Paper DV07

Clinical design - Visualization and analysis using the interactive application by R Shiny

Mahesh Divakaran, Genpro Research, Trivandrum, India
Athira Sudhakar, Genpro Research, Trivandrum, India

ABSTRACT

Shiny is an open-source R package, provides a great framework for interactive analysis and visualization of data using R. It has great power to make the design and interactive analysis, so it acquired popularity among R users. These applications are used by non-programmers to perform analysis, therefore, utilize the statistical capacities of R by naive clicks. So, its utilization has been on the rise over a couple of years. Through the paper, we demonstrate a Shiny application example in clinical trial design and data review. The application is aimed to visualize and analyze clinical trial-related tasks such as calculate sample size, calculating the probability of study success based on all kinds of information in a Bayesian framework, and visualization of ADaM datasets in a more efficient manner. In the future, Shiny apps will help in dynamic dataset creation, report making as well as visualization of clinical trial data.

1. INTRODUCTION

Clinical trials are used to determine the safety and efficacy of a new drug. Clinical data is the most flexible and useful of all the data obtained. During the duration of a study, clinicians and statisticians constantly examine the clinical results to determine the optimal treatment. A graphical depiction of clinical data is frequently used to highlight safety or anomalies discovered in research, as well as to offer visual information on trends or to compare clinical tests among various treatment groups.

In clinical trials, laboratory samples are collected and analyzed on a regular basis to verify that patients are not experiencing any unexpected toxicities. Because of their impartiality, connection to organ function, and as a preliminary hint to impending clinical signs or symptoms, these data are significant. Laboratory safety evaluations are important inputs that aid investigators and sponsors in making informed treatment decisions.

Tables, listings, and figures are used to present common clinical data. Tables may indicate the proportion of patients with each sort of clinical abnormality, as well as the frequency with which patients' status changes from normal to abnormal or abnormal to normal. The data may be sorted by analysis days and visits in the listings. Line graphs depicting analytical values over time and boxplots depicting the overall distribution and skewness of the data are the most used figures.

Statistical programming has historically relied on printable TLFs providing all the study's information. We can create stunning static graphs with SAS ODS graphics and SGPLOT, but these plots lack the capability of interactivity. A good visualization may reduce hundreds of pages of data/listings into a single figure, and adding interactivity makes the results much more entertaining and usable for reviewers. Reviewers can investigate trends and anomalies in the data using interactive visualization, which helps them better grasp the data's messages. Many other software products, such as Tableau/TIBCO Spotfire, give interactive visualization choices but are limited in their capacity to analyze data. We can use R Shiny1 to create interactive web applications without having to know JavaScript, HTML, or CSS. R Shiny is built for completely interactive visualization, with real-time updates whenever the user changes something. R's analytical capability mixed with Shiny's interactive visualization skills make them an excellent future choice.

This paper describes a R Shiny application that combines the ability to see single or numerous clinical tests with the ability to pick either a single subject or a user-selected sample of subjects to quickly find patterns, trends, and outliers. This allows reviewers and stakeholders to make early decisions about the trial's fate or any necessary course corrections. In a word, statistical programmers, statisticians, and investigators can utilize R statistical software to analyze, visualize, and swiftly deep dive into data in a user-friendly style.

2. APPLICATION LAYOUT

This application contains five main modules: Clinical Viz, Data Viz, Statistical Inference, Logistic Regression, and Power Calculation. **Figure 1** shows the landing page of the proposed application and using the file browser, we can upload the SAS datasets (with extension .sas7bdat) to create TFLs. Nine imaginary datasets were created in SAS, to mimic the standard ADaM datasets for the demonstration of the dashboard and after uploading the files the dashboard provide two options to view the raw data and generate TFLs (**Figure 2**). The ADSL dataset is shown in **Figure 3** is the view of the raw dataset, which is providing the options for customization. We can select how many rows and columns to display

and select which are the variables to display in the panel. The application also provides the option to download the data in CSV or PDF format using the power of R language.

