Paper TT15

PhUSE US Connect 2020

Good Cooperation Between SAS and R for Regulatory Submission

Jingyuan Chen, Hoffman-La Roche Limited, Mississauga, Canada

ABSTRACT

Recently there has been a trend for using R ™ in pharmaceutical industry, though SAS ™ has been a dominant software over the years. In this paper, I am going to illustrate how SAS and R can complement each other for creating health authority submission package. R was involved in preparing for eSUB Module 5 in one of the submission I worked on before. This approved that R can make the products with submission ready quality. Additionally, making use of the advantages from each tool can provide an excellent work for Health Authority's requests. Recently, I was involved in an EMA request for exploratory analysis, where SAS was used to implement the part of step-wise modelling selection and R markdown was used for the rest part of exploratory analysis. Therefore, detecting the right tool for right situation can provide us the good quality, clear documentation and efficient manner for submission.

1. BACKGROUND

There are more and more featured software and tools for implementing data analysis and it can be both benefit and challenge for analyst to pick the right tool for right situation. SAS with stable, verified and powerful data analyzing procedures dominants the work of submission to Health Authorizes in Pharma Industry. Meanwhile, open source becomes noticeable for its free, fast, flexible and equally powerful data analyzing features. With the contribution from users from all over the world, open source is abstracting more and more attention from people. Being aware of the changes and chances in industry, I would like to make use of the advantages from both to improve the efficiency of my submission work to Health Authorities.

However, how much contribution does open source can make to pharma industry? How much effort it would take to do standard submission? To get the answers, I decided to involve open source in the real case to my routine work.

2. INTRODUCTION

In this paper, I will use real case to illustrate that R can provide submission ready standard products, and making use of SAS and R can provide an efficient way to implement and document exploratory analysis. One of the case I want to share is to involve R in eSUB module 5 for an oncology study post marketing submission. The other case is to use R markdown along with SAS to execute and document the exploratory request from EMA.

3. ILLUSTRATION

3.1 INVOLVE R FOR ESUB MODULE 5

3.1.1 BRIEF:

The task we worked on is to prepare materials for e-submission to FDA second post marketing committee for a phase III oncology study with 516 patients. SAS contributed the majority part of Clinical Study Reports (CSR). To pilot with R, we generated two types of graphs, one table and one dataset by R. Both Clinical Study Reports and datasets are in a standard format that comply with FDA and our internal company rules. We got the reply from FDA of clear submission in the next year.

3.1.2 WORKING ENVIRONMENT FOR R:

For our analysis, we used RStudio™ Professional, version 0.98.1091, hosted on an internal R server at Roche. The RStudio Professional IDE is integrated with the file systems in our Biometrics Computing Environment over a NFS import. The underlying server runs Red Hat Enterprise Linux 6.3, Intel E5-2670. It has 8 cores with a RAM size of 64GB. All R packages installed on the server went through a standard vetting process by the Systems Development and Support Group, within Roche Biometrics. The version of R used for this submission was 3.1.2

3.1.3WORKING PROCESS FOR R:

There is no pre-defined functions from other resource except for reading in SAS dataset; the analysis done by R and was programed from scratch. Figure 1 shows the workflow. Analysis started with reading in sas7bdat dataset, and we used the functions from BASE to derive variables and data manipulation. Analysis results were in ASCII or rtf format and handover for medical writer colleagues for further procedures.

