

ABSTRACT

Biomarker data is important in oncology trials, as it can help diagnose disease or help indicate if there is a change in a patient's condition due to treatment. Biomarker data analysis often includes conducting standard post-hoc exploratory analysis to delve into the disease mechanism and identify an appropriate patient population. Various types of biomarker data, including high-dimensional genomic data, as well as PDL1, TMB, etc., are collected at screening and analyzed to assess their prognostic and/or predictive property.

The analysis includes clinical outcomes (Progression-free Survival, Overall Survival, Objective Response Rate, Odds Ratio, and Disease Control Rate), baseline characteristics, and biomarker data from different platforms. Patient-level clinical data and a binary form of biomarker data are required for analysis. Manual preparation of data is time-consuming, and an automation process is needed to improve efficiency. A variety of formats of source data from different platforms is a challenge for automation process development. A Biomarker Analysis Data Centralization Platform was developed to automate the clinical data preparation and biomarker data standardization and management process.

INTRODUCTION

Biomarker data analysis plays a pivotal role in clinical research, significantly enhancing our understanding of disease mechanisms, guiding therapeutic interventions, and improving patient outcomes. Here's a brief overview of its applications and significance:

Applications of Biomarker Data Analysis in Clinical Research:

Disease Diagnosis and Prognosis: Biomarkers serve as indicators of normal biological processes, pathological conditions, or responses to therapeutic interventions. They can aid in early disease detection, enabling clinicians to initiate treatment promptly. Furthermore, biomarkers can predict disease progression and outcomes, providing valuable information for patient management strategies.

Patient Stratification: By analyzing biomarker data, researchers can identify distinct patient subgroups based on their molecular profiles. This stratification enables the development of personalized treatment plans tailored to individual patient needs, maximizing treatment efficacy, and minimizing adverse effects.

Drug Development and Evaluation: Biomarkers are essential in evaluating the pharmacological effects of drugs, including pharmacodynamics and pharmacokinetics. They can help identify potential responders to specific therapies, reducing the cost and duration of clinical trials. Additionally, biomarkers can inform dose selection and treatment duration, optimizing clinical trial design and outcomes.

Surrogate Endpoints: Biomarkers often serve as surrogate endpoints, predicting clinical benefit without the need for long-term follow-up. This approach accelerates drug development by shortening the time required to demonstrate efficacy and safety.

Significance of Biomarker Data Analysis:

Improved Clinical Decision-Making: Accurate biomarker data analysis enables clinicians to make informed decisions regarding diagnosis, treatment selection, and patient monitoring. This leads to more targeted and effective interventions, improving patient outcomes.

Enhanced Drug Development Efficiency: By leveraging biomarkers to stratify patients and predict treatment responses, drug developers can reduce the failure rate of clinical trials. This not only saves time and resources but also accelerates the delivery of novel therapies to patients in need.

Personalized Medicine: The ability to analyze biomarker data facilitates the development of personalized medicine, where treatments are tailored to individual patient characteristics. This approach maximizes treatment efficacy, minimizes side effects, and enhances patient quality of life.

Advancements in Scientific Understanding: Biomarker research contributes to our fundamental understanding of disease mechanisms and treatment responses. Insights gained from biomarker data analysis can lead to the discovery of novel therapeutic targets and approaches, further advancing medical science.

In conclusion, biomarker data analysis is a critical component of modern clinical research, enabling more accurate diagnoses, personalized treatments, and efficient drug development. Its significance lies in its ability to transform healthcare by improving patient outcomes and advancing medical knowledge.

RShiny is a powerful web application framework built on top of R, a widely used programming language for statistical computing and graphics. It allows users to create interactive web interfaces for data visualization, analysis, and reporting directly from R. With RShiny, developers can effortlessly transform their R scripts into engaging and user-friendly web applications that can be accessed by anyone with a web browser.

One of the key advantages of RShiny is its seamless integration with R's extensive ecosystem of packages and functions. This enables users to leverage the full capabilities of R for data manipulation, statistical modeling, and machine learning within their web applications. Additionally, RShiny provides a rich set of pre-built widgets and layouts, making it easy to design visually appealing and functional interfaces without extensive web development knowledge. Another notable feature of RShiny is its ability to support real-time data analysis and visualization. Users can interact with the web application, modify inputs, and immediately see the updated results or visualizations, creating a dynamic and immersive experience. This capability is particularly valuable for exploratory data analysis, where users can quickly test different hypotheses and visualize the outcomes.

RShiny applications are also highly scalable and can be easily deployed on various platforms, including local servers, cloud services, or even embedded within other web applications. This flexibility makes RShiny an ideal choice for organizations and researchers who want to share their data insights and analyses with a broader audience or integrate them into their existing workflows.

