Paper SM15

Kraken: A Shiny Application for Visualizing SDTM Data in Early Phase Clinical Trials

Pravin Lakkaraju, Servier Pharmaceuticals, USA
Samiar Ashtiany, Phastar, United Kingdom
Rutu Patel, The Lotus Group, USA

Abstract

Kraken is a cutting-edge Shiny application designed to revolutionize how clinicians and statisticians visualize SDTM data in early-phase portfolio studies. The application provides a user-friendly interface for generating summary tables and graphs from SDTM data, that help clinicians explore and understand investigational drug performance and safety profiles with the click of a button. This capability allows for faster decision-making and supports the innovative exploration of trial data. Built on R's Shiny framework, kraken's visualizations are clickable, customizable, downloadable, and come with supporting listings for each analysis. As an essential tool for early-phase studies, Kraken promotes efficiency and progress in pharmaceutical research, empowering researchers to confidently make data-driven decisions. This paper explores the capabilities, methodologies, and challenges in the context of early-phase clinical research.

Introduction

Early-phase clinical trials play a pivotal role in the drug development lifecycle by assessing the safety and efficacy of investigational drugs. These studies generate large volumes of data that require comprehensive analysis to extract meaningful insights. However, traditional methods for analyzing and reporting clinical trial data often involve static reports and manual processes, which can be time-consuming and inflexible. Researchers and decision-makers require dynamic data exploration and customization tools to meet their specific needs.

In recent years, the demand for interactive and user-friendly tools to analyze clinical trial data has grown. Researchers and clinicians require platforms that allow them to generate insights quickly and efficiently, without the restrictions of static, pre-defined reports. The need for such tools is particularly pronounced in early-phase studies, where rapid feedback can influence the direction of the trial.

To address this challenge, we developed the Kraken Shiny application, a platform that leverages the power of R to create interactive and customizable visualizations and tables. The application is designed to help clinical researchers generate customized outputs efficiently and accurately. The app bridges the gap between complex data processing and user-friendly interfaces by integrating features such as dynamic filtering and multiple download options, enabling a seamless workflow from data to decision-making.

This paper provides a detailed overview of the Kraken Shiny application's design and functionality, highlighting its role in improving the efficiency and flexibility of clinical trial data analysis.

Kraken: An Overview

Kraken is designed to simplify the visualization and analysis of clinical trial data, particularly for early-phase studies. Built on the robust R Shiny framework, it combines advanced data

visualization capabilities with an intuitive and interactive interface. This dual focus on technical sophistication and usability makes Kraken a powerful tool for clinicians and statisticians seeking insights into their data.

The application was created with the users in mind, ensuring that even those with minimal technical expertise can explore their early phase data with ease. Kraken's dynamic navigation and customizable outputs enable users to quickly access, filter, and manipulate data in real-time.

Key Features

- **Interactive Dashboard Layout:** Kraken provides a streamlined dashboard that organizes key functionalities in an intuitive way. The app allows users to easily navigate between visualizations, filters, and other tools, making the application accessible to those with minimal technical expertise.
- **Dynamic Inputs and Filtering Options:** The application updates in real-time, dynamically populating dropdown menus for selecting studies, visualization categories, data extraction dates, and other variables.
- **Customizable and Exportable Outputs:** Users can generate a wide range of interactive visualizations and download them in a range of formats, including rtf for tables listings and figure. Each visualization is paired with supporting listings, providing more detailed subject-level data.
- **Reactive Outputs:** All outputs, including visualizations and summary tables, respond reactively to user inputs and adjustments. This feature allows for immediate feedback and exploration of data without the need for manual intervention.
- **Performance and Error Handling:** Kraken is designed to handle large SDTM datasets efficiently, with built-in mechanisms to address issues like missing data or inconsistent formats. This ensures a future-proof application.
- **User Feedback System:** An integrated feedback mechanism allows users to report issues, request features, or provide suggestions directly through the platform.

Methodology and Technical Framework

This section outlines the methodology and frameworks that underpin Kraken, highlighting its data processing capabilities and visualization framework.

Data Processing Workflow

The data preparation process forms the backbone of Kraken, ensuring accurate, reliable, and customized outputs. The workflow involves two key steps:

- **DSL Dataset Creation:** The Derived Subject-Level (DSL) dataset is created by processing SDTM data. This intermediate dataset consolidates information, providing a structured and consistent foundation for further analysis. It ensures that data is traceable and aligned with clinical reporting standards.

- **Analysis Dataset Creation:**
  The DSL dataset is combined with additional SDTM domains as needed to generate tailored analysis datasets for specific outputs. For example, merging the DSL dataset with domains such as RS (Response), DS (Disposition), or TR (Tumor Response) allows for deeper analysis and a more comprehensive understanding of the data. This flexible approach ensures that the resulting datasets are well-suited to meet diverse analytical needs.

