PharmaSUG 2026 - Paper TT-118

Automated Quality Checks for SDTM and ADaM Datasets Using R Shiny

Danny Hsu, Pfizer Inc.; Wei Qian, Pfizer Inc.

ABSTRACT

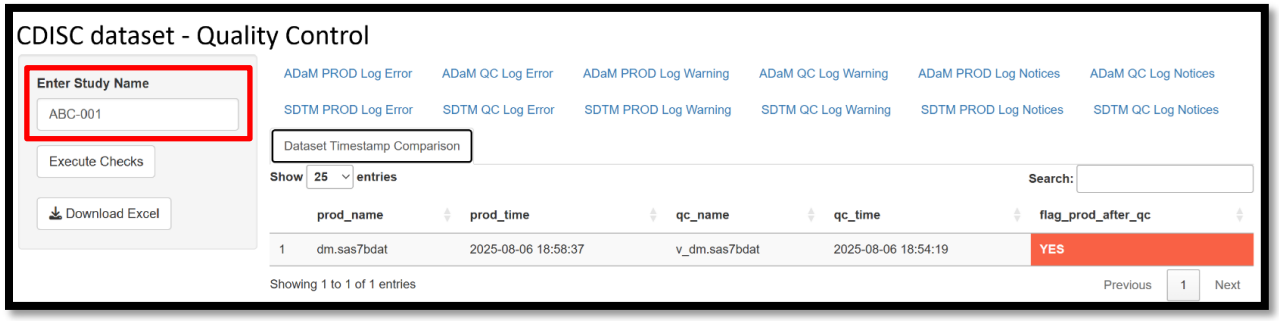
In clinical trial analysis programming, SDTM and ADaM data sets are an essential component in the process of ensuring accuracy and reliability of clinical trial data reporting; managing a project that involves numerous SDTM and ADaM datasets along with their corresponding log files can be complex and resource-intensive. Identifying data quality issues—such as inconsistent timestamps, missing values, or log errors—often requires manual inspection across multiple files, which is time-consuming and prone to oversights. To address this challenge, we developed an R Shiny application that automates overall dataset quality checks and presents the results in a centralized, interactive dashboard. Users interact with the application by specifying a folder path containing SDTM and ADaM datasets along with associated log files. Upon submission, the application performs key validations including timestamp consistency checks, log issue detection, and structural integrity assessments across both SDTM and ADaM datasets. With a single click, programming leads are presented with a visualized report that highlights potential issues and can be filtered by dataset or domain and display detailed summaries for further investigation. This tool not only streamlines the review process but also enhances transparency, traceability, and reproducibility in clinical data workflows. By integrating R Shiny's dynamic interface with robust and customizable back-end logic, the application empowers teams to proactively monitor data quality and reduce the risk of downstream reporting errors. This presentation will showcase the application's design, core functionalities, and real-world impact on improving efficiency and accuracy in clinical programming review cycles.

INTRODUCTION

Clinical trial programming often involves managing dozens of SDTM and ADaM datasets, each accompanied by log files that require validation. Manually reviewing these files for timestamp mismatches, missing variables, and log errors is time-consuming and error-prone. Programming leads need a streamlined way to monitor data quality across multiple domains without diving into each dataset individually.

SOLUTION OVERVIEW

To address these challenges, we developed an R Shiny application that automates quality checks for SDTM and ADaM datasets. The application consolidates timestamp comparisons, log error and warning detection, and structural validations into a single, interactive dashboard. Programming leads can identify issues with just one click, significantly improving review efficiency and reducing the risk of oversight. For instance, the lead programmer can enter the study or delivery name in the designated field (i.e., the red column in Display 1) and click the "Execute Checks" button. Instantly, all quality checks are summarized in the right-hand panel, providing a clear overview of dataset status.



Display 1. Example of R-shiny main page

All quality checks are customizable to meet departmental standards and needs. Key features include:

- **Timestamp Validation:** Compares file modification times across PROD and QC folders to detect inconsistencies.
- **Log File Parsing:** Extracts and categorizes errors, warnings, and unwanted notices from SAS log files.
- **Interactive Dashboard:** Uses R Shiny tabs to separate views for SDTM and ADaM domains, with dynamic filtering and download options.
- **Customizable Paths:** Users can input folder paths to run checks on specific studies or environments.
- **Excel Export:** Allows downloading of results for documentation or further analysis.