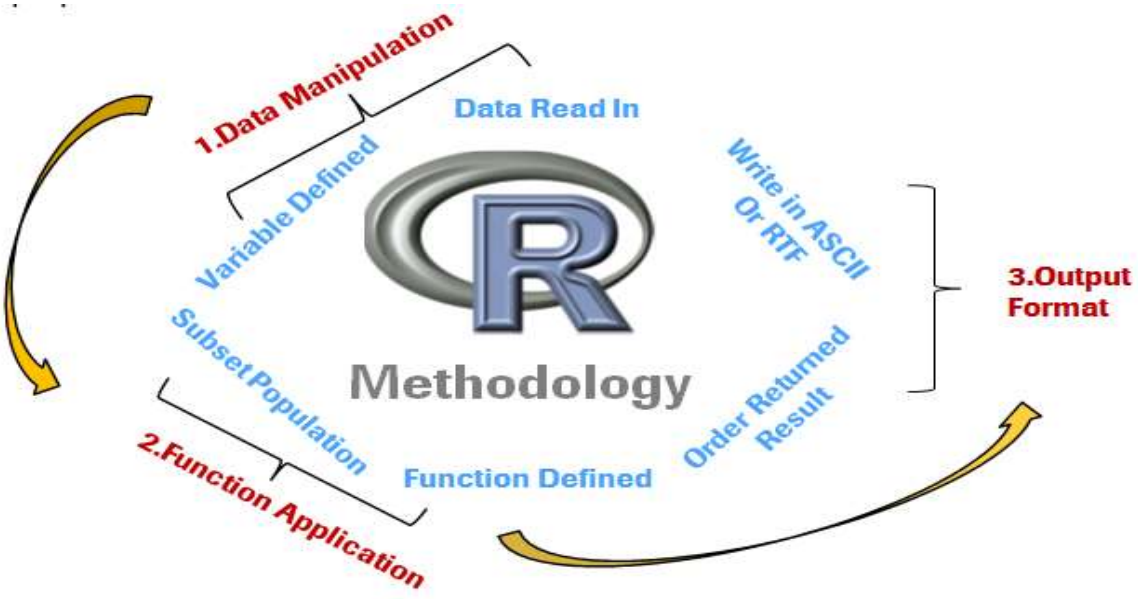

Figure 1: R Process workflow

Double programming in SAS follow standards from Roche as the way for quality control. For outputs with high risk, we compared the numbers calculated from R and SAS and got consistent results; for dataset, we used procedure of proc compare in SAS to compare the dataset generated from R and SAS, and did not find any discrepancy.

3.1.4 TASKS MADE BY R AND COMPARISON WITH SAS OUTPUT

Kaplan Meier Plot:

Figure 2 was generated by R, and Figure 3 was made by SAS. When compare these two figures, Figure 2 is very similar as Figure 3, and with the same calculated numbers.

