International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 07 | July - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

# Streamlining Clinical Trial Data from Raw to Regulatory Submission with R Shiny and Pharmaverse

Arvind Uttiramerur
Programmer Analyst at Thermofisher Scientific, USA

## ABSTRACT

In recent years, the pharmaceutical and biotech industries have increasingly adopted open-source tools to enhance the efficiency, transparency, and reproducibility of clinical trial data analysis. **R Shiny**, a web application framework for R, has become a key component in creating interactive dashboards and visualizations that support real-time data exploration and decision-making. As part of the **Pharmaverse** ecosystem, R Shiny plays a vital role in regulatory submissions, data analysis, and reporting, offering dynamic, user-friendly interfaces for clinical trial monitoring, patient safety reporting, and efficacy analysis.

This abstract explores the application of R Shiny within the Pharmaverse, emphasizing its seamless integration with tools such as **Tplyr** and **admiral** for generating CDISC-compliant **SDTM** and **ADaM** datasets. **Metadata** management and validation are facilitated through **OAK**, ensuring consistency and traceability. The automated creation of **Define.xml** using tools like **defineR** further enhances transparency and regulatory compliance.

Through case studies and practical examples, this abstract highlights how R Shiny and related tools improve data transparency, accelerate decision-making, and support regulatory compliance in clinical trials. By leveraging these open-source technologies, clinical trial teams can streamline data workflows, enhance collaboration, and improve outcomes, ultimately driving innovation within the pharmaceutical industry. The integration of these tools into a cohesive framework significantly enhances the submission process to regulatory bodies like the FDA and EMA, ensuring that clinical data meets both industry and regulatory standards.

## INTRODUCTION

**R Shiny** is an open-source R package that enables users to build interactive web applications directly from R. Its user-friendly framework allows the creation of dynamic dashboards and data visualizations without requiring extensive knowledge of web development. By combining the power of R's statistical computing with the flexibility of a web-based interface, R Shiny enables real-time interaction with data, offering a highly responsive environment for exploring, analyzing, and presenting data in a visually compelling manner. Shiny apps are used across various industries, and in the context of healthcare and clinical trials, they provide crucial capabilities for stakeholders to engage with complex data in an intuitive and interactive format.

Within the pharmaceutical and biotechnology sectors, R Shiny plays a pivotal role in the emerging **Pharmaverse** ecosystem. **Pharmaverse** is a collaborative initiative designed to promote the use of open-source R packages in the preparation, analysis, and reporting of clinical trial data. The initiative aligns with the increasing demand for transparency, reproducibility, and regulatory compliance in clinical research. R Shiny, as part of the Pharmaverse, serves as a vital tool for clinical programmers, data managers, and statisticians by enabling the development of customized, interactive applications that streamline the visualization and monitoring of clinical trial data.

In clinical trials, R Shiny applications can be used to build interactive tools that allow for real-time patient safety monitoring, endpoint evaluation, and exploratory data analysis. By leveraging Shiny apps, clinical trial teams can dynamically adjust parameters, visualize trends, and generate reports on-demand, enhancing decision-making and communication across interdisciplinary teams. Furthermore, R Shiny seamlessly integrates with other **Pharmaverse**

packages, such as admiral for ADaM data generation and Tplyr for table creation, making it a critical component of end-to-end workflows in clinical research.

**Process Flow in the Pharma Industry: From Raw Data to Regulatory Submission**

In the pharmaceutical industry, the process of transforming raw clinical trial data into standardized datasets for regulatory submission follows a structured and well-defined workflow. This process ensures that clinical trial data is presented in a consistent, reproducible, and transparent manner, which is critical for regulatory review and approval. Below is an outline of this process, highlighting the key steps, R packages, and tools commonly used in the industry.

**1. Raw Data Collection and Management**

- **Source**: Raw data in clinical trials typically comes from case report forms (CRFs), electronic health records (EHRs), laboratory reports, and other patient-related data collection systems.
- **Tools**: Data is often managed using databases or Clinical Data Management Systems (CDMS).

**2. Data Ingestion and Preprocessing**

- **Goal**: The raw data must be cleaned and preprocessed before further analysis.
- **R Packages**:
    - **haven**: Used for reading SAS datasets (sas7bdat) into R for preprocessing.
    - **dplyr**: For data wrangling and manipulation.
    - **tidyr**: To transform and clean the data for consistency.
    - **lubridate**: For handling dates and times in clinical data.

