PharmaSUG China 2023 - Paper AA-168

Using R Shiny to Streamline and Visualize Prostate-Specific Antigen (PSA) Analysis in Prostate Cancer Trials

Tongtong Weng, MSD China

ABSTRACT

Prostate-specific antigen (PSA) is a glycoprotein which is found almost exclusively in normal and malignant prostate cells. Many prostate cancer trials use PSA as an important biomarker and observe an association between changes in PSA level and survival. PSA-related analysis, including but not limited to PSA response rate and time to PSA progression, is widely used for assessing efficacy in prostate cancer trails. This paper provides a basic introduction of PSA analysis concepts following the guideline from Prostate Cancer Working Group 3 (PCWG3), and addresses the design of a Shiny app for PSA check.

R shiny is used to develop a user-friendly interface that contains several embedded requests, which can be easily executed for quick PSA navigation. A demonstration of the R shiny app is provided. The app demonstrates how the Analysis Data Model (ADaM) ADSL and Study Data Tabulation Model (SDTM) LB domain are uploaded and analyzed, and how these requests perform specified actions in a statistical computing environment and return detailed PSA results to be viewable in the R Shiny dashboard. In the end, the paper discusses the challenges during development and some considerations for the future enhancement of this R Shiny tool.

INTRODUCTION

Prostate cancer is one of the most frequently diagnosed malignancies and is the fifth leading cause of cancer-related deaths in men worldwide. Prostate-specific antigen (PSA) level is widely used as an important tumor marker in monitoring patients with prostatic cancer. Changes in PSA level often antedate the detection of bone metastases or subsequent disease progression. If a patient's PSA level begins to rise during treatment, it may be the sign of a recurrence. PSA-related analysis is usually used for assessing efficacy. This paper specifically focuses on PSA response rate and time to PSA progression, which is typically conducted in prostate cancer trials.

With the increasing need for exploring PSA data in an efficient and visualized manner, developing an R Shiny app is a smart choice. The web app enables users to review the PSA data and do basic tidy-up like sorting, filtering, and variable selection. Users can also dive deeper to the PSA analysis outcomes and plots. It only requires few requisite input datasets, which can save time from rerunning multiple SAS program of analysis datasets, tables, and figures.

This paper gives an overview of PSA workflow from Analysis Data Model (ADaM) datasets to outputs in analysis and reporting process, and the streamlined design of the R Shiny app for quick PSA navigation.

PSA-RELATED ANALYSIS

PSA is a glycoprotein produced by cells of the prostate gland and releases into the bloodstream. PSA measurements are collected through laboratory tests and mapped to the Study Data Tabulation Model (SDTM) Laboratory Test Result (LB) domain. The results are usually reported as nanograms of PSA per milliliter of blood (ng/mL).

With the compliance of Clinical Data Interchange Standards Consortium (CDISC) Analysis Data Model, two analysis datasets, ADPSA and ADTTP, are created for PSA analysis in regular analysis and reporting process.

The figure 1 outlines the flowchart of PSA-related analysis in statistical programming in prostate cancer study. The subject-level ADaM ADSL and SDTM LB datasets provide basic demographic information and PSA input data to derive the two efficacy datasets. The ADPSA dataset serves as a source dataset to create PSA response rate table; the ADTTP is used for analysis of time-to-event progression and creation of Kaplan-Meier plot.