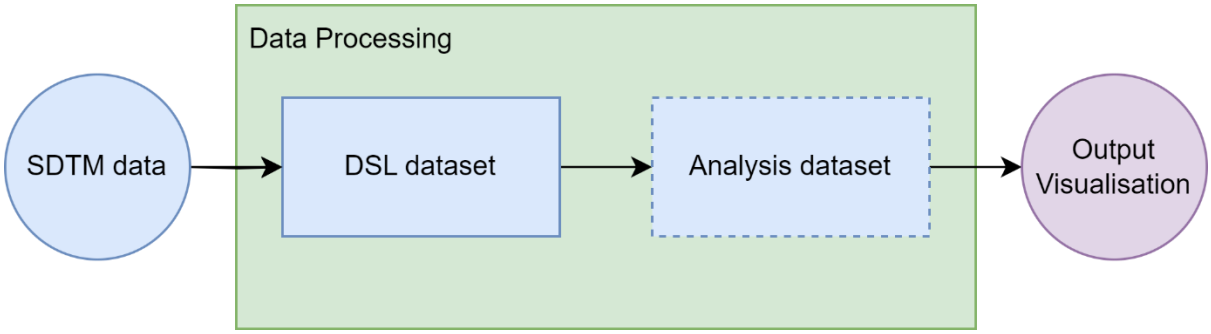


Figure 1: Data processing workflow diagram

Application Workflow

Kraken's application workflow is designed to guide users seamlessly from data upload to outputs. Each step of the workflow supports interactivity, allowing users to refine their analysis and customize outputs to their specific needs.

The workflow begins with the input of SDTM datasets, which serve as the foundation for Kraken's analysis. These datasets, including domains like DM, RS, DS, and TR, are uploaded into Kraken. The system validates the data to ensure consistency, traceability, and compliance with clinical standards.

UI Selection: Users interact with Kraken through a user-friendly interface that allows them to define their analysis parameters. The following inputs are selected dynamically:

- **Study:** The clinical study of interest.
- **Category:** The analysis focus, such as efficacy, safety, or demographics.
- **Type:** The specific visualization or analysis type, e.g., swimmer plots or bar charts.
- **Extraction Date:** The cutoff date for data to be included in the analysis.

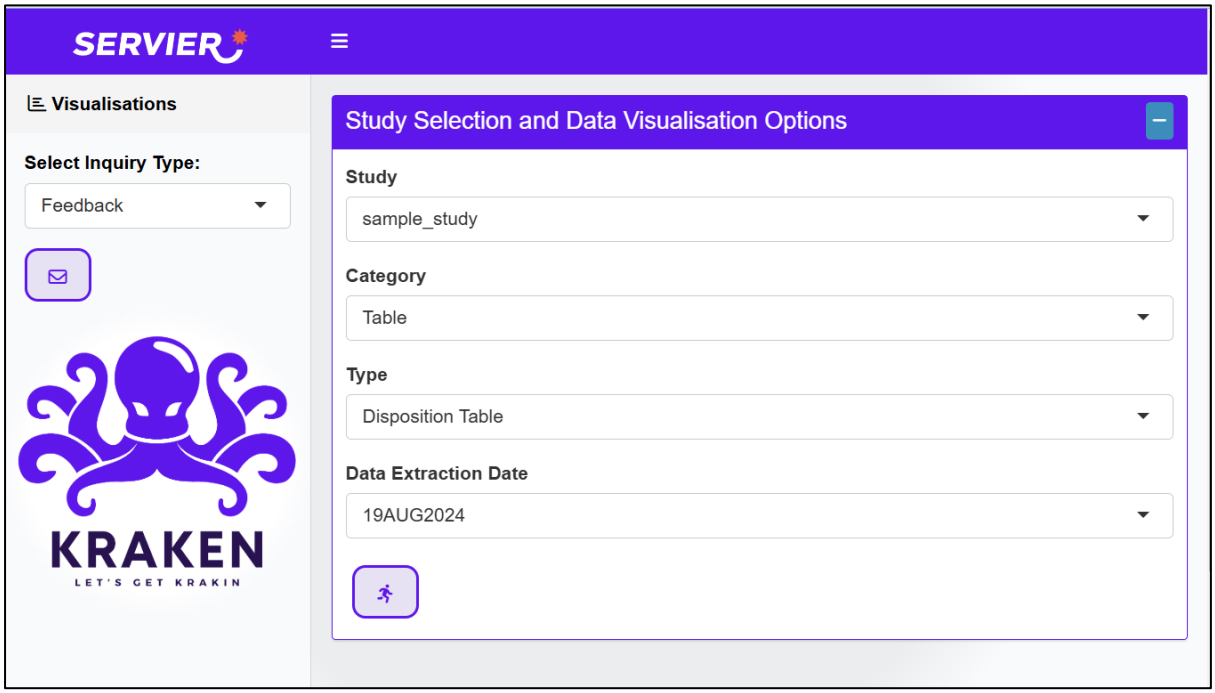


Figure 2: Front page UI

Data Processing: Once the parameters are set, Kraken processes the input data. This step is outlined in the previous subsection.

Filtering: Once the outputs have been generated, users can further refine their analysis by applying dynamic filters to focus on specific subsets of the data. Filters may include:

- Treatment arms or dose levels.
- Demographic variables such as age, gender, or biomarkers.
- Specific time points or clinical events.

