Visualization & interactive application in design and analysis by using R Shiny

Baoyue LI, Eli Lilly & Company, Shanghai, China
Boyao SHAN, Eli Lilly & Company, Shanghai, China

ABSTRACT

Shiny is an open source R package that provides an elegant and powerful web framework for building web applications using R. It has great power to make the design and analyses visible and interactive with very user-friendly interface, therefore has gained extensive popularity among R users.

In this paper, we will show a few Shiny application examples in clinical trial design and data review. The first application is to calculate sample size of showing similar trend for China joining global pivotal study. The second application calculates probability of study success based on all kinds of information in a Bayesian framework. The final application helps quick review the ADaM datasets in a more efficient manner. Given all its great features, we believe Shiny will keep shining in many areas, within and out of drug development.

INTRODUCTION

Shiny is an open source R package that provides an elegant and powerful web framework for building web applications using R. It has great capability to make the design and analyses visible and interactive with very user-friendly interface, therefore has gained extensive popularity among R users. We listed here serval of the features mentioned from its developers:

- Build useful web applications with only a few lines of code – no JavaScript required.
- Shiny applications are automatically "live" in the same way that spreadsheet are live. Outputs change instantly as users modify inputs, without requiring a reload of the browser.
- Shiny user interfaces can be built entirely using R, or can be written directly in HTML, CSS, and JavaScript for more flexibility.
- Develop and redistribute your own Shiny widgets that other developers can easily drop into their own applications.

In drug development area, more and more pharmaceutical companies have started using Shiny for various purposes along the whole drug development cycle including from pre-clinical research to drug approval on market. The developed Shiny applications, with necessary instructions, can be easily used by both statisticians and non-statisticians with simple input to get reliable results. It is therefore becoming a convenient tool for statisticians to communicate with each other and with other functions, e.g. physicians. In this paper, we will demonstrate in the next section a couple of Shiny applications we developed within company that cover the design and analysis phases of a clinical trial.

CASE EXAMPLE SHARING

CASE EXAMPLE 1: SAMPLE SIZE CALCULATION OF SHOWING SIMILAR TREND FOR CHINA JOINING GLOBAL PIVOTAL STUDY

Joining global pivotal study is one of the most efficient paths of simultaneous development of new drugs for multinational pharmaceutical companies, and recent policies released from CFDA have shown a favor to this path. There are basically two scenarios in such circumstances: 1) all Chinese population that support local submission can join global pivotal study and 2) not all Chinese population could join global study due to e.g. narrow enrollment window, global overall strategy, etc, and an extended cohort has to be conducted to further enroll Chinese population. We also call scenarios 2 maximized extended enrollment (ME2) design. For either scenario, we need to show similar trend between Chinese and global populations in terms of efficacy, safety and other aspects. There is then a need for a statistically sound approach to calculate sample size in such circumstances, for design purpose as well as for regulatory purpose. Among quite a few other methods, our Shiny application uses the Method 1 from the "Basic Principles on Global Clinical Trials" guidance released by Ministry of Health, Labour and Welfare of Japan (MHLW) back in 2007. Specifically, to show similar trend according to this method, the sample size for Chinese population should satisfy

$$\Pr(D_C > \rho\, D_G) \geq 1 - \beta',$$

meaning that the treatment effect from Chinese population D_C should retain a proportion of ρ (usually 50%) of global treatment effect D_G , and the probability should be at least $1 - \beta'$ (usually 0.8). A more systematic and comprehensive discussion on sample size calculation was elaborated by Quan et al (2010).

