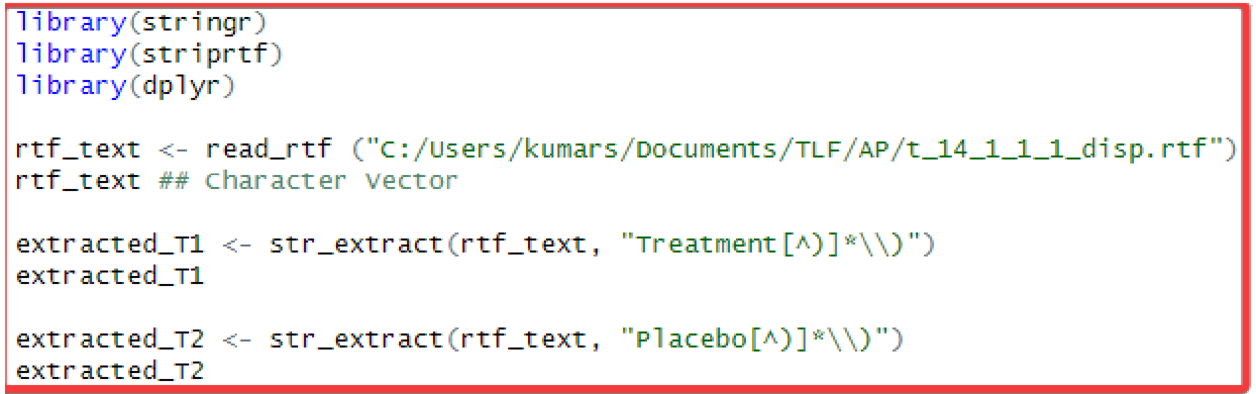
```
> library(unrtf)
> unrtf <- unrtf ("C:/Users/kumars/Documents/TLF/t_pgic_phuse.rtf", format="text")
> unrtf
[1] "###  Translation from RTF performed by UnRTF, version 0.21.9 \r\n### font table
9.4 SAS System Output\r\n\r\nAUTHOR: SAS Version 9.4\r\n### creation date: 3 June 202
0:19 \r\n### total pages: 13\r\n### total words: 1810\r\n### total chars: 10321\r\n\r
\tTreatment\r\n(N=71)\tPlacebo\r\n(N=101)\t\r\n\t\t\r\n\tVisit 2\t\t\t\t\r\n\t Very M
erately Better\t\t 13 ( 22%)\t 3 ( 6%)\t\r\n\t A Little Better\t\t 14 ( 24%)\t 4 ( 8%
\t\r\n\t A Little Worse\t\t 10 ( 17%)\t 12 ( 24%)\t\r\n\t Moderately Worse\t\t 3 ( 5%
3%)\t 1 ( 2%)\t\r\n\t\t\r\n\tVisit 3\t\t\t\t\r\n\t Very Much Better\t\t 10 ( 19%)\t 1
5%)\t 5 ( 8%)\t\r\n\t A Little Better\t\t 8 ( 15%)\t 12 ( 20%)\t\r\n\t No Chagne\t\t
\t\t 9 ( 17%)\t 9 ( 15%)\t\r\n\t Moderately Worse\t\t 2 ( 4%)\t 2 ( 3%)\t\r\n\t Very
\n\tVisit 4\t\t\t\t\r\n\t Very Much Better\t\t 14 ( 31%)\t 0\t\r\n\t Moderately Bette
ter\t\t 7 ( 16%)\t 6 ( 17%)\t\r\n\t No Chagne\t\t 7 ( 16%)\t 15 ( 43%)\t\r\n\t A Litt
```

```
> head(cleaned_text)
[1] "### Translation from RTF performed by UnRTF, version 0.21.9 ### font
tem Output AUTHOR: SAS Version 9.4 ### creation date: 3 June 2022 18:42 #
### total words: 1810 ### total chars: 10321 ------------------- Visit Trea
10%) 2 ( 4%) Moderately Better 13 ( 22%) 3 ( 6%) A Little Better 14 ( 24%
( 17%) 12 ( 24%) Moderately Worse 3 ( 5%) 1 ( 2%) Very Much Worse 2 ( 3%)
ately Better 13 ( 25%) 5 ( 8%) A Little Better 8 ( 15%) 12 ( 20%) No Chag
oderately Worse 2 ( 4%) 2 ( 3%) Very Much Worse 1 ( 2%) 2 ( 3%) Visit 4 V
( 6%) A Little Better 7 ( 16%) 6 ( 17%) No Chagne 7 ( 16%) 15 ( 43%) A Li
( 14%) Very Much Worse 0 1 ( 3%) Visit Treatment (N=71) Placebo (N=101) V
er 9 ( 21%) 5 ( 12%) A Little Better 5 ( 12%) 6 ( 15%) No Chagne 7 ( 17%)
rse 3 ( 7%) 5 ( 12%) Very Much Worse 0 2 ( 5%) Visit 6 Very Much Better 7
ter 10 ( 26%) 7 ( 20%) No Chagne 6 ( 16%) 15 ( 43%) A Little Worse 7 ( 18
rse 0 1 ( 3%) Visit 7 Very Much Better 11 ( 26%) 0 Moderately Better 12 (
e 9 ( 21%) 17 ( 49%) A Little Worse 3 ( 7%) 5 ( 14%) Moderately Worse 1 (
```

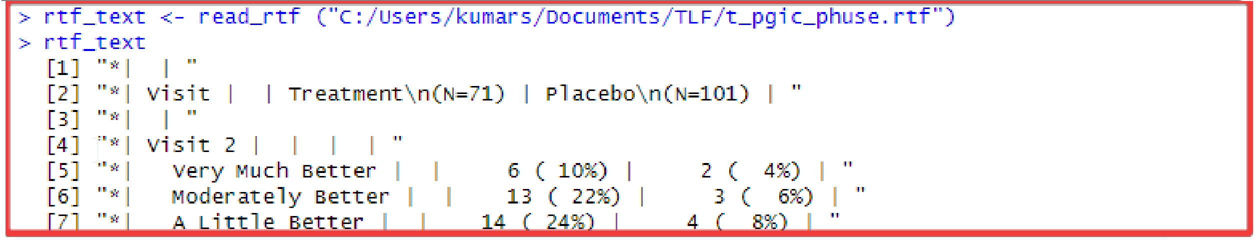
SCENARIO 1 – POPULATION COUNTS

BIG N

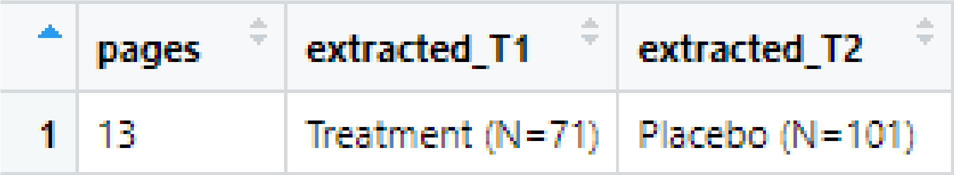
- Population counts (Big N) are assigned during the PROC Report stage in SAS
- Potential Issue : Incorrect calculation of values
- Incorrect assignment of values, such as placebo counts appearing in the treatment group and vice versa.
- Automation Using R: R packages can extract Big N values from the RTF outputs to ensure accurate population counts.