This customization enables users to narrow their focus and generate precise insights.

Output Visualization: Kraken generates interactive visualizations and analysis outputs based on the refined data. These outputs include:

- Figures such as swimmer, waterfall and spider plots.
- Summary tables and detailed listings for supporting data transparency.
  All outputs are fully interactive, allowing users to explore and drill down into specific data points.

The final outputs can be customized and exported in various formats for reporting, regulatory submission, or presentation purposes. Export options include:

- Images (e.g., PNG, PDF) for visual reports.
- Data files (e.g., CSV, Excel) for detailed reviews or submissions.
- RTF files for tables, listings and figures.

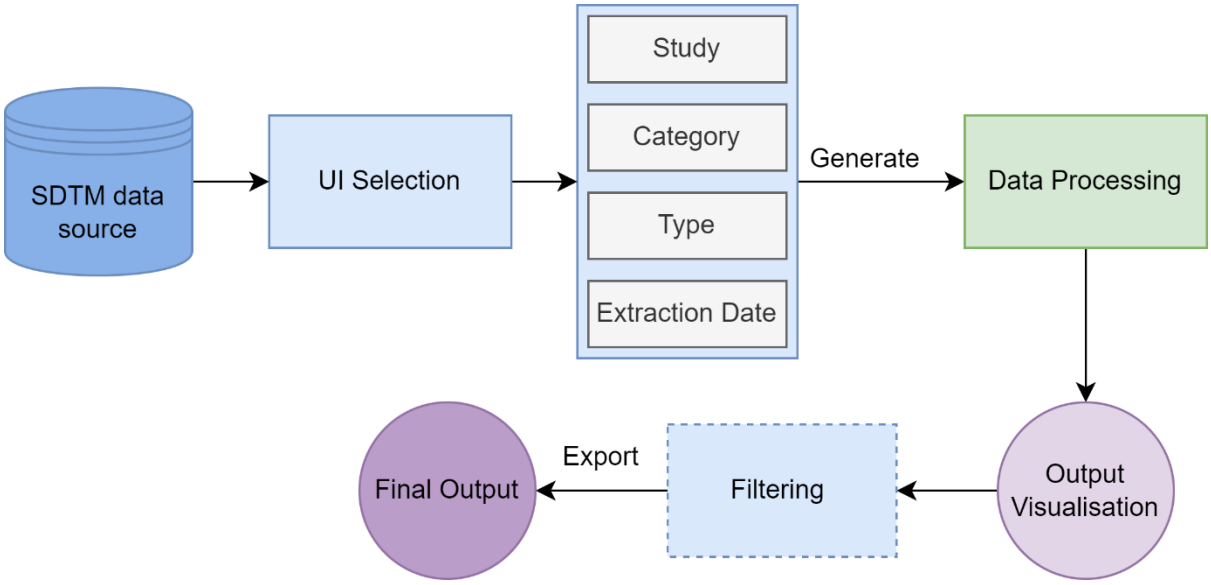


Figure 3: Application workflow diagram

Kraken in Action: Application Demonstration

In this section, we present two working examples of the Kraken application, showcasing its functionality and outputs. The process begins by selecting the study, category, and other parameters, as illustrated in Figure 2.

Example 1: Table

The first example demonstrates the disposition table output generated by Kraken.

Disposition

Participant Disposition
(Treated Set)

	Part 1a						Part 1b					
	A 10 mg Q2W (N = 1) n (%)	A 30 mg Q2W (N = 1) n (%)	A 100 mg Q2W (N = 4) n (%)	A 300 mg Q2W (N = 3) n (%)	A 750 mg Q2W (N = 8) n (%)	A 1500 mg Q2W (N = 4) n (%)	A 30 mg Q2W + B 200 mg Q2W (N = 3) n (%)	A 100 mg Q2W + B 200 mg Q2W (N = 4) n (%)	A 300 mg Q2W + B 200 mg Q2W (N = 4) n (%)	A 750 mg Q2W + B 200 mg Q2W (N = 5) n (%)	A 1500 mg Q2W + B 200 mg Q2W (N = 4) n (%)	Overall (N = 41) n (%)
Ongoing treatment	0 (0.0)	0 (0.0)	0 (0.0)	1 (33.3)	1 (12.5)	0 (0.0)	0 (0.0)	1 (25.0)	0 (0.0)	0 (0.0)	0 (0.0)	3 (7.3)
Discontinued treatment	1 (100.0)	1 (100.0)	4 (100.0)	2 (66.7)	7 (87.5)	4 (100.0)	3 (100.0)	3 (75.0)	4 (100.0)	5 (100.0)	4 (100.0)	38 (92.7)
Reason for treatment discontinuation												
Confirmed radiographic disease progression	1 (100.0)	0 (0.0)	3 (75.0)	1 (33.3)	5 (62.5)	0 (0.0)	2 (66.7)	2 (50.0)	4 (100.0)	4 (80.0)	4 (100.0)	26 (63.4)
Clinical progressive disease	0 (0.0)	0 (0.0)	1 (25.0)	0 (0.0)	2 (25.0)	2 (50.0)	1 (33.3)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	6 (14.6)
Adverse event	0 (0.0)	1 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (25.0)	0 (0.0)	1 (25.0)	0 (0.0)	0 (0.0)	0 (0.0)	3 (7.3)
Lost to follow-up	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (25.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (2.4)
Physician decision	0 (0.0)	0 (0.0)	0 (0.0)	1 (33.3)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (2.4)
Withdrawal of consent	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (20.0)	0 (0.0)	1 (2.4)

