Paper SS02

Combine multiple R outputs into one HTML file with the bookmarks

Eric Zhang, Janssen Pharmaceuticals, Pennsylvania, USA
Baoqing Li, Janssen Pharmaceuticals, Pennsylvania, USA

ABSTRACT

Documents in HTML format are popular in pharmaceutical companies. Statistical programmers might wish to combine multiple tables, listings and/or figures in RTF or HTML generated by R into one HTML file for reviews, topline reports, submissions, or other purposes. R markdown is one good programming language to make reports in high quality document. We create R programs to generate the R markdown scripts, which achieve the combination of multiple R outputs with bookmarks into one HTML file. The generated R markdown scripts also can be edited by statisticians as needed before combining multiple R outputs. Then we run the R markdown scripts to generate the HTML file. The paper describes the method, capabilities, features, and usage of the combination of multiple R outputs.

R OBJECTS

A R program usually contains three sections: the data manipulations, the creation of R object, which presents table or figure, with R package, and the generation of output from R object with R function. The data manipulation arranges data structure, selects variables and records, performs statistics calculations, like frequencies and summaries. R packages are used to create R objects, like huxtable or ggplot2. R output functions are used to generate outputs with the format of RTF, HTML or PDF, like quick_rtf, quick_html, quick_pdf. For example, the quick-output functions generate the outputs with R objects, which are created by the R package huxtable. We save R objects at R object files in one place and use them to generate R markdown files. The following R codes are used to save R object at R object files in the subfolder Objects. One output file has one R object saved at one R object file.

```
pgmobj <- as.name(pgmname)
eval(call("<-", pgmobj, tbl))
eval(call("save", pgmobj, file=paste("Objects/", pgmname, ".Rds", sep="")))
```

In the above sample, pgname is the vector with the program name, tbl is the R object presenting table or figure, the file with the extension Rds is the R object file. For example, the R object LSIDM1 is saved at the R object file LSIDM1.Rds.

R MARKDOWN SCRIPTS

Markdown is a popular programming language. R markdown is developed in R and can be used to document R outputs. A R markdown file contains one header, tabs, and chunks. The following codes are a R markdown header, which generates a HTML file.

```
---
title: 'Table of Contents'
output:
  html_document:
    toc: true
    toc_float: true
    toc_collapsed: true
    toc_depth: 2
    number_sections: true
    theme: united
    highlight: tango
    fig_width: 9
    fig_height: 6
    css: 'MYSTYLE.css'
---
```

The sections at the R markdown file are handled by tabs with the symbols # and/or ## and/or others. The symbol # at the tab presents the sections and the symbol ## presents the tables, listings or figures at the following sample. The sample has three sections: Subject Information, Efficacy, and Safety. Each section has tables, listings, or figures. For example, the section Subject Information has one listing, the section Efficacy has two tables, and the section Safety has one figure. The chunk starts

with ```{r} and ends with ```. At the sample below, LSIDM1, TEFRESP1A, TEFPFS1A and GSFLAB1A at the chunks are the R objects, which are stored in R object files LSIDM1.Rds, TEFRESP1A.Rds, TEFPFS1A.Rds and GSFLAB1A.Rds, respectively. The following codes load the R objects from R object files and put the R objects at the chunks.

````
```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
```

```{r library, include=FALSE}
library(huxtable)
```

```{r object files, include=FALSE}
load('../Objects/GEFLAB1A.Rds')
load('../Objects/LSIDM1.Rds')
load('../Objects/TEFRESP1A.Rds')
load('../Objects/TEFPFS1A.Rds')
```

# - Subject Information
## - LSIDM1: Listing of Demographics and Disease Characteristics at Enrollment; All Enrolled Analysis Set
```{r, echo=FALSE}
LSIDM1
```

# - Efficacy
## - TEFRESP1A: Overall Best Response Based on Independent Review Committee (IRC) Assessment; All Treated Analysis Set
```{r, echo=FALSE}
TEFRESP1A
```

## - TEFPFS1A: Progression-Free Survival Based on Independent Review Committee (IRC) Assessment; All Treated Analysis Set
```{r, echo=FALSE}
TEFPFS1A
```

# - Safety
## - GSFLAB1A: Mean (+/-SE) and Median Hemoglobin (g/L) Over Time; All Treated Analysis Set
```{r, echo=FALSE}
GSFLAB1A
```
````