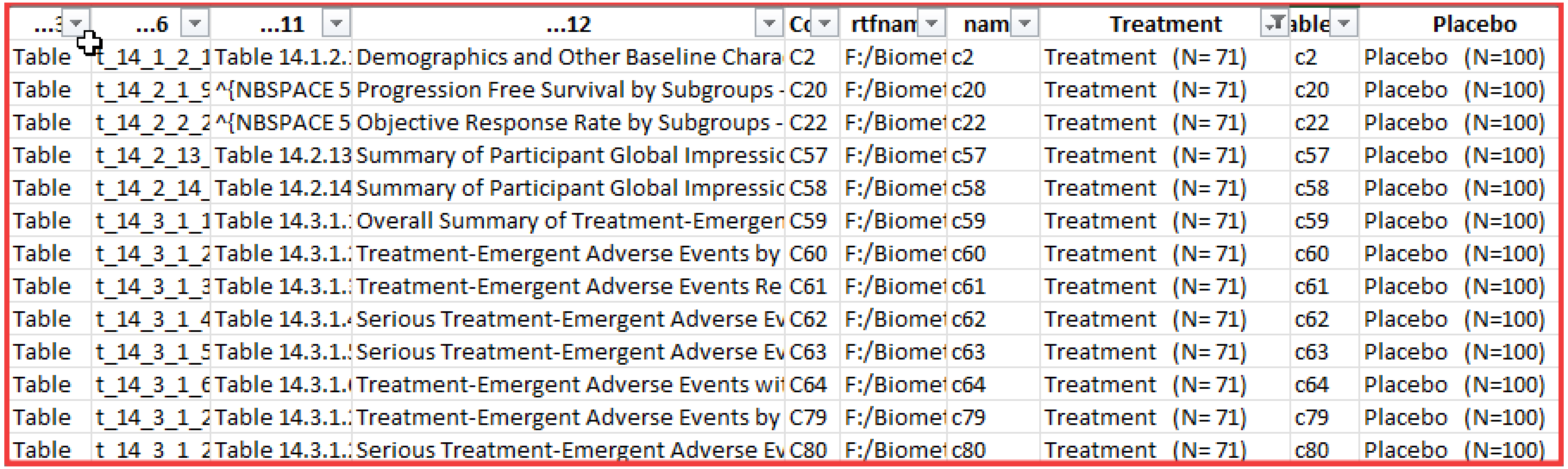
```
library(stringr)
library(striprtf)
library(dplyr)

rtf_text <- read_rtf ("C:/Users/kumars/Documents/TLF/AP/t_14_1_1_1_disp.rtf")
rtf_text ## Character Vector

extracted_T1 <- str_extract(rtf_text, "Treatment[^)]*\\)")
extracted_T1

extracted_T2 <- str_extract(rtf_text, "Placebo[^)]*\\)")
extracted_T2
```



```
> rtf_text <- read_rtf ("C:/Users/kumars/Documents/TLF/t_pgic_phuse.rtf")
> rtf_text
  [1] "*|  | "
  [2] "*| Visit |  | Treatment\n(N=71) | Placebo\n(N=101) | "
  [3] "*|  | "
  [4] "*| Visit 2 |  |  |  | "
  [5] "*|   Very Much Better |  |     6 ( 10%) |     2 (  4%) | "
  [6] "*|   Moderately Better |  |    13 ( 22%) |     3 (  6%) | "
  [7] "*|   A Little Better |  |    14 ( 24%) |     4 (  8%) | "
```



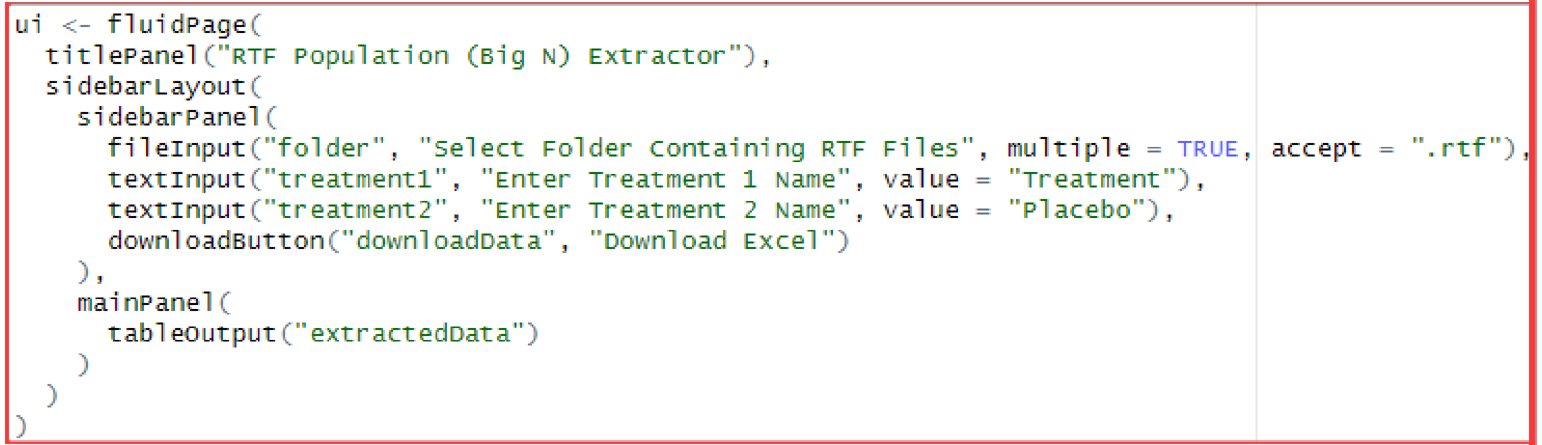
	pages	extracted_T1	extracted_T2
1	13	Treatment (N=71)	Placebo (N=101)



