PharmaSUG China 2025 - Paper 178X-F5A3P6C7H3

Pre-clinical internal process optimization platforms introduction

Yusi Liu, BeOne Medicines

ABSTRACT

An overview of pre-clinical research projects: toxicological and drug synergy analyses. The presentation begin with introduction to the significance and methodology of toxicity and synergy studies in early-stage drug development, highlights key concepts and challenges faced during data collection. A central emphasis is placed on the data analysis workflow, examined from the perspective of a programmer. This includes approaches to data cleaning, statistical modeling, and visualization techniques tailored for high-throughput pre-clinical datasets.
Two interactive web applications were developed using the R Shiny framework. The first application focuses on toxicological data analysis, providing automated pipelines for dose-response modeling, outlier detection, and visualization of toxicity profiles across different compounds and animals.
The second application is designed for synergy analysis, integrating multiple synergy scoring models to evaluate the combinatorial effects of drug pairs, and offering users a dynamic interface to explore interaction landscapes and summary metrics.

INTRODUCTION

In the early phases of drug development, pre-clinical studies play a critical role in assessing the safety and efficacy of candidate compounds before progressing to human trials. (Figure 1.) Among these studies, IND-enabling toxicology (tox) studies are of particular importance. These investigations are designed to evaluate the potential toxicity risks associated with new drug candidates and provide the data necessary to support an Investigational New Drug (IND) application to regulatory authorities such as the FDA. The outcomes from IND-enabling tox studies help determine safe starting doses and dosing regimens for first-in-human clinical trials, thereby serving as a regulatory and scientific foundation for clinical development.

In parallel, drug synergy analysis has emerged as a valuable strategy in pre-clinical research, particularly in oncology and infectious disease models. Synergy analysis assesses the combined effects of two or more drugs to determine whether their joint effect exceeds the sum of their individual effects (synergistic), is equal to the sum (additive), or is less than the sum (antagonistic). Identifying synergistic drug pairs can enhance therapeutic efficacy, reduce required dosages, and minimize side effects, making it an essential component of combination therapy development.

Despite their importance, both toxicological and synergy studies pose several challenges, particularly in terms of data complexity, variability, and high-throughput nature. This paper presents an overview of pre-clinical research projects with a focus on toxicological and synergy data analysis workflows from a programmer's perspective. Central to this work is the development of two interactive R Shiny web applications, each designed to streamline and standardize data analysis in these domains. The first application addresses toxicity data analysis, while the second supports in-depth exploration of drug synergy through integrated scoring models and interactive visualization tools.

By developing these two tools, I use two different method of interactive data visualization (plotly and ggiraph**)** within R shiny. Shiny is a free package for R programming language that let you turn your data analyses into interactive web applications. Plotly is cross-language graphing library built on plotly.js, while ggiraph is a R package extending ggplot2 to create interactive SVG graphics.

TOX DATA VISUALIZATION PLATFORM

In the Investigational New Drug (IND) process, pre-clinical toxicity studies—also known as IND-enabling

studies—help estimate safe starting doses and identify potential adverse effects across organ systems and species. These studies generate vast datasets, often acquired from external vendors, each with different formats, nomenclature systems, and levels of granularity.

One of the first challenges addressed in this app is data standardization. To ensure consistency, we adopted the SEND (Standard for Exchange of Nonclinical Data) data model developed by CDISC. This structure maps toxicological results into clearly defined domains such as: a). LB (Laboratory Test Results) – e.g., ALT, AST, BUN, creatinine. b). BW (Body Weight) – animal weights over time. (LB and BW domain are the main domain used in this app)
This standardized format allows for easier downstream processing and aligns with future regulatory expectations. For example, ALT values across different datasets are normalized and coded into the LB domain, with harmonized units, species tagging, and time points.

Once data are standardized, the app allows users to select one or multiple lab parameters via this interface. (Figure 2.) Based on the selection, the application automatically detects treatment groups and assigns them distinct colors for creating figures. This mapping is critical when comparing across multiple groups or when experimental designs involve crossover or repeated-measures protocols.