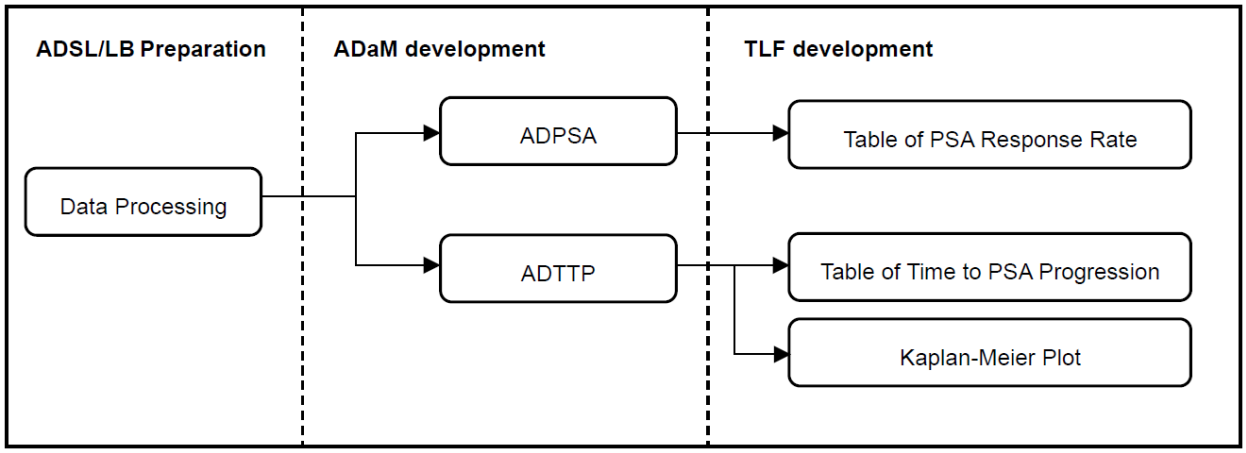


Figure 1. Flowchart of PSA-related analysis

ADPSA: PSA ANALYSIS DATASET

The ADPSA is a Basic Data Structure (BDS) analysis dataset with key analysis parameter variables (PARAM and PARAMCD), analysis value (AVAL and AVALC), baseline flag (ABLFL) and baseline value (BASE), change and percent change from baseline values (CHG and PCHG). The ADPSA contains all valid PSA records retrieved from SDTM LB dataset, and derived PSA response with protocol-specified decline from baseline.

In the example study for this paper, PSA response rate is defined as the proportion of subjects in the randomized population with PSA decline of >=50% from baseline, followed by a consecutive value that also has a decrease >= 50% from baseline and is conducted at least 3 weeks later. Note that one subject is likely to have multiple PSA measurements that satisfy the above definition. To simplify the PSA response rate analysis, every subject is counted a single time for applicable PSA confirmed assessment. The analysis date (ADT) for the derived PSA response parameter is the initial confirmed PSA decline date.

Table 1 is an example of subject data in ADPSA dataset. PSA records (PARAMCD="PSA") are directly from SDTM LB dataset for traceability. The closest PSA measurement prior to the randomization date is flagged as the baseline (ABLFL="Y"). CHG and PCHG are imputed for post-baseline records. The confirmed PSA decline response (PARAMCD="PSA50CNF") is derived for PSA response rate analysis.

	USUBJID	NCTXSDT	RANDDT	PARAMCD	ADT	ADY	AVAL	BASE	ABLFL	CHG	PCHG
1	9999-999_000000000	.	2021-07-09	PSA	2021-07-02	-7	0.8	0.8	Y	.	.
2	9999-999_000000000	.	2021-07-09	PSA	2021-08-07	30	0.5	0.8		-0.3	-37.5
3	9999-999_000000000	.	2021-07-09	PSA	2021-08-28	51	0.3	0.8		-0.5	-62.5
4	9999-999_000000000	.	2021-07-09	PSA	2021-09-15	69	0.3	0.8		-0.5	-62.5
5	9999-999_000000000	.	2021-07-09	PSA	2021-10-06	90	0.5	0.8		-0.3	-37.5
6	9999-999_000000000	.	2021-07-09	PSA	2021-10-13	97	0.5	0.8		-0.3	-37.5
7	9999-999_000000000	.	2021-07-09	PSA	2022-01-06	182	0.2	0.8		-0.6	-75
8	9999-999_000000000	.	2021-07-09	PSA	2022-04-21	287	0.3	0.8		-0.5	-62.5
9	9999-999_000000000	.	2021-07-09	PSA	2022-06-23	350	0.2	0.8		-0.6	-75
10	9999-999_000000000	.	2021-07-09	PSA	2022-09-15	434	0.2	0.8		-0.6	-75
11	9999-999_000000000	.	2021-07-09	PSA50CNF	2022-01-06	182	1.1	0.8		0.3	37.5

Table 1. Sample subject data in ADPSA dataset

From programming's perspectives, the confirmed PSA decline response needs to satisfy the three criteria below:

1. The PSA measurement should be assessed before the start date of new anti-cancer therapy (NCTXSDT), if NCTXSDT is valued.

