PharmaSUG 2024 - Paper DV-396

Piloting data visualization and reporting with Rshiny apps

Yifan Han, Boehringer Ingelheim (China) Investment Co., Ltd.
Yun Ma, Boehringer Ingelheim (China) Investment Co., Ltd.

ABSTRACT

The pharmaceutical industry is increasingly adopting open-source R and shiny for web-based data visualization and reporting, given its numerous advantages such as easy comprehension, interactive analysis, real-time updates, etc. This paper will provide some real practical examples to illustrate these benefits, such as Exploratory Biomarker Analysis in Statistics, efficacy analysis plots in oncology, real-time safety data monitoring. In addition, we will share the learnings that our team went through in the last two years since we established a team of R programmers within our biometrics function. Our goal was to leverage the R open-source software language for data visualization and reporting across the company. Over this period, we've worked closely with various stakeholders within the medical functions to align our goals and gather user requirements. Starting from single app to an application warehouse, known within the company as Data Access and Dynamic Visualization for Clinical Insights (DaVinci). This warehouse hosts over 15 different applications, ranging from heavily statistical apps for early clinical development trial planning to apps with intuitive features for safety monitoring and oncology efficacy analysis. Our team learned from many challenges when started few years back, such as IT Infrastructure, Online Systems, user experience, etc. We will share the above points with more details in the presentation. These challenges highlighted the need for careful planning, robust design, and thorough testing when developing data visualization modules.

INTRODUCTION

Statistical programmers in the pharma industry have been using the SAS software language to handle clinical trial analysis and static reporting activities for more than two decades as good practice and to meet regulatory requirements. Until recently the hot topic of AI/ML and the popularity of open-source software, such as R, Python, etc, gained a lot more attention both academically and within the health care industry. It is foreseeable that in the future, we could evolve into language agnostic software developers or data scientists to embrace all latest development in technology of big data, AI/ML. The data analysis and reporting could come to a true modernized stage that enhances the efficiency, accuracy, and comprehensibility of data analysis in clinical trials and healthcare research, providing a significant upgrade over traditional methods and bring the unmet medical needs to patient faster and better.

That been said, it also means that the traditional statistical programmer will need learn together with our stake holders from medicine functions to fully understand what and how the new technologies, such as interactive graphics and tables will change the current working habit or process to fully meet GxP requirement in trial conduct and reporting activities. Change management should put into place and carefully plan to mitigate the potential risks at piloting stage and the future scale up goal. This paper will provide some real cases on data visualization in oncology therapeutic area and our lesson learned in the last few years to overcome the challenge in cloud platform, user experience/interface design and more.

DATA ACCESS AND DYNAMIC VISUALIZATION FOR CLINICAL INSIGHT (DAVINCI)

We all know that Leonardo da Vinci, the famous Italian painter, craftsman, sculptor, architect, and his famous paint “*Last Supper*” and “*Mona Lisa*”. With a nice wish to achieve the artistic touch in our new way of delivery patient data and analysis to our stake holders, and ultimately bring the innovative treatment to patient faster and safer, we named our data visualization team, “DaVinCi” and formed the team more than three years ago. The starting point was some basic data plots and patient profile listings to test the water and to gather user feedbacks. This was achieved with a single dedicated web server to host the apps, then moved to the openshift container platform on AWS with companywide cloud initiatives. With the

mindset and vision to automate current manual legacy system setups, we have set three long term goals for the team: 1. Enable real time data access; 2. Integrate data visualization and analysis with user practice process; 3. Module extendable. The following are some real examples from oncology trial data monitoring and analysis based on inputs from our biometrics and medicine colleagues who are also passionate to revolutionize away from the old fashioned static and tons of pages in the data review process.

EXAMPLES FROM ONCOLOGY TRIAL

Apps

Multiple Apps were developed and deployed under DaVinci ecosystem. Then they were put into practice and served teams for different purpose.

EBAS

The first one is exploratory biomarker analysis in statistics (EBAS). You can see some features from **Figure 1**. It can provide basic statistics, line plots and association heatmap to translational medicine colleagues to visualize and analysis the vast amount of biomarker data. The team receives biomarker data in multiple batches as the trial is ongoing. After each batch, the team performs “interim” biomarker analyses. EBAS can help the team to determine how much and which biomarker data are available. Plus, EBAS can directly analyze the biomarker and clinical data, such as, can we potentially predict response based on baseline values? (e.g., Identify enriched populations?) Is there a treatment effect on the biomarkers?