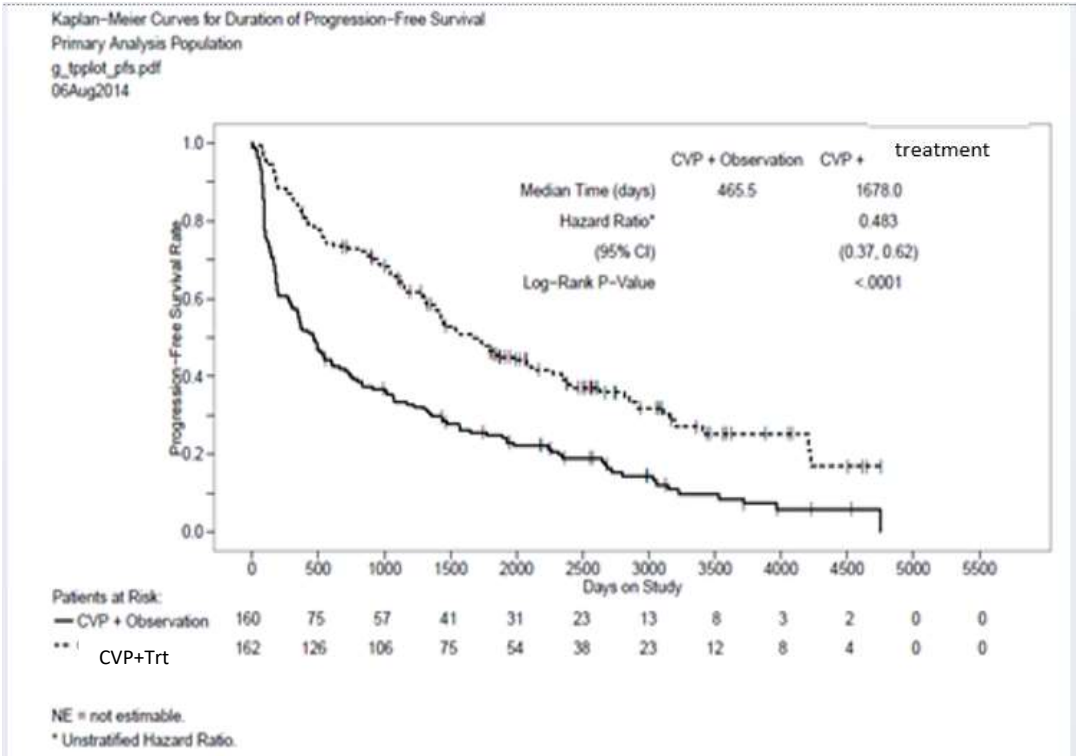

Figure 2. Kaplan Meier Plot created using R

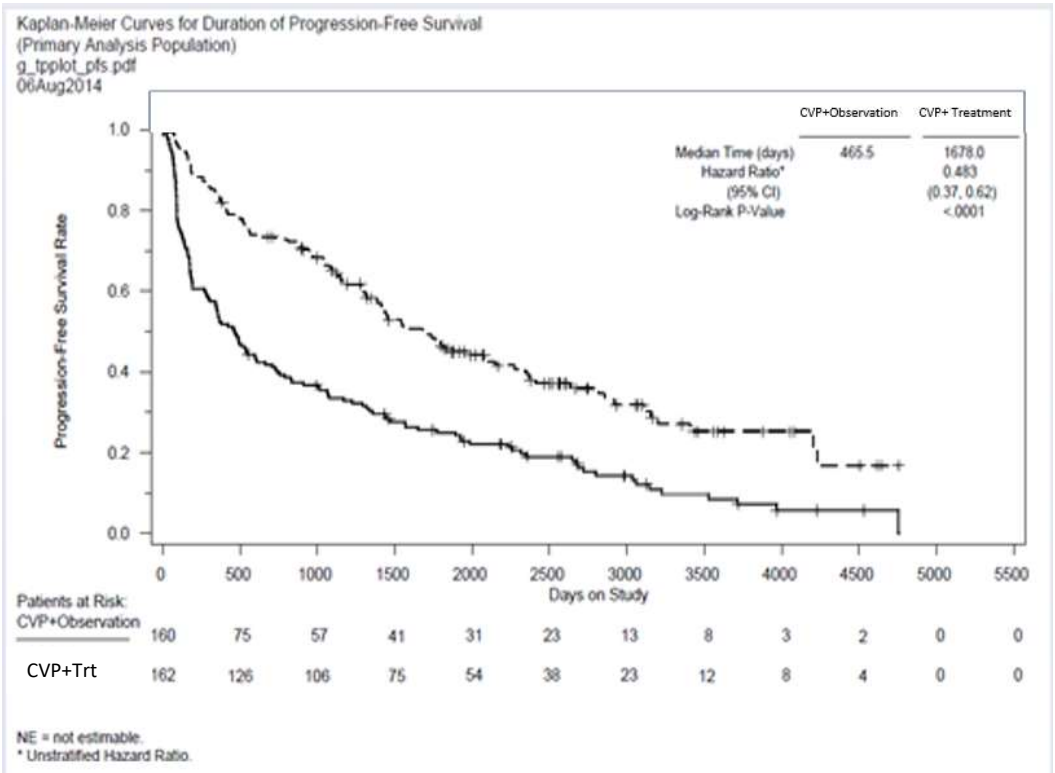

Figure 3. Kaplan Meier Plot created using SAS

Forest Plot:

Figure 4 was generated by R, while Figure 5 was made by SAS. Figure 4 and 5 are with the very close layout and same calculated statistics.

