Paper AS13

Leveraging Data Visualization and R Shiny for Efficient Subgroup Analysis

Srinidhi Ramachandran, Eli Lilly and Company, Bengaluru, India

ABSTRACT

Presenting adverse event data across timepoints and clinically meaningful subgroups is essential in understanding patient safety, particularly in long-term or adaptive clinical trials. While Tables and Listings remain foundational for regulatory and reporting purposes, their extensive use in complex study designs can lead to lengthy outputs. These can mask essential patterns, limit visual understanding of trends and distributions, hindering quick assessments. This is especially true in trials involving dynamic dosing strategies based on biomarkers or lab parameters, where swift interpretation of safety data is critical. To address these challenges, a dynamic approach with adverse event data across subgroups (e.g., prior corticosteroid usage and varying treatment based on BMI) and timepoints are visualized through interactive heatmaps, plots implemented through R Shiny. These when used in tandem with traditional methods supported by metadata and traceability principles that complement formats accepted by regulatory agencies can be used as a tool for efficient understanding.

INTRODUCTION

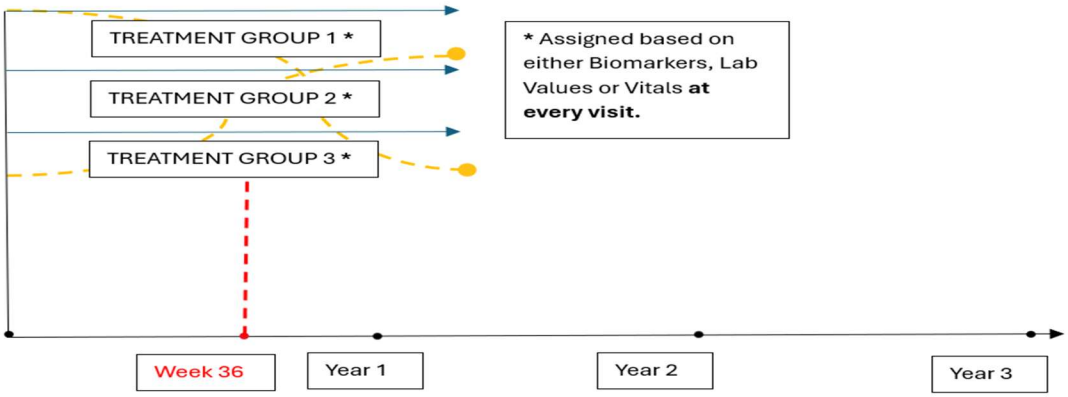
As clinical trials evolve in scale and complexity, the datasets they generate have grown exponentially, particularly in long-term extension studies that span multiple years. Traditional approaches that rely heavily on aggregate statistics and extensive tables might fail to reveal subtle but clinically meaningful patterns within specific subgroups, especially in study designs involving dynamic dosing strategies or biomarker-driven treatment adjustments. (e.g., in studies that use biomarker-based precision dose finding or in dose escalation studies). Tools such as R Shiny enable real-time, dynamic exploration of adverse event data across timepoints and subgroups, presenting complex information in formats that enhance interpretability and accelerate insight generation. R Shiny not only reduces analytical burden but also promotes efficient cross-functional collaboration and evidence-based decision-making.

In long-term extension and adaptive design studies, adverse event counts may reach into the thousands, and treatment groups may shift over time. visualizations such as heatmaps help reveal safety trends that may otherwise remain hidden in traditional listings. When coupled with metadata-driven, traceable structures aligned with regulatory expectations, these visual tools complement conventional Tables, Figures, and Listings, enriching both scientific understanding and submission readiness. This paper explores how integrating R Shiny into the clinical reporting workflow can streamline subgroup analyses, improve interpretation of longitudinal safety data, and enhance collaborative review and data monitoring eventually contributing to more informed decision-making.

METHODOLOGY

To illustrate the core problem, let us consider an example using the dummy schema shown below.

Figure 1: A Dummy Study Schema



Let us consider three treatment groups, the treatment groups are decided at every visit based on Biomarkers or Lab, Vitals values (let us consider these as assessing values). As assessing values might differ over time, subjects would and are expected to shift between treatment groups. This means that the number of subjects in a particular treatment group may not be constant at every visit. E.g., Number of subjects in treatment group 1 at Year 1 can differ from Number of subjects in treatment group 1 at the end of Year 2.
Re-iterating, number of subjects in one treatment group does not remain static and the counts keep changing based on their parameters, i.e., the denominator that is usually used to calculate summary statistics keeps changing, adding the complexity to represent the exact treatment group for AEs in a static tabular format.