COMBINATION OF R OUTPUTS

If we would like to put multiple R outputs into one R markdown document, i.e., combine multiple R outputs into one HTML file, it's better to develop a R program, which puts multiple R objects of R outputs at one R markdown file. We create an Excel file, which stores the information we need, like the sections, orders, titles, and names of tables, listing and figures. We create a R program, which gets the records at the Excel file and R objects at R object files and put all of them at a R markdown file. The following codes create the R markdown scripts, which load R objects from R objects files in the subfolder Objects:

```
cat("```{r object files, include=FALSE}\n")
objectfiles <- list.files(path = 'Objects', pattern='.Rds')
for (val in objectfiles) {
  batchobjectfile <- paste('../Objects/', val, sep='')
  cat("load('", eval(batchobjectfile), "')\n", sep="")
}
cat("```\n")
```

The following codes create the R markdown scripts, which get the records at the Excel file TOC.xlsx in the subfolder Utility:

```
objects <- read_excel("Utility/TOC.xlsx")
objects$ROWN = 1:nrow(objects)
objects <- mutate(objects,
                  OBJECTFILE=paste(OBJECTS, '.Rds', sep=''))
objects <- mutate(objects,
                  SECTIONOBJECT=ifelse(!is.na(SECTION),
                                       paste("#  -  ",  SECTION,  ",",  OBJECTS,  ":  ",  TITLES,
sep=""),
                                       paste(",", OBJECTS, ": ", TITLES, sep="")))
objects <- objects[order(objects$OBJECTFILE),]
```

The following codes create the R markdown scripts, which put each R object at each chunk and put the chunks under the tabs of the sections with the orders.

```
objectfiles <- list.files(path = 'Objects', pattern='.Rds')
objectdf  <- as.data.frame(objectfiles)
colnames(objectdf) <- "OBJECTFILE"

objects <- group_by(objects, OBJECTFILE)
objectdf <- group_by(objectdf, OBJECTFILE)
objectdf <- merge(objects, objectdf, by="OBJECTFILE")
objectdf <- objectdf[order(objectdf$ROWN),]
objectfiles  <- as.vector(objectdf$SECTIONOBJECT)

for (val in objectfiles) {
  sectionname <- substr(val, 1, str_locate_all(pattern = ',', val)[[1]][1] - 1)
  objectname  <- substr(val, str_locate_all(pattern = ',', val)[[1]][1] + 1,
                        str_locate_all(pattern = ':', val)[[1]][1] - 1)
  objecttitlename  <- substr(val, str_locate_all(pattern = ',', val)[[1]][1] + 1, 10000)
  cat("\n")
  cat(eval(sectionname), "\n", sep="")
  cat("## - ", eval(objecttitlename), "\n", sep="")
  cat("```{r, echo=FALSE}\n")
  cat(eval(objectname), "\n", sep="")
  cat("```\n")
}
```

We name the R program and the R markdown file as GENTOC.R and TOC.Rmd. GENTOC.R gets R objects at the R object files in the subfolder Objects, gets the records at the Excel file TOC.xlsx in the subfolder UTILITY, creates the R markdown scripts, and saves the R markdown scripts at the R markdown file TOC.Rmd.