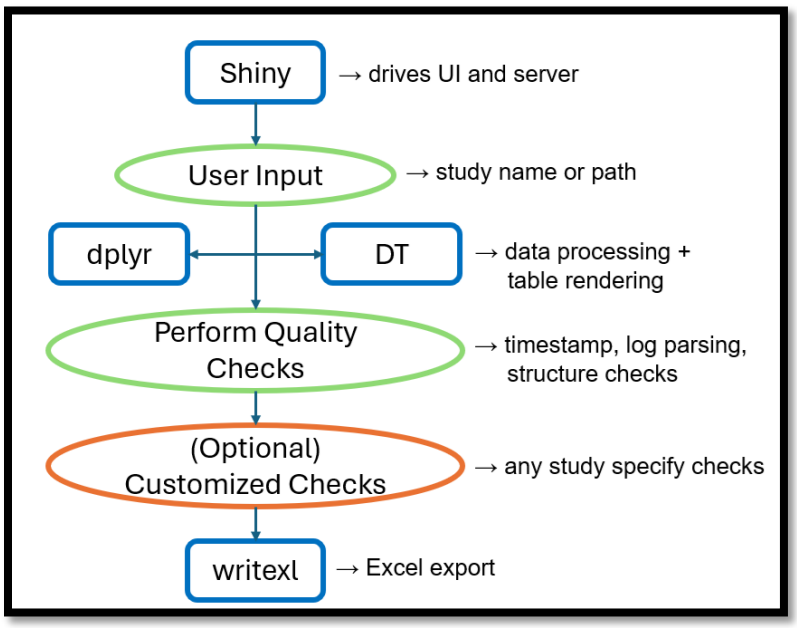
TECHNICAL IMPLEMENTATION

The application leverages R packages such as shiny, dplyr, DT, and writexl. It reads .sas7bdat and .log files, performs timestamp comparisons, and displays results in structured tables. The modular logic allows for easy extension to new domains or validation rules. Example UI components in Table 1 include tabs for dataset timestamp comparison, SDTM PROD log errors, and ADaM QC log warnings.

```
tabPanel("Dataset Timestamp Comparison", DTOutput("checktime"))
tabPanel("SDTM PROD Log Error", tableOutput("proderr2"))
tabPanel("ADaM QC Log Warning", tableOutput("qcwar"))
```

Table 1. Example UI snippet

Display 2 presents the end-to-end workflow of the quality-checking app. The diagram distinguishes R packages (shown in *square-corner boxes*) from application components (shown in *rounded boxes*). User inputs captured through the **Shiny** interface initiate data processing in **dplyr** and table rendering in **DT**, while the central Perform Quality Checks module coordinates timestamp validation, log parsing, and structural checks across SDTM and ADaM. Optional customized checks may be added to support study-specific needs. Results can be exported to Excel using **writexl** for documentation and hand-off.