Figure 1. EBAS

RENOVATE

RE*CIST (v1.1) based data mo****N****it****O****ring* ***V****isualization and* ***A****nalytics Tool for* ***E****fficacy*

RENOVATE is an effective tool to generate evidence from POCP (proof of clinical principle) studies. It can help fast decision making for late phase development. Team uses it to visualize the response data, monitor study data during the trial conduct and detect early efficacy signal, etc.

Edit check is also an important part in RENOVATE. Major types of edit check is related to dates, response, and examination method.

In Figure 2., it showed the interactive plots of tumor response, which seems as comparable to our legacy figure products, but with click features to greatly improve data review efficiency.

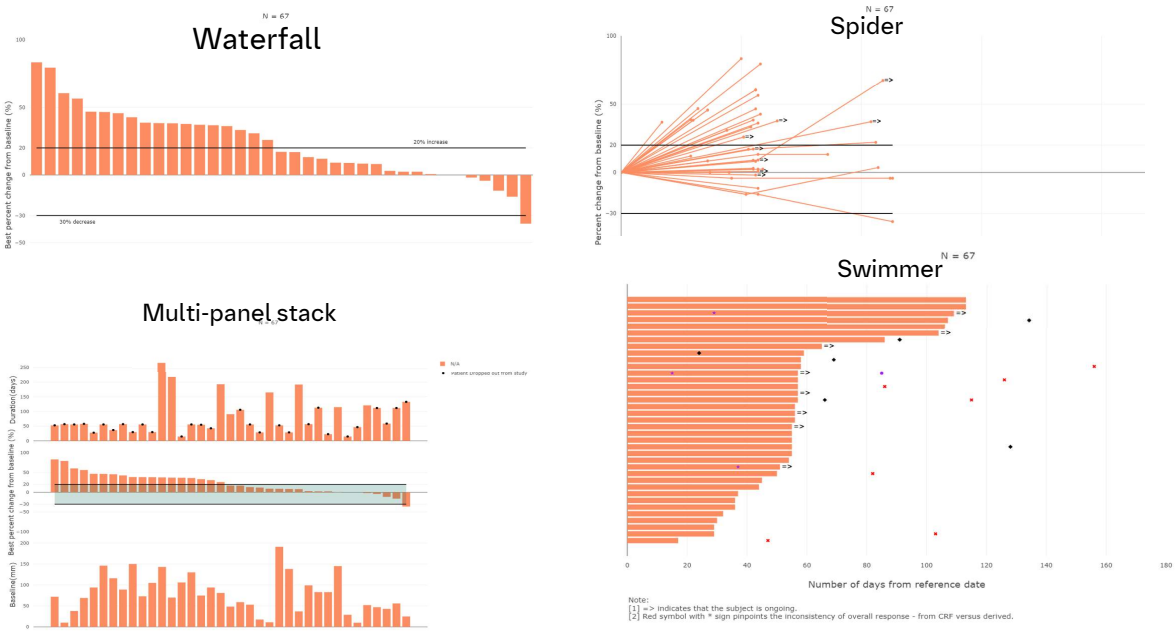


Figure 2. Tumor response plots

MODULE DAVINCI

As all of us in the statistical programming profession knew that good programming practice with SAS macros are efficient ways to build standards and be flexible for the users. The same concept can be applied to R programming in open-source environment. With rich resource from github and other places online, such as from Pharmaverse, DaVinci also shares the “Module” mindset. In our vision, when fully developed and integrated, users are free to pick any modules of their interest and integrate them into one app or use them separately but fit in their general practice.