In summary, RShiny is a versatile and user-friendly web application framework that enables R users to create interactive and visually appealing data analysis tools. Its seamless integration with R, real-time analysis capabilities, and scalability make it a valuable tool for data scientists, researchers, and organizations seeking to harness the power of R in a web-based environment.

BIOMARKER ANALYSIS DATA CENTRALIZATION PLATFORM

The Biomarker Analysis Data Centralization Platform was developed by Rshiny which includes two modules: the Clinical Data Preparation Automation module and the Biomarker Data Centralization module. The Clinical Data Preparation Automation module will automate the process of preparing clinical data. The Biomarker Data Centralization module is designed for the standardization and centralization of biomarker data, thereby accelerating the data preparation and analysis procedures.

Biomarker colleagues typically require clinical patient-level data for exploratory analysis. Due to data security considerations, biomarker scientists are unable to access study folders directly and are unaware of the exact names of the variables needed. Consequently, programmers must communicate extensively with biomarker scientists to confirm the required variables and accurately identify their exact names in the corresponding studies when preparing the clinical data. This process is often time-consuming, resource-intensive, and may delay project progress.

With the Clinical Data Preparation Automation module, biomarker scientists can view and select the labels and exact names of variables for their concerned studies through a user-friendly visual interface after obtaining Amazon Web Services (AWS) access (Fig1). Once the required variables are chosen, the system automatically generates a downloadable specification document and a template code to also create patient-level data (Fig2). This automated process simplifies the preparation of clinical patient-level

data, reduces communication costs between Biomarker scientists and programmers, and speeds up project progression.

☐ Select All

	Study ID	Variable Label	Value List
	All	All	All
1		Study Identifier	
2		Unique Subject Identifier	
3		Subject Identifier for the Study	
4		Study Site Identifier	
5		Age	
6		Age Units	YEARS
7		Pooled Age Group 1	< 60 years ; >= 60 years
8		Pooled Age Group 1 (N)	
9		Sex	F ; M

Fig1 Labels and exact names of variables

[Main Panel](#) [Preview Panel](#)

[Download Spec](#)

	Study ID	Variable Label	Value List
	All	All	All
1		Study Identifier	
2		Unique Subject Identifier	
3		Subject Identifier for the Study	
4		Age	
5		Sex	F ; M

Fig 2 Downloadable specification document

Currently, biomarker data from different platforms are collected for exploratory analysis. The Biomarker Data Centralization module provides an interface, which will facilitate the process of data uploading (Fig3). This module also offers a standardized biomarker template file to which biomarker scientists can refer. A summary table containing log check information will be generated after uploading data to allow preliminary validation and quality control. A template code to standardize biomarker data will be created when the data passes quality control, and this code will be used for subsequent data processing.

	STUDYTP	STUDYID	USUBJID	VISIT	SampleID	PLATFORM	BIOTYPE
1	SINGLE			Screening			Gene
2	SINGLE			Screening			Gene
3	SINGLE			Screening			Gene
4	SINGLE			Screening			Gene
5	SINGLE			Screening			Gene
6	SINGLE			Screening			Gene
7	SINGLE			Screening			Gene
8	SINGLE			Screening			Gene
9	SINGLE			Screening			Gene
10	SINGLE			Screening			Gene

Fig 3 Biomarker data upload interface

As a robust and comprehensive solution catering to a broad spectrum of biomarker data analysis initiatives, it provides seamless clinical and biomarker data management and analysis capabilities, empowering biomarker projects.

CONCLUSION

The Biomarker Analysis Data Centralization Platform represents a substantial achievement in the quest for automation process and efficiency by reducing the communication cost in the process of data preparation, further accelerating the process of data analysis. For the next step, the dedicated teams will further promote the process of automation, such as preparing patient-level clinical data sets automatically. These future improvements are aimed at addressing evolving needs and challenges, ensuring that the platform remains at the forefront of data management innovation.

REFERENCES

1. <https://shiny.posit.co/>
2. <https://pubmed.ncbi.nlm.nih.gov/27010052/>
3. <https://www.cde.org.cn/main/news/viewInfoCommon/321ca4648e2e2dfc8ac05e9ba28d6de4>

ACKNOWLEDGMENTS

Thanks to our managers (Leo Li, Cindy Song and Tony Guo) for their great support on the conference and other colleagues for their comments on this shell script.

CONTACT INFORMATION

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

Jinzhi Zhou
BeiGene (Beijing) Co., Ltd.
jinzhi.zhou@beigene.com

Jie Liu
BeiGene (Beijing) Co., Ltd.
jie1.liu@beigene.com

Jiazheng Zhang
BeiGene (Beijing) Co., Ltd.
jiazheng.zhang@beigene.com

Linxin Zhang
BeiGene (Beijing) Co., Ltd.
linxin.zhang@beigene.com