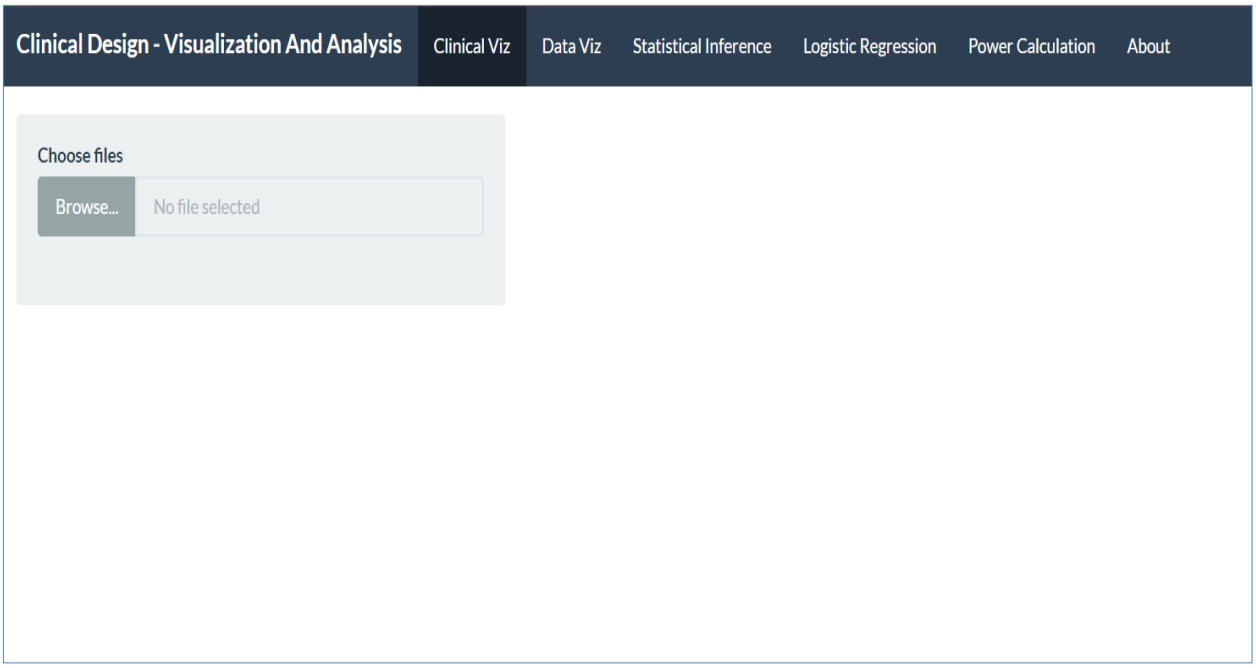


Figure 1: Landing page of the application

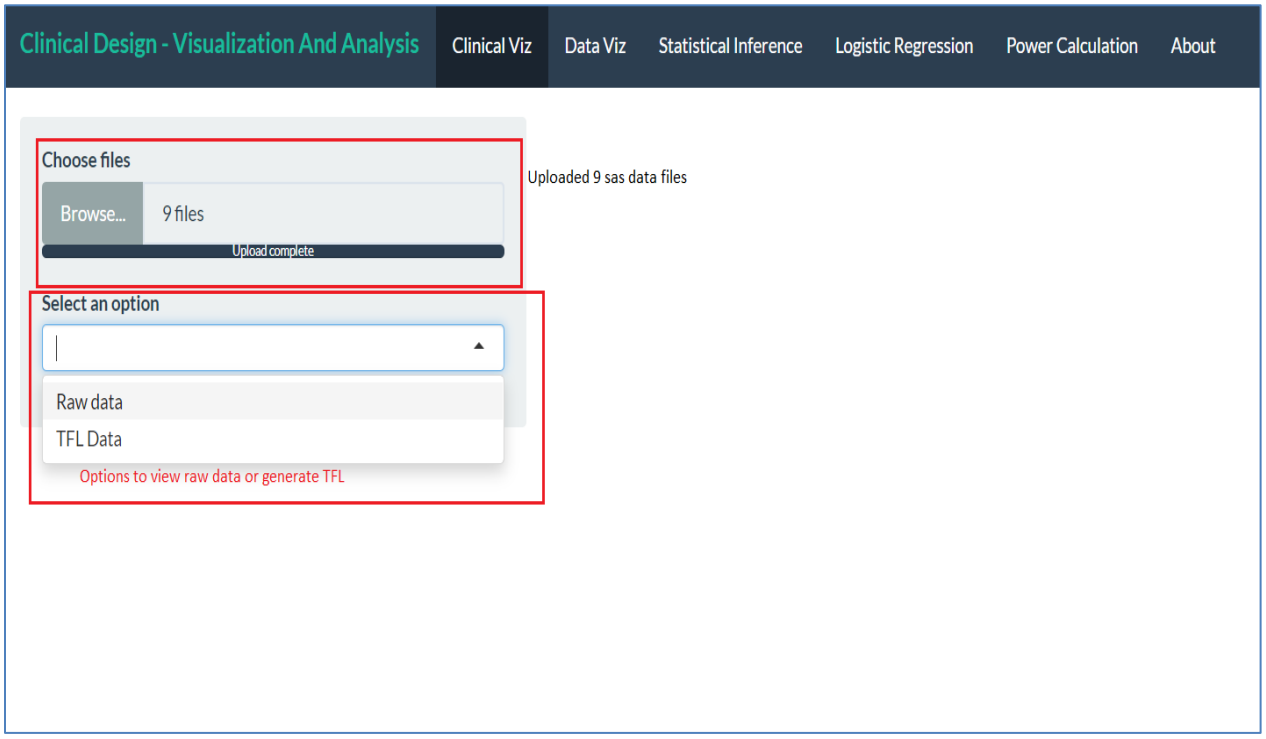


Figure 2: Options for viewing Raw data and TFLs

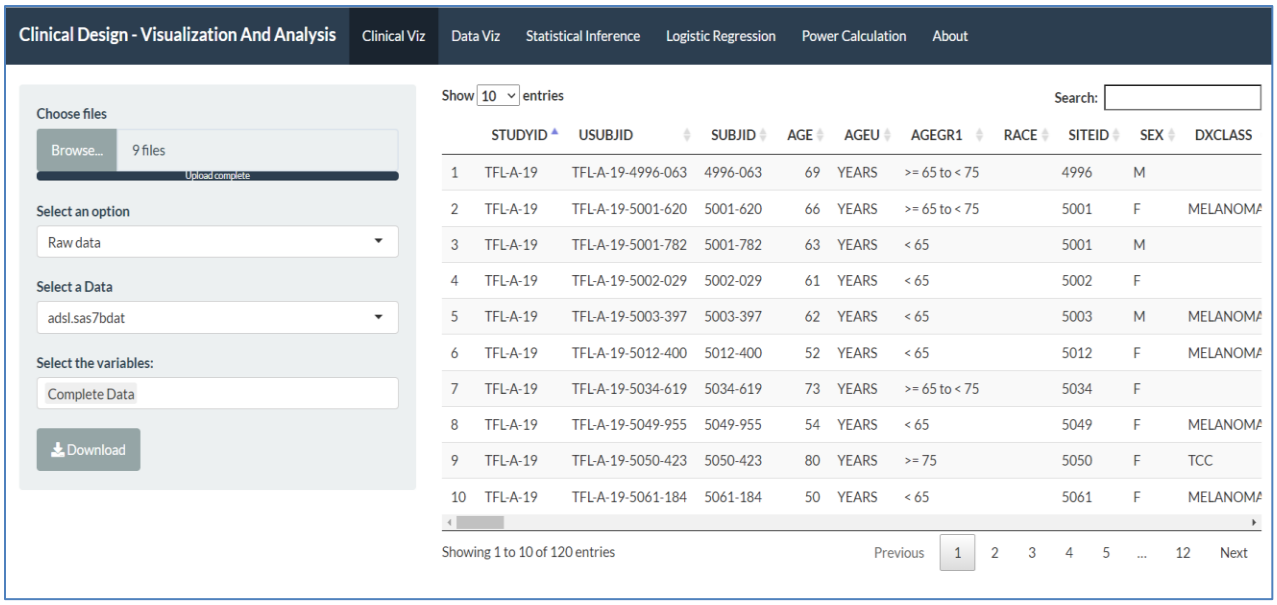


Figure 3: ADSL dataset

Once you are good with ADaM datasets for your study, we can simply create our TFLs with naive clicks. No need for lengthy SAS programs! If you select the TFLs from the drop-down, you can choose, tables, listings, or figures you want to generate. And from the selection, you will get more options to select which table, listing, or graph need to display. Also, we can download the outputs in supportive formats like PDF, Excel, Image, or DOC. In this visualization, the dedicated search bar provides the ability to search among the tables or listing instead of applying filter or writing conditions.