Figure 1 shows the web interface of this application. This is for the time-to-event endpoint only, but applications for binomial and continuous endpoints are also available. The left two columns contain all necessary input information and the right part contains three outputs: elapsed time, a summary table and a plot. Except for a few cases that we can find analytical solutions, most of the calculations are based on simulations, which may take 20-30 minutes. Thanks to the HPC (high performance computing) framework, the running time could be dramatically shortened to around half a minute.
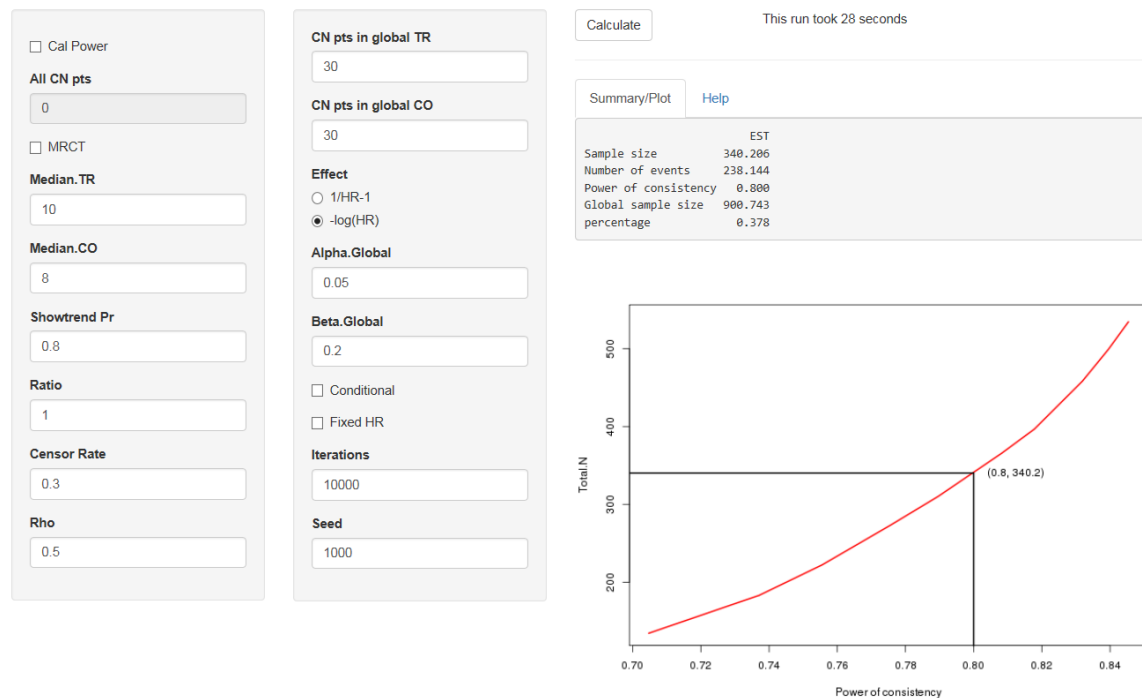

Figure 1. Web interface of show trend sample size/power calculation application

CASE EXAMPLE 2: CALCULATE PROBABILITY OF STUDY SUCCESS BASED ON ALL KINDS OF INFORMATION IN A BAYESIAN FRAMEWORK

Probability of study success (PrSS), also called averaged conditional power, assurance, etc., has been used assisting decision making in several circumstances, for instance commercial decision and estimating probability of technical success.

Suppose we have phase II study results comparing a new drug vs standard of care, with hazard ratio (HR) the primary efficacy measure. We now need to make a commercial decision whether we continue with a larger phase III registration trial. An intuitive question to help make this decision is "what is the probability that the phase III trial be positive based on currently available data?" which is actually PrSS. Note that this is different from the concept of "power" in that power calculation is under the assumption of a single value of true HR therefore does not take into account the uncertainty that the true HR is unknown. PrSS, on the other hand, calculates an averaged power over a distribution of HR based on currently available data. In the above situation, to calculate PrSS we first need to draw a distribution of HR reflecting the uncertainty towards the true HR based on the phase II study. Then for each possible value in this distribution, we can calculate a power and PrSS is the averaged power over the entire distribution. Not only the phase II data, other related information could be also incorporated to calculate PrSS, e.g. data of similar drug from other companies, information from literature, etc.

We developed a Shiny application for this purpose using a Bayesian framework, which is a natural and power approach to combine current and previous data together. This application can also update PrSS after interim analyses have been performed, e.g. the independent data monitoring committee (IDMC) recommend the trial to continue based on the futility/efficacy interim. In such case, the "observed" HR is actually censored, either left, right or interval censored, and the interim and final data are correlated.

Figure 2 shows the web interface of this Shiny application. The left-hand-side contains all input parameters and the right-hand-side contains a summary table and a density plot with various distributions including prior, likelihood, posterior, posterior predictive distributions. PrSS is calculated from posterior predictive distribution.
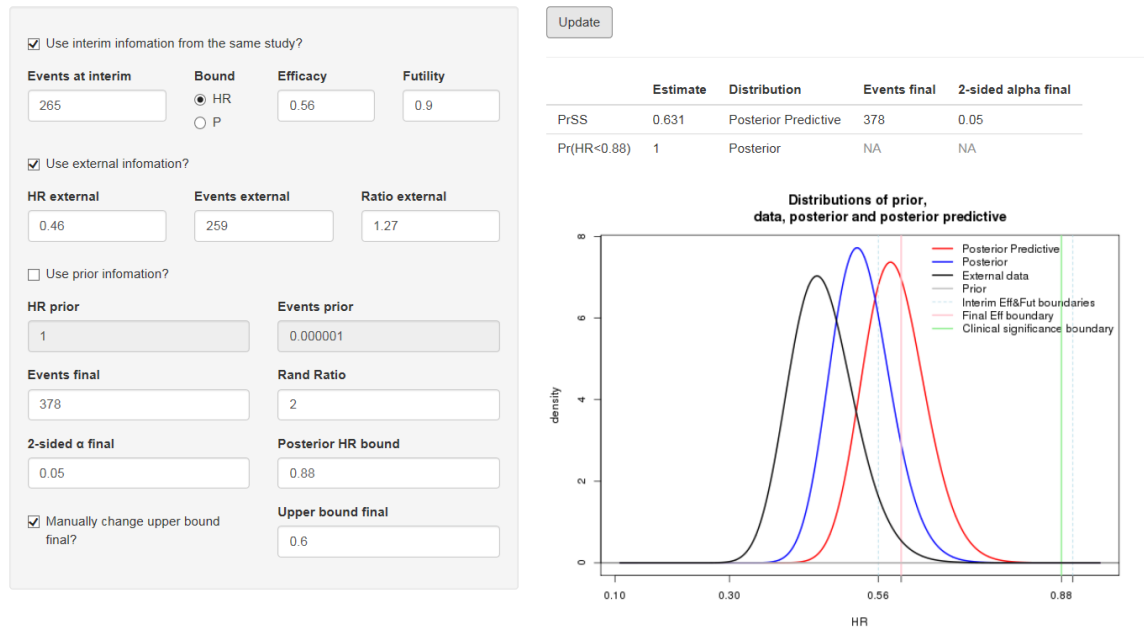