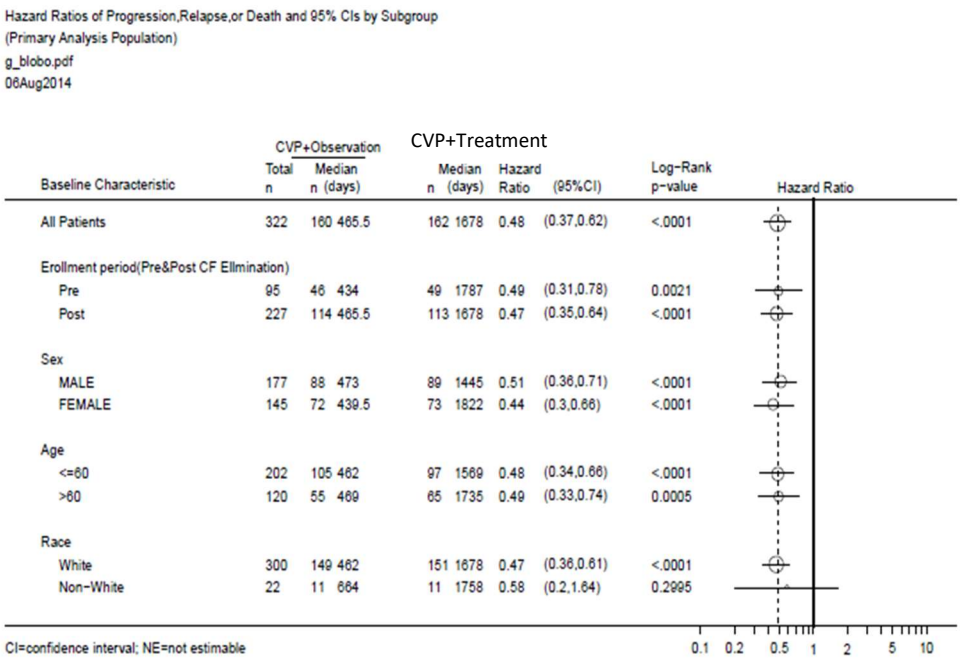

Figure 4. Forest Plot created using R

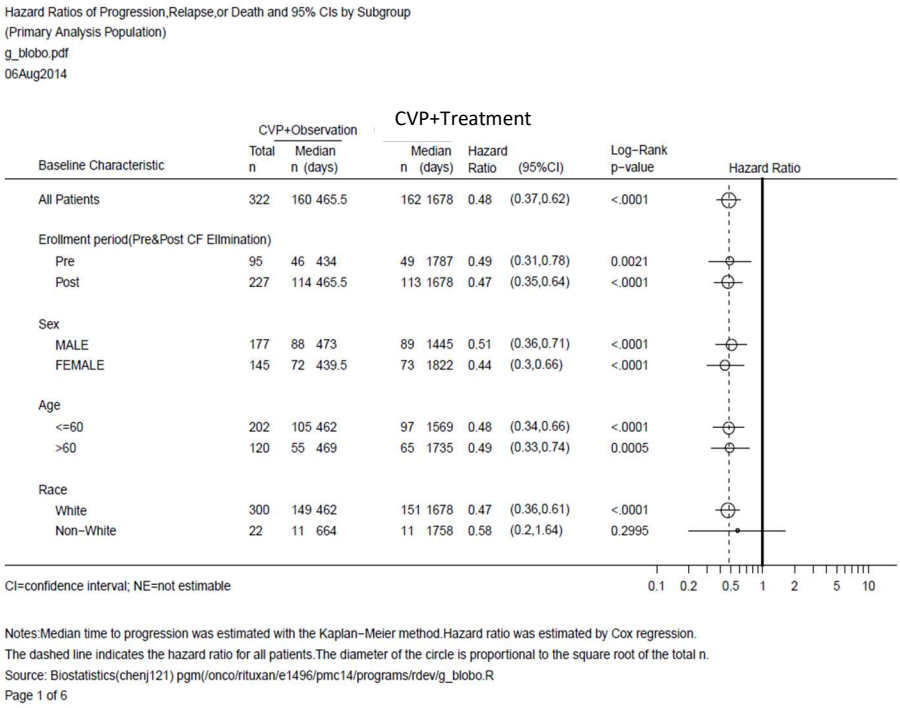

Figure 5. Forest Plot created using SAS

Survival Table:

Figure 6 was generated by R; Figure 7 was made by SAS. Again, we got same numbers.