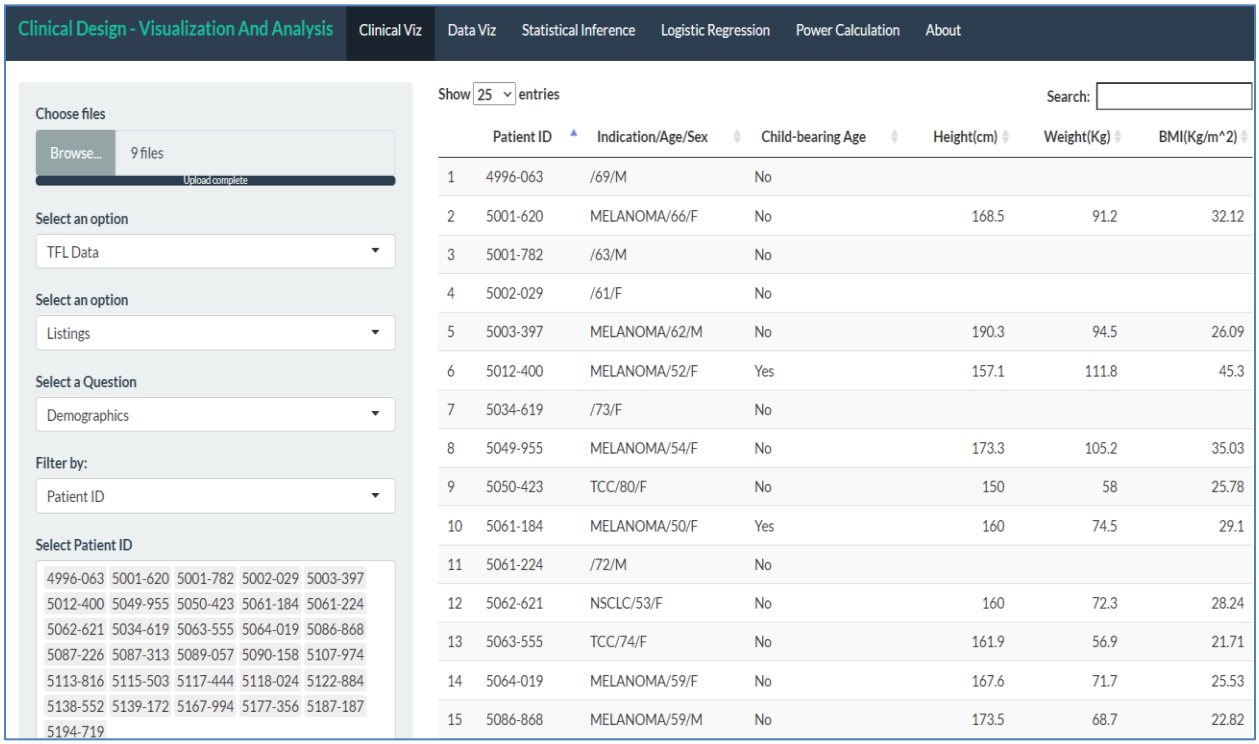
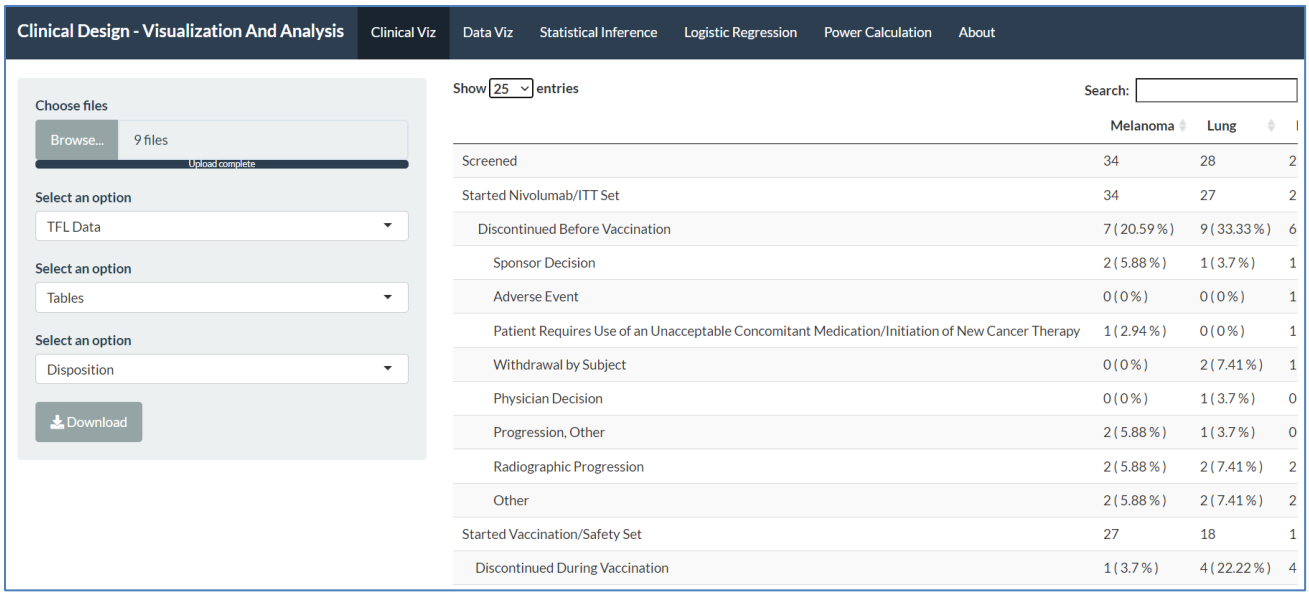