Figure 2. Web interface of PrSS calculation application

## CASE EXAMPLE 3: ADAM QUICK REVIEW

In clinical trials, ADaM datasets come after SDTM and are followed by tables, figures and listings (TFLs). Sometimes statisticians want to have a quick look at a very simple summary about the blinded data to check the enrollment information as well as data quality, while at that time point maybe just several key ADaM datasets are ready and no TFLs are generated. Therefore programmers may need to generate these simple summaries. Not complicated though, they are duplicated work of what programmers will do in later TFLs generating.

We build this Shiny application to give a quick review about ADaM datasets. It aims to save the time for the repeated checking and give a better show to answer the questions in study team discussion.
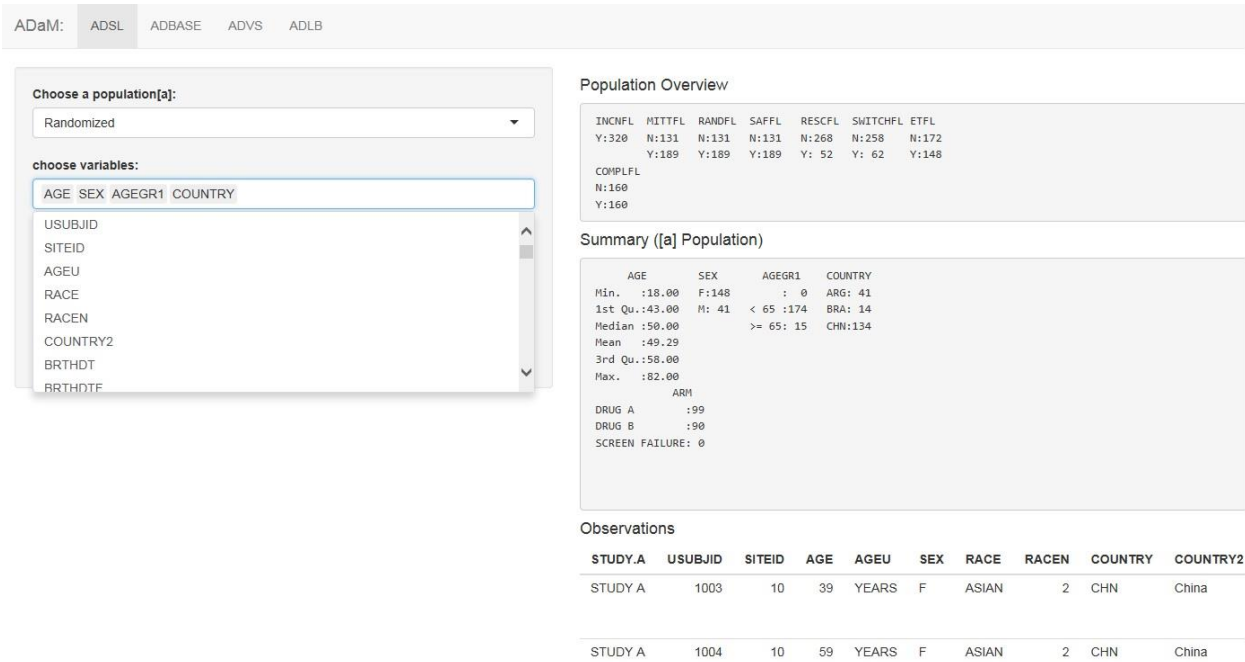
Figure 3. Web interface of ADaM review application