**3. Creation of SDTM Datasets**

- **Objective**: The raw data is standardized according to the **Study Data Tabulation Model (SDTM)**, a standard format required by regulatory authorities like the FDA.
- **R Packages**:
    - **admiral**: A key R package in the **Pharmaverse** used for mapping raw data into SDTM format.
    - **metacore**: Helps manage metadata and define mapping rules for SDTM datasets.
- **SDTM Domains**: Key domains include DM (Demographics), AE (Adverse Events), LB (Laboratory Data), etc.
- **SDTM Metadata**: Ensuring proper metadata creation is crucial for the SDTM process. Tools like **OAK (Open-Access Knowledge)** help maintain consistency with metadata definitions and standards.

**4. Creation of ADaM Datasets**

- **Objective**: SDTM datasets are further processed into **Analysis Data Model (ADaM)** datasets, which are designed to support statistical analysis and regulatory submissions.
- **R Packages**:
    - **admiral**: Extends support for creating ADaM datasets, such as ADSL (Subject-Level Analysis) and ADAE (Adverse Events Analysis).
    - **Hmisc**: For managing variable labels, which is essential for regulatory documentation.

- **tibble**: Simplifies the handling of data frames in ADaM creation.

- **ADaM Metadata**: Similar to SDTM, metadata management is essential for ADaM datasets.

**5. Define.xml Generation**

- **Define.xml** is a critical document that describes the content and structure of SDTM and ADaM datasets. It is a required component of regulatory submissions. **R** is an R-based tool within the Pharmaverse ecosystem that automates the creation of **Define.xml**, a critical component required in regulatory submissions to the FDA and other regulatory bodies. Define.xml is a machine-readable file that details the structure, attributes, and metadata of SDTM and ADaM datasets. It describes variables, controlled terminology, and data derivations, making it essential for documenting the clinical trial datasets and ensuring transparency and traceability.

**Key Features of definer and package:**

- **Metadata Management**: Helps create metadata for SDTM and ADaM datasets, ensuring that variables and attributes are properly documented.
- **XML Schema Generation**: Automatically generates the Define.xml schema based on the clinical trial datasets.
- **Validation**: Supports validation against CDISC standards to ensure regulatory compliance.
- **Integration**: Works with admiral, metacore, and other Pharmaverse packages to streamline the end-to-end data flow.

**Steps for Using defineR:**

1. **Prepare Metadata**: Use metacore or similar packages to create structured metadata, including variable attributes, labels, and controlled terminology.
2. **Generate Define.xml**: Use defineR to generate the Define.xml file based on the structured metadata and data attributes from SDTM or ADaM datasets.
3. **Validate**: Ensure that the Define.xml complies with CDISC standards by performing a validation step (optional: using external tools like Pinnacle 21).

**6. Creation of TLGs (Tables, Listings, and Graphs)**

- **Objective**: After the creation of ADaM datasets, statistical analyses are performed, and **Tables, Listings, and Graphs (TLGs)** are generated for reporting.
- **R Packages**:
  - **Tplyr**: Provides an efficient way to generate regulatory-compliant tables and listings for clinical trial reporting.
  - **gt** and **ggplot2**: For creating customizable tables and graphs for visualization and analysis.
- **Output Formats**: Reports are generated in **RTF** or **PDF** formats using packages such as rtf and rmarkdown.

| **Key R Packages and Tools** | |
|---|---|
| **Shiny Applications** | **shiny, shinydashboard** |
| **Data Manipulation** | **dplyr, tidyr, lubridate** |
| **SDTM Creation** | **admiral, metacore** |
| **ADaM Creation** | **Hmisc,admiral** |
| M | **metatools, OAK** |

| | |
|---|---|
| **Define.xml** | **metatools, Pinnacle 21** |
| **TLG Creation** | **Tplyr, ggplot2, gt** |
| **Report Generation** | **rmarkdown,rtf** |

**7.Interactive Exploration and Review Using Shiny**

- **Objective**: To facilitate real-time exploration and visualization of clinical trial data for stakeholders (clinical programmers, statisticians, regulatory reviewers).
- **R Packages**:
  - **Shiny**: An interactive web framework for creating applications that allow dynamic data exploration. Clinical teams can develop **Shiny dashboards** to visualize data trends, safety signals, and other key metrics.
  - **shinydashboard**: For structuring interactive web dashboards with sidebar menus and tabs.
- **Use Case**: Shiny apps are often used for interactive safety monitoring, efficacy analysis, and decision-making during the clinical trial process.

**8.TidyCDISC (Shiny App: Data Reviewer, Table Generation)**

The **Pharmaverse** is a collaborative ecosystem of open-source R packages designed to streamline the analysis, reporting, and submission of clinical trial data, often replacing or complementing traditional SAS-based workflows. Among these packages, **TidyCDISC** plays a pivotal role, especially in leveraging **Shiny apps** for interactive data exploration and reporting in clinical trials.