Data Extraction Date: 19AUG2024

Reason

ADVERSE EVENT
CONFIRMED RADIOGRAPHIC DISEASE PROGRESSION
CLINICAL PROGRESSIVE DISEASE
PHYSICIAN DECISION
WITHDRAWAL OF CONSENT
LOST TO FOLLOW-UP

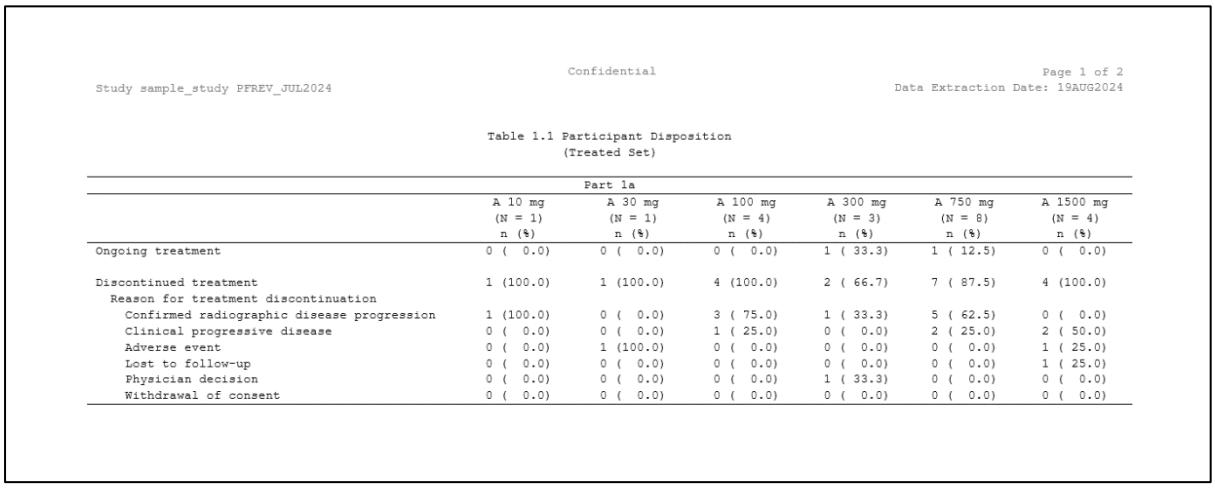
Select Status

- [x] COMPLETED
- [x] DISCONTINUED
- [x] ONGOING

.xlsx .rtf

Figure 4: Disposition table output

The table displayed in Figure 4 can be customized using the available filters. It can be downloaded as an RTF or Excel file, with the output reflecting the selected filters. A snippet of the RTF output is shown below.



Confidential

Study sample_study PFREV_JUL2024 Page 1 of 2
Data Extraction Date: 19AUG2024

Table 1.1 Participant Disposition
(Treated Set)

	Part 1a					
	A 10 mg (N = 1) n (%)	A 30 mg (N = 1) n (%)	A 100 mg (N = 4) n (%)	A 300 mg (N = 3) n (%)	A 750 mg (N = 8) n (%)	A 1500 mg (N = 4) n (%)
Ongoing treatment	0 (0.0)	0 (0.0)	0 (0.0)	1 (33.3)	1 (12.5)	0 (0.0)
Discontinued treatment	1 (100.0)	1 (100.0)	4 (100.0)	2 (66.7)	7 (87.5)	4 (100.0)
Reason for treatment discontinuation						
Confirmed radiographic disease progression	1 (100.0)	0 (0.0)	3 (75.0)	1 (33.3)	5 (62.5)	0 (0.0)
Clinical progressive disease	0 (0.0)	0 (0.0)	1 (25.0)	0 (0.0)	2 (25.0)	2 (50.0)
Adverse event	0 (0.0)	1 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (25.0)
Lost to follow-up	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (25.0)
Physician decision	0 (0.0)	0 (0.0)	0 (0.0)	1 (33.3)	0 (0.0)	0 (0.0)
Withdrawal of consent	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)

Figure 5: Disposition table rtf output

The RTF output has been designed carefully to align with regulatory templates. Additionally, the application includes a detailed listing that provides individual subject-level information for further insights.