R MARKDOWN DOCUMENT

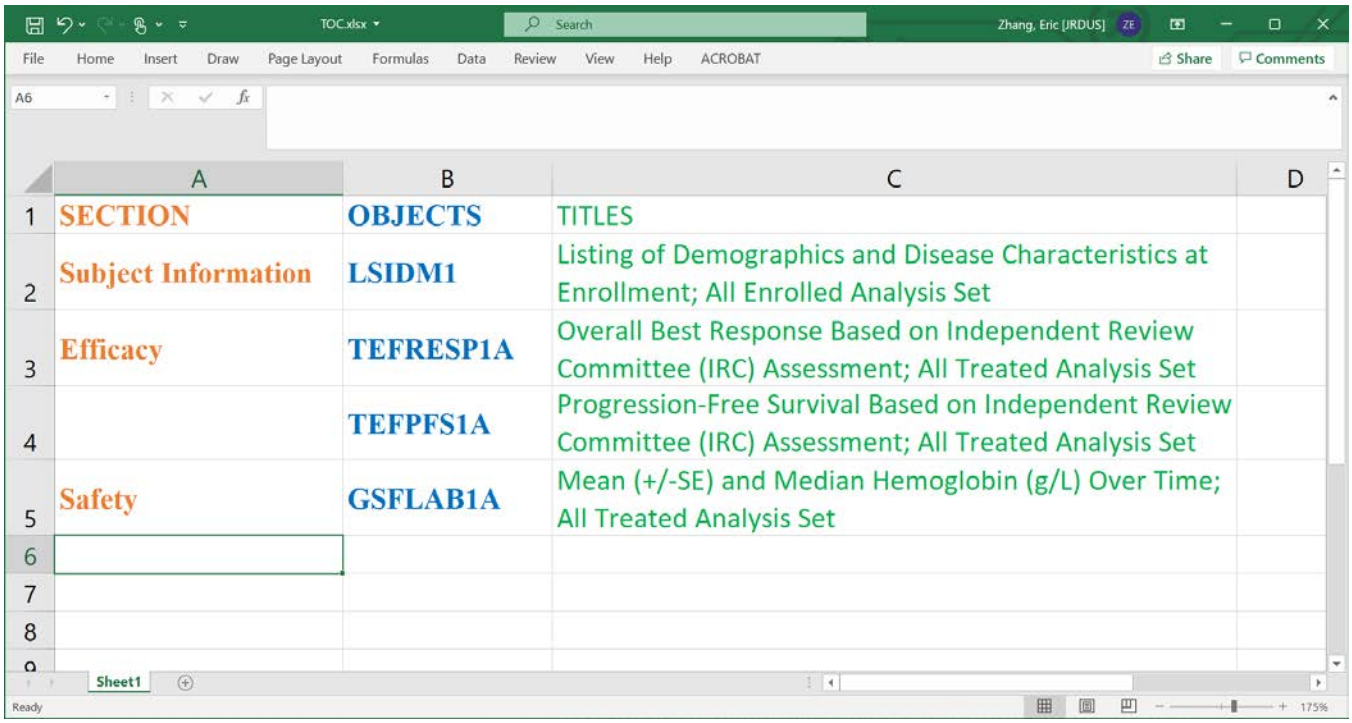
We create another R program TOC.R, which executes the R markdown scripts at the R markdown file TOC.Rmd and generates the R markdown document, i.e., one HTML file TOC.html. TOC.html contains multiple R outputs with the bookmarks. R program TOC.R has the codes below:

```
rmarkdown::render('Utility/TOC.Rmd', output_file = '../Utility/TOC.html')
```

We also extended the capability. The statisticians can add the descriptions/explanations of table, listing and figure at the column ‘DESCRIPTIONS’ next to the column ‘TITLES’ at TOC.xlsx. These descriptions/explanation will be displayed under tables, listings and figures at the HTML file accordingly.