Log-Rank Test for Equality and Hazard Ratios of Progression, Relapse, or Death
(Primary Analysis Polulation)
t_surv_test1.out
06AUG2014

CVP + Observation versus CVP+Treatment	Log-Rank Test	Hazard Ratio	Cox Regression 95% CI	p-value[ ]
No stratification	<0.0001	0.483	(0.375,0.624)	<0.0001
Stratified by residual disease (minimal vs. gross) and histology (follicular vs. other)	<0.0001	0.481	(0.372,0.622)	<0.0001

Note: Censoring time based upon date last known to be in remission/SD or date of last contact if date last known to be in remission/SD is missing.
[ ] The p-value is from PROC PHREG Wald Test using Efron`s method for handling ties.

Source:Biostatistics(chenj121) pgm(/onco/rituxan/e1496/pmc14/programs/t_surv-test)

Figure 6. Survival Table created using R

Log-Rank Test for Equality and Hazard Ratios of Progression, Relapse, or Death
(Primary Analysis Population)
t_surv-test1.out
06AUG2014

CVP + Observation versus CVP+Treatment	Log-Rank Test	Cox Regression Hazard Ratio	Cox Regression 95% CI	Cox Regression p-value[*]
No stratification	<.0001	0.483	(0.375, 0.624)	<.0001
Stratified by residual disease (minimal vs. gross) and histology (follicular vs. other)	<.0001	0.481	(0.372, 0.622)	<.0001

Figure 7. Survival Table created using R

Laboratory Analysis Dataset:

Output data set is in .xpt format (SAS version 5). We used ‘Proc compare’ in SAS to compare the data produced by R and SAS, which showed a consistent result.

To compare the performance from both tool, Table 1 shows a brief comparison between SAS and R for some cases we met in data manipulation.

	R	SAS
Condition Calculation	Loop Take long time	If .. then in data step Easy and fast
Change Data Structure (wide format to long format)	'reshape' function Easy and fast	Transpose Take some time
Select 'last' record	'order' and 'aggregate' function Two lines of code	Sort and last in steps Two data steps
Date format	'as.Date' function	Input function
Format and Label	'SASformat' and 'label' function	Format and label in data step
Xpt file	'write.xport' function One line code, but take long	Libname xport and proc copy Two steps

Table 1. Comparison of Program Efficiency between SAS and R

3.2 R MARKDOWN AND SAS COLLABORATION FOR EXPLORATORY ANALYSIS REQUEST FROM EMA

3.2.1 BACKGROUND

There is an observed trend (initial crossing of OS K-M curves) towards an increased risk of early death for subjects treated with anti-P (dummy agent) from a number of controlled clinical trials in different cancer indications(tumor types/line of treatment), especially in monotherapy trials against standard chemotherapy regimens or in subgroups with low P expression. Such findings suggest that there could be subgroups of patients for whom P targeting agents may not be the preferred treatment option. The Committee for Proprietary Medicinal Products (CHMP) from European Medicines Agency (EMA) requested to conduct exploratory analyses aiming to identify factors (baseline characteristics/biomarkers) that could predict the likelihood of not benefiting from immunotherapy.

3.2.2 STATISTICAL ANALYSIS PLAN

The issue is quite complex and no established predictive factors have been identified so far. The goal is to find a way to identify patients who are at risk of early progression and death with anti-P antibodies. Taking into account the complexity of the data, including different products (different marketing authorization holders) and different indications, different lines of treatment, different biomarkers and diagnostic assays, and different cut-offs used for study inclusion or as stratification factors. Early deaths defined as all patients who die before the instantaneous hazard rates (time varying hazard rate estimated) show a higher risk in the point inhibitor arm vs control arm.

We started with the all the potential factors that may predict the risk of early death, and possible interaction between treatment and potential relevant risk factors. Logistic regressions were applied with taking earth death as response. We applied univariate regression and univariate regression with treatment interaction to fishing the covariates with p-value <=0.05, and a multi-variate regression is applied with those significant covariates. Stepwise logistic model selection was applied, in order to select factors for the next step modeling. The last step modeling came up with factors from stepwise selection plus each significant covariate from univariate regression. Table 2 below shows the analysis plan.