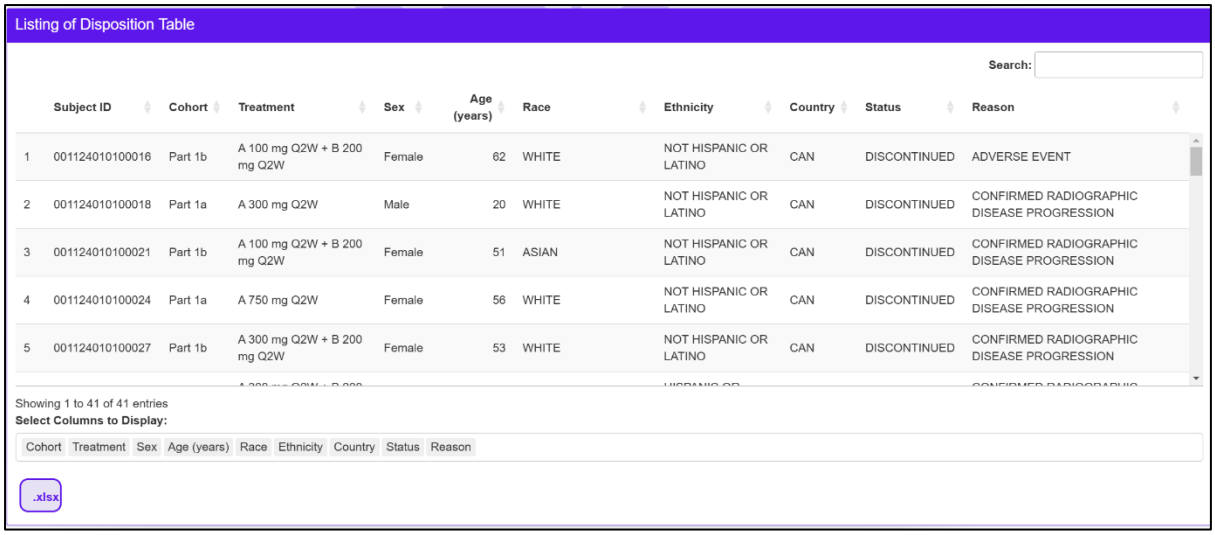


Figure 6: Disposition table listing

Example 2: Figure

In the second example, changing the category to "Figure" unlocks the available figure types. Here, we select the "Swimmer" plot.

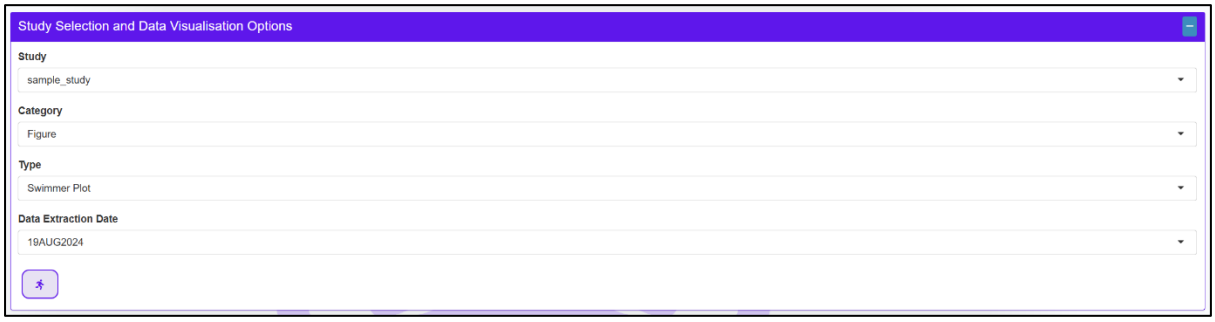


Figure 8: UI swimmer plot selection

We then click "Generate" to create the selected visualization.