Figure 4: Sample Demographics Listing



	Melanoma	Lung	
Screened	34	28	2
Started Nivolumab/ITT Set	34	27	2
Discontinued Before Vaccination	7 (20.59 %)	9 (33.33 %)	6
Sponsor Decision	2 (5.88 %)	1 (3.7 %)	1
Adverse Event	0 (0 %)	0 (0 %)	1
Patient Requires Use of an Unacceptable Concomitant Medication/Initiation of New Cancer Therapy	1 (2.94 %)	0 (0 %)	1
Withdrawal by Subject	0 (0 %)	2 (7.41 %)	1
Physician Decision	0 (0 %)	1 (3.7 %)	0
Progression, Other	2 (5.88 %)	1 (3.7 %)	0
Radiographic Progression	2 (5.88 %)	2 (7.41 %)	2
Other	2 (5.88 %)	2 (7.41 %)	2
Started Vaccination/Safety Set	27	18	1
Discontinued During Vaccination	1 (3.7 %)	4 (22.22 %)	4

Figure 5: Sample Disposition Table

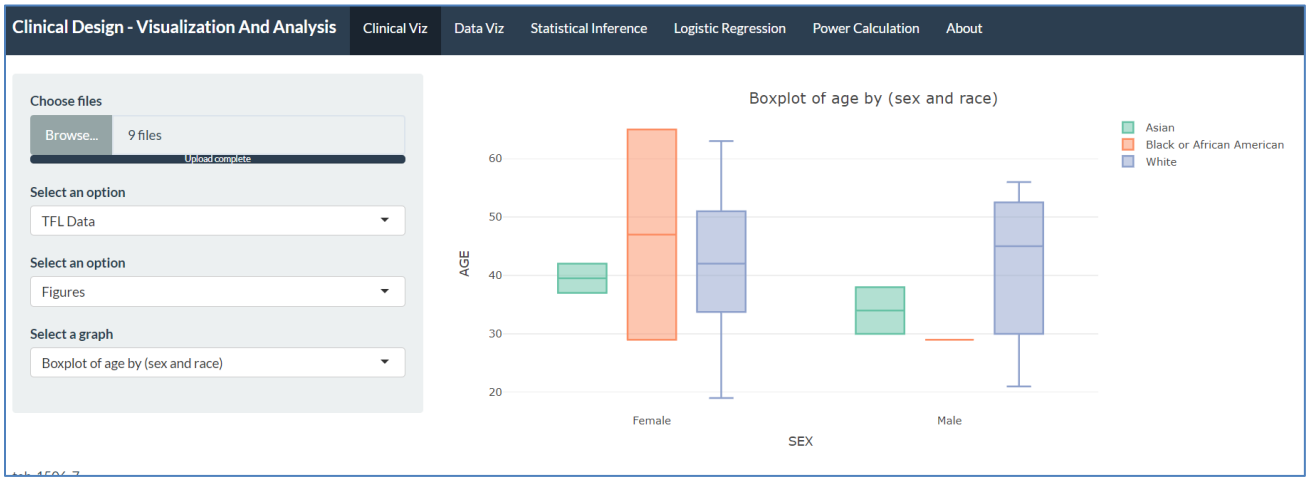


Figure 6: Sample Boxplot

All the figures generated by this shiny dashboard are created with the Plot.ly[2] package in R. Plot.ly's R graphing library makes interactive, publication-quality graphs. Examples of how to make line plots, scatter plots, area charts, bar charts, error bars, box plots, histograms, heatmaps, subplots, multiple-axes, and 3D (WebGL based) charts. Plot.ly R is free and open source, and you can view the source, report issues, or contribute on GitHub. The plot.ly package provides interactive visualization. So, all the plots generated using this application are interactive in nature. While hover over the graphs, we can see the corresponding values and we can also select a part of graph by selecting the area.