Modules could work independently in general but sometimes were linked per user preference. For instance, when you overview the patient journey in timeline module, finding one specific case very interesting and want to know more details, you could simply click that patient timeline and then it will go to corresponding patient profile modules. The following module examples include Figure 3 SDTM data listing, Figure 4 Clinical timeline, Figure 5 Patient profile as well as Figure 6 Abnormal lab figure, which helps reviewer

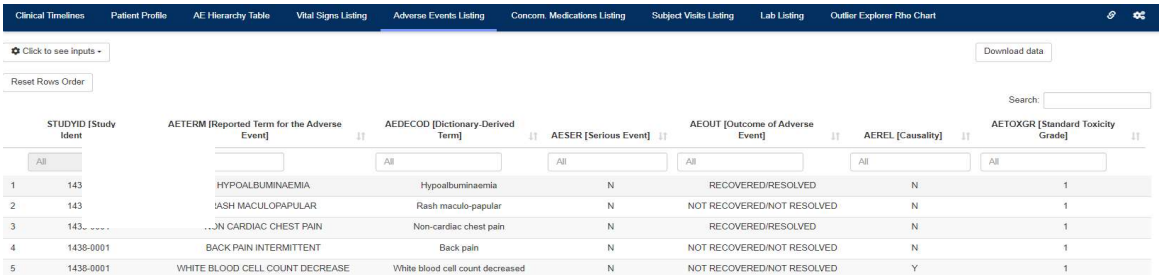


Figure 3 SDTM data listing

to quick see the outliers in the figure.