Let us call the Denominator = Number of Subjects in one treatment group.
When doses change at visits (or timepoints), N becomes N_i where $i = 1,2 \ldots (visits)$.
Here is an example of a dummy table that displays the concept of Dynamic N,

Figure 2: Dummy Template to represent AE data

Overview of Adverse Events at timepoint i			
	TREATMENT GROUP 1 N_i	TREATMENT GROUP 2 N_i	TREATMENT GROUP 3 N_i

One way to map the AE to the correct Treatment can be to split up the outputs based on necessary timepoints. To facilitate continuous monitoring, this paper explores a visual representation of such data using R Shiny.

The problem statement comes down to the following four points:

1) Adverse Event Data
2) Timepoints
3) Changing Treatment Groups
4) Subgroups

DATA PREPARATION

Referring to the above points, Adverse Event data from ADAE, visit information from SV, or ADVISIT and treatment assignments, and subgroup indicators from ADSL were identified as key parameters for analysis.

As AE data is represented in the Occurrence Data Structure (OCCDS) format, a custom dataset was created to accommodate visit information based on dates. AE start dates (ASTDT), visit start dates (SVSTDTC) were considered, along with a window period 7 of days, which means, based on the SVSTDTC, an interval of (SVSTDTC - 7, SVSTDTC + 7) was considered & if any AE started within this period, it was mapped to the corresponding visit. We can also consider two more cases,

Case1: When ASTDT is present, AENDT is missing, and the subject has discontinued the study. This can be managed by using imputation algorithms based on disposition date.

Case2: When ASTDT and AENDT occur between two visits, the window period can be extended as (visit i start date, visit(i+1) – 1day).

As multi-centre studies sometimes require region or country specific analyses delivered, a cohort flag depicting such information was created. (e.g., flags that indicate US population) Subgroups can also be filtering conditions such as the age, gender, use of a medication prior to enrolment in the study or baseline characteristics. These flags were included in the analysis.

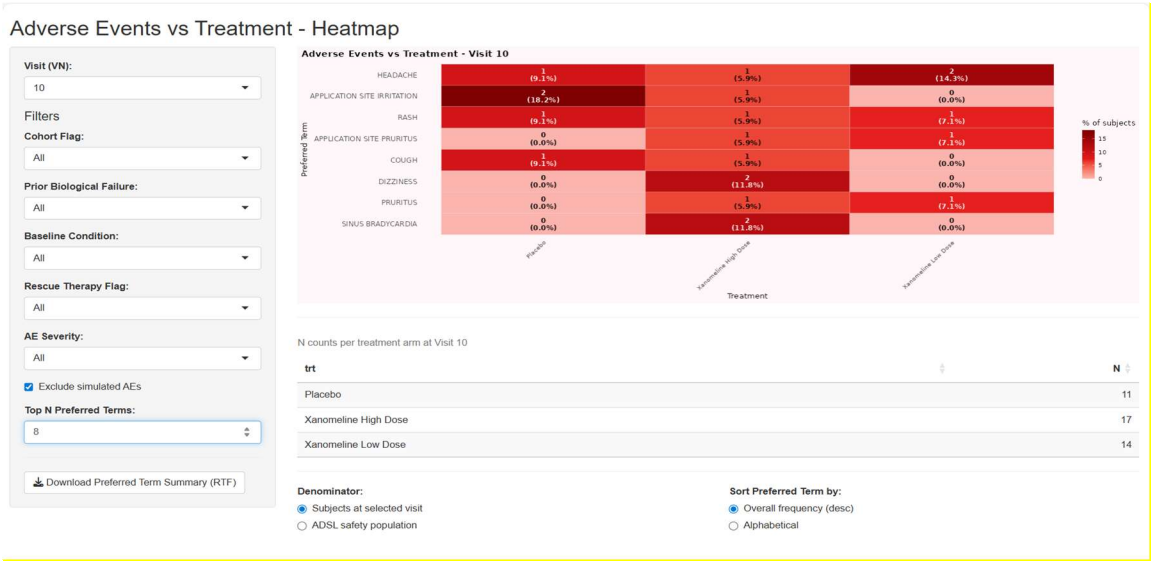
VISUALIZATION TECHNIQUES

Heatmaps were implemented using **R Shiny** to facilitate real-time exploration of AE data across treatments and visits. These provide an overview of AE distribution, with rows representing adverse event preferred terms (PTs) and columns denoting treatment arms. Each cell displayed both subject counts and percentages, while colour gradients encoded intensity.

Users can filter by visit, study-specific parameters, and adjust the number of AE terms displayed for clarity.

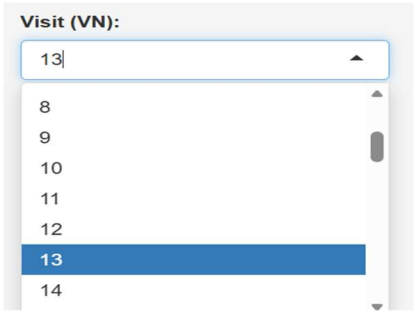
There is also a table that displays the number of subjects at a particular visit by treatments. Moreover, includes an option to choose the denominator for displaying the summaries and an option to have export a .rtf listing displaying the counts and percentages of AEs by treatments in a tabular format.