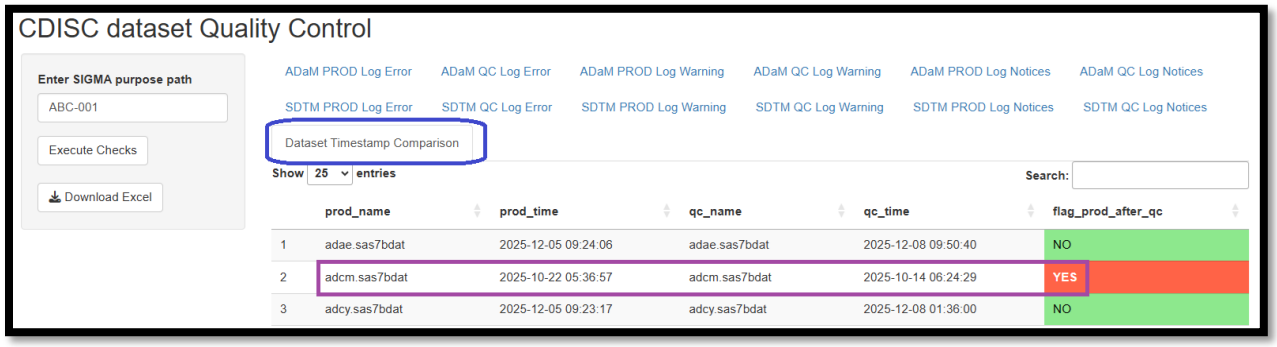


Display 2: Workflow Diagram Showing R Package Utilization

EXAMPLE 1: DATASET TIMESTAMP COMPARISON

In this scenario, the lead programmer enters the study name (e.g., "ABC-101") and initiates the quality check. The dashboard immediately displays a table comparing the modification times of datasets in the PROD and QC folders. Each row represents a dataset, showing the production and QC timestamps side by side. If a discrepancy is detected—such as the QC file being updated before the PROD file—a flag is raised in the "Status" column, clearly indicating the need for further review.

For instance in Display 3, if the QC dataset was modified at “2025-10-14 06:24:29” while the corresponding PROD dataset was last updated at “2025-10-22 05:36:57”, the dashboard highlights this inconsistency with a "YES" flag. This visual cue allows the programming lead to quickly identify and address potential issues before they impact downstream processes. The dashboard also supports exporting the results to Excel for documentation or further analysis. This streamlined approach eliminates the need for manual timestamp checks and ensures that all datasets are synchronized and ready for regulatory submission.

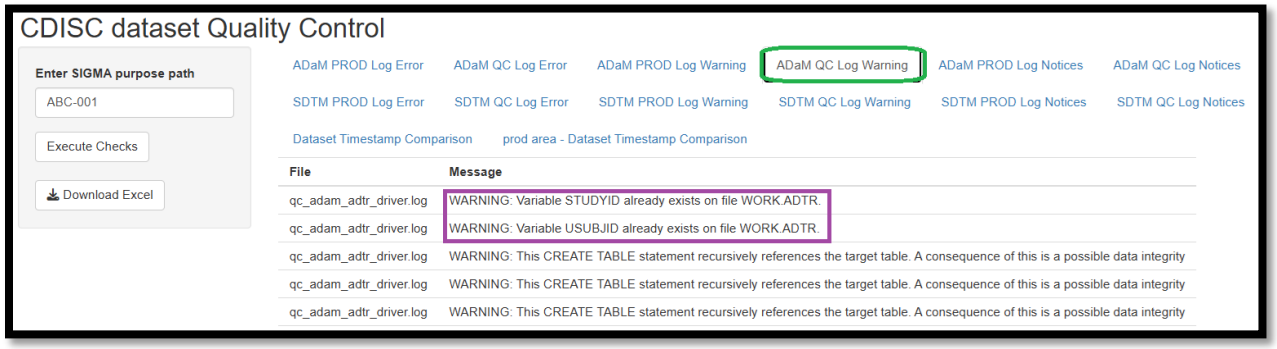


Display 3: example 1 – Dataset Timestamp Comparison

EXAMPLE 2: LOG FILE WARNING DETECTION

In this example, Display 4, the application parses SAS log files associated with ADaM QC datasets. After entering the study name and executing the checks, the dashboard presents a detailed list of warnings and errors extracted from the log files. Each message is categorized and displayed with context, such as the affected variable or table.

For example, the dashboard may show warnings like "Variable STUDYID already exists in WORK.ADTR" or "Variable USUBJID already exists in WORK.ADTR." These messages are highlighted, enabling the programming lead to quickly pinpoint issues that could affect data integrity. The dashboard’s filtering options allow users to focus on specific types of warnings or errors, and the export feature facilitates sharing findings with team members or including them in project documentation. By automating log file parsing and categorization, the application saves significant time and ensures that no critical warnings are overlooked during the review process.



Display 4: Log checking

CONCLUSION

This R Shiny tool has improved how programming teams validate SDTM and ADaM datasets by automating key checks and centralizing results in an interactive dashboard, and it has been successfully integrated into internal workflows, supporting deliverables such as patient profiles and data transfer packages. It has significantly reduced manual validation time and increased confidence in dataset readiness for regulatory submission. Customizable features and easy export options help teams adapt the tool to their needs, supporting efficient workflows and reliable clinical data management.

CONTACT INFORMATION

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

Author Name: Danny Hsu; Wei Qian
Company: Pfizer Inc.
Address: 66 Hudson Blvd E, New York, NY 10001
Email: danny.hsu@pfizer.com; wei.qian@pfizer.com
Website: https://www.pfizer.com