2. There are two consecutive records with PSA decline of >=50% from baseline (PCHG <= -50).
3. The consecutive analysis relative day (ADY) difference should be at least 3 weeks (21 days) apart. Therefore, the PSA measurement occurred on January 6, 2022, is considered the first confirmed PSA decline response for the subject.

ADTTP: TIME TO EVENT ANALYSIS DATASET FOR PSA

The ADTTP is a Time-to-Event Analysis dataset deriving PSA progression endpoint. The ADTTP includes variables to support time to PSA progression derivation: using AVAL for the length of time from the start of the at-risk period to the event of interest (PSA progression), ADT for the date of event or censoring, and CNSR for event or censoring status.

In this paper, time to PSA progression is defined as the time from randomization to PSA progression. The PSA progression date is defined as the date that:

1. >=25% increase and >=2 ng/mL above the nadir, and which is confirmed by a second value >=3 weeks later if there is PSA decline from baseline;
2. Or >=25% increase and >=2 ng/mL increase from baseline beyond 12 weeks if there is no PSA decline from baseline.

Of note, nadir is defined as the lowest value prior to the current assessment, and it is derived separately for each time point. For the subject who has a reduction in PSA from baseline (the original AVAL from ADPSA where PARAMCD = “PSA” < BASE), PSA progression uses nadir as the reference base. Alternatively, for the subject without PSA decline (AVAL >= BASE where PARAMCD = “PSA” in ADPSA), baseline value is used to verify PSA progression criteria. It is suggested to create custom variables (NADIR, NCHG, and NPCHG for nadir value, change from nadir, and percent change from nadir, respectively), which makes data more explicit to distinguish in different scenarios. If PSA progression is confirmed, the censor indicator (CNSR) is set to 0 with EVNTDESC (Event or Censoring Description) = “PSA Progression” in ADTTP dataset.

The start date of new anti-cancer therapy (NCTXSDT) is also taken into consideration in PSA progression derivation. If PSA meets progression criteria but occurs after NCTXSDT, it is still considered censoring. Participants without PSA progression or new anti-cancer therapy are censored at the last PSA assessment date. For those subjects who have no baseline or post-baseline PSA values, the date of censoring is the randomization date.

R SHINY APP FOR PSA CHECK

R is a very popular programming language for statistical analysis and data visualization. R shiny, with its various packages, provides the framework to explore and analyze clinical data in a more efficient way. In this paper, R Shiny is used to develop a user-friendly interface that contains several embedded requests for quick PSA navigation in prostate cancer trials.

The R Shiny app is built with its essential user interface (UI) and server R components. The front end is designed and controlled by the UI module and the functional coding logic is incorporated in the server module. The application is designed to run with the Shiny generated UI to load the input datasets, perform data manipulation and interactive visualization, based on the user requirements.

The design of the R Shiny app and SAS programming in prostate cancer study run in the same groove. A layout function *sidebarLayout()* is called to organize panels and elements into one page. The *sidebarPanel()* creates a sidebar “Data Processing” on the left, which enables users to import datasets and do pre-processing selection. The main area “PSA Navigation” using *mainPanel()* contains tabs to direct users to view data and check PSA-related outcomes. *TabsetPanel()* subdivides the user-interface into five discreate sections: ADPSA, ADTTP, PSA Response Rate, Time to PSA Progression, and Kaplan-Meier (KM) plot. Two action buttons *actionButton()* are designed. The “Load Data” button triggers coding logic to run the two efficacy datasets, ADPSA and ADTTP. The “Ready to Process” button allows the PSA-related analysis outputs to be viewable in the R Shiny dashboard.

The screenshot of the R Shiny app illustrates the page layout design in Figure 2. Tabs for different functionalities are introduced in the following step sections.