**TidyCDISC** (Tidy Clinical Data Integration System) is a Shiny-based application tailored for managing and visualizing clinical trial datasets. It integrates with **Pharmaverse** packages like admiral, metacore, and Tplyr, enabling users to navigate SDTM and ADaM datasets in a user-friendly interface. By adopting **tidycdis**, clinical programmers, statisticians, and data managers can benefit from:

- Interactive exploration of clinical datasets (SDTM and ADaM).
- Visualization of safety signals and efficacy outcomes.
- Real-time filtering and analysis of data across patient demographics and trial endpoints.

This interactive approach to clinical data analysis allows trial teams to visualize trends and generate tables, listings, and graphs (TLGs) dynamically, improving decision-making and communication.

**Pharmaverse Packages and Key Advantages**

The **Pharmaverse** packages are designed to provide a comprehensive, open-source solution for clinical trial data handling, replacing or enhancing traditional SAS-based workflows. Here's why you should consider using them:

1. **Cost-Effectiveness**: As open-source tools, Pharmaverse packages like admiral, Tplyr, and metacore eliminate licensing costs associated with proprietary software such as SAS, significantly reducing the overhead for clinical trial sponsors.
2. **Interoperability**: Pharmaverse packages are built with industry standards (e.g., CDISC for SDTM and ADaM datasets) in mind, ensuring compatibility with regulatory submission requirements from authorities like the FDA and EMA.

3. **Efficiency and Scalability**: The R ecosystem is highly efficient for large-scale data processing and statistical analysis. Packages like dplyr and admiral provide robust tools to quickly manipulate, analyze, and format clinical trial data.
4. **Transparency and Reproducibility**: Being open-source, Pharmaverse packages allow greater transparency in data manipulation and statistical analysis, ensuring that all processes are reproducible and auditable, a key requirement for regulatory submissions.
5. **Active Community Support**: The Pharmaverse community, including contributions from experts in clinical programming, biostatistics, and data management, is continuously enhancing these packages, providing regular updates, bug fixes, and new features.

## CASE STUDIES AND EXAMPLES

- **Data Monitoring: Enables clinical teams to monitor trial progress, explore adverse events, and track patient outcomes dynamically.**
- **On-Demand Report Generation: Allows generation of customized tables and listings based on real-time data, reducing the need for static reports and enabling flexible data review**

**Case Study 1: Transforming Raw Data into SDTM and ADaM Using admiral**

**Problem**: A mid-sized pharmaceutical company was conducting a multi-site clinical trial and needed to standardize raw data into CDISC-compliant SDTM and ADaM datasets for regulatory submission. The traditional SAS-based process was slow and costly.

**Solution**: The team adopted the admiral package from the Pharmaverse to handle SDTM and ADaM dataset creation. By using admiral:

- Raw data from multiple sites was imported into R using haven and processed with dplyr and tidyr.
- The SDTM datasets, including DM (Demographics), AE (Adverse Events), and LB (Laboratory Data), were generated using built-in functions from admiral.
- ADaM datasets like ADSL (Subject-Level Analysis Dataset) and ADAE (Adverse Events Analysis) were created seamlessly, with consistent metadata handling using metacore.

**Outcome**: The team reported a 30% reduction in time to submission, thanks to the speed and flexibility of the R-based workflow, while eliminating the need for expensive SAS licenses.

**Case Study 2: Interactive Data Review and Safety Monitoring with tidycdis**

**Problem**: A CRO (Contract Research Organization) needed to provide real-time safety monitoring and patient data exploration for a large-scale oncology trial. Traditional static reports were inadequate for the project's complexity, as the team needed to filter patient data dynamically and assess adverse event trends.

**Solution**: The team built an interactive Shiny dashboard using **tidycdis**, allowing stakeholders to:

- Interactively filter clinical trial data by patient demographics, sites, and treatment arms.
- Monitor safety signals by visualizing adverse event rates and trends across subgroups.
- Generate and export customized reports directly from the Shiny app in RTF and PDF formats using rmarkdown.

**Outcome**: The dynamic data exploration enabled faster decision-making during safety reviews, helping the team quickly respond to emerging trends. The Shiny app provided a transparent and interactive way to explore the data, which improved communication across clinical teams.

**Case Study 3: Automated Table and Listing Generation Using Tplyr**

**Problem**: A biotech company faced challenges in generating regulatory-compliant tables and listings for clinical trials using manual processes in SAS. This resulted in inconsistencies and delays in reporting.