Below is a snapshot of the Shiny App:



Details about the filtering conditions:

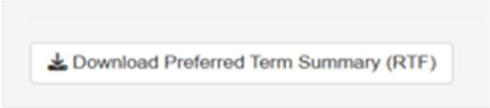
1. Heatmap of AE Preferred Terms versus Treatment Groups at visit ten.
2. The table below the heatmap depicts the number of subjects in each treatment group at visit ten.
3. The Denominator options allows users to choose either N (if it exists in ADAM.ADSL) or N_i *at* i^{th} *visit*.
4. Sort Preferred Term option allows users to choose if the terms are required to be displayed in an order of overall descending frequency or alphabetical order.
5. Coming to the drop-down options to the left, Visit (VN) allows users to filter the required visit.



6. Subgroup flags help us in different combinations of filtering are also included.

    - Cohort Flag: Filtering by Countries, Region (e.g.: Subjects in the US have value Y.
    - Prior Biological Failure: Marked as Y or N based on subject status.
    - Baseline Condition: E.g.: If Baseline values for certain lab parameters are normal.
    - Rescue Therapy Flag: Marked as Y if subject has received Rescue Medication.
    - AE Severity: Can filter based on ADAM.ADAE. AESEV values of 'MILD,' 'MODERATE' or 'SEVERE'

7. Users can also download a .rtf file with the filters that have been applied to generate a table depicting the same information.



Future Enhancements

This R Shiny app currently is limited to working on a custom dataset which needs to be created. Going forward, it can be optimized to handle directly using SV/ADVISIT & ADAE as inputs.
Different summary statistics like exposure adjusted incidence rates (EAIR), odds ratios (OR) & corresponding p-values can be implemented to better represent AE data along with visualizations.
This can be extended to include other safety data such as Lab & Vital Signs where we can use Boxplots, e-DISH plots and graphs to visualize and monitor data.
Continuing to download the rtf report from the app, when the rtf generated and shiny app values are validated, users can explore the scope of using the table to aid the submission process. There is scope for enhancing the application which could eventually act as supporting documentation and used in the review process and not completely replace the existing way of reporting and submitting outputs to regulatory agencies. *

*Referring to a PHUSE White Paper [POS_PP16.pdf] pilot project published in 2022, link to the HTML plot along with CSS (cascading style sheets) & the JavaScript were included for submission and in 2024, R Submissions Working Group successfully submitted Pilot 4 project where the submission package involved a Web Assembly component through FDA's Electronic Common Technical Document (eCTD) [R:fda-pilot4].

CONCLUSION

- Implementing visualizations for adverse event data when multiple parameters are involved serve as an efficient tool for regular data monitoring of adverse events.
- The Shiny App can also be used to monitor and review AE data for studies involving dynamic dosing.
- Collaboration with different stakeholders can be improved and interpretation of subject data can be optimized.
- This app can help the programmers to verify counts generated in Tables, Listings and Figures.
- Inclusion of options that allow subgroup filtering and generating listings can reduce the repetition of work and if the output generated using the tool is validated for counts, contribute to the submission process. This also explores the way forward on the topic of using interactive visualizations in the submission process.

REFERENCES

- https://lexjansen.com/css-us/2020/POS_PP16.pdf
- https://lexjansen.com/phuse-us/2020/dv/DV13_ppt.pdf
- https://www.lexjansen.com/phuse/2018/dv/DV11.pdf
- Archive Page | PHUSE
- Heatmap | Introduction to Statistics | JMP
- R Consortium - Using R to Submit Research to the FDA: Pilot 4 Successfully Submitted to FDA Center for Drug Evaluation and Research
- GitHub - phuse-org/TestDataFactory: All files related to the PhUSE Test Data Factory (TDF) project
- ADaM Structure for Occurrence Data (OCCDS) v1.0 | CDISC

ACKNOWLEDGMENTS

I would like to extend my sincere thanks to Team Lilly and the session chairs for giving me the opportunity to present this topic and for supporting my learning throughout the process.

CONTACT INFORMATION

AS13 – PHUSE APAC Connect 2026

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

Author Name: Srinidhi Ramachandran
Company: Eli Lilly and Company
Address: 1st Floor 7B, Wing B, Primrose, Eli Lilly Services India Private Limited,
Embassy Tech Village, Devarabisanahalli, Bengaluru, Karnataka – 560103

Work Phone: NA
Email: srinidhi.ramachandran@lilly.com
Website: NA