



PhUSE US Connect 2019

Paper DV11

Interactive Exploratory Analysis and Visualization for Late Phase Clinical Data

Sheng-Wei Wang, Janssen R&D, Raritan, NJ, US
Casey Yang, Janssen R&D, Shanghai, China

ABSTRACT

The challenge of numerous post-hoc analysis requests for clinical trial data often arises after database lock. These analysis requests are mostly non-preplanned and exploratory. For example, Kaplan-Meier (K-M) plot and Forest plot for different endpoints and subgroups of the time-to-event data are often requested.

For addressing this challenge, we present an interactive exploratory analysis template using TIBCO Spotfire connecting with R. This template can quickly visualize the analysis results for exploratory purpose. It is built through: (1) loading the SAS analysis datasets (i.e. ADaM datasets), (2) building the TERR (TIBCO Enterprise Runtime for R) data functions which utilize the loaded ADaM datasets as inputs for analysis requested, and (3) interactive visualizations of the analysis results. We will present a case study of visualizing survival analysis results (i.e. K-M plot and Forest plot) and safety analysis results (e.g. AE summary and Lab summary overtime) using a late phase clinical trial data.

INTRODUCTION

Extensive volume of creation for post-hoc analysis requests (ad-hoc and exploratory analysis) often arise after database lock of a clinical study. Clinical team usually expect to see these analysis results from statistician and programmer team in a timely manner. The purpose of this paper is to address this challenge by presenting an interactive exploratory analysis tool using TIBCO Spotfire connecting with R.

The commonality of the post-hoc exploratory analysis requests is to tweak the statistical analyses pre-defined by Statistical Analysis Plan (SAP) with different endpoints/parameters, analysis sets, and/or sub-groups of interest. The conventional way for programming team to work on these requests is to manually modify the SAS programs/macros, which can be tedious, repetitive, and time-consuming with respect to many requests. To relieve this programming burden, this paper presents an effective approach of building an interactive exploratory analysis platform, which can allow agile exploration of analysis results based on the interactive user inputs (i.e. through simple clicking to specify different inputs). Therefore, exploratory analysis results can be visualized interactively before requesting the TLFs (Table, Listing, and Figures) from the programming team.

OVERVIEW OF INTERACTIVE EXPLORATORY ANALYSIS



Figure 1: A high-level data process flow for interactive exploratory analysis and visualization



TIBCO Spotfire provides a data analytics and visualization platform with interactive capability, which enable quick and easy exploration of analysis results dynamically [1]. TIBCO Spotfire has been used to build interactive platforms in the pharmaceutical industry for clinical data review, safety reporting, and efficacy analysis [2,3,4]. In this paper, we use TIBCO Spotfire connecting with R to build the interactive exploratory analysis dashboards for post-hoc efficacy analysis (i.e. survival analysis) and safety analysis (i.e. AE summary and Lab summary) using a late phase clinical trial data [5]. Figure 1 above provides a high-level data process flow in building the interactive exploratory analysis dashboards. Further details in each process are provided below.

Data Access: There are different options of importing data into Spotfire such as loading data files (e.g. text, csv, excel, SAS datasets), querying data tables from databases and data connectors. We are using the option of loading the SAS datasets directly into Spotfire, since our focus here is to perform exploratory statistical analysis for clinical study data. The SAS datasets are the ADaM datasets developed for creating the Clinical Study Report (CSR) TLFs. The development of the exploratory analysis dashboards may proceed in parallel with the CSR programming activities. Once the ADaM datasets are refreshed, we would need to reload them into Spotfire. However, the data reloading process need to be done only once after database lock.

Data Analysis: Spotfire has an embedded statistical engine called TIBCO Enterprise Runtime for R (TERR), which is compatible with open-source R. This enable us to develop customized TERR data functions that can reproduce the statistical analysis results created by SAS programs in supporting CSR TLFs. An important feature of the TERR data function is the capability of linking the data inputs with the interactive user interface dynamically, which may include dynamic data filtering, on-demand execution of the data functions, and so on. This allows us to create the dynamic and interactive visualization of the analysis results based on the inputs defined by users, which only requires a few mouse clicks in the interactive interface.

Data Visualization: Spotfire provides user friendly tool bar for creating different data visualizations such as bar chart, pie chart, line plot, scatter plot, box plot, and cross table etc. Graphical attributes (e.g. plotting axes, line/marker attributes, color, legend, etc.) can be easily adjusted through the drop-down dialog menu of visualization properties. We use the Spotfire toolbar to create data visualizations based on the statistical analysis results from running TERR data functions. Furthermore, Spotfire dynamic data filters are utilized to drive the interactive visualizations of the analysis results. Both data visualization and data filters are incorporated for building the interactive exploratory analysis dashboard along with an analysis panel created by Spotfire text area tool.

