PharmaSUG 2024 - Paper SD- 258

An approach to make Data Validation and Reporting tool using R Shiny for Clinical Data Validation

Madhavi Gundu and Vivek Jayesh Mandaliya, Ephicacy Lifescience Analytics

ABSTRACT

Data validation in clinical trials plays a critical role in ensuring the integrity, reliability, and validity of the data collected during the study. Clinical trials are essential for evaluating the safety and efficacy of new drugs, medical devices and results of these trials can have significant implications for patient care, regulatory approvals, and public health.

Traditionally, data validation has relied heavily on proprietary software such as **SAS®** to generate reports, which may come with limitations in terms of flexibility and accessibility. With the advent of Open-source tools like **RStudio and Python**, we have developed a **Data validation tool using R Shiny**.

The tool introduces a dynamic user-friendly interface with the option to select data points and logic blocks that a user can customize per the validation criteria and generate reports without any programming skills. Our tool empowers Data management teams to conduct efficient and accurate data validation.

INTRODUCTION

Clinical trials play a crucial role in evaluating the safety and efficacy of medical interventions. The results of these trials impact patient care, regulatory approvals, and statistical analysis. Ensuring high-quality and trustworthy data is essential for accurate decision-making.

Timely data validation Timely data validation helps to identify errors or discrepancies in patient data promptly. It ensures that accurate patient information is available to healthcare professionals, minimizing the risk of medical errors, enhancing patient safety and contributes accuracy of the clinical database.

The data manager is responsible for overseeing the quality and integrity of the clinical database. To achieve this, they rely on several types of validation processes like Data Review Listings, External data Reconciliation, Logical and Integrity Checks, Audit Trial reports and follow a systematic approach to identify and address data issues. Most of the pharmaceutical companies and their partner CROs use **SAS® Software** for generation of these Data Validation reports.

The increasing importance of **RStudio** in the pharmaceutical sector, specifically in areas of **clinical analytics and statistical programming**, reflects a broader trend in the industry.

This paper explores the implementation of various validation checks using open-source frameworks to enhance versatility and efficiency. An interactive web-based application referred to as the **Edit Check Tool**. The tool has been developed internally as a project to facilitate Data Managers in effortlessly setting up edit checks for studies and generating reports, even without a programming background, through provided training on its usage.

TOOL FEATURES

Built on RStudio, this graphical user interface (GUI) enables users to:

- Set-up Edit Checks for Identifying Discrepant Data. These rules can be customized based on the specific requirements of the study.
- Create Intermediate Tables for Complex Checks: In cases where data checks involve multiple data tables this feature enables users to create intermediate tables to facilitate the checking process.
- Expression Builder for Complex Conditions: The expression builder feature enables users to create edit checks with complex conditions involving multiple criteria or logical operations. It

provides a user-friendly interface for building and customizing these conditions without needing to write code.

- Central Library for Reusable Checks: This feature provides a repository where commonly used checks can be stored and reused across multiple studies.
- Generate Reports and Combine Comments: The tool allows users to generate reports summarizing the results of data checks and issues identified. It also supports the aggregation of comments or feedback from previous reports, aiding data managers in addressing data issues and tracking their resolution over time.

EDIT CHECK TOOL

Edit Check tool is the tool developed with RStudio using packages like **shiny, shinyWidgets, dplyr, haven, stringr, openxls, rhandsontable, tidyverse** and others. The tool selects the SAS® datasets and sets-up validation rules on the datasets and variables as per the Study requirement and generates the discrepancy report.

HIGH LEVEL PROCESS FLOW

Data study data validation checks are configured in an Edit Check Tool based on a Data Validation Specification (DVS), and then reports are generated from these checks.

Data Validation Specification (DVS)

Data Validation Specification document outlines the specific validation checks needed for the study. Include information on the expected ranges, formats, consistencies, and other criteria for each data element.

Data Edit Check Tool

The Edit Check Tool, a platform designed for creating, managing, and reporting data validation checks. The user will import the study data into the tool and configure the tool to add the edit checks based on the data rules outlined in the Data Validation Specification through a graphical user interface (Figure 1).

Run the configured validation checks automatically on the imported data. The tool compares the actual data against the defined rules to identify discrepancies, outliers, or issues and generate detailed reports.