Figure 4 Clinical timeline

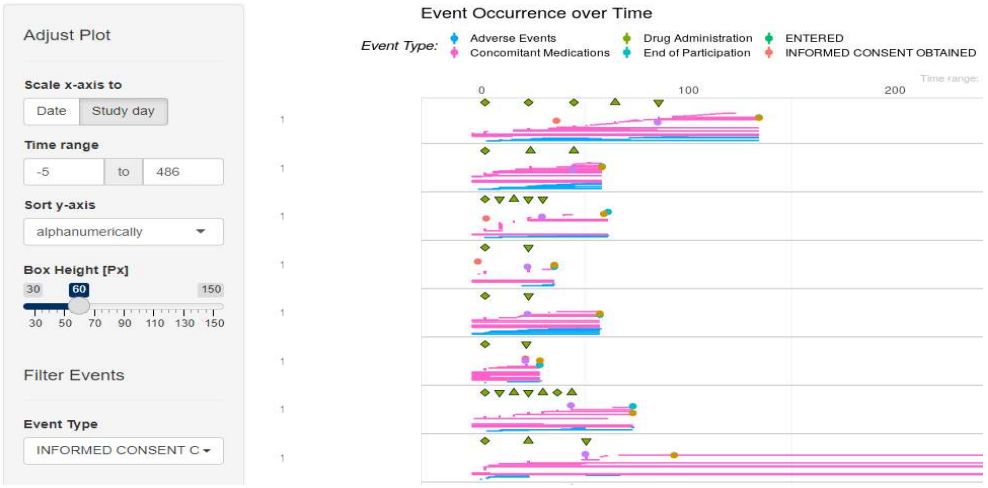


Figure 5 Patient profile

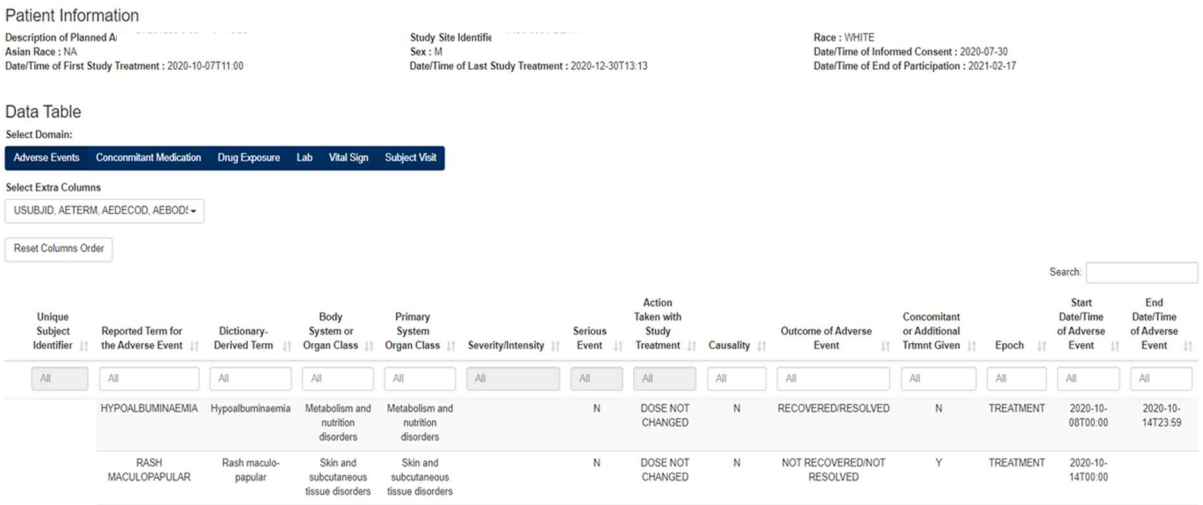
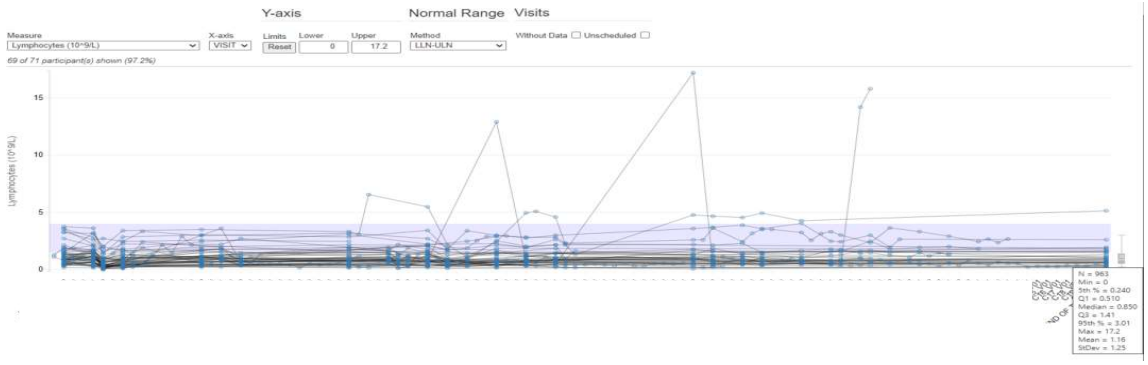


Figure 6 Abnormal lab figure



Finding the Best Practice in Early Phase Oncology Studies

With the above modules as bases to build flexible yet innovative apps and with multiples trial teams' inputs, oncology TA initially led the way of introducing new way of data review in trial practice and the other TAs followed.

Usually, when you have a good compound in the TA of oncology demonstrating promising cancer treatment effects, you could have multiple studies are simultaneously enrolling patients and need trial data review quite frequently. In the phase I stage, these studies shared very similar designs such as dose escalation and expansion, which enable us to build up App for each study sharing common modules and, in the meantime, to preserve the flexibility in adaptation of individual trial specific needs.

Our intention is to promote these Apps to the entire compound medicine team, not limiting its use within biometrics department. Our target users take different roles in clinical trial and have their own focus. Clinical Program Leads generally stay at high level and have the overview. Clinical timeline would fit into them purpose more so that they can have a quick glance at what happened overall. Study physicians especially care about adverse events and lab abnormalities. It makes modules 'adverse events tables and lab values outliers' valuable in our App.

CHALLENGE AND SOLUTION

No pain, no gain. That's true for our team since the beginning. We dealt not only with technical challenges, but also the usual working environment change as we all experienced during COVID period. What's not breaking us makes us stronger. We've learned quite a lot together as a team in different phases of the projects and achieved nicely to move forward. Here are some discussing points worth to consider.

Transformation Challenges

IT Infrastructure:

1. **Data Management:** Handling large volumes of data can be challenging. The infrastructure must be robust enough to process and visualize large datasets without performance degradation.
2. **Security:** Ensuring the security of sensitive data is a major concern. Need robust security measures to protect data from unauthorized access and breaches.
3. **System Performance:** Capable of handling complex computations and delivering visualizations quickly. Slow performance can hinder the user experience and reduce efficiency.

Online Systems:

4. **Data Accessibility:** Ensuring that data is easily accessible to authorized users can be a challenge. This includes providing access to real-time data and enabling users to interact with the data.
5. **System Integration:** The data visualization module should be able to integrate seamlessly with existing systems and data sources. This can be challenging if the systems use different formats or protocols.
6. **Scalability:** The system should be able to handle an increasing amount of data and users without performance issues.

User Experience:

7. **Usability:** The data visualization module should be easy to use, even for non-technical users. This includes intuitive navigation, clear labels, and helpful tooltips.