USER DESIGNED SPOTFIRE DASHBOARD

Spotfire text area tool provides a user-friendly interface to create analysis panel, where users can customize analytical flow and provide instruction to conduct analysis. Figure 2 below depicts the snapshot of a Spotfire exploratory analysis dashboard for forest plot. It consists of two components: visualization area (display of analysis results) and analysis panel (provide instruction to conduct interactive exploratory analysis).

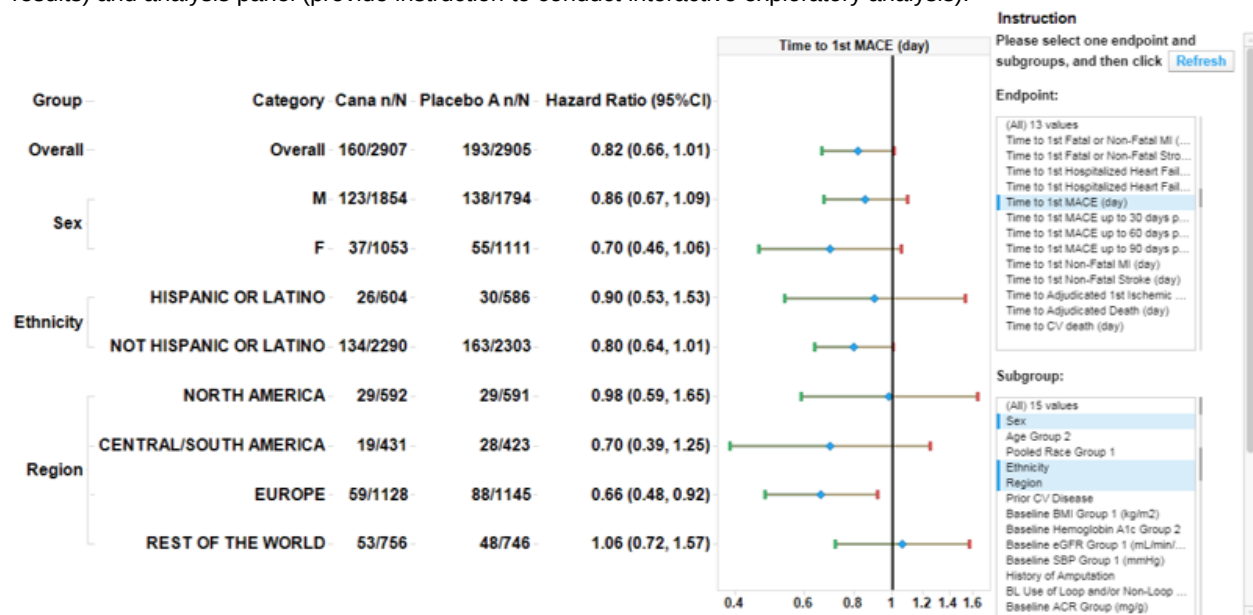


Figure 2: The snapshot of a Spotfire exploratory analysis dashboard for forest plot



A case study of visualizing survival analysis results (i.e. K-M plot and Forest plot) and safety analysis results (e.g. AE summary and Lab summary overtime) are illustrated in following sections using a late phase clinical trial data [5].

SURVIVAL ANALYSIS VISUALIZATION

The most common visualization tools for survival analysis include Kaplan-Meier (K-M) plot and forest plot. Time-To-Event ADaM datasets created by SAS are loaded in Spotfire for building the interactive exploratory analysis dashboard of survival analysis. In the SAS programs, the procedures of PROC LIFETEST and PROC PHREG are used to calculate K-M estimates and hazard ratios with confidence intervals for creating K-M plot and forest plot respectively through the GTL (Graph Template Language) templates. In Spotfire, we install the survival package in the TERR console from CRAN (Comprehensive R Archive Network), and then use the survfit and coxph functions to develop TERR data functions to obtain the survival analysis results that can match the SAS results for the pre-defined input conditions (this is the validation step for building the exploratory analysis dashboard). The input specifications of the TERR data functions are connected to the user designed interface (i.e. analysis panel), where end-users can interactively select different input conditions to run the TERR data functions and the updated analysis results are displayed in the visualization area on the fly.

KAPLAN-MEIER (K-M) PLOT

The survival probabilities (starting at 100%) by treatment groups are usually plotted over time in K-M plot for assessing treatment effect over the study period. We plot the cumulative event rates (100% minus survival probability) over time in the K-M plot here, which present similar assessment of treatment effect.