**Solution**: The team implemented Tplyr for automated table creation:

- The package allowed them to build customizable, regulatory-compliant tables and listings (TLGs) based on ADaM datasets.
- The team used Tplyr to dynamically generate tables that were automatically updated with changes in the input datasets, saving time and ensuring data accuracy.

**Outcome**: This approach led to an 80% reduction in the time required to generate TLGs, while improving consistency and traceability in the data.

**CDISC Pilot Project**

**Overview**

The **CDISC Pilot Project** was initiated by the Clinical Data Interchange Standards Consortium (CDISC) to demonstrate the practical application of CDISC standards (e.g., SDTM, ADaM, and Define.xml) in clinical trials. The project aimed to show how standardizing clinical data formats across trials can improve consistency, data integration, and regulatory submission processes.

**Key Objectives:**

- **Standardization**: To demonstrate the utility of CDISC standards (SDTM, ADaM) in clinical trials, improving data interoperability.
- **Submission Readiness**: To showcase how these standards prepare datasets for regulatory submission, streamlining the review process for agencies like the FDA and EMA.
- **Reproducibility**: Ensuring that clinical data is reproducible and consistent across different trials, facilitating multi-study analyses and data pooling.

**CDISC Pilot Project Process Flow:**

1. **Data Collection**: Collect raw data from multiple sites and harmonize it based on CDISC SDTM standards.
2. **Data Conversion**: Convert raw data into SDTM domains such as DM (Demographics), AE (Adverse Events), LB (Laboratory Data), etc., using admiral or similar tools.
3. **Analysis-Ready Datasets (ADaM)**: Create ADaM datasets to facilitate statistical analysis and reporting, ensuring all variables are traceable and standardized.
4. **Define.xml Generation**: Use Define.xml (with defineR) to describe dataset structure, variables, and derivations, providing metadata for regulatory bodies.
5. **Submission Package**: Compile SDTM, ADaM, Define.xml, and associated tables, listings, and graphs (TLGs) for regulatory submission.

**Key Contributions of the CDISC Pilot Project:**

- **Proof of Concept**: Demonstrated that applying CDISC standards improves efficiency, transparency, and consistency in clinical trials.
- **Regulatory Alignment**: Highlighted how regulatory agencies could more effectively review standardized data, shortening review times.
- **Tools and Resources**: Provided tools (like R-based workflows) to help industry professionals adopt and apply CDISC standards in their trials.

**9. Regulatory Submissions**

- **Objective**: The final datasets (SDTMs, ADaMs), Define.xml, and TLGs are compiled and prepared for submission to regulatory bodies such as the **FDA** or **EMA**.
- **Tools**:
    - **PharmaR**: A suite of tools for regulatory submissions in the R ecosystem.
    - **rmarkdown**: For producing submission-ready documents in PDF or HTML format.
    - Submission packages also include datasets, metadata, and validation reports.

## CONCLUSION:

The integration of open-source tools and structured workflows has revolutionized the management, analysis, and reporting of clinical trial data. **Metadata management** ensures datasets are consistently defined and compliant with regulatory requirements, with tools like **OAK** offering robust metadata handling and validation that seamlessly integrate into clinical workflows.

**Admiral**, as part of the **Pharmaverse**, automates the creation of **CDISC-compliant SDTM and ADaM datasets**, reducing manual effort and enhancing consistency across clinical trials. The generation of **Define.xml** files using tools like **defineR** provides essential documentation of dataset structures, variables, and derivations, improving transparency and traceability for regulatory submissions.

**Tables, Listings, and Graphs (TLGs)** in formats such as **RTF** and **PDF** are efficiently created using tools like **Tplyr**, ensuring outputs meet both clinical and regulatory standards. Coupled with standardized datasets and metadata, these outputs form a comprehensive submission package for regulatory bodies, simplifying the submission process.

**Shiny applications**, such as **TidyCDISC**, further enhance real-time data exploration and visualization, empowering stakeholders to interactively review clinical data, monitor safety signals, and dynamically generate reports.

Together, this ecosystem—from metadata to dataset creation, reporting, and visualization—provides a transparent, efficient, and compliant approach to clinical trial data management. By leveraging these open-source tools, pharmaceutical companies can streamline workflows, reduce costs, ensure regulatory compliance, and maintain the integrity and reproducibility of their data.

## REFERENCE

1. https://www.rstudio.com/assets/img/validation-shiny-rmd.pdf
2. https://cosa.cdisc.org/directory/admiral
3. https://www.rstudio.com/assets/img/validation-tidy.pdf
4. https://www.r-project.org/doc/R-FDA.pdf