Interactive plots are generated using the plotly R package. This choice provides several user-friendly features: a) Hover tooltips display summary statistics like median, sample size, standard deviation. b) Plots support zooming, panning, and exporting to HTML or static image formats. c) Color and group recognition is automatically inferred from the dataset (Figure 3.)

In addition to visualization, statistical hypothesis testing is embedded within the app. Users can choose from t-tests, ANOVA, or non-parametric tests, and results (including p-values and confidence intervals) are shown directly on the plot or adjacent to it. Another valuable feature is the download function, which allows users to export their customized figures. This export system was carefully designed to preserve formatting, color schemes, and interactivity, allowing seamless integration into reports and regulatory documents.

Crucially, the app significantly reduces the time and manual steps required to generate clean, informative visualizations. By combining SEND mapping, statistical analysis, and dynamic plotting, it transforms what was once a labor-intensive task into an intuitive and efficient process.

SYNERGY ANALYSES PLATFORM

Combination therapy is a growing focus in drug development, especially in oncology, infectious disease, and neurology. Synergy is a greater pharmacological effect for a two-drug combination than what would be predicted for “no interaction” from the knowledge of the effects of each drug individually (Greco et al., 1995). Synergy analysis evaluates whether the effect of combining two drugs is greater (synergistic), equal (additive), or less than (antagonistic) the sum of their individual effects. Identifying synergistic combinations can enhance efficacy, reduce required dosages, and prevent resistance—making it a critical part of pre-clinical screening.

Unlike single-agent testing, synergy experiments often produce multi-dimensional, noisy datasets, typically arranged in response matrices (dose vs. dose for each compound pair). These experiments are susceptible to various types of technical noise—edge effects, pipetting errors, low signal-to-noise ratios—which must be filtered out before scoring.

Our synergy app was developed with a reactive programming structure, meaning each part of the UI responds dynamically to changes elsewhere in the pipeline. This allows seamless transitions from data upload to analysis and visualization.

2.1 Quality Control Workflow and data visualization data flow

Figure 4. shows working flow of this shiny app. A major emphasis of this app is a three-step quality control

(QC) process, which allows users to inspect, filter, and refine their input data.
There are three main features in each step:

a. Raw Data Visualization (Figure 5.)
   In this part, raw data from microplate are shown as a real microplate: an interactive heatmap, identifying irregular patterns or outliers.
b. Automated Filtering
   The app applies rule-based filters such as removing negative Z values, extreme standard deviations, or low signal wells. These rules are adjustable and defined by statisticians.
c. Manual Deletion
   Users can further remove or retain specific data points by clicking directly on the interactive plot. This dual-stage filtering balances automation with scientific oversight.

2.2 Interactive Plotting with ggplot2 + ggiraph

This app utilizes ggplot2 for its rich grammar of graphics and extends it with the ggiraph package, enabling interactivity without sacrificing aesthetic control. Key visualizations include: a). Dose-response curves for individual drugs and combinations. b) Interaction heatmaps colored by synergy score. c) mono curve with reactive points

With geom_line_interactive() , geom_point_interactive() and scale_fill_gradientn_interactive(), users can hover to display dose, effect, replicate ID, and other metadata. These plots maintain full compatibility with ggplot themes, faceting, and layering, which is especially valuable when users need customized layouts for presentations or publications.

Except for raw data visualization and data qc, this app also calculates and visualizes multiple synergy scores (Figure 6.)

PLOTLY VS. GGPLOT2+GGIGRAPH

Both Shiny applications rely heavily on interactive visualizations, yet they use different plotting engines tailored to their specific needs. Below is a comparison of the two approaches:

Feature	plotly	ggplot2 + ggiraph
Interactivity	Excellent (hover, zoom, drag, export)	Good (hover, tooltip, selection)
Customization	Moderate	High
Performance	Fast on medium-sized datasets	Slower for large datasets
Aesthetic control	Limited	Extensive (supports all ggplot2 features)
Ease of use in Shiny	Seamless	Requires manual linking
Ideal for	Dashboards, exploratory plots	Publication-quality plots with interactivity

Recommendation Use Cases:
Choose plotly when you need fast, responsive plots for quick data exploration or interactive dashboards with minimal setup. Choose ggplot2 + ggiraph when high-level customization, scientific clarity, or layered presentation is necessary—such as for publications, reports, or in-depth analysis workflows.