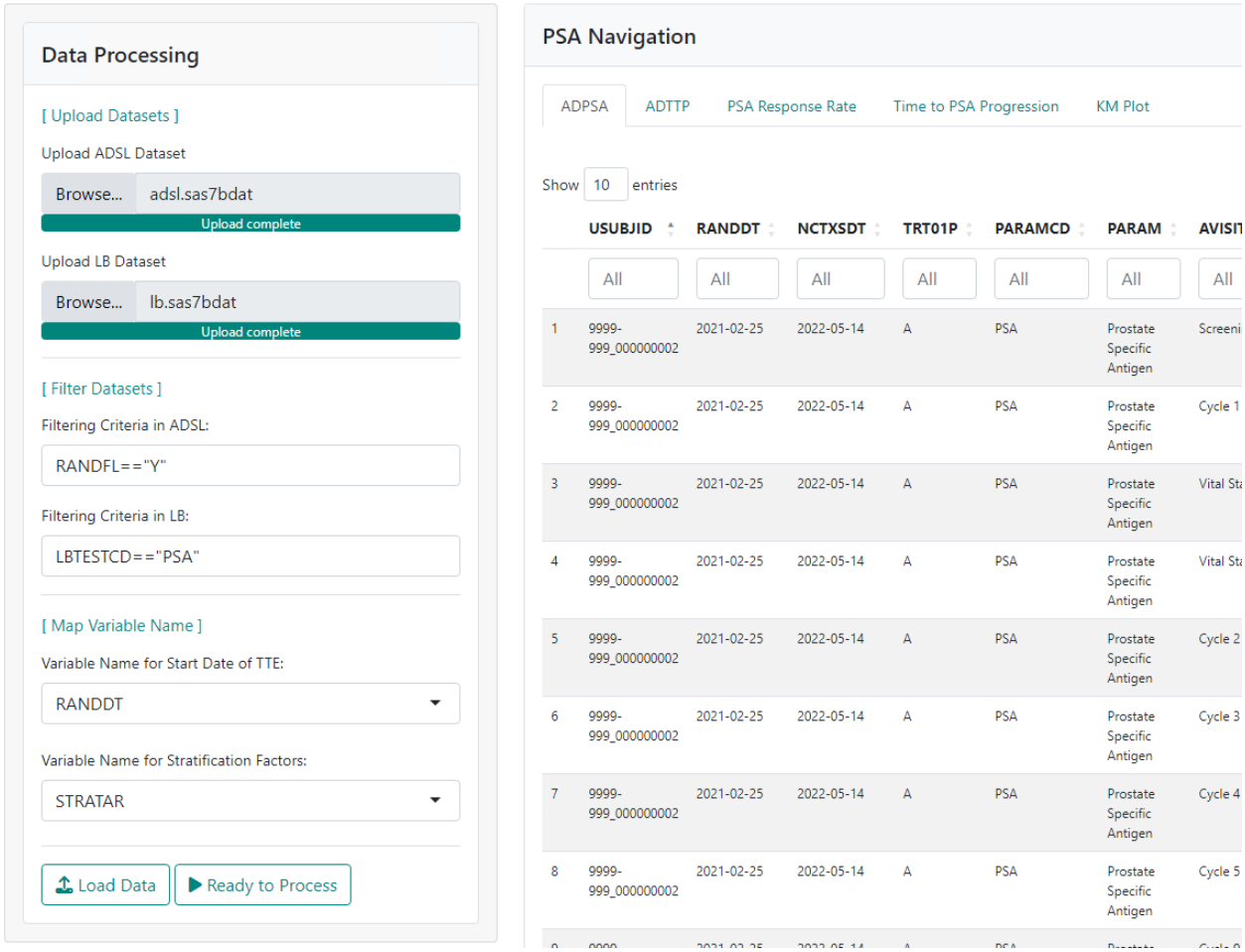


Figure 2. R Shiny app dashboard

STEP 1: IMPORT SAS DATASETS

The sidebar on the left is developed to import SAS datasets and define filtering criteria and critical variables for data processing. The subject-level dataset ADSL and SDTM LB dataset are requisite for the R Shiny app. The *fileInput()* function with *accept* argument creates a file upload control and identifies the unique file type as '.sas7bdat'. Considering the two datasets are always saved under different paths, the app enables users to upload sas7bdat files one after another. The ADSL dataset provides information on core variables, which consists of treatment group, randomization date, stratification factor, or any other demographic variables of interest; the SDTM LB dataset carries all original PSA measurements, including specimen collection date, result or finding in standard units, etc.