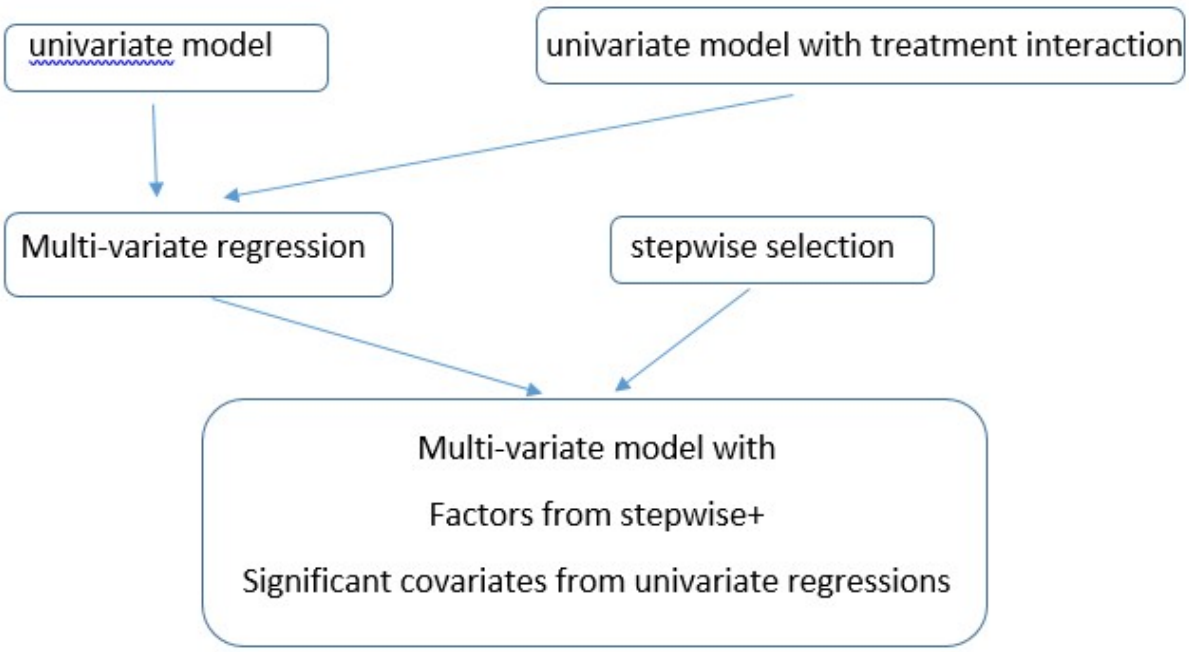

Table 2. Modeling Plan

3.2.3 ANALYSES EXECUTION

Being aware of the complexity of analyses, we decided to use R markdown to execute the analysis from univariate model fishing to final models. Since R is very powerful at applying functions for analysis, especially for those with repeated pattern. We built a function with the help from "tidyverse" package to implement univariate regressions with each of the candidate. Unfortunately, we did not find a good in R to do stepwise model selection; however, SAS is good functioned for model selection with easily implemented features. Therefore, we decided to collaborate SAS and R to finish the whole plan, and R markdown would be served as an executable documentation for the whole analysis plan. Meanwhile, we achieved the centralized reporting.

4. CONCLUSION

From the cases of collaborating SAS and R for eSUB module5 and exploratory analyses from Health Authorities, we can get the conclusion that R is capable to make submission standard products. Making the advantages from both open source and SAS provided a more efficient way to speed up the filling activity. By learning the features from different tools and using the right tool in the right situation, we can always connect to new concepts, advanced theories, and mature skills, which helps us to bring medications and therapies to patients earlier.

References

FDA presentation at the 2012 UseR Conference

http://blog.revolutionanalytics.com/2012/06/fda-r-ok.html

FDA Module 5 Study Data Specifications:

http://www.fda.gov/downloads/ForIndustry/DataStandards/StudyDataStandards/UCM312964.pdf

CTD Guidance Document:

http://www.regprofessional.com/resources/eCTD%20Tips%2010%20Mar%202011.pdf

ACKNOWLEDGMENTS

Laura Harris

Aditi Qamra

Will Harris

CONTACT INFORMATION

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

Jingyuan Chen

Hoffman-La Roche Limited

7070 Mississauga Road,

Mississauga, ON L5N 5M8 Canada

Email:jingyuan.chen@roche.com