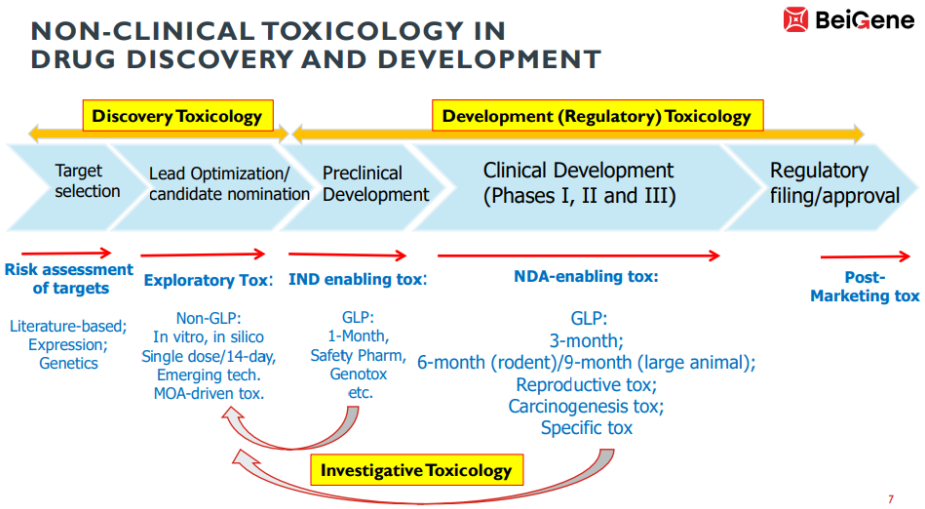
Both frameworks serve essential roles in data visualization, and the decision often depends on the balance between speed and visual complexity.

CONCLUSION

This paper presented two R Shiny applications for pre-clinical data analysis: one designed for toxicological study visualization, and the other for drug synergy analysis. Both tools were built with an emphasis on handling complex, variable, and high-throughput datasets, common in early-stage pharmaceutical research.

The toxicology app leverages SEND-compliant data structures, interactive plotly-based visualizations, and real-time statistical testing to support regulatory readiness and exploratory research. Meanwhile, the synergy app combines a structured QC process, real-time reactivity, and customized interactive plots to support robust evaluation of drug combinations.

Together, these tools demonstrate how modern data visualization frameworks—when thoughtfully integrated with domain-specific logic—can empower researchers to make faster, more reliable, and more transparent decisions. As the scale and complexity of pre-clinical data continue to grow, such tools will become increasingly essential in both discovery and regulatory contexts.



错误!未找到引用源。. Non-clinical Toxicology in drug discovery and development.