To define filtering criteria in ADSL and LB datasets, there are two *textInput()* functions for entry of unstructured text values. The default value for ADSL selection criteria is "RANDFL=='Y'" (randomized population selected). It's assumed that PSA data is documented as "PSA" in Laboratory Test Code (LBTESTCD) in SDTM LB dataset. Users can modify the selection criteria following R language rule based on the study design.

SelectInput() function creates a select list of variables that can be chosen for further survival analysis use. Users can identify the start date of time-to-event analysis (RANDDT by default or TRTSDT) and stratification variable (STRATRA specified in the sample study for this app).

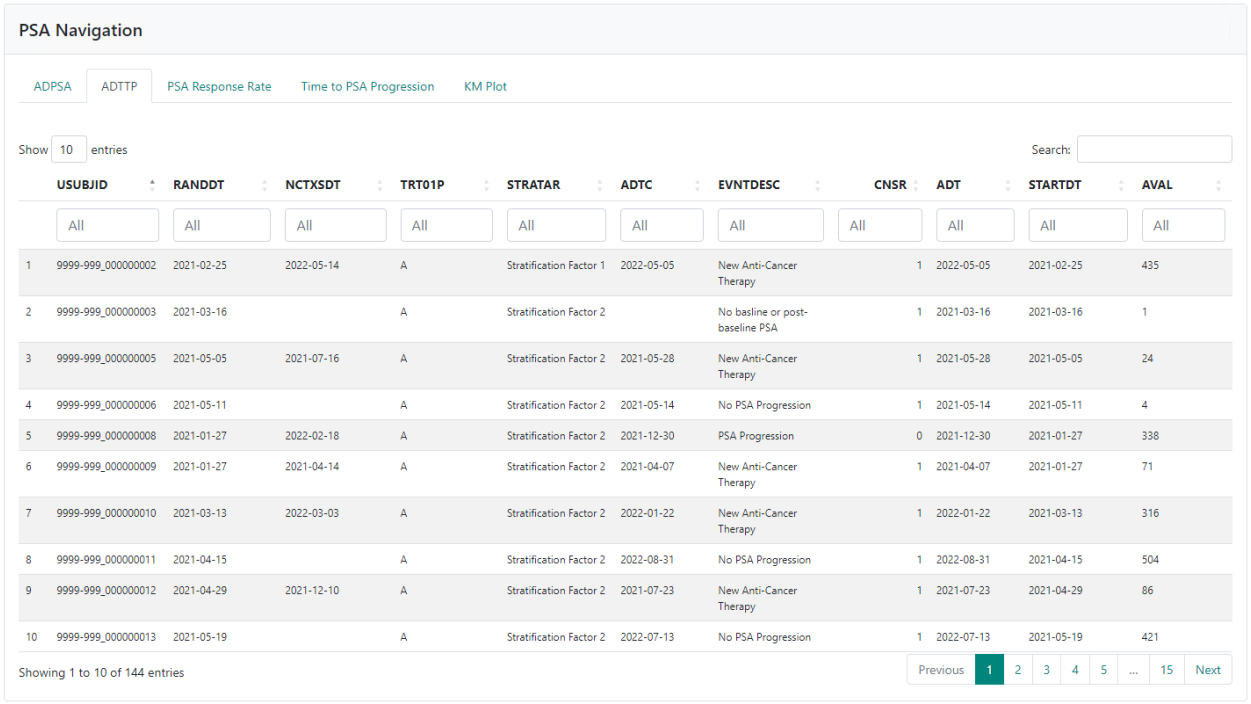
STEP 2: DATA PROCESSING

Before obtaining any meaningful results from the Shiny app, derivation and data processing are necessary. Once the ADSL and LB datasets are successfully loaded and the “Load Data” button is clicked, the app displays content of derived PSA data in ADPSA and ADTTP tabs.