When we look at the second module data Viz, the application will display the five number summary and the basic plot of the uploaded dataset (CSV, XLSX, JSON or sas7bdat). **Figure 7** shows the simple visualization of a csv dataset. There are two tabs for a summary table and summary plot. Tabulations of summary statistics are an essential part of any clinical data analysis. The summary table gives the five number summary of the mean, standard deviation, median, and minimum and maximum for a variable of interest. The summary plot, **Figure 7**, allows the user to graphically view the summary variable of interest.

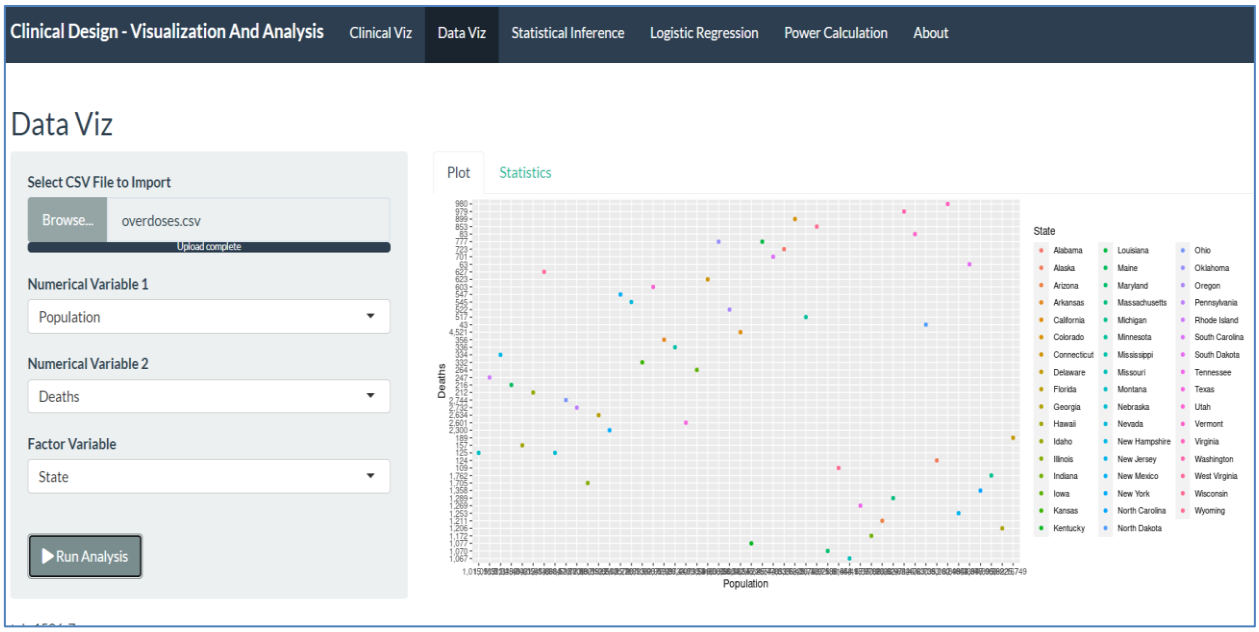


Figure 7: Data Viz Module Example

The third module in the application helps to perform the basic statistical inference tests. Inferential statistics uses a random sample of data taken from a population to make inferences, i.e., to draw conclusions about the population. In other words, information from the sample is used to make generalizations about the parameter of interest in the population. The two major tools in inferential statistics are confidence intervals and hypothesis tests. The module provides one sample tests, two sample tests, paired t-test, one variable and two variable population proportion tests and the variance tests. In this module you can specific the data and tune all the parameters for the test. The results will print in detail and describes with the density plot of the distribution.