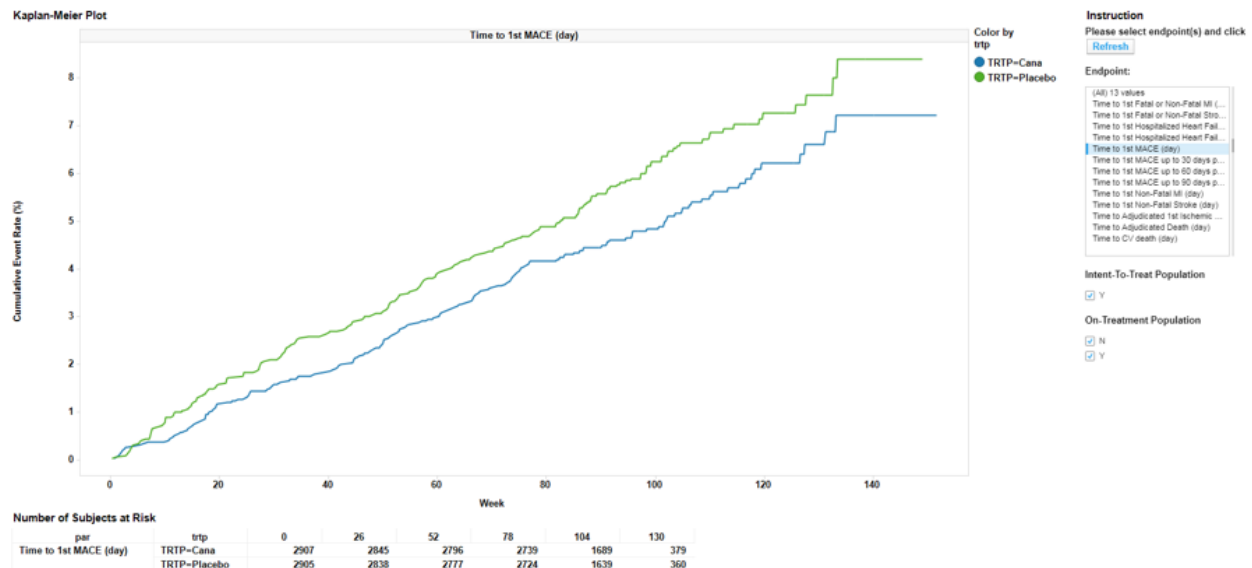


Figure 3: Kaplan-Meier plot of first occurrence of MACE

Figure 3 shows the K-M plot of the primary endpoint (time to the first occurrence of Major Cardiovascular Adverse Events(MACE)) for a late phase clinical study. The number of subjects at risk across different analysis visits are displayed under the K-M plot. For conducting exploratory analysis, the end-user can select different endpoint in the list box of the analysis panel that is not specified for CSR TLFs, and then simply click the refresh button to run the analysis for updating the K-M plot on the fly.

Another important feature in the exploratory analysis dashboard of K-M plot is the capability of selecting multiple endpoints and display multiple K-M plots by trellis panel as shown in Figure 4 below. This is particularly useful for the composite endpoint such as MACE, which is defined by three cardiovascular events (non-fatal MI, non-fatal stroke, and cardiovascular death). The multiple panel of K-M plots for MACE and its three component endpoints provides holistic view of K-M curves for the primary composite endpoint of this study, which can be useful for the clinical team to gain further insight of the treatment effect.