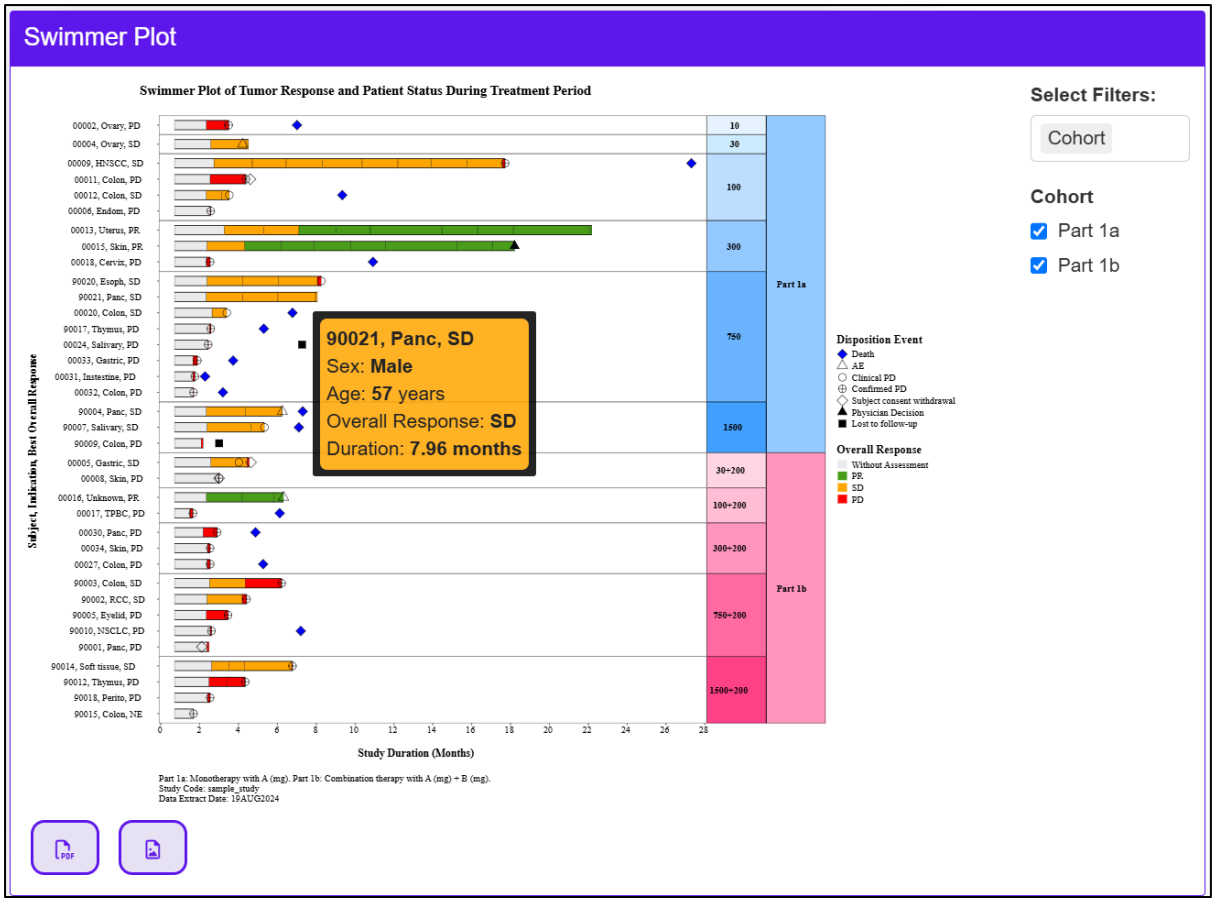


Figure 9: Swimmer plot output

The figure is interactive, allowing users to hover over different elements to explore their features. Customization options, such as tooltips, can be tailored in R to enhance the user experience. The filters provided are both dynamic and selectable. For instance, the snippet below highlights the selected filters and their available options. The dynamic functionality ensures that if a cohort is removed, any unavailable options in the other filters are automatically greyed out or deselected.