The panel design is consistent with SAS programming in regular analysis and reporting process. The tab ADPSA includes all valid PSA measurements and confirmed PSA decline response; the tab ADTTP contains time to PSA progression parameter. To improve the reusability of code and to facilitate its functioning, the app processes data following CDISC guideline, which can be comparable with SAS datasets, such as converting Date/Time of specimen collection (LBDTC) to analysis date (ADT) and deriving analysis value (AVAL).

The R package *DT* provides an R interface to the JavaScript library *DataTables*. The *RenderDataTable()* function uses R to process the data object on the server side, and makes a reactive version that returns a data frame, which is rendered with the *DataTables* library. R data object can be displayed as tables and provides filtering, pagination, sorting, and many other features in the table.

Figure 3 shows example data records from ADTTP supporting the time-to-event analysis.



	USUBJID	RANDDT	NCTXSDT	TRT01P	STRATAR	ADTC	EVNTDESC	CNSR	ADT	STARTDT	AVAL
	All	All	All	All	All	All	All	All	All	All	All
1	9999-999_000000002	2021-02-25	2022-05-14	A	Stratification Factor 1	2022-05-05	New Anti-Cancer Therapy	1	2022-05-05	2021-02-25	435
2	9999-999_000000003	2021-03-16		A	Stratification Factor 2		No basline or post-baseline PSA	1	2021-03-16	2021-03-16	1
3	9999-999_000000005	2021-05-05	2021-07-16	A	Stratification Factor 2	2021-05-28	New Anti-Cancer Therapy	1	2021-05-28	2021-05-05	24
4	9999-999_000000006	2021-05-11		A	Stratification Factor 2	2021-05-14	No PSA Progression	1	2021-05-14	2021-05-11	4
5	9999-999_000000008	2021-01-27	2022-02-18	A	Stratification Factor 2	2021-12-30	PSA Progression	0	2021-12-30	2021-01-27	338
6	9999-999_000000009	2021-01-27	2021-04-14	A	Stratification Factor 2	2021-04-07	New Anti-Cancer Therapy	1	2021-04-07	2021-01-27	71
7	9999-999_000000010	2021-03-13	2022-03-03	A	Stratification Factor 2	2022-01-22	New Anti-Cancer Therapy	1	2022-01-22	2021-03-13	316
8	9999-999_000000011	2021-04-15		A	Stratification Factor 2	2022-08-31	No PSA Progression	1	2022-08-31	2021-04-15	504
9	9999-999_000000012	2021-04-29	2021-12-10	A	Stratification Factor 2	2021-07-23	New Anti-Cancer Therapy	1	2021-07-23	2021-04-29	86
10	9999-999_000000013	2021-05-19		A	Stratification Factor 2	2022-07-13	No PSA Progression	1	2022-07-13	2021-05-19	421

Figure 3. Example data from ADTTP

STEP 3: ANALYZE PSA RESPONSE RATE

In the PSA Response Rate tab, a summary table is designed to display basic statistics such as frequency count by treatment group, number of PSA responses, PSA response rate (%), difference in PSA response rate and its confidence interval.

The stratified Miettinen and Nurminen method is used for comparison of the PSA response rate between two treatment groups using participants with at least one PSA measurement at baseline. The difference in PSA response rate and its 95% confidence interval from the stratified Miettinen and Nurminen method with strata weighting by sample size is reported. A customized R package is developed to run the analysis of equivalence.

Figure 4 displays a summary table of PSA response rate from derived ADPSA dataset.

PSA Navigation

ADPSA | ADTTP | PSA Response Rate | Time to PSA Progression | KM Plot

Show 10 entries Search:

	Treatment	ITT population	PSA with Baseline Value	Number of PSA Response	PSA Response Rate (%)(95% CI)
1	A	71	69	58	84.1 (75.5, 92.7)
2	B	73	72	47	65.3 (54.3, 76.3)
3					
4	Difference in % vs. Treatment B				
5					Estimate (95% CI): 8.3 (6,10.6)
6					p-Value: 0.14124

Showing 1 to 6 of 6 entries Previous 1 Next

Figure 4. Summary table of PSA response rate

STEP 4: ANALYZE TIME TO PSA PROGRESSION

In the data processing step, the time to event analysis dataset is available with endpoint of interest, event description and censor description. This step returns an analysis table to display time-to-event statistics, including number of events (PSA progression, %), the magnitude of the treatment difference, hazard ratio (HR), and its confidence interval.

The magnitude of the treatment difference HR between the treatment groups is assessed using a stratified Cox proportional hazard model with Efron's method of tie handling. The stratification factors used for randomization is applied to the stratified Cox model.

The *survival* package is the cornerstone of R survival analysis edifice. *Surv(time, status)* builds the basic survival analysis data structure, containing failure time and censoring information. *Time* records survival time; *status* indicates whether the subject's event is observed or the survival time is censored. The function *coxph(formular, data, method)* is used to compute the Cox proportional hazards regression model in R. An analysis table of time to PSA progression is shown in Figure 5.

PSA Navigation

ADPSA | ADTTP | PSA Response Rate | Time to PSA Progression | KM Plot

Show 25 entries Search:

	var	A	B
1	ITT Population	71	73
2			
3	Number of Events	3 (4.2)	8 (11)
4	Event: PSA Progression	3 (4.2)	8 (11)
5	Number of Censors	68 (95.8)	65 (89)
6			
7	Person-months	454.1	566.5
8	Event Rate / 100 Person-months	0.7	1.4
9	vs Treatment B		
10	Hazard Ratio (95% CI)	0.43(0.11, 1.6)	
11	p-value	0.097	

Showing 1 to 11 of 11 entries Previous 1 Next

Figure 5. Analysis table of time to PSA progression

STEP 5: DISPLAY KAPLAN-MEIER PLOT

Survival data incorporates data from subjects who have the event of interest and those being censored to estimate the survival probabilities at various time points. Kaplan-Meier curves in each treatment group are generated for time to PSA progression.

The *survfit()* function is used to produce the Kaplan-Meier estimates of the probability of survival over time. The *survminer* R package provides functions for facilitating survival analysis and visualization.

Figure 6 shows the KM curve by using *ggsurvplot()* function, which is called to draw survival curves by treatment with the 'Number at risk' table.

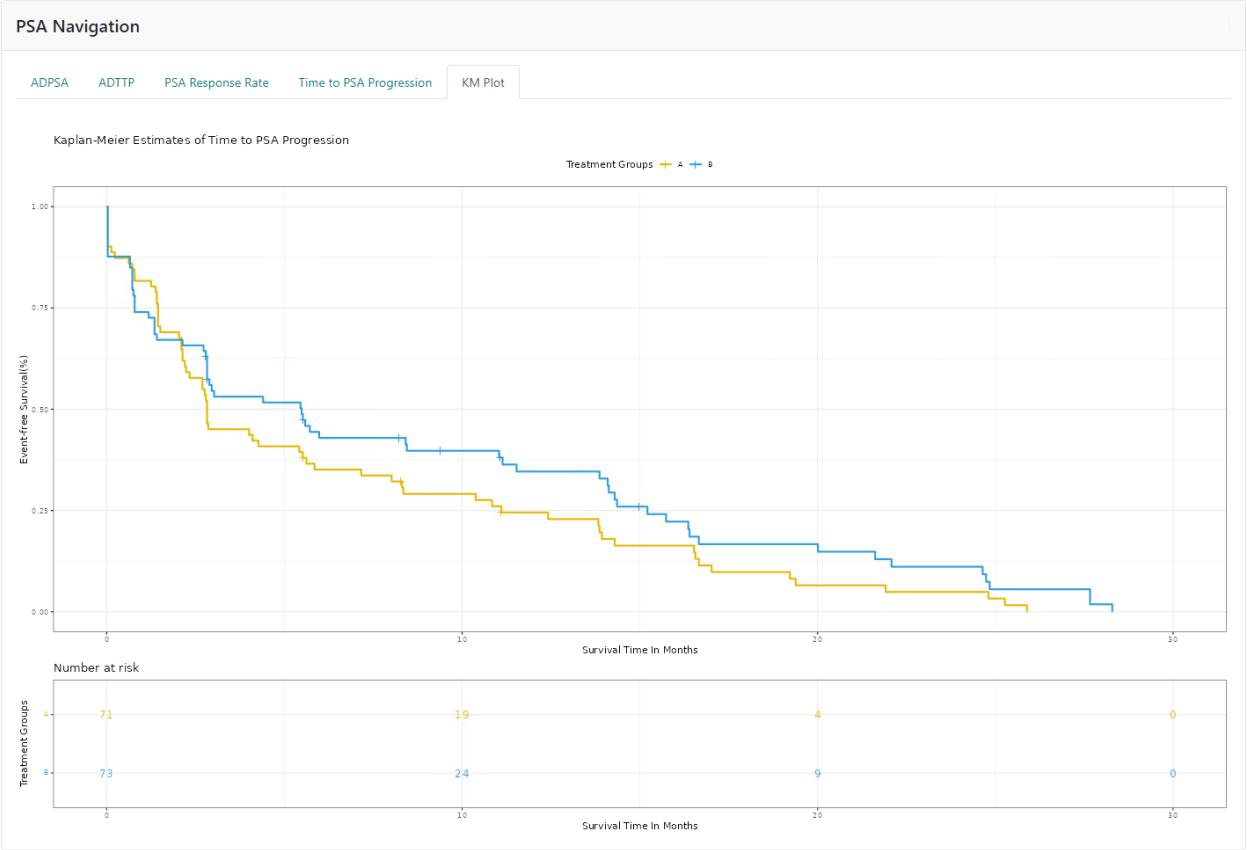


Figure 6. KM plot using ggsurvplot()

CHANLLENGES AND LIMITATIONS

Building up a Shiny app to visualize PSA data is feasible and has lots of strengths. However, there is still a lot of room for this R Shiny app improvement. As of now, this Shiny app is designed only for few prostate cancer trials internally. New and enhanced features, and design optimization are necessary for wider and more comprehensive application.

The R Shiny tool can be improved by increasing flexibility and agility. For instance, currently the R Shiny tool supposes PSA response rate as a minimum PSA decline of at least 50% and it is confirmed by a second PSA value 3 or more weeks later. Investigators may require higher measures of PSA changes (i.e., 75% decline, 90% decline) or longer response duration (i.e., at least 4 weeks apart).

In addition to the basic PSA analysis, more functionalities of statistical summaries and data visualization, including subgroup table and graph, waterfall plot of the best percentage change in PSA, can be implemented to the Shiny app.

CONCLUSION

In prostate cancer trials, PSA results by laboratory is measured and monitored throughout the study. R Shiny makes it easy to build an interactive web app to streamline the analysis and reporting process, and to explore PSA data. If users have access to the analysis dataset ADSL and SDTM LB domain, they can use this R Shiny app and quickly conduct PSA analysis without re-running complex SAS code each time, and the results can be displayed in a more visualized manner.

REFERENCES

Rawla P. 2019. “Epidemiology of Prostate Cancer”. *World J Oncol*. 2019 Apr;10(2):63-89. doi: 10.14740/wjon1191.

U.S. Cancer Statistics Working Group. “U.S. Cancer Statistics Data Visualizations Tool, based on 2022 submission data (1999-2020): U.S. Department of Health and Human Services”. *Centers for Disease Control and Prevention and National Cancer Institute.* June 2023. Available at https://www.cdc.gov/cancer/dataviz.

National Cancer Institute. “Prostate-Specific Antigen (PSA) Test.” March 11, 2022. Available at https://www.cancer.gov/types/prostate/psa-fact-sheet#how-is-the-psa-test-used-in-men-who-have-been-treated-for-prostate-cancer

Miller PD, Eardley I, Kirby RS. 1992. Prostate specific antigen and bone scan correlation in the staging and monitoring of patients with prostatic cancer. *Br J Urol*, 70(3):295-8. doi: 10.1111/j.1464-410x.1992.tb15734.x.

ACKNOWLEDGMENTS

Author is highly indebted to Nicole Zhang for her inspiring and extraordinary support on the paper. The author would also like to express great appreciation to Amy Gillespie for her professional guidance and constructive recommendations on the example data. Special thanks are extended to Jinchun Zhang and Danfeng Fu for their valuable advice and assistance in R programming.

CONTACT INFORMATION

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

Tongtong Weng
MSD China
tong.tong.weng@merck.com