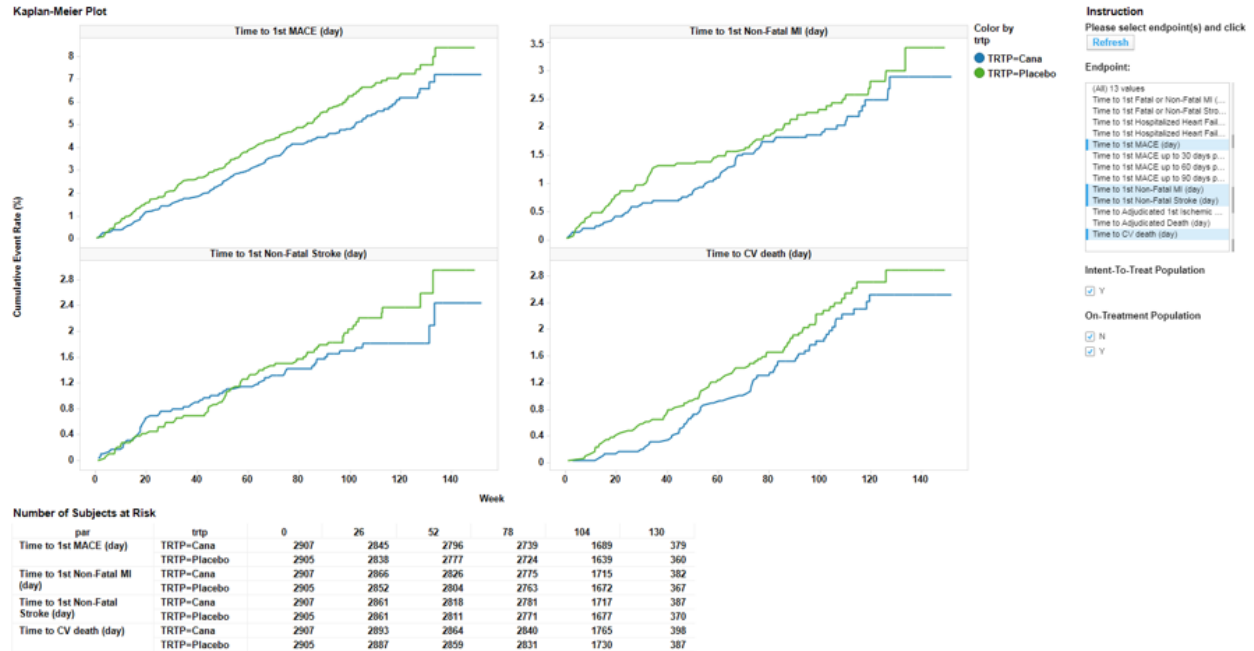


Figure 4: Trellis panel of multiple K-M plots for MACE and its 3 component endpoints

FOREST PLOT

Forest plot is traditionally used in meta-analysis to show the variability across studies. For a clinical study, forest plot is used to illustrate the variation of treatment effects across different sub-groups of the study population. As shown in the forest plot below, the overall treatment effect is placed on top of the y axis followed by the list of sub-groups (i.e. age group < 65 and age group ≥ 65), while the corresponding hazard ratios (represented by diamond markers) and 95% confidence intervals are shown in horizontal lines in the x axis. A vertical line of 1 represents the null hypothesis of the treatment effect. Forest plot provides an effective way for quick interpretation of sub-group analysis results.

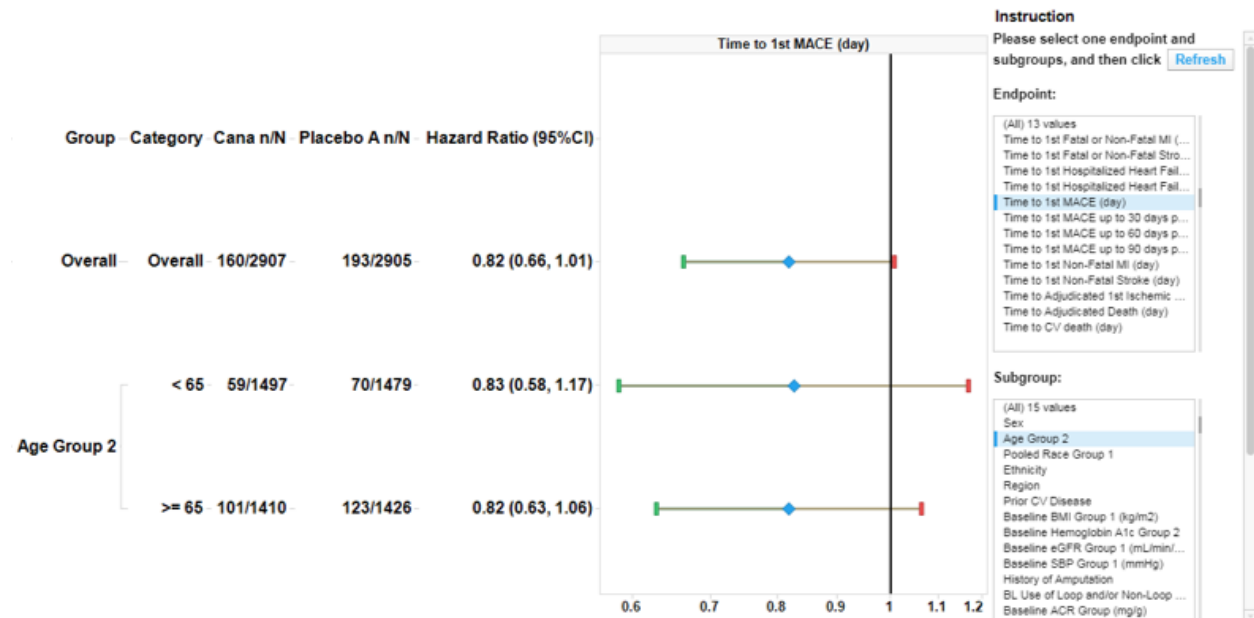


Figure 5: Forest plot of MACE with age sub-groups

The analysis panel provides a user-friendly interface to interactively create customized forest plot. Let's assume that the clinical team has a request to create a forest plot for the primary endpoint of MACE with 4 sets of sub-groups characterized by age group, prior cardiovascular disease, baseline eGFR level, and baseline use of insulin. The interactive instruction in the analysis panel can be provided as following:

- In the listbox of Endpoint, click on 'Time to 1st MACE (day)'.
- In the listbox of Subgroup, click on the 4 subgroup variables specified by the clinical team.
- In the dropdown menu of Ties method for Cox proportional hazard model, there are 3 options: efron, exact, and Breslow. The default is efron, and that is what we want to use here also.
- The default analysis set is intent-to-treat population. If the clinical team like to see the analysis results for on-treatment population, we can just check off the check box of 'N' under on-treatment population.
- After making all the selections, click on the 'Refresh' button on top of the analysis panel and then the requested forest plot will be delivered to the visualization area of the dashboard on the fly, which is shown in Figure 6 below.