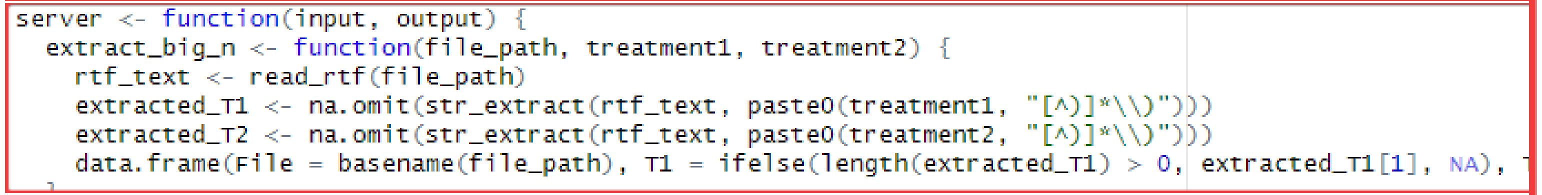
...3	...6	...11	...12	C	rtfnam	nam	Treatment	able	Placebo
Table	t_14_1_2_1	Table 14.1.2.	Demographics and Other Baseline Chara	C2	F:/Biome	c2	Treatment (N= 71)	c2	Placebo (N=100)
Table	t_14_2_1_9	^{NBSPACE 5	Progression Free Survival by Subgroups -	C20	F:/Biome	c20	Treatment (N= 71)	c20	Placebo (N=100)
Table	t_14_2_2_2	^{NBSPACE 5	Objective Response Rate by Subgroups -	C22	F:/Biome	c22	Treatment (N= 71)	c22	Placebo (N=100)
Table	t_14_2_13_	Table 14.2.13	Summary of Participant Global Impressi	C57	F:/Biome	c57	Treatment (N= 71)	c57	Placebo (N=100)
Table	t_14_2_14_	Table 14.2.14	Summary of Participant Global Impressi	C58	F:/Biome	c58	Treatment (N= 71)	c58	Placebo (N=100)
Table	t_14_3_1_1	Table 14.3.1.	Overall Summary of Treatment-Emergen	C59	F:/Biome	c59	Treatment (N= 71)	c59	Placebo (N=100)
Table	t_14_3_1_2	Table 14.3.1.	Treatment-Emergent Adverse Events by	C60	F:/Biome	c60	Treatment (N= 71)	c60	Placebo (N=100)
Table	t_14_3_1_3	Table 14.3.1.	Treatment-Emergent Adverse Events Re	C61	F:/Biome	c61	Treatment (N= 71)	c61	Placebo (N=100)
Table	t_14_3_1_4	Table 14.3.1.	Serious Treatment-Emergent Adverse Ev	C62	F:/Biome	c62	Treatment (N= 71)	c62	Placebo (N=100)
Table	t_14_3_1_5	Table 14.3.1.	Serious Treatment-Emergent Adverse Ev	C63	F:/Biome	c63	Treatment (N= 71)	c63	Placebo (N=100)
Table	t_14_3_1_6	Table 14.3.1.	Treatment-Emergent Adverse Events wi	C64	F:/Biome	c64	Treatment (N= 71)	c64	Placebo (N=100)
Table	t_14_3_1_2	Table 14.3.1.	Treatment-Emergent Adverse Events by	C79	F:/Biome	c79	Treatment (N= 71)	c79	Placebo (N=100)
Table	t_14_3_1_2	Table 14.3.1.	Serious Treatment-Emergent Adverse Ev	C80	F:/Biome	c80	Treatment (N= 71)	c80	Placebo (N=100)

R SHINY

```r
library(shiny)
library(dplyr)
library(readxl)
library(writexl)
library(striprtf)
library(stringr)
```



```r
ui <- fluidPage(
  titlePanel("RTF Population (Big N) Extractor"),
  sidebarLayout(
    sidebarPanel(
      fileInput("folder", "Select Folder Containing RTF Files", multiple = TRUE, accept = ".rtf"),
      textInput("treatment1", "Enter Treatment 1 Name", value = "Treatment"),
      textInput("treatment2", "Enter Treatment 2 Name", value = "Placebo"),
      downloadButton("downloadData", "Download Excel")
    ),
    mainPanel(
      tableOutput("extractedData")
    )
  )
)
```



```r
server <- function(input, output) {
  extract_big_n <- function(file_path, treatment1, treatment2) {
    rtf_text <- read_rtf(file_path)
    extracted_T1 <- na.omit(str_extract(rtf_text, paste0(treatment1, "[^)]*\\)")))
    extracted_T2 <- na.omit(str_extract(rtf_text, paste0(treatment2, "[^)]*\\)")))
    data.frame(File = basename(file_path), T1 = ifelse(length(extracted_T1) > 0, extracted_T1[1], NA),
```