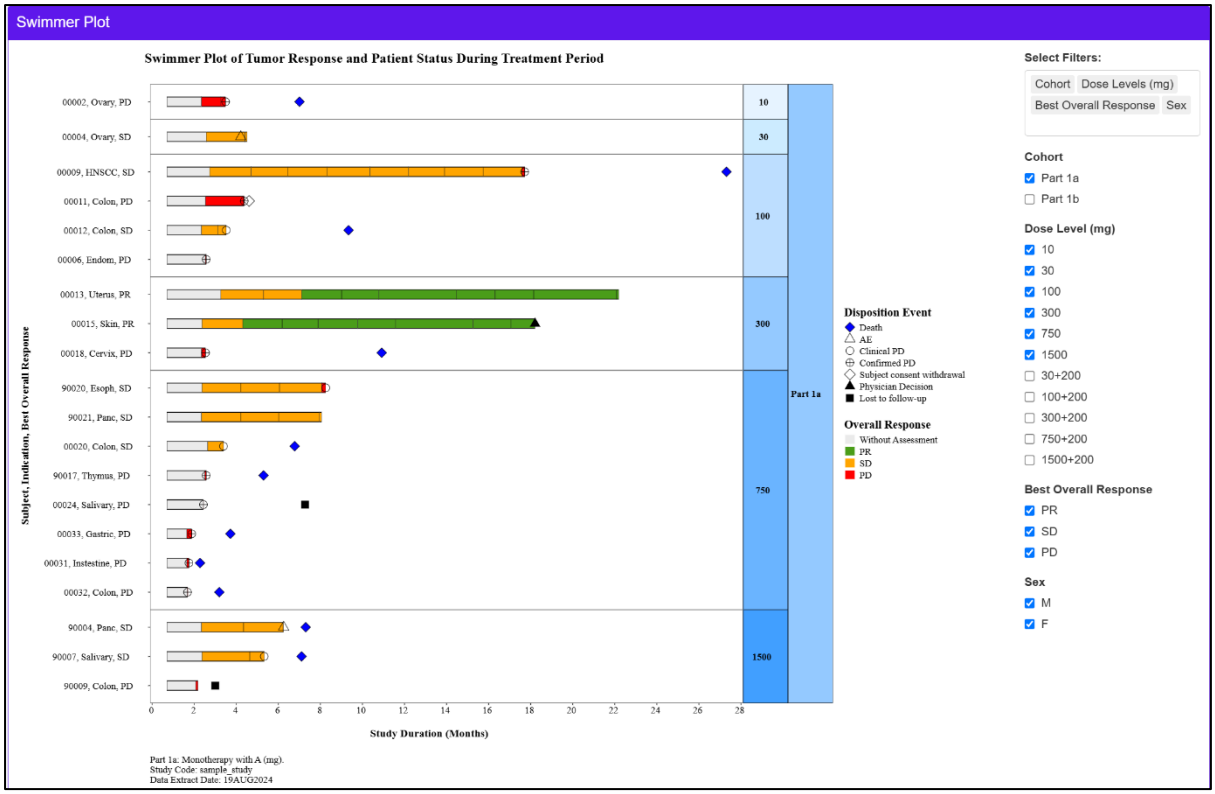


Figure 10: Swimmer plot output with filters

Figure 10 illustrates an example of filters applied within the Kraken application. Here, specific cohorts and parameters have been selected, demonstrating how the dynamic filtering system updates in real-time. This functionality ensures users can focus on relevant data subsets while maintaining flexibility to adjust their analysis as needed.

Note that excel outputs for subject listings are designed to follow a detailed and well-designed template in excel. Below is an example listing for a waterfall plot.

	A	B	C	D	E	F	G	H
1								
3	**Sample_study**							
4	**Data Extraction Date: 19AUG2024**							
5								
6							**Listing of Waterfall Plot**	
7								
8	**Cohort**	**Dose Level (mg)**	**Subject ID**	**Cancer Type**	**BO**	**SoD at Baseline (mm)**	**Best Change % SoD from Baseline**	**Treatment Ongoing (Yes/No)**
9	Part 1a	A 10 mg Q2W	00002	Ovarian Cancer	PD	113	22.12	No
10	Part 1a	A 30 mg Q2W	00004	Ovarian Cancer	SD	45	4.76	No
11	Part 1a	A 100 mg Q2W	00006	Other Cancer	PD	111	20.12	No
12	Part 1a	A 100 mg Q2W	00009	Other Cancer	SD	124	0.80	No
13		A 100 mg Q2W	00011	Colorectal Cancer	PD	248	17.34	No
14	Part 1a	A 100 mg Q2W	00012	Colorectal Cancer	SD	194	9.7	No
15	Part 1a	A 300 mg Q2W	00013	Other Cancer	PR	52	-50.00	Yes
50								
51	BOR = Best Overall Response, EOT = End of Treatment, SoD = Sum of Longest Diameter.							
52	Part 1a: Monotherapy with A (mg). Part 1b: Combination therapy with A (mg) + B (mg)							
53	Listing is sorted by ascending order of Cohort, Dose level and Subject ID.							
54								
55	Applied Filters:							
56	Dose Level (mg): 10, 30, 100, 300, 750, 1500, 30+200, 100+200, 300+200, 750+200, 1500+200							
57	Cancer Type: Breast Cancer, Cervical Cancer, Colorectal Cancer, Gastric Cancer, Ovarian Cancer, Pancreatic Cancer, Renal Cell Cancer, Skin Cancer, Other Cancer							
58	Best Overall Response: PR, PD, SD, NE							
59	Sex: M, F							
60								

Figure 7: Excel output of waterfall plot subject listings

Challenges and Future Enhancements

There were a few key challenges during the development of Kraken which were overcome using innovative solutions and back-end coding. One key challenge faced was generating RTF outputs for tables and listings as clinical reporting requires detailed formatting that adheres to regulatory standards. While the r2rtf package in R offered a starting point, it lacked the flexibility to meet all formatting requirements, particularly for complex tables and multi-page listings, which also have embedded text in headers and footers. To address this, a custom RTF generation tool was developed using Python. The docx library was used, and the final code was integrated into the Shiny framework through the Reticulate package. This approach allowed greater control over table structures and formatting but introduced additional complexities, such as maintaining compatibility between R and Python environments.

Creating customisable, interactive plots while meeting the detailed requirements of clinical standards was another significant challenge. Initially, the Plotly package was used, but it proved inadequate for handling the high level of detail needed, such as custom legends, axis labels, and multi-panel layouts. To overcome these limitations, the ggiraph package was adopted, extending the functionality of ggplot2 to include interactivity like tooltips and hover effects. While this approach enhanced user engagement, it posed challenges in balancing interactivity with the strict formatting and consistency. Extensive customization was required to ensure compliance while maintaining a seamless user experience, highlighting the trade-offs between interactivity and regulatory needs.

Looking forward, Incorporating AI into Kraken offers significant potential to streamline the process of mounting new studies, reducing the time and effort needed to rewrite code for each analysis. By leveraging LLM's, Kraken could automatically adapt its workflows and visualizations to new datasets by recognizing patterns and mapping variables within the uploaded SDTM datasets. This capability would allow the system to intelligently configure plots, tables, and filters specific to each study's structure and focus area, minimizing manual intervention.