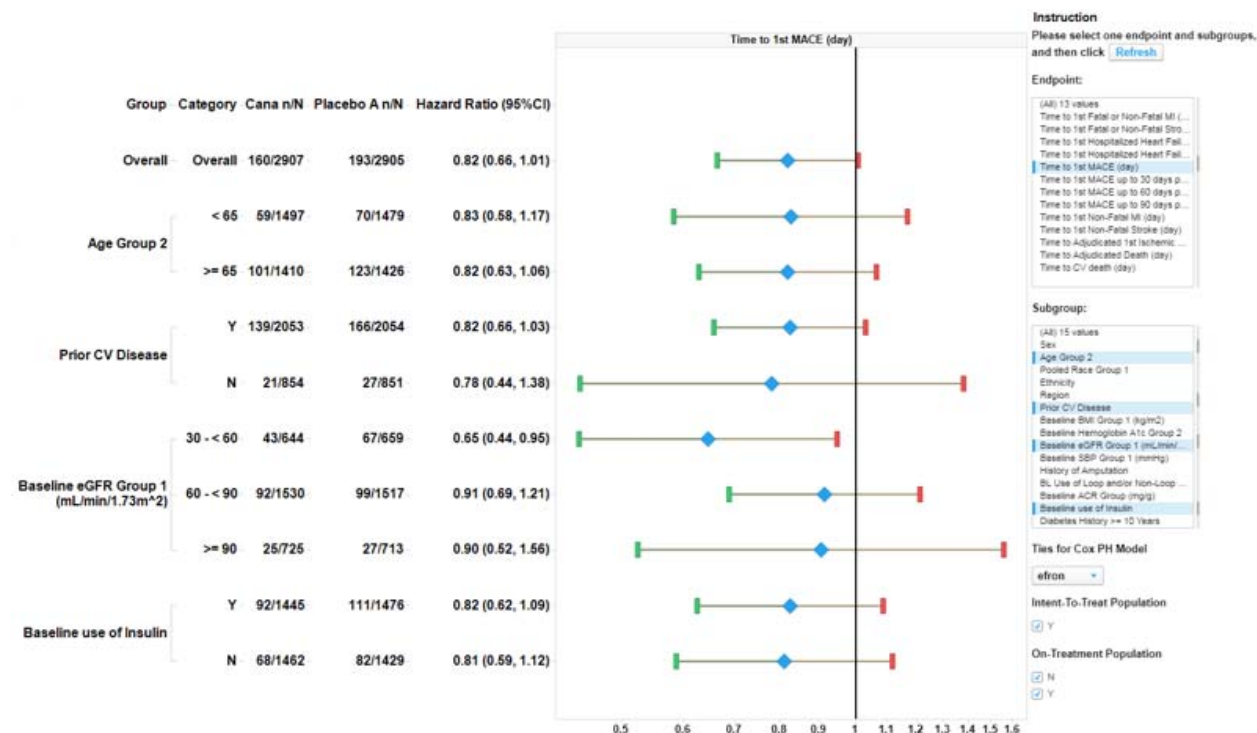


Figure 6: Forest plot of MACE with 4 sets of subgroups characterized by age group, prior CV disease, baseline eGFR level, and baseline insulin use

AE SUMMARY ANALYSIS VISUALIZATION

AE summary tables usually display frequency counts and standard percentages that are calculated by dividing the number of subjects experiencing the AEs with the total number of subjects at risk by treatment groups. For the clinical studies with long-term follow-up, the standard percentages may not be appropriate since the follow-up duration between treatment groups could be different. A more appropriate measure under this scenario is the exposure-adjusted incidence rate, where the total exposure time of all subjects at-risk is used as the denominator to calculate the percentage by treatment group. Both the standard percentages and the exposure-adjusted rates are built into our exploratory analysis dashboard of AE summary.

AE SUMMARY TABLE BY PREFERRED TERMS

In this paper, we focus on creating the AE summary tables for special interest AEs. Special interest AEs is a group of AEs of specific medical concern for the study drug. The customized query (CQNAM) flag variables are derived in the AE analysis dataset to facilitate the creation of AE summary tables for special interest AEs. A subset of the CQ flags are usually used in SAS programs for CSR TLFs. However, clinical team may have the exploratory analysis requests to create other special interest AE tables that are not covered by CSR TLFs.