8. **User Interface Design:** The design of the module should be visually appealing and should effectively communicate the data. Poor design can make the module difficult to use and understand.
9. **User Interaction:** The module should allow users to interact with the data, such as by filtering, sorting, and drilling down into the data. Implementing these features in a user-friendly way can be challenging.
10. **Customization:** Users may have different needs and preferences, so the module should be customizable. However, providing customization options can add complexity to the design process.

With technology evolving so fast in the last few years, by doing due diligent work internally cross functions and collaborating with external partners on providing latest business solutions, the team has achieved significant milestones along the way and find proper solutions on some of the challenges and progressed well in the overall pictures. As one example illustrated in the Figure 7 System Process Improvement perspective. We reduced the app go live time from two weeks with uncertainties to almost one click.

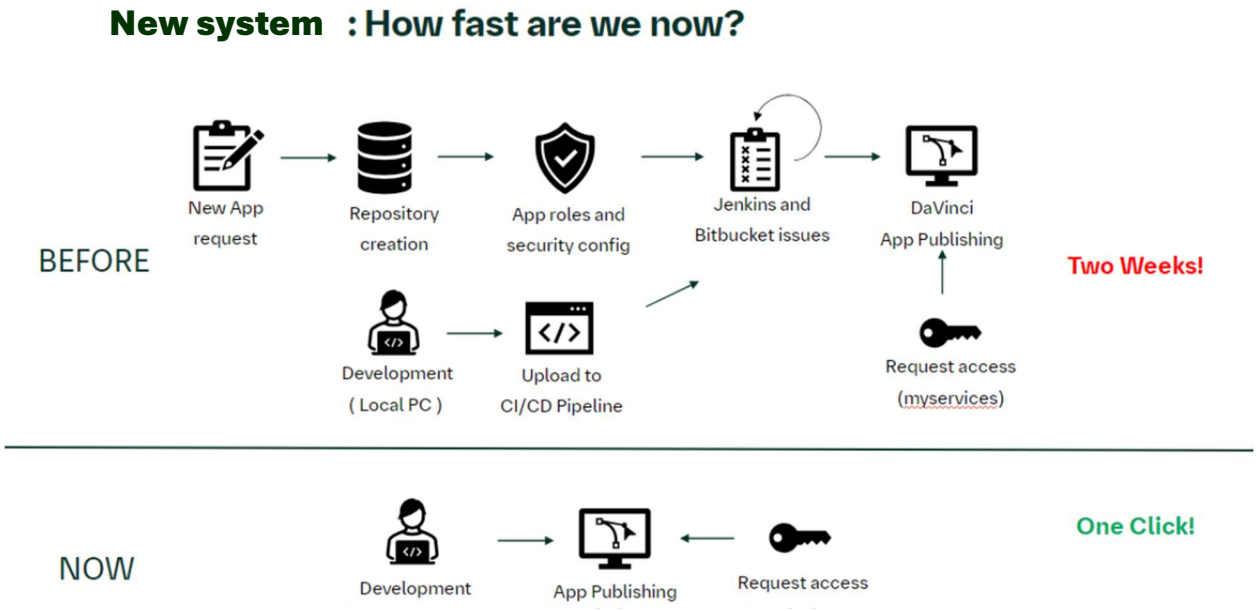


Figure 7 System Process Improvement

CONCLUSION

Many of us within the industry believe that we can significantly enhance drug discovery and development and shorten the timeline using technology, such as AI/ML and open-source software. However, it's crucial not to overlook the importance of human knowledge in decision-making, especially when introducing new technologies to colleagues who may not be as tech-savvy. Effective and frequent cross function communication and understanding from each other are necessary. As the famous saying goes, 'Rome wasn't built in a day.' Progress may seem slow at first, but every firm step forward counts. Sometimes, even a few steps back can serve as positive indicators of learning and growth.

REFERENCES

ACKNOWLEDGMENTS

The authors would like to acknowledge all colleagues from DaVinCi team, user community (Global Biostatistics & Data Sciences & medicine functions, and more) within Boehringer Ingelheim. Their work in DaVinCi and regular work made the preparation of this paper possible.

CONTACT INFORMATION <HEADING 1>

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

<Yun Ma>
<Boehringer Ingelheim (China) Investment Co., Ltd.>
< yun_1.ma@boehringer-ingelheim.com>

<Yifan Han>
<Boehringer Ingelheim (China) Investment Co., Ltd.>
< yifan.han@boehringer-ingelheim.com>

Any brand and product names are trademarks of their respective companies.