Tech Stack

Kraken is developed using a R primarily with some Python for RTF generation. Below is a categorized overview of the technologies and libraries used:

Core Libraries

R is primarily used for creating the user interface, data handling, and visualization:

- **User Interface and Dashboard:** `shiny, shinydashboard, shinydashboardPlus, shinyFiles, shinyalert, shinyWidgets, shinyscreenshot`
- **Visualizations:** `ggplot2, plotly, ggthemes, ggh4x, ggnewscale, ggiraph, cowplot, grid, png, htmlwidgets, bbcShiny`
- **Data Handling and Manipulation:** `dplyr, tidyverse, purrr, reshape2, data.table, haven, openxlsx`
- **Document Generation:** `r2rtf`

Python

Python is utilized specifically for RTF document creation:

- **RTF Document Creation:**
  `pandas, docx`

Integration

The application leverages a hybrid approach where Python is focused on document creation tasks, while R handles the user interface and visualization. This cohesive combination ensures the application is efficient, interactive, and capable of handling complex workflows.

Conclusion

Kraken is a practical solution designed to meet the need for interactive and efficient tools in early-phase clinical trial data analysis. Built on the R Shiny framework, Kraken provides clinicians and statisticians with an easy-to-use platform for visualizing, exploring, and reporting SDTM data. Its features, such as customizable outputs, interactive dashboards, and reliable error handling, simplify the process of generating insights needed to evaluate drug safety and performance. With real-time outputs and flexible export options, Kraken supports faster decision-making in clinical research.

Kraken transforms the traditionally static and time-consuming process of clinical data analysis into a more efficient and user-friendly experience. This tool demonstrates how existing technologies can be effectively integrated to meet the specific needs of early-phase studies, helping researchers explore their data with greater ease and precision.

References

1. **Chang, W., Cheng, J., Allaire, J. J., Xie, Y., & McPherson, J. (2021).** *Shiny: Web Application Framework for R.* R package version 1.7.3. Retrieved from https://CRAN.R-project.org/package=shiny
2. **Chang, W., Borges, B., & Basile, V. (2021).** *shinydashboard: Create Dashboards with 'Shiny'.* R package version 0.7.2. Retrieved from https://CRAN.R-project.org/package=shinydashboard
3. **Wickham, H. (2016).** *ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag New York. ISBN: 978-3-319-24277-4. Retrieved from https://ggplot2.tidyverse.org/
4. **Wickham, H., & Grolemund, G. (2017).** *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data.* O'Reilly Media. Retrieved from https://r4ds.had.co.nz/
5. **Plotly Technologies Inc. (2015).** *Collaborative Data Science.* Retrieved from https://plot.ly
6. **RStudio, PBC. (2022).** *htmlwidgets: HTML Widgets for R.* R package version 1.6.2. Retrieved from https://CRAN.R-project.org/package=htmlwidgets
7. **RStudio, PBC. (2022).** *DT: A Wrapper of the JavaScript Library 'DataTables'.* R package version 0.27. Retrieved from https://CRAN.R-project.org/package=DT
8. **Jin, H., & Crow, T. (2020).** *r2rtf: Create Rich Text Format (RTF) Tables for Reporting.* R package version 1.0.4. Retrieved from https://CRAN.R-project.org/package=r2rtf
9. **Python Software Foundation. (2023).** *pandas: Powerful Python Data Analysis Toolkit.* Retrieved from https://pandas.pydata.org/
10. **Python Software Foundation. (2023).** *python-docx: Create and Modify Word Documents.* Retrieved from https://python-docx.readthedocs.io/
11. **R Core Team. (2023).** *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org/
12. **Allaire, J. J., & Wickham, H. (2022).** *shinyalert: Create Alert Modals in 'Shiny'.* R package version 3.0.0. Retrieved from https://CRAN.R-project.org/package=shinyalert
13. **Wickham, H., François, R., Henry, L., & Müller, K. (2023).** *dplyr: A Grammar of Data Manipulation.* R package version 1.1.2. Retrieved from https://CRAN.R-project.org/package=dplyr
14. **Xie, Y., Cheng, J., & Tan, X. (2020).** *shinyWidgets: Custom Inputs Widgets for Shiny.* R package version 0.6.4. Retrieved from https://CRAN.Rproject.org/package=shinyWidgets
15. **Hester, J., Wickham, H., Chang, W., RStudio, & Team, T. (2021).** *haven: Import and Export 'SPSS', 'Stata' and 'SAS' Files.* R package version 2.4.3. Retrieved from https://CRAN.R-project.org/package=haven
16. **RStudio, PBC. (2023).** *shinyFiles: A Server-Side File System Viewer for Shiny.* R package version 0.9.3. Retrieved from https://CRAN.R-project.org/package=shinyFiles

Contact Information

- **Pravin Lakkaraju**
  Company: Servier Pharmaceuticals
  Address: 200 Pier 4 Blvd., Boston, MA 02210
  Email: pravinlakkaraju1@gmail.com
- **Samiar Ashtiany**
  Company: Phastar
  Address: 2D Bollo Ln, Chiswick, London W4 5LE
  Email: samiar.ashtiany@phastar.com
- **Rutu Patel**
  Company: The Lotus Group
  Address: 330 Changebridge Rd., Pine Brook, NJ 07058
  Email: rpatelstatistics@gmail.com