Figure 7 shows the exploratory analysis dashboard for special interest AEs. In the analysis panel, we have a dropdown menu that includes all the CQNAM flag variables in the AE analysis dataset for querying special interest AEs. Many CQNAM flags are usually built into the AE analysis dataset, and it's difficult to remember which special interest AEs each flag represents. A handy search tool is built in the analysis panel for easy finding of the desired CQNAM flag. As shown in Figure 7, the word of 'Renal' is typed in the search box of CQ type, and 3 types of special interest AEs containing 'Renal' are displayed in the listbox below. We click on the desired special interest AEs of 'Acute Renal Failure (Narrow)', and the corresponding CQNAM flag (i.e. CQ02NAM) is revealed in the CQNAM listbox. Then we go to the top dropdown menu of CQNAM variables and select CQ02NAM, and the AE summary table of Acute Renal Failure (Narrow) is created in the visualization area.

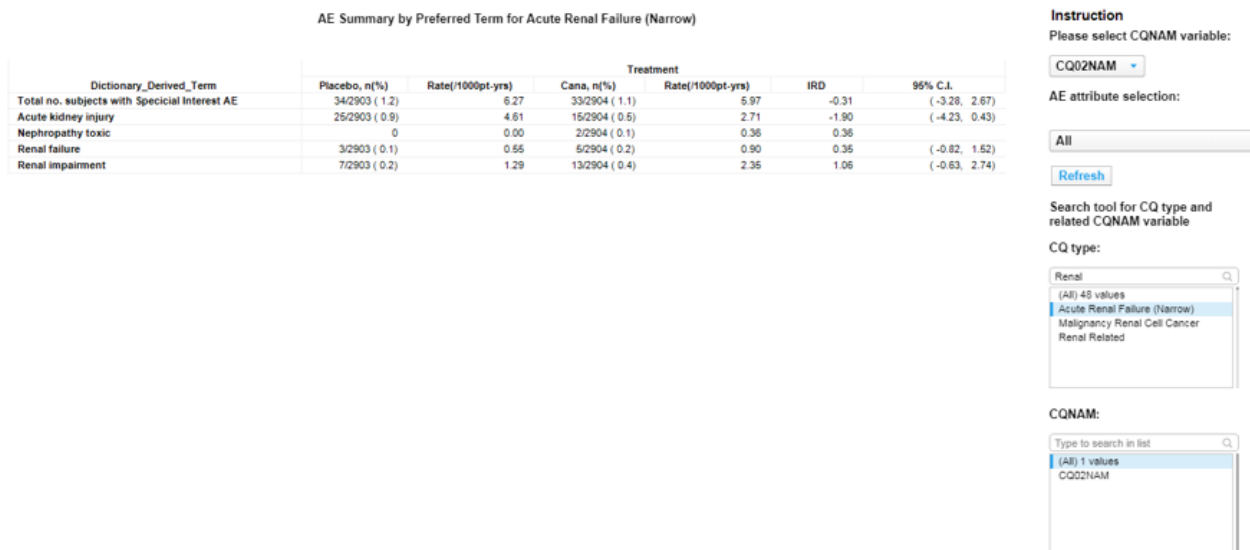


Figure 7: Exploratory analysis dashboard of AE summary by preferred terms for special interest AEs

The AE summary table can be further refined by imposing additional AE attribute such as serious AE, AE leading to drug discontinuation, etc. This is also built into a dropdown menu in the analysis panel for the user to explore. For a more comprehensive summary of the AE attributes related to special interest AEs, we create another summary table shown in Figure 8 below.

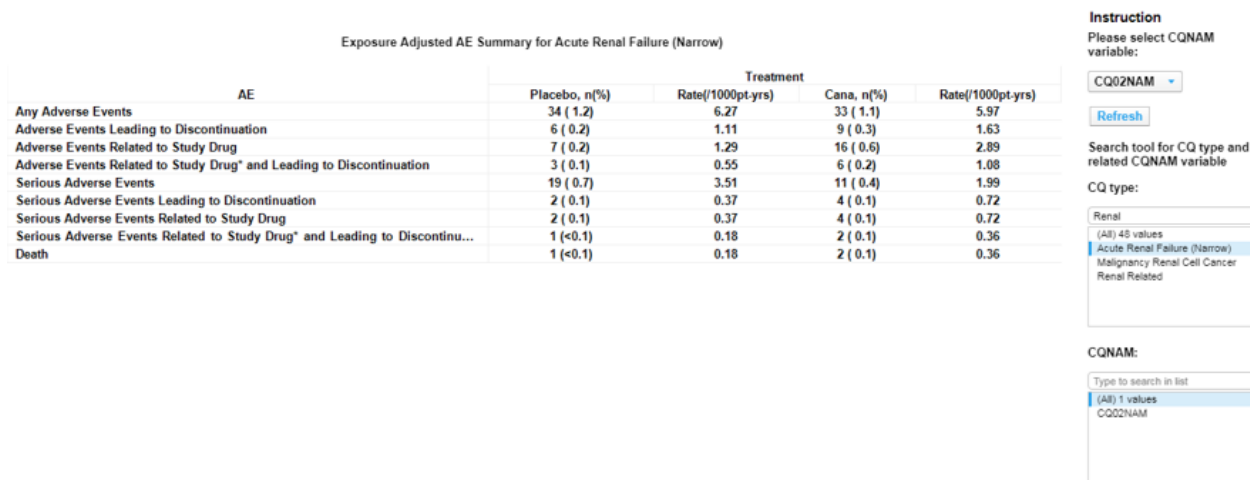


Figure 8: Exploratory analysis dashboard of AE attribute summary for special interest AEs