Figure 3 shows the interface of the ADaM review application. Left input part includes a list of ADaM datasets, population selection and interested variables. Right output part contains three parts. "Population Overview" gives the patients disposition based on different population flags. "Summary" gives descriptive summary of each selected variable based on selected population. "Observations" list all records in the ADaM dataset. From the ADSL dataset selected in figure 3, you can obtain patient level information including patient population, age, sex, race, country, tobacco use, etc. Similar summaries about baseline information such as weight, height and BMI could be obtained by selecting ADBASE dataset.
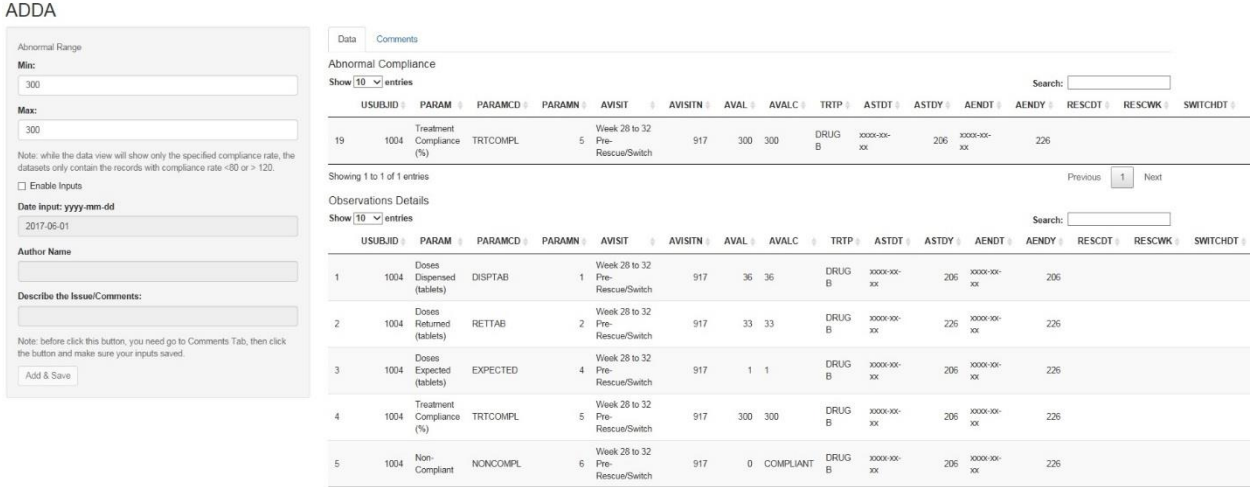
Figure 4. Web interface of ADDA compliance check application

Figure 4 shows another simple Shiny application to help us identify the potential data issues by listing the patients with high/ low compliance rate according to the specified range input. This Shiny application uploads the required ADDA ADaM dataset directly from the working space, which guarantees the synchronization with current work.

CONCLUSION

Shiny tool provides an efficient way of making the design and analysis visible and interactive. Users can deploy their own applications either in PC or on a Shiny server. The latter allows much convenience such as generating a web link

for broader use of the applications without requiring any analytical software/packages installed in the users' end. Further, it is also possible to implement the HPC environment into Shiny server to accelerate the running procedure of an application especially for that involves huge amount of simulations.

In this paper we showed three Shiny apps we developed, covering sample size calculation, probability of study success and AdaM data analyses, respectively. We have received positive feedbacks from within the company and will foresee more applications of Shiny throughout all drug development phases, as well as many other areas.

REFERENCE

Shiny Home Page : http://shiny.rstudio.com/
Shiny tutorial at github: http://rstudio.github.io/shiny/tutorial/
Shiny dash board information: http://rstudio.github.io/shinydashboard/
High Performance Computing Documentation, Enterprise RStudio Server, at Lilly inner website
Ministry of Health, Labour and Welfare of Japan. Basic Principles on Global Clinical Trials, September 28, 2007.
Rufibach, Kaspar, Paul Jordan, and Markus Abt. Sequentially updating the likelihood of success of a Phase 3 pivotal time-to-event trial based on interim analyses or external information. Journal of biopharmaceutical statistics 26.2 (2016): 191-201.
Quan, Hui, et al. Sample size considerations for Japanese patients in a multi-regional trial based on MHLW guidance. Pharmaceutical Statistics 9.2 (2010): 100-112.

CONTACT INFORMATION

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

    Name: Baoyue LI
    Enterprise: Eli Lilly & Company (China)
    Address: HKRI Taikoo Hui, Tower 1, 19F
    City, State ZIP: Shanghai, China
    E-mail: li_bao_yue@lilly.com

    Name: Boyao SHAN
    Enterprise: Eli Lilly & Company (China)
    Address: HKRI Taikoo Hui, Tower 1, 19F
    City, State ZIP: Shanghai, China
    E-mail: shan_bo_yao@lilly.com