Edit Check Report

The report generated by the Edit Check Tool is in the excel format, which includes summaries of validation errors, specifics on flagged data points, and recommendations for resolution. The Data Managers investigate the report for flagged data points and resolve discrepancies as needed.

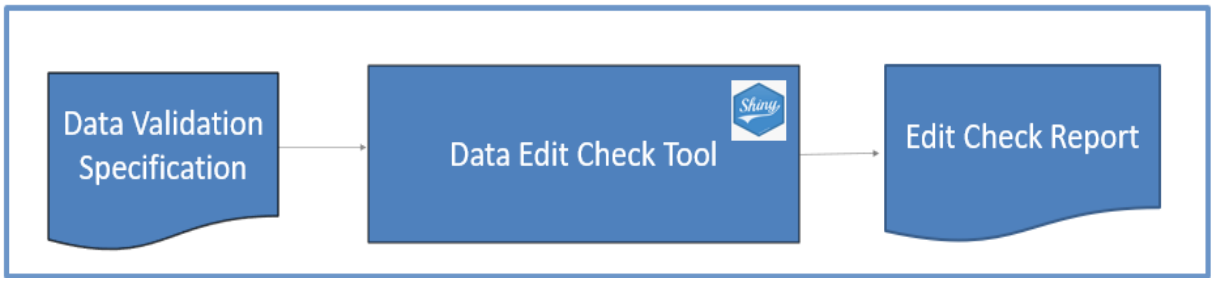

Figure 1. Process Flow

TYPES OF DATA VALIDATION CHECKS

Here are types of data validation checks that can be setup or defined in the tool.

- Range Checks: These checks verify that numerical data, such as vital signs and other quantitative measurements, fall within predefined acceptable ranges.
- Completeness Checks: These checks Ensure that all required data fields are completed.
- Logical Checks: These checks verify the logical relationships between different data points. For instance, checking that start dates precedes end dates.

ARCHITECTURE

Figure 2 outlines the architecture of the application. It is designed with a modular architecture, comprising four interconnected modules, each serving distinct functionalities. The interconnected nature of these modules promotes seamless communication and collaboration within the tool.

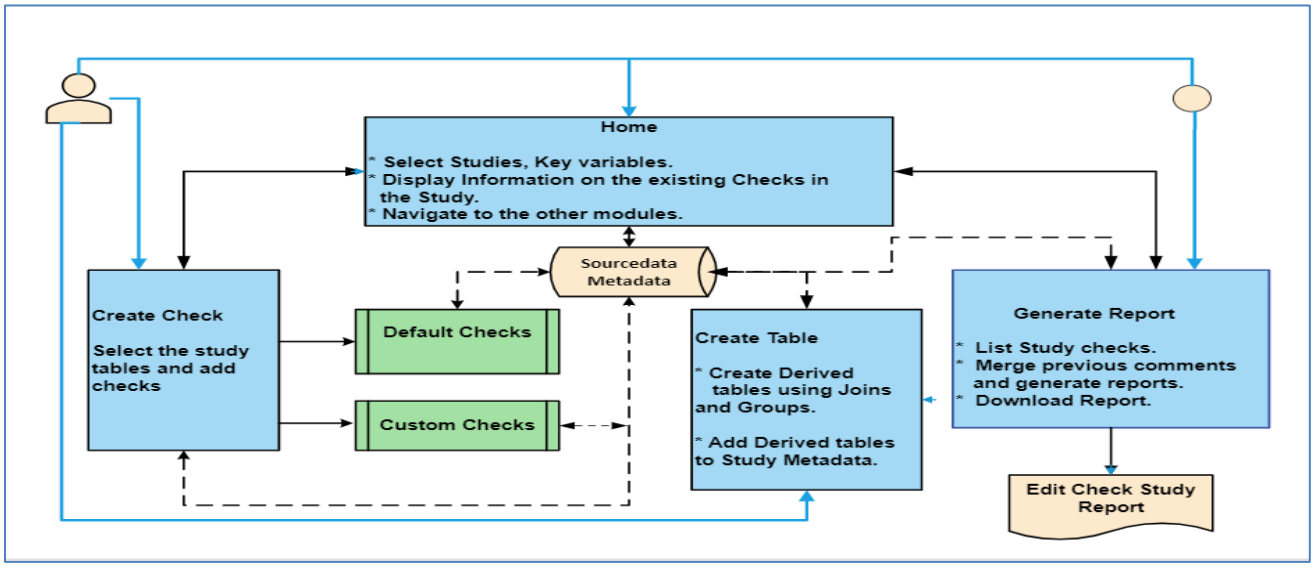

Figure 2. Edit Check Tool Architecture

Source Data and Metadata

The tool utilizes a file system structure to organize study data. Each study within this system has its own set of specific sub-folders to accommodate the source data, edit check tool metadata files, and generate reports by the tool. The metadata relative to checks, derived tables, check status is stored in the Edits folder.