AE SUMMARY PLOT OVER TIME

Another type of AE summary analysis is to create the occurrence summaries over different time periods of the study. Besides the summary tables, the graph of bar charts over time provides a clear picture of the AE occurrence trend over time. Figure 9 presents the AE exploratory analysis dashboard for the over-time summary by plotting bar charts (representing percentage of AE occurrence) over different time periods of the study. The summary table is also shown below the graph of bar charts. The denominators of total number of subjects at risk are different at different time periods, since subjects may discontinue from the study at different time points of the study.

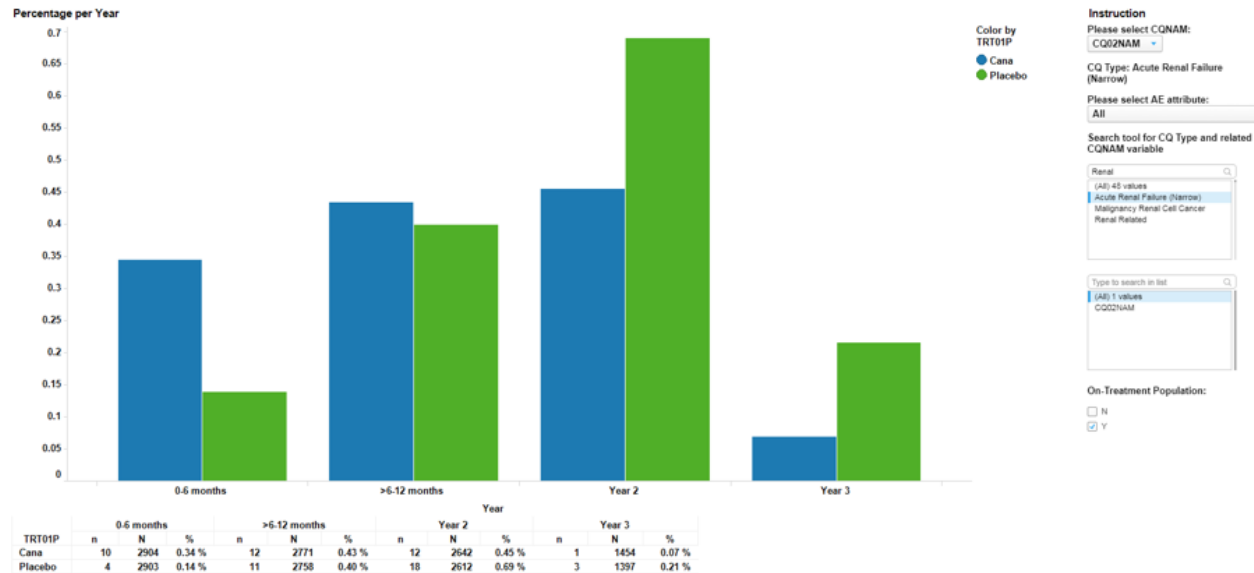


Figure 9: Percentage of AE occurrence over time for special interest AEs

LONGITUDINAL ANALYSIS VISUALIZATION

LAB SUMMARY STATISTICS OVER TIME

The over-time plots of summary statistics (e.g. mean $\pm$ SE, mean of change from baseline $\pm$ SE) of laboratory measurements are often among the post-hoc requests for different lab parameters not covered by CSR TLFs. For addressing this type of requests, the exploratory analysis dashboard of lab summary over time is created as shown in Figure 10.