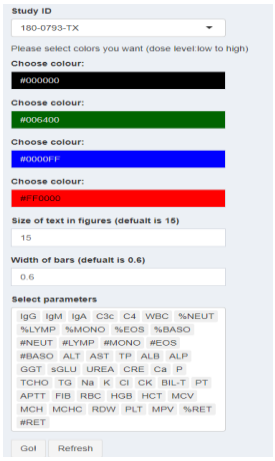
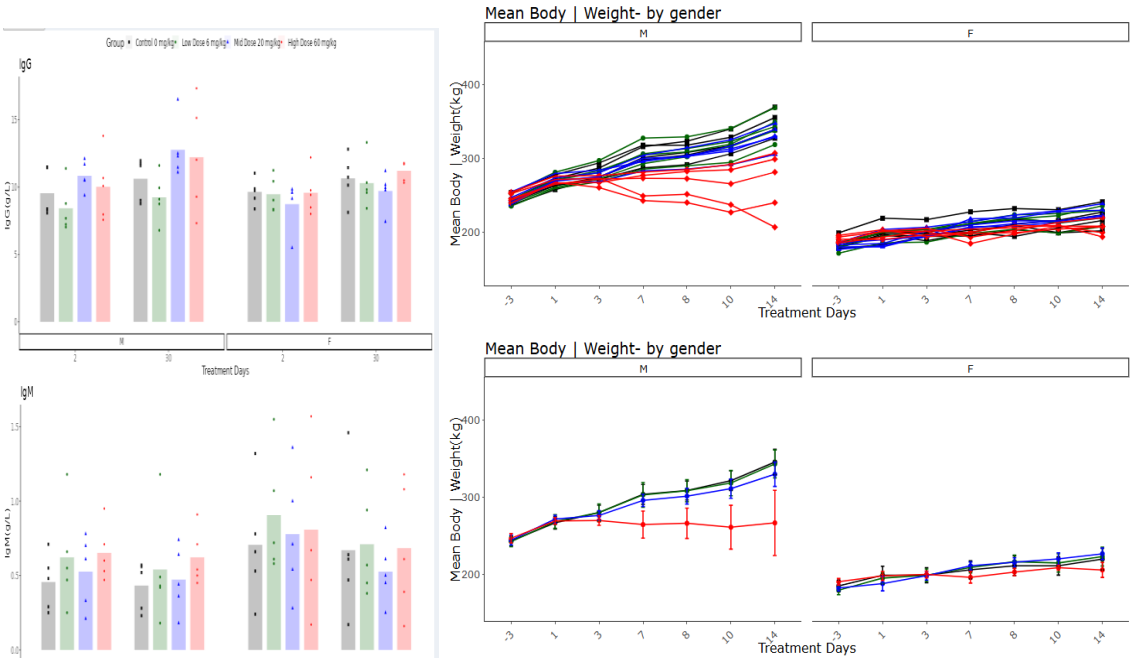


Figure 2. Interface of Tox data visualization platform



a) b)

Figure 3. Output created by plotly. a) Bar plot for LB domain. b) Line plot for BW domain

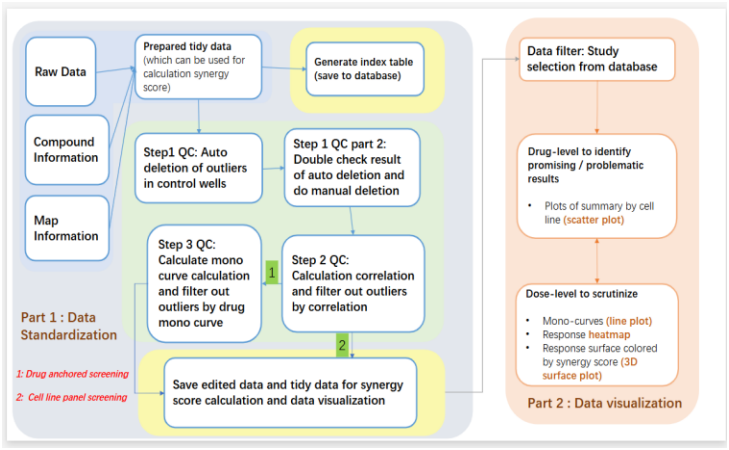


Figure 4. Working flow of Synergy analysis platform

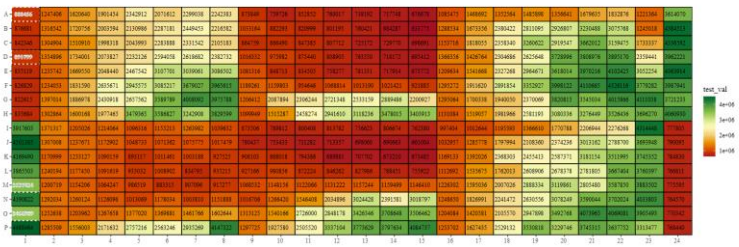


Figure 5 . Raw data visualization (A heatmap of Z value)

Figure 6. Synergy score visualization illustrate.