RTF Population (Big N) Extractor

Enter Folder Path Containing RTF Files

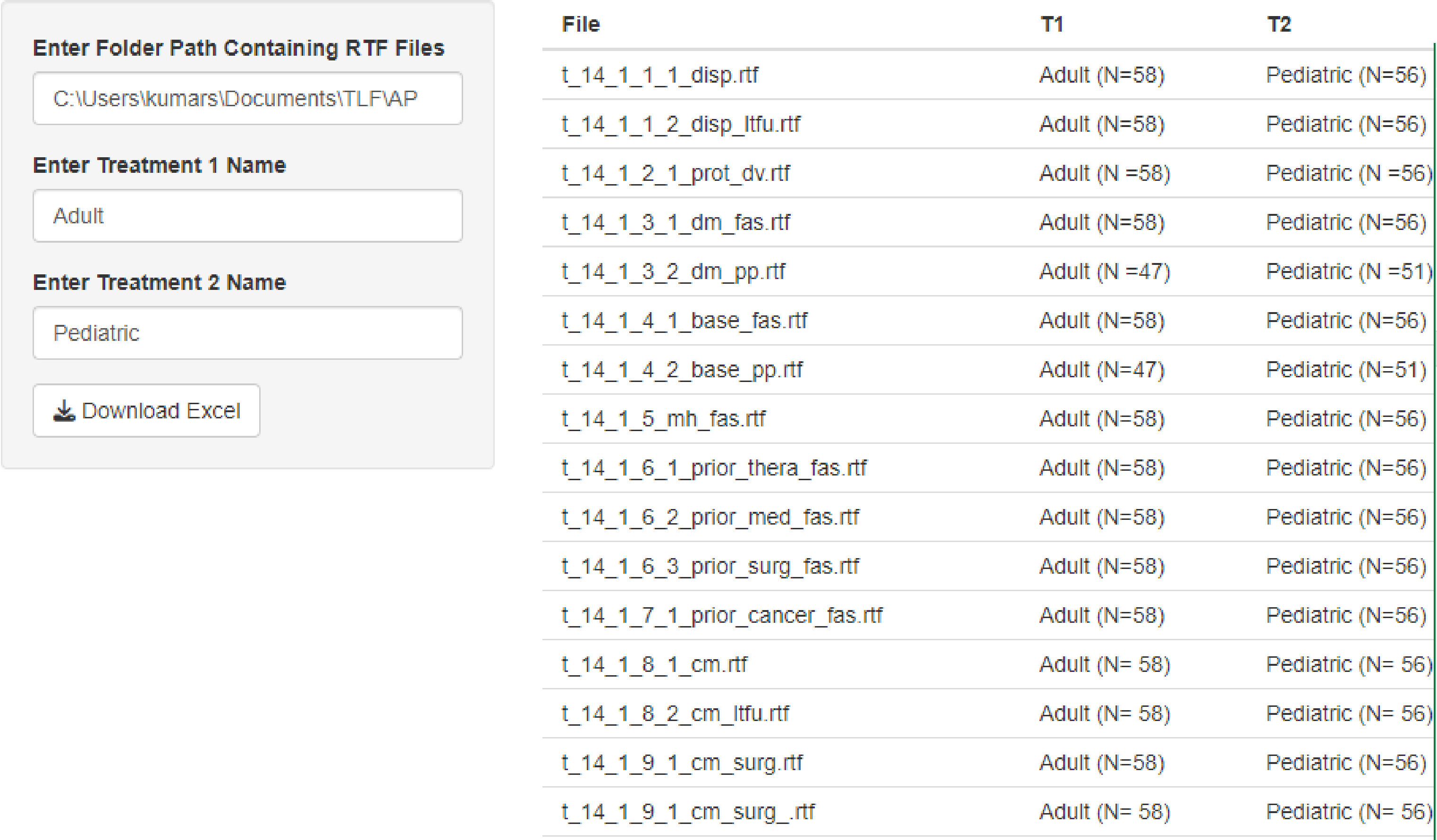
Enter Treatment 1 Name

Treatment

Enter Treatment 2 Name

Placebo

RTF Big N Extractor



Enter Folder Path Containing RTF Files

C:\Users\kumars\Documents\TLF\AP

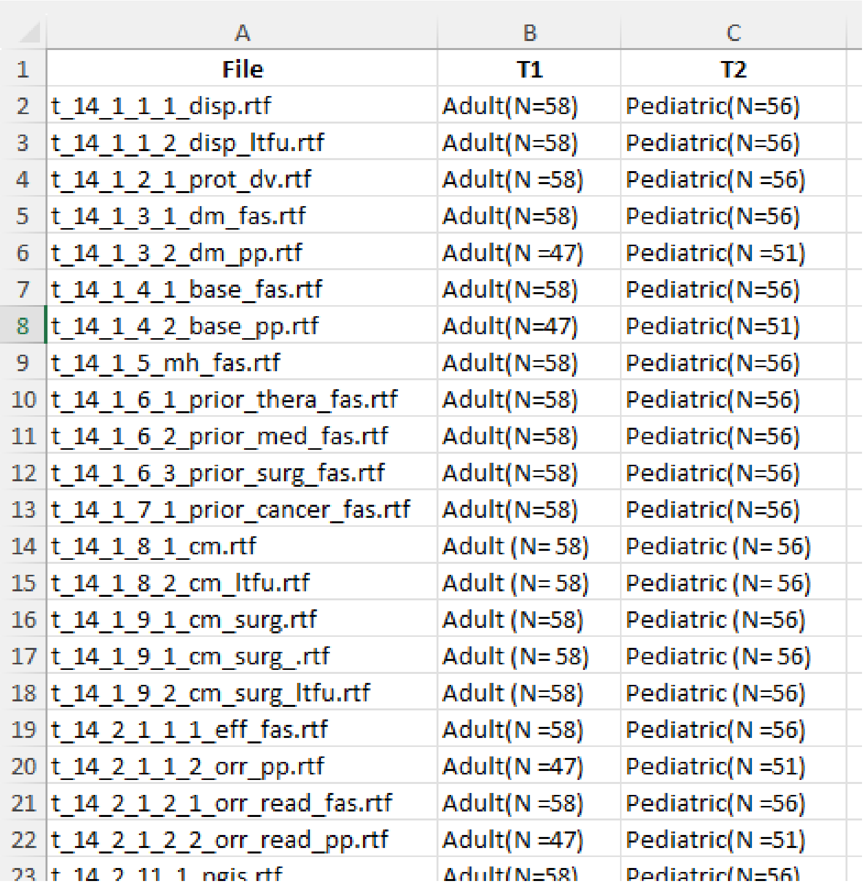
Enter Treatment 1 Name

Adult

Enter Treatment 2 Name

Pediatric

File	T1	T2
t_14_1_1_1_disp.rtf	Adult (N=58)	Pediatric (N=56)
t_14_1_1_2_disp_ltfu.rtf	Adult (N=58)	Pediatric (N=56)
t_14_1_2_1_prot_dv.rtf	Adult (N =58)	Pediatric (N =56)
t_14_1_3_1_dm_fas.rtf	Adult (N=58)	Pediatric (N=56)
t_14_1_3_2_dm_pp.rtf	Adult (N =47)	Pediatric (N =51)
t_14_1_4_1_base_fas.rtf	Adult (N=58)	Pediatric (N=56)
t_14_1_4_2_base_pp.rtf	Adult (N=47)	Pediatric (N=51)
t_14_1_5_mh_fas.rtf	Adult (N=58)	Pediatric (N=56)
t_14_1_6_1_prior_thera_fas.rtf	Adult (N=58)	Pediatric (N=56)
t_14_1_6_2_prior_med_fas.rtf	Adult (N=58)	Pediatric (N=56)
t_14_1_6_3_prior_surg_fas.rtf	Adult (N=58)	Pediatric (N=56)
t_14_1_7_1_prior_cancer_fas.rtf	Adult (N=58)	Pediatric (N=56)
t_14_1_8_1_cm.rtf	Adult (N= 58)	Pediatric (N= 56)
t_14_1_8_2_cm_ltfu.rtf	Adult (N= 58)	Pediatric (N= 56)
t_14_1_9_1_cm_surg.rtf	Adult (N=58)	Pediatric (N=56)
t_14_1_9_1_cm_surg_.rtf	Adult (N= 58)	Pediatric (N= 56)