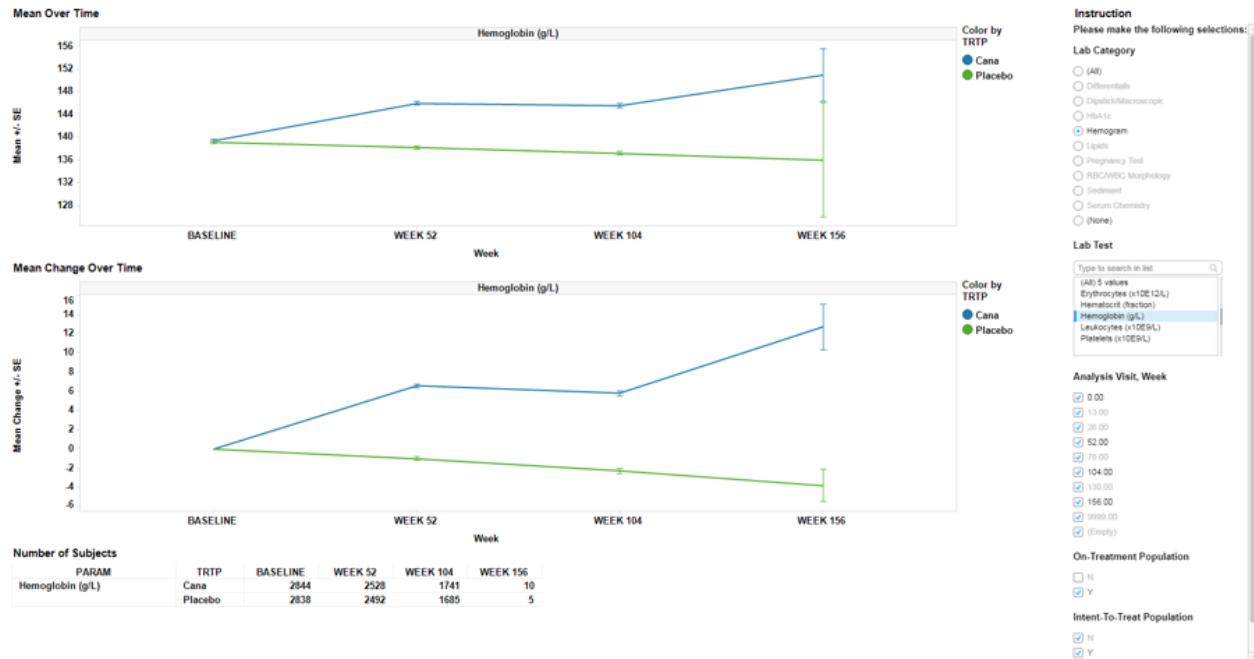


Figure 10: Mean and mean change over time for hemoglobin (g/L)

In the analysis panel, users need to select the lab category first by clicking the radio button of the desired category (Hemogram is selected in Figure 10). Then all the available lab tests below to the selected lab category will appear on the listbox of lab test (there are 5 lab tests for Hemogram). The next step is to select the desired lab test(s) and the plots of mean over time and mean change over time will be displayed on the visualization area of the dashboard as shown in Figure 10.

If the user doesn't know which lab category of the desired lab test belongs to, the user can type in the lab test directly in the search field on top of the lab test listbox for making the selection. The table of number of subjects who have the available measurements over the time points is displayed below the plots. This provides a way to assess the unusual variability shown in the plots. In Figure 10, there are only very few subjects who have available hemoglobin measurements at week 156, thus causing the large SE. Therefore, we can check off the check box of week 156 in the analysis panel to get a better visualization of the over-time plots.

CONCLUSION

We have shown the use cases of an interactive exploratory analysis tool developed through TIBCO Spotfire connecting with R for survival analysis and safety (AE summary and lab summary) analyses of a late phase clinical study. The Spotfire exploratory analysis dashboards shown above can be published onto Spotfire Consumer (previously called as Web Player) for sharing within an organization [1]. It is a web browser application without any installation, and the system administrator has the control over the user access to the web application. Web player allows end-users to perform the interactive actions on the analysis panel that we designed for exploratory analysis and export the analysis results.

The intended use of the Spotfire exploratory analysis dashboard is for the statistician team to perform quick and easy exploration of the post-hoc analysis requests for quick review by the clinical team. Therefore, the unnecessary analysis requests can be quickly filtered out by the clinical team before sending over to the programming team for creating the formal TLFs, which improve the efficiency of exploratory analysis and relieve the burden of programming team.

REFERENCES

1. TIBCO Spotfire® Business Author and Consumer User's Guide.
2. Busa B. and Pothuri K. 2018, Building Dashboards for Real-Time Insights into Clinical Data, paper DV17, PhUSE US Connect 2018.
3. Gupta A. 2018, Advanced Visualization using TIBCO Spotfire and SAS, paper DV-04, PharmaSUG 2018.



4. Lu Z., Graham H., and Chen J. 2018, Spotfire Clinical Visualization from SAS and R, paper HS-142, MWSUG 2018.
5. Neal et al: Canagliflozin and Cardiovascular and Renal Events in Type 2 Diabetes, N Engl J Med 2017; 377:644-657.

ACKNOWLEDGMENTS

We would like to acknowledge our CVM colleagues (Lisa Lyons, Yingshan You, James Pan, Hsiaowei Deng, Wentao Lu, George Capuano, and Tao Sun) and a PDW colleague (Sandy Lei) for their feedbacks to the work presented in this paper.

CONTACT INFORMATION

Contact the authors at:

Sheng-Wei Wang
Janssen R&D
920 US 202 South
Raritan, NJ 08869, USA
Email: swang69@its.jnj.com

Casey Yang
Janssen R&D
65 Guiqing Road
Shanghai 200231 China
Email: wyang41@its.jnj.com

Brand and product names are trademarks of their respective companies.