Figure 3 represents the Study File System used by the Tool.

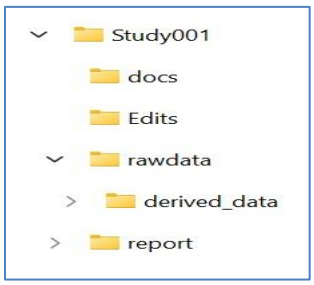

Figure 3. File Structure

EDIT CHECK TOOL WORKFLOW

When the application is executed the User land on the Home tab. The home tab screen of the Edit check tool is shown in Figure 4. It contains tabs to direct the user to access various functionalities of the application. The main tabs of the app are.

- Home
- Create Check
- Create Table
- Generate Report

Each of the tabs are explained in detail in the rest of the paper.

HOME TAB

The study validation setup begins by selecting a study from the Home tab, as illustrated in Figure 4. Upon selection, the application activates three other modules and seamlessly loads the study data. Within the Home tab, users can access a dashboard displaying the edit checks previously configured for the study. This dashboard provides insights, including the count of past checks and total derived tables.

For new study set-up, users can choose the list of variables that uniquely identify the observation within a table. These variables will subsequently be included in the validation report (Figure 9) along with the review variables. Default variables such as Review Date and DM Comments are automatically included in the report layout. However, for further customization, users have the option to add additional review variables tailored to specific validation needs.

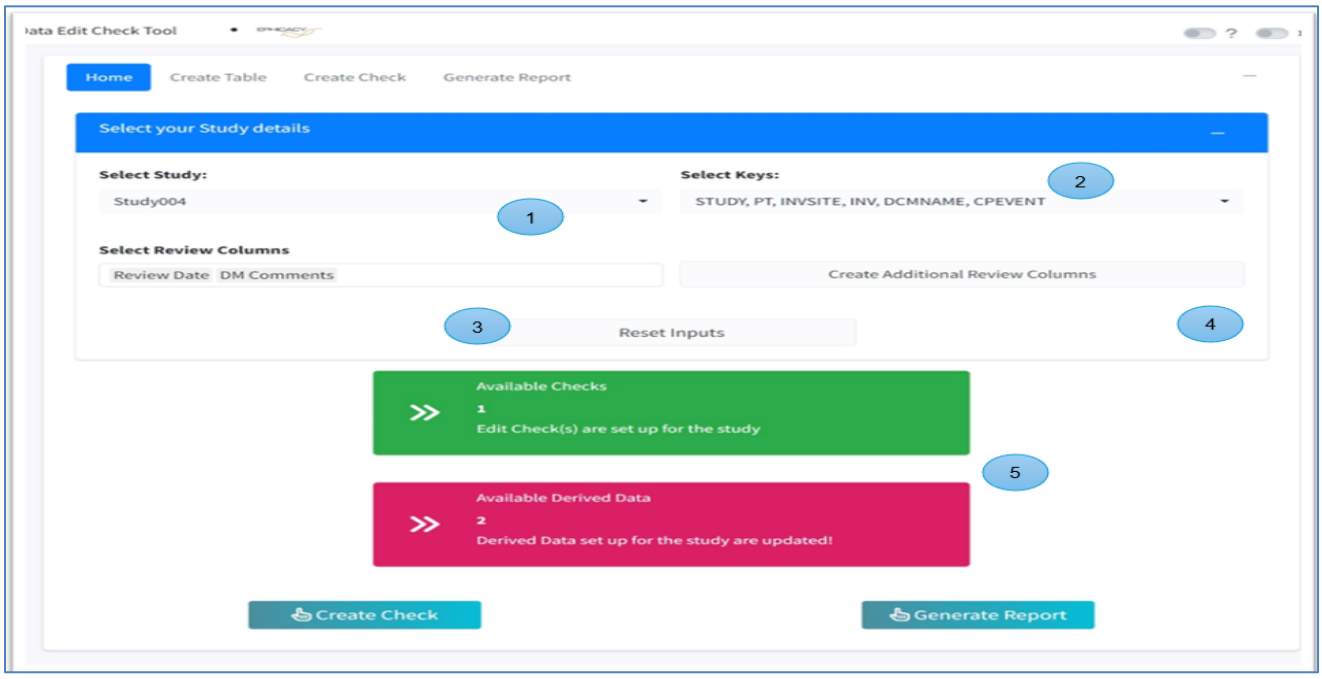

Figure 4. Screen capture of Home Tab

CREATE CHECK

Enables users to add checks to the selected study. Users have an option to create edit checks in three ways using Default Checks, Custom Checks or Fetch Library. Once the edit checks are created, the functionality of the checks is stored in the Edits folder (Figure 3) which is referenced later for other functionalities.

Default Checks

Allows user to use Pre-programmed checks with configurable parameters. The tool has functionality for two checks.

- Missing Check: Creates checks for missing datapoint of a table.
- Date Comparison Check: Create checks which compare Date datapoints or constant date value using a date picker.

After navigating to Default Check section, the user selects one of the default check options tailored to their specific validation needs as illustrated in Figure 5. Subsequently, choose the data table, desired data variables that require validation and assign a unique check number for reference purpose. Following this, the check is seamlessly added to the current study and the interface presents observations from the table based on the applied check, highlighting up to five discrepant rows for review.

To enable easy retrieval and reuse, users have the option to add the check to the library, ensuring accessibility for future studies.

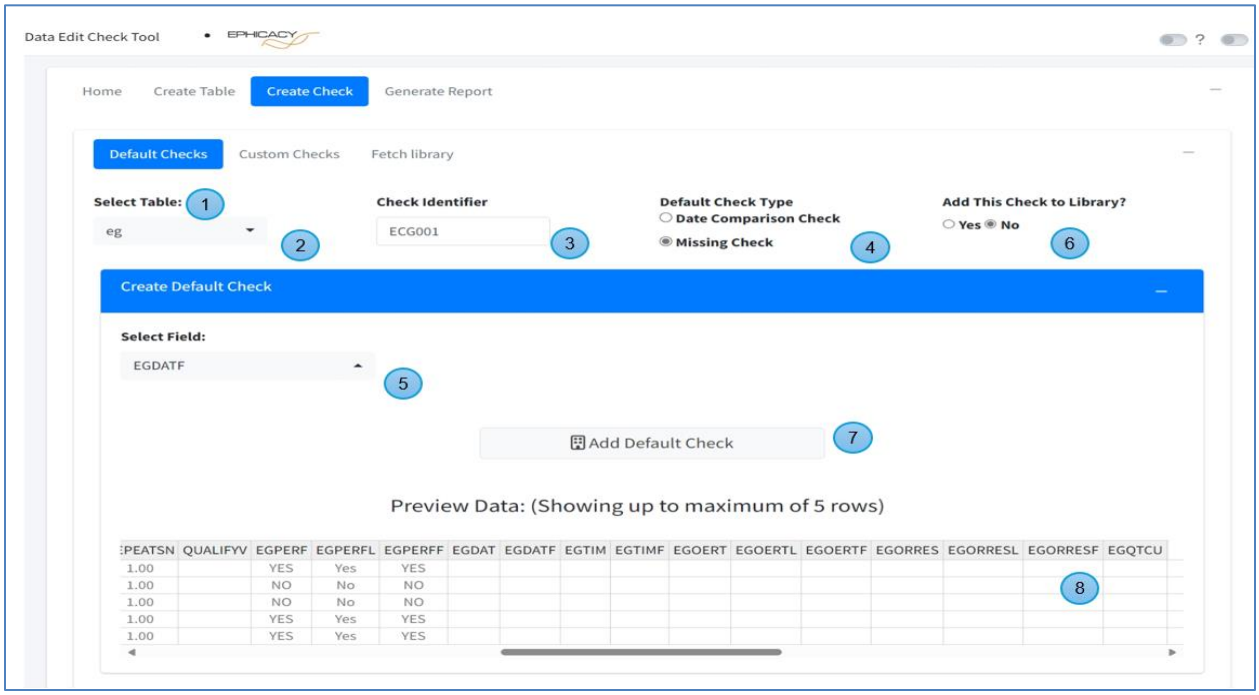

Figure 5. Screen Capture of Create Default Check

Custom Checks

Users can create custom checks to create data rules which involve multiple conditions based on a single table.

The Custom Check option serves as a solution for scenarios where edit checks involve multiple conditions or are different from the default checks provided by the tool. Users have the flexibility to select the relevant data table, specify desired data variables, and choose necessary operators to formulate the validation rules. This section (Figure 6) empowers users to incorporate multiple conditions and combine them using logical operators, thereby constructing validation expression. Conditions can be conveniently grouped and identified by unique numbers.

The tool can automatically populate a Default Query Text based on the generated expression. Additionally, user has ability to fine-tune their validation rules and modifying the Query text, allowing for the insertion of additional data variables as needed to ensure precise Query Text as in Figure 9.

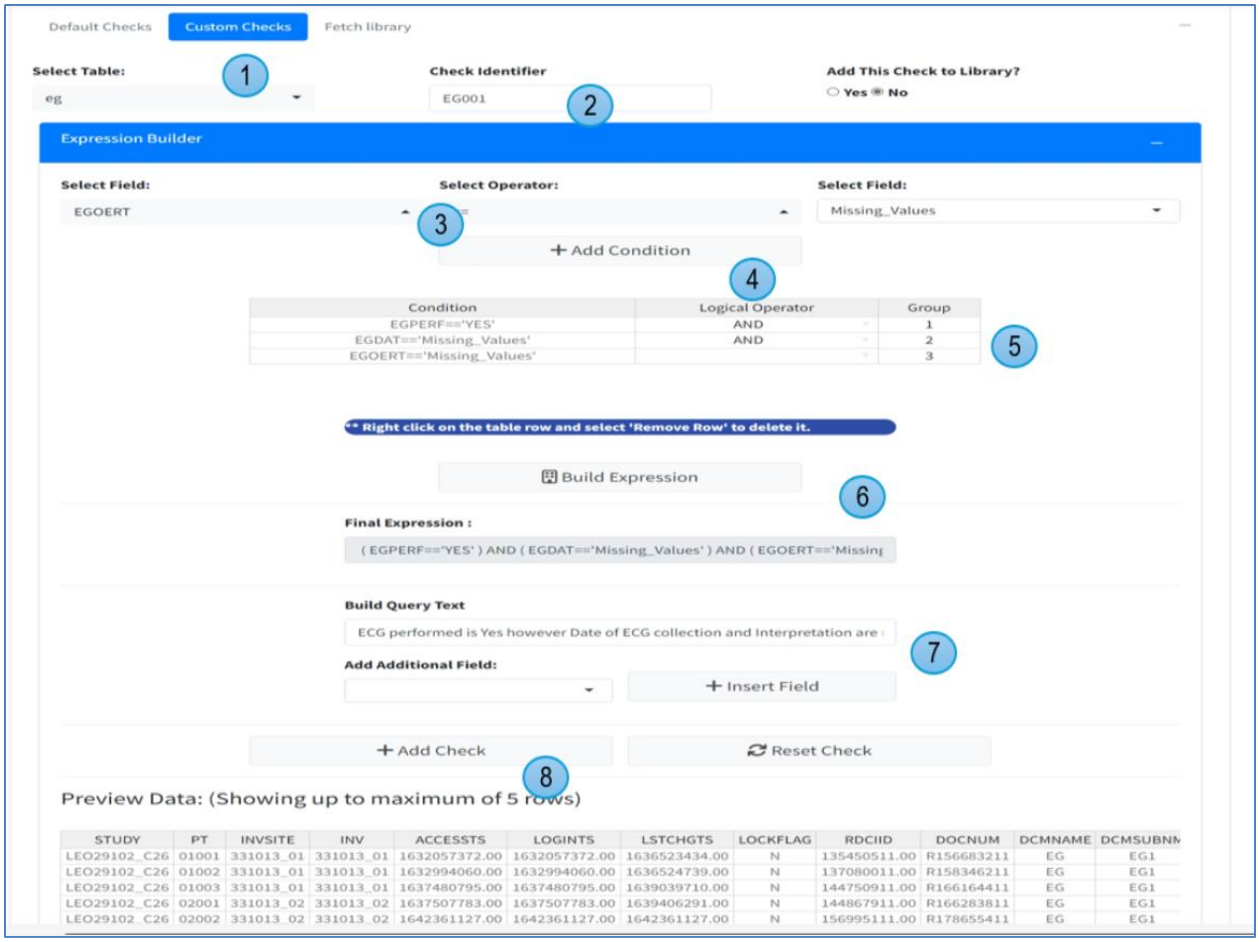

Figure 6. Screen capture of Create Custom Check

Fetch Library

The tool has a Central repository to store the commonly used checks across studies.

- This feature suggests User list of all compatible checks available in the library as per the metadata of the selected study.
- Users can directly copy the checks from the library. This enables the re-usability feature of the tool.

CREATE TABLE

Allows users to derive new tables using join and grouping. The derived datasets will be stored in the study rawdata -> derived_data folder (Figure 3). This allows the user to set-up checks involving more than one data table.

To accommodate edit check based on multiple tables, the Edit Check Tool provides options to create intermediate data tables using Group and Join options.

User can navigate to Create Table tab and select Join option. Here, they have the flexibility to choose two tables along with the desired variables, while also specifying key variables for joining and select the type of joins to be used as in Figure 7 to create intermediate table.

Following the joining process, the user can conveniently review the resulting table and add it to the study.

Tool also facilitates user to create tables by grouping variables and generating the first row, last row from the group of variables selected also created basic statistics.

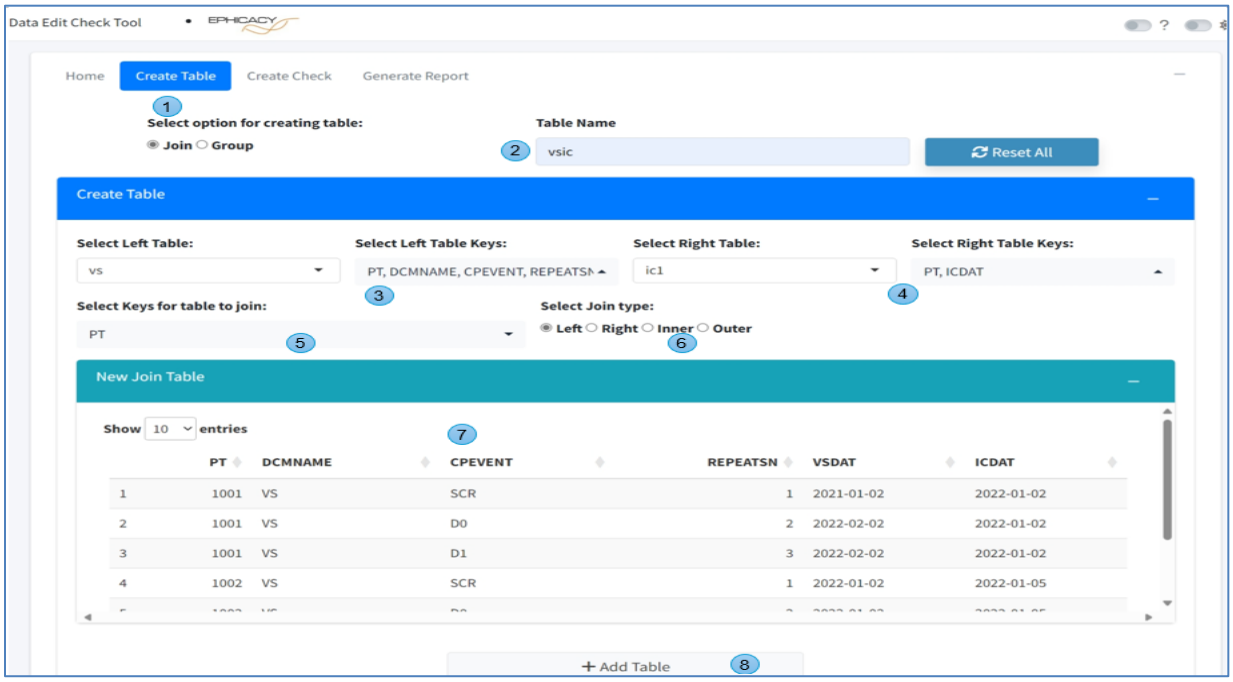

Figure 7. Screen capture of Create Table

GENERATE REPORT

The Module provides options to generate and download reports for the selected study. Enable or disable study checks. Data Manager can update the report to add the DM comments and combine the Comments from the previous reports and generate report to track the discrepancies. The two sections are.

- Report Generation
- Edit Comment