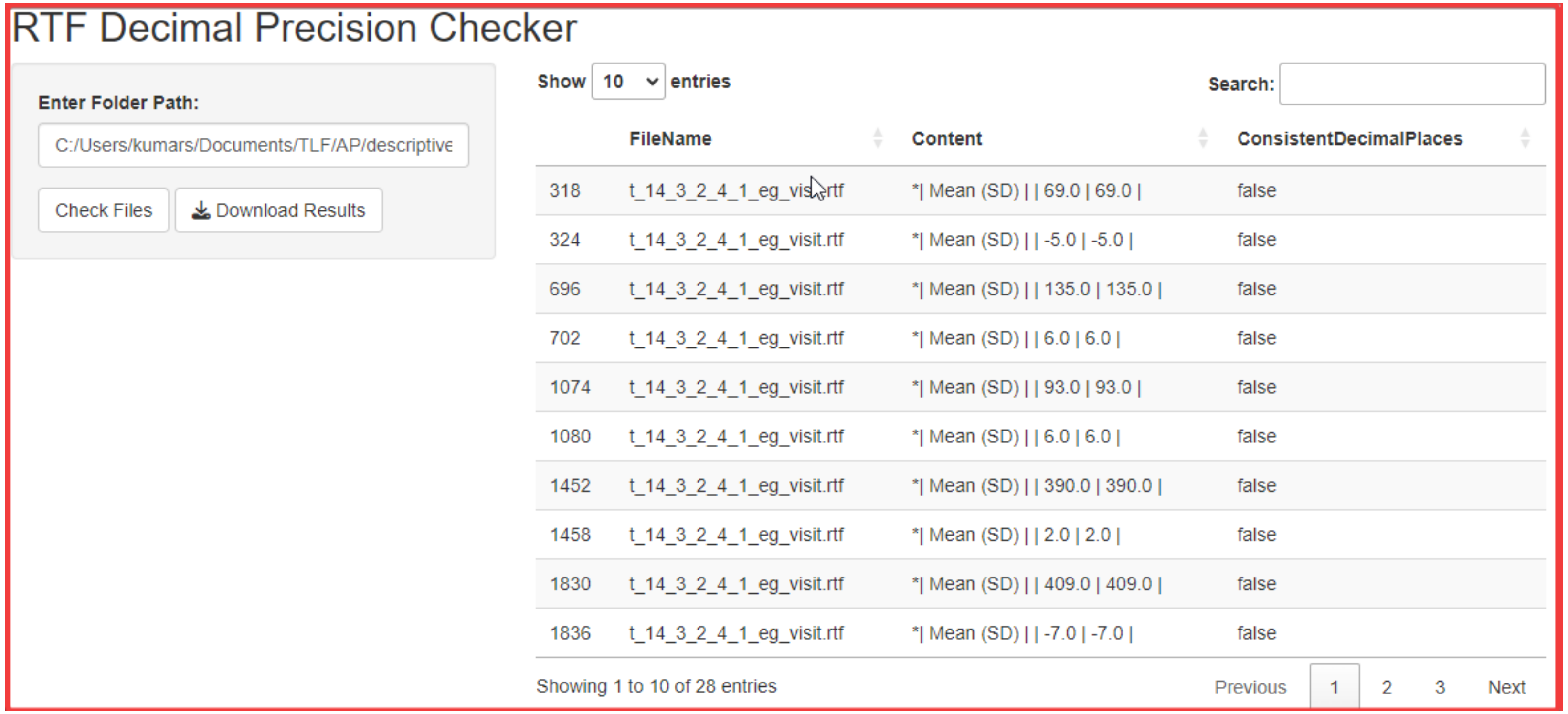
	A	B	C
1	**File**	**T1**	**T2**
2	t_14_1_1_1_disp.rtf	Adult(N=58)	Pediatric(N=56)
3	t_14_1_1_2_disp_ltfu.rtf	Adult(N=58)	Pediatric(N=56)
4	t_14_1_2_1_prot_dv.rtf	Adult(N =58)	Pediatric(N =56)
5	t_14_1_3_1_dm_fas.rtf	Adult(N=58)	Pediatric(N=56)
6	t_14_1_3_2_dm_pp.rtf	Adult(N =47)	Pediatric(N =51)
7	t_14_1_4_1_base_fas.rtf	Adult(N=58)	Pediatric(N=56)
8	t_14_1_4_2_base_pp.rtf	Adult(N=47)	Pediatric(N=51)
9	t_14_1_5_mh_fas.rtf	Adult(N=58)	Pediatric(N=56)
10	t_14_1_6_1_prior_thera_fas.rtf	Adult(N=58)	Pediatric(N=56)
11	t_14_1_6_2_prior_med_fas.rtf	Adult(N=58)	Pediatric(N=56)
12	t_14_1_6_3_prior_surg_fas.rtf	Adult(N=58)	Pediatric(N=56)
13	t_14_1_7_1_prior_cancer_fas.rtf	Adult(N=58)	Pediatric(N=56)
14	t_14_1_8_1_cm.rtf	Adult (N= 58)	Pediatric (N= 56)
15	t_14_1_8_2_cm_ltfu.rtf	Adult (N= 58)	Pediatric (N= 56)
16	t_14_1_9_1_cm_surg.rtf	Adult (N=58)	Pediatric (N=56)
17	t_14_1_9_1_cm_surg_.rtf	Adult (N= 58)	Pediatric (N= 56)
18	t_14_1_9_2_cm_surg_ltfu.rtf	Adult (N=58)	Pediatric (N=56)
19	t_14_2_1_1_1_eff_fas.rtf	Adult(N =58)	Pediatric(N =56)
20	t_14_2_1_1_2_orr_pp.rtf	Adult(N =47)	Pediatric(N =51)
21	t_14_2_1_2_1_orr_read_fas.rtf	Adult(N =58)	Pediatric(N =56)
22	t_14_2_1_2_2_orr_read_pp.rtf	Adult(N =47)	Pediatric(N =51)
23	t_14_2_11_1_ngis.rtf	Adult(N=58)	Pediatric(N=56)

SCENARIO 2 – DECIMAL PRECISION

COUNTS

- Ensuring decimal consistency across all outputs, as defined in the SAP, is essential.
- The lead programmer is responsible for verifying that all outputs adhere to the specified decimal precision.
- The program captures actual decimal precision in outputs, compares it against SAP-defined precision, and flags discrepancies.
- The R package Striprtf is used to read RTF files, with further processing using stringr, dplyr, dt, and Shiny.
- The Shiny interface allows users to input file locations and review flagged inconsistencies efficiently.

striprtf → stringr → dplyr → dt → writexl → shiny



RTF Decimal Precision Checker

Enter Folder Path: C:/Users/kumars/Documents/TLF/AP/descriptive

<table>
<tr><th></th><th>FileName</th><th>Content</th><th>ConsistentDecimalPlaces</th></tr>
<tr><td>318</td><td>t_14_3_2_4_1_eg_vis rtf</td><td>*| Mean (SD) | | 69.0 | 69.0 |</td><td>false</td></tr>
<tr><td>324</td><td>t_14_3_2_4_1_eg_visit.rtf</td><td>*| Mean (SD) | | -5.0 | -5.0 |</td><td>false</td></tr>
<tr><td>696</td><td>t_14_3_2_4_1_eg_visit.rtf</td><td>*| Mean (SD) | | 135.0 | 135.0 |</td><td>false</td></tr>
<tr><td>702</td><td>t_14_3_2_4_1_eg_visit.rtf</td><td>*| Mean (SD) | | 6.0 | 6.0 |</td><td>false</td></tr>
<tr><td>1074</td><td>t_14_3_2_4_1_eg_visit.rtf</td><td>*| Mean (SD) | | 93.0 | 93.0 |</td><td>false</td></tr>
<tr><td>1080</td><td>t_14_3_2_4_1_eg_visit.rtf</td><td>*| Mean (SD) | | 6.0 | 6.0 |</td><td>false</td></tr>
<tr><td>1452</td><td>t_14_3_2_4_1_eg_visit.rtf</td><td>*| Mean (SD) | | 390.0 | 390.0 |</td><td>false</td></tr>
<tr><td>1458</td><td>t_14_3_2_4_1_eg_visit.rtf</td><td>*| Mean (SD) | | 2.0 | 2.0 |</td><td>false</td></tr>
<tr><td>1830</td><td>t_14_3_2_4_1_eg_visit.rtf</td><td>*| Mean (SD) | | 409.0 | 409.0 |</td><td>false</td></tr>
<tr><td>1836</td><td>t_14_3_2_4_1_eg_visit.rtf</td><td>*| Mean (SD) | | -7.0 | -7.0 |</td><td>false</td></tr>
</table>

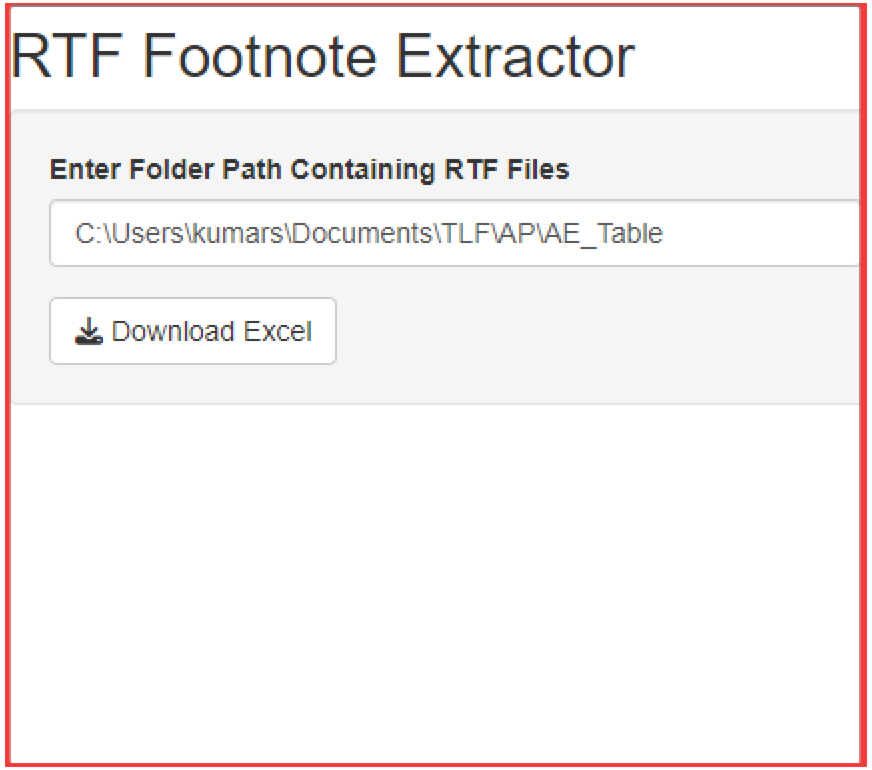
Showing 1 to 10 of 28 entries

SCENARIO 3 – TITLE & FOOTNOTES

TLF METADATA

- Ensuring that titles and footnotes in RTF outputs are consistent, accurate and error-free
- Identifying discrepancies such as unresolved macros, spell checks, formatting inconsistencies are key during manual review.
- additional R package hunspell used
- Parsing Data – Extract and parse data from the RTF files for further analysis.


striprtf → stringr → dplyr → hunspell → shiny → writexl

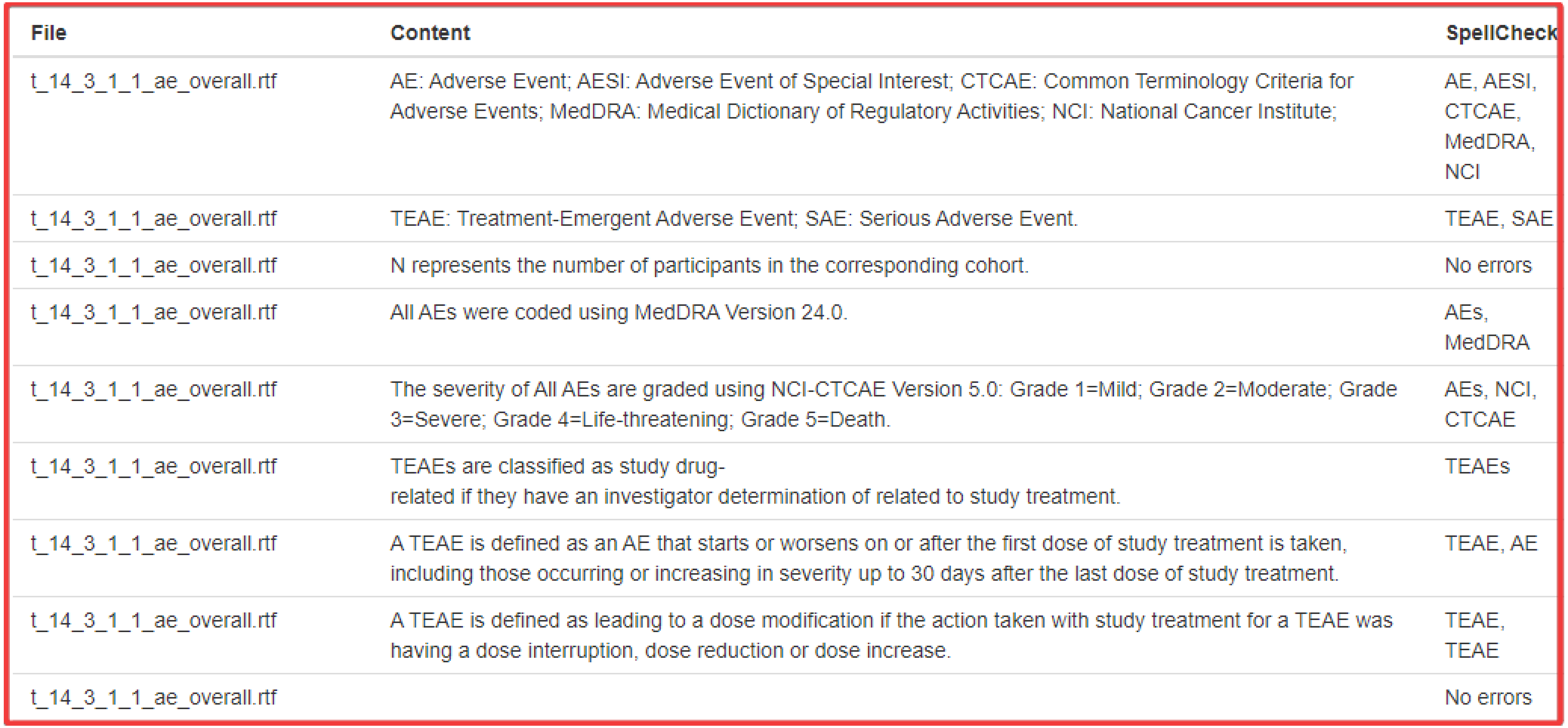

RTF Footnote Extractor

Enter Folder Path Containing RTF Files

C:\Users\kumars\Documents\TLF\AP\AE_Table

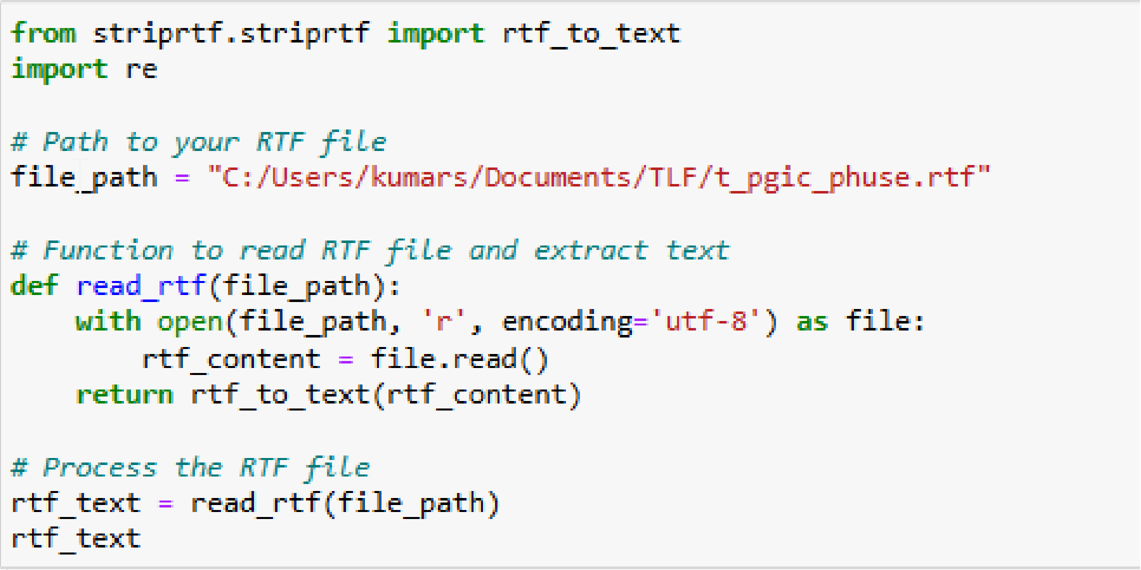
SCENARIO 3 – TITLE & FOOTNOTES

striprtf → stringr → dplyr → hunspell → shiny → writexl

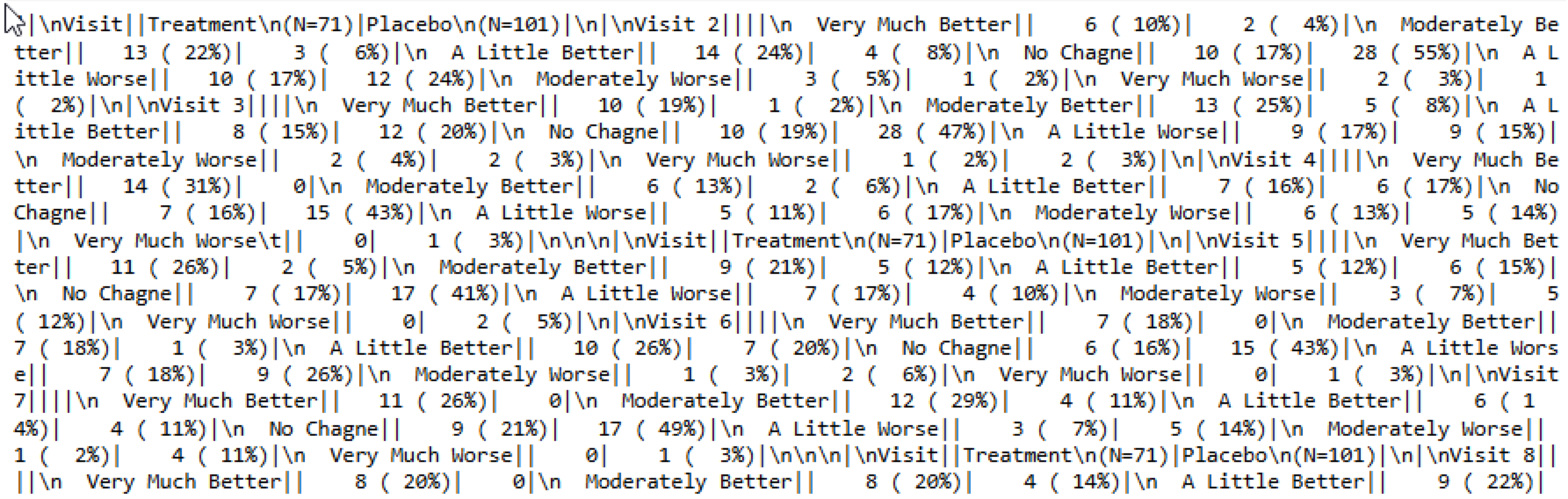


File	Content	SpellCheck
t_14_3_1_1_ae_overall.rtf	AE: Adverse Event; AESI: Adverse Event of Special Interest; CTCAE: Common Terminology Criteria for Adverse Events; MedDRA: Medical Dictionary of Regulatory Activities; NCI: National Cancer Institute;	AE, AESI, CTCAE, MedDRA, NCI
t_14_3_1_1_ae_overall.rtf	TEAE: Treatment-Emergent Adverse Event; SAE: Serious Adverse Event.	TEAE, SAE
t_14_3_1_1_ae_overall.rtf	N represents the number of participants in the corresponding cohort.	No errors
t_14_3_1_1_ae_overall.rtf	All AEs were coded using MedDRA Version 24.0.	AEs, MedDRA
t_14_3_1_1_ae_overall.rtf	The severity of All AEs are graded using NCI-CTCAE Version 5.0: Grade 1=Mild; Grade 2=Moderate; Grade 3=Severe; Grade 4=Life-threatening; Grade 5=Death.	AEs, NCI, CTCAE
t_14_3_1_1_ae_overall.rtf	TEAEs are classified as study drug-related if they have an investigator determination of related to study treatment.	TEAEs
t_14_3_1_1_ae_overall.rtf	A TEAE is defined as an AE that starts or worsens on or after the first dose of study treatment is taken, including those occurring or increasing in severity up to 30 days after the last dose of study treatment.	TEAE, AE
t_14_3_1_1_ae_overall.rtf	A TEAE is defined as leading to a dose modification if the action taken with study treatment for a TEAE was having a dose interruption, dose reduction or dose increase.	TEAE, TEAE
t_14_3_1_1_ae_overall.rtf		No errors

PYTHON PACKAGE STRIPRTF & HUNSPELL



```python
from striprtf.striprtf import rtf_to_text
import re

# Path to your RTF file
file_path = "C:/Users/kumars/Documents/TLF/t_pgic_phuse.rtf"

# Function to read RTF file and extract text
def read_rtf(file_path):
    with open(file_path, 'r', encoding='utf-8') as file:
        rtf_content = file.read()
    return rtf_to_text(rtf_content)

# Process the RTF file
rtf_text = read_rtf(file_path)
rtf_text
```



```
|\nVisit||Treatment\n(N=71)|Placebo\n(N=101)|\n|\nVisit 2||||\n  Very Much Better||    6 ( 10%)|    2 (  4%)|\n  Moderately Be
tter||   13 ( 22%)|    3 (  6%)|\n  A Little Better||   14 ( 24%)|    4 (  8%)|\n  No Chagne||   10 ( 17%)|   28 ( 55%)|\n  A L
ittle Worse||   10 ( 17%)|   12 ( 24%)|\n  Moderately Worse||    3 (  5%)|    1 (  2%)|\n  Very Much Worse||    2 (  3%)|    1
(  2%)|\n|\nVisit 3||||\n  Very Much Better||   10 ( 19%)|    1 (  2%)|\n  Moderately Better||   13 ( 25%)|    5 (  8%)|\n  A L
ittle Better||    8 ( 15%)|   12 ( 20%)|\n  No Chagne||   10 ( 19%)|   28 ( 47%)|\n  A Little Worse||    9 ( 17%)|    9 ( 15%)|
\n  Moderately Worse||    2 (  4%)|    2 (  3%)|\n  Very Much Worse||    1 (  2%)|    2 (  3%)|\n|\nVisit 4||||\n  Very Much Be
tter||   14 ( 31%)|    0|\n  Moderately Better||    6 ( 13%)|    2 (  6%)|\n  A Little Better||    7 ( 16%)|    6 ( 17%)|\n  No
Chagne||    7 ( 16%)|   15 ( 43%)|\n  A Little Worse||    5 ( 11%)|    6 ( 17%)|\n  Moderately Worse||    6 ( 13%)|    5 ( 14%)
|\n  Very Much Worse\t||    0|    1 (  3%)|\n\n\n|\nVisit||Treatment\n(N=71)|Placebo\n(N=101)|\n|\nVisit 5||||\n  Very Much Bet
ter||   11 ( 26%)|    2 (  5%)|\n  Moderately Better||    9 ( 21%)|    5 ( 12%)|\n  A Little Better||    5 ( 12%)|    6 ( 15%)|
\n  No Chagne||    7 ( 17%)|   17 ( 41%)|\n  A Little Worse||    7 ( 17%)|    4 ( 10%)|\n  Moderately Worse||    3 (  7%)|    5
( 12%)|\n  Very Much Worse||    0|    2 (  5%)|\n|\nVisit 6||||\n  Very Much Better||    7 ( 18%)|    0|\n  Moderately Better||
7 ( 18%)|    1 (  3%)|\n  A Little Better||   10 ( 26%)|    7 ( 20%)|\n  No Chagne||    6 ( 16%)|   15 ( 43%)|\n  A Little Wors
e||    7 ( 18%)|    9 ( 26%)|\n  Moderately Worse||    1 (  3%)|    2 (  6%)|\n  Very Much Worse||    0|    1 (  3%)|\n|\nVisit
7||||\n  Very Much Better||   11 ( 26%)|    0|\n  Moderately Better||   12 ( 29%)|    4 ( 11%)|\n  A Little Better||    6 ( 1
4%)|    4 ( 11%)|\n  No Chagne||    9 ( 21%)|   17 ( 49%)|\n  A Little Worse||    3 (  7%)|    5 ( 14%)|\n  Moderately Worse||
1 (  2%)|    4 ( 11%)|\n  Very Much Worse||    0|    1 (  3%)|\n\n\n|\nVisit||Treatment\n(N=71)|Placebo\n(N=101)|\n|\nVisit 8||
||\n  Very Much Better||    8 ( 20%)|    0|\n  Moderately Better||    8 ( 20%)|    4 ( 14%)|\n  A Little Better||    9 ( 22%)|
```

- Striprtf and hunspell can also be used as python packages.
- Python packages provides capabilities to output to SAS datasets
- Python packages reads rtf and pdf files efficiently

SUMMARY

Targeted QC – Efficiently validating key parameters during review and ensuring the accuracy and reliability of TLF outputs.

Generative AI-powered app could further enhance efficiency. This process is highly beneficial for sponsors seeking to perform a final sanity check before review and submission. It also serves as a valuable safeguard for CROs, allowing them to conduct a final verification before delivering outputs to sponsors.

Within a statistical programming team, this represents the last quality control checkpoint before the data is shared for cross-functional review, **reinforcing confidence in the integrity of the results**.

FINAL QUALITY CHECK: ENHANCING ACCURACY, CONFIDENCE, AND EFFICIENCY IN TLF REVIEW

ENSURING ACCURACY AND RELIABILITY IN TERTIARY QC

Validating key parameters is critical for maintaining the integrity of TLF outputs

FINAL VERIFICATION FOR SPONSORS AND CROs

A crucial safeguard before submission, enhancing confidence in deliverables

OPTIMIZING QUALITY CONTROL WITH REVIEW TOOLS

Strengthening data integrity while improving efficiency and reducing errors

POWERING POSSIBILITIES

IN DRUG DEVELOPMENT

Navitas is a knowledge-driven, tech-enabled CRO offering bespoke solutions across the drug development life cycle. From strategic advisory services that navigate complex regulatory landscapes to clinical development ensuring compliant, data-driven operations, and post-marketing solutions enhancing product value with safety surveillance, life cycle management, and real-world evidence—we integrate deep expertise, cutting-edge technology, and industry insights to drive success.

Siddharth Kumar

Director – Clinical Programming Solutions
E: siddharth.kumar@navitaslifesciences.com

Website

LinkedIn