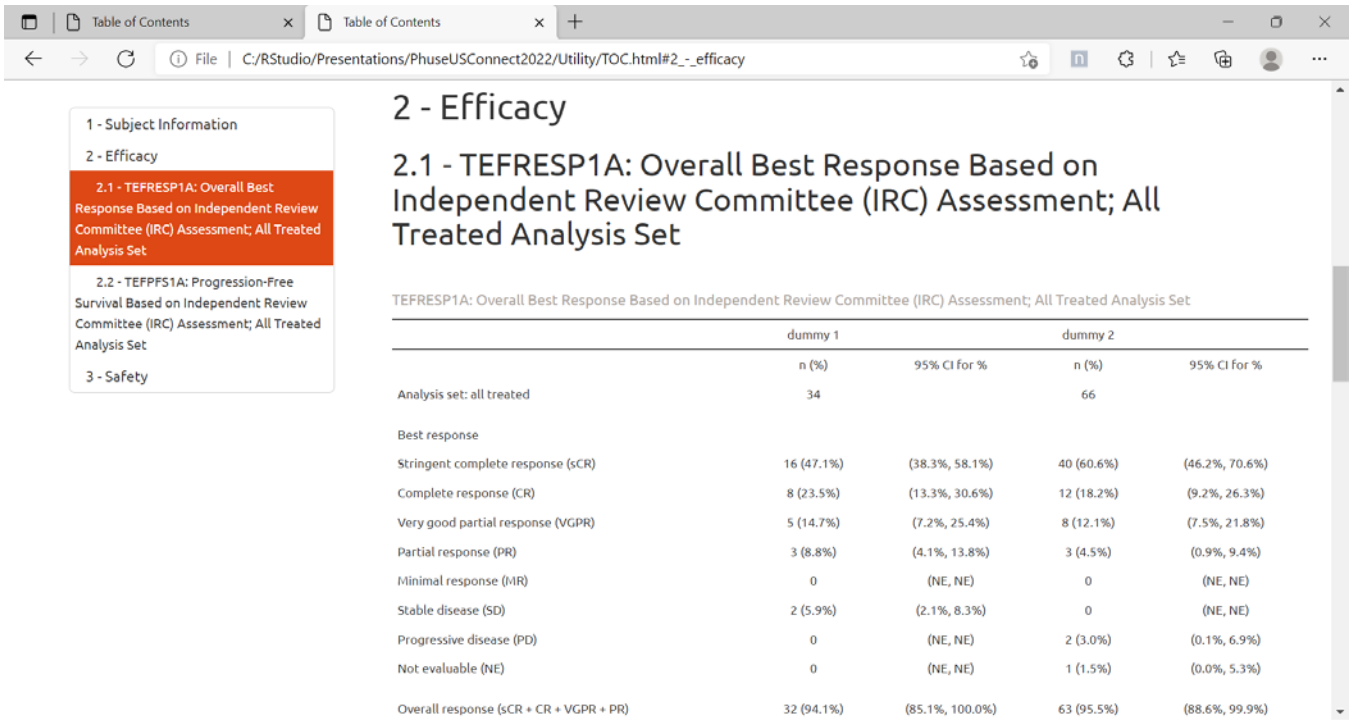
REFERENCE

The screenshot of the sample of the Excel sheet TOC.xlsx is below:



	A	B	C	D
1	SECTION	OBJECTS	TITLES	
2	Subject Information	LSIDM1	Listing of Demographics and Disease Characteristics at Enrollment; All Enrolled Analysis Set	
3	Efficacy	TEFRESP1A	Overall Best Response Based on Independent Review Committee (IRC) Assessment; All Treated Analysis Set	
4		TEFPFS1A	Progression-Free Survival Based on Independent Review Committee (IRC) Assessment; All Treated Analysis Set	
5	Safety	GSFLAB1A	Mean (+/-SE) and Median Hemoglobin (g/L) Over Time; All Treated Analysis Set	
6				
7				
8				

The screenshot of the sample of R markdown document TOC.html, which combines multiple R outputs with the bookmarks, is below:



	dummy 1		dummy 2	
	n (%)	95% CI for %	n (%)	95% CI for %
Analysis set: all treated	34		66	
Best response				
Stringent complete response (sCR)	16 (47.1%)	(38.3%, 58.1%)	40 (60.6%)	(46.2%, 70.6%)
Complete response (CR)	8 (23.5%)	(13.3%, 30.6%)	12 (18.2%)	(9.2%, 26.3%)
Very good partial response (VGPR)	5 (14.7%)	(7.2%, 25.4%)	8 (12.1%)	(7.5%, 21.8%)
Partial response (PR)	3 (8.8%)	(4.1%, 13.8%)	3 (4.5%)	(0.9%, 9.4%)
Minimal response (MR)	0	(NE, NE)	0	(NE, NE)
Stable disease (SD)	2 (5.9%)	(2.1%, 8.3%)	0	(NE, NE)
Progressive disease (PD)	0	(NE, NE)	2 (3.0%)	(0.1%, 6.9%)
Not evaluable (NE)	0	(NE, NE)	1 (1.5%)	(0.0%, 5.3%)
Overall response (sCR + CR + VGPR + PR)	32 (94.1%)	(85.1%, 100.0%)	63 (95.5%)	(88.6%, 99.9%)

ACKNOWLEDGEMENT

The authors would like to thank Nicholas Masel and Sheng-Wei Wang, who reviewed the ideas and gave the comments during the development of the programs. The authors also would like to thank people, who supported the development of the programs.

CONTACT INFORMATION

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

Eric Zhang
Janssen Pharmaceuticals
1400 McKean Road
Spring House, PA 19477
USA
E-mail: ezhang5@its.jnj.com

Baoqing Li
Janssen Pharmaceuticals
1400 McKean Road
Spring House, PA 19477
USA
E-mail: bli6@its.jnj.com