Figure 8: Statistical Inference Module

The fourth module is Logistic Regression, a statistical model that in its basic form uses a logistic function to model a binary dependent variable, although many more complex extensions exist.[5] [Tolles et.al. 2016]. This Logistic

Regression model allows the user to select both the independent and dependent variables and perform the regression analysis. In the module, user can upload the dataset and the dataset can be seen in the right pane. In the next tab user can select the variables which need to be regressed. So once you finalize the dependent and independent variables, the analysis will perform and the detailed output will be displayed in the right pane.

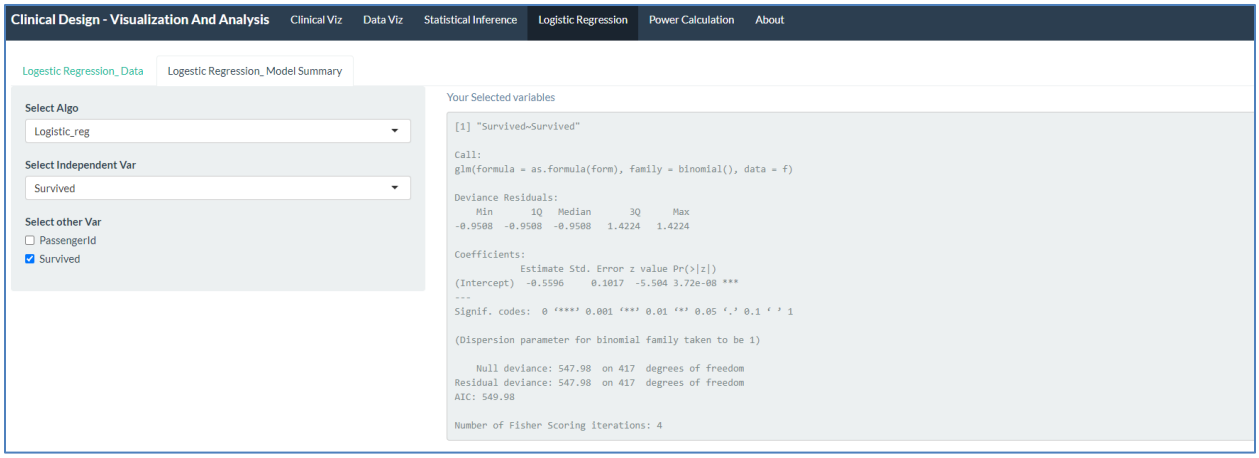


Figure 9: Logistic Regression

The final Module of the application is power calculation. As everyone knows, power calculation is one of the critical tools used in the life sciences. Understanding the direct relationship between sample size and power is important in interpreting the conclusions that can be drawn from a study. In this module you can compute the require sample size by tuning the parameters using simple scrolls.

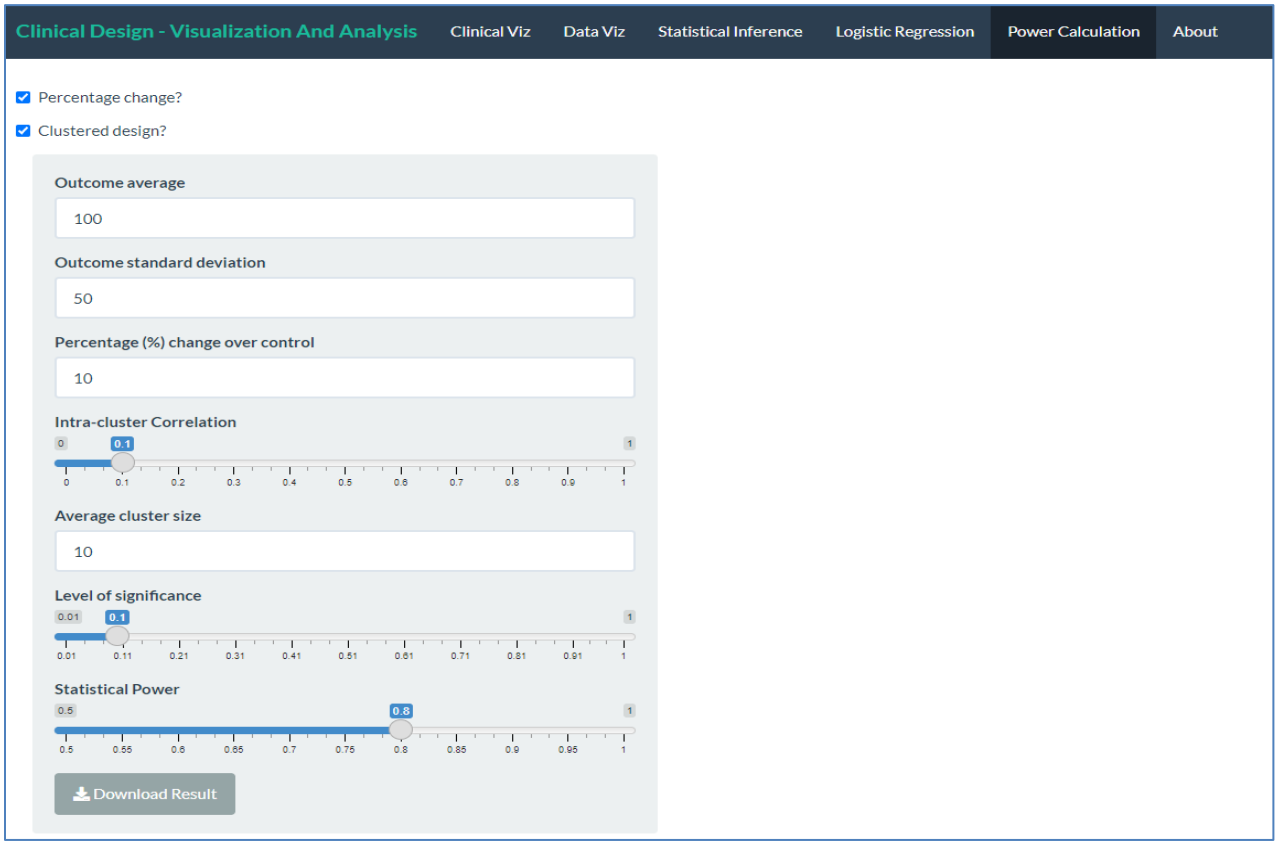


Figure 10: Power Calculation

3. CONCLUSION

Almost all clinical data analysis is exploratory in nature, and unanticipated data behaviors typically lead to new questions that are difficult to address using standard outputs. Data visualization technologies' ability to interactively choose both measures and subgroups allows researchers to swiftly answer these essential questions and make vital clinical trial decisions. Because of their dynamic nature, shiny apps are incredibly powerful. They can automate route duties and achieve great efficiency while still being simple to modify. Shiny is a user-friendly framework that allows R users to deliver data insights to decision-makers without the need for additional tools. In conclusion, the R Shiny application can be considered as a future clinical trial data analytics tool.

REFERENCES

1. R Shiny, https://cran.r-project.org/web/packages/shiny/shiny.pdf
2. Plot.ly, https://plotly.com/r/
3. shiny: Web Application Framework for R, https://cran.r-project.org/web/packages/shiny/index.html
4. The Shiny application is hosted at https://clinicalviz.shinyapps.io/ClinViz/
5. Tolles, Juliana; Meurer, William J (2016). "Logistic Regression Relating Patient Characteristics to Outcomes". JAMA. 316 (5): 533–4. doi:10.1001/jama.2016.7653. ISSN 0098-7484. OCLC 6823603312. PMID 27483067
6.

ACKNOWLEDGMENTS

We gratefully acknowledge the inputs and the support by our Limna Salim (Director – Biometrics, Genpro Research, TVM), Ajith Nair (Project Manager, Genpro Research, TVM) and Anoop P Ambika (Chief Executive Officer, Genpro Research, TVM).

CONTACT INFORMATION (HEADER 1)

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

Author Name	:	Mahesh Divakaran
Company	:	Genpro Research
Address	:	Second Floor, Nila Building Technopark Campuss
City / Postcode	:	Thiruvananthapuram , Kerala – 695581
Work Phone	:	+91.4714.026.700
Email	:	mahesh.divakaran@genproresearch.com
Web	:	www.genproresearch.com

Co-Author Name	:	Athira Sudhakar
Company	:	Genpro Research
Address	:	Second Floor, Nila Building Technopark Campus
City / Postcode	:	Thiruvananthapuram , Kerala – 695581
Work Phone	:	+91.4714.026.700
Email	:	athira.sudhakar@genproresearch.com
Web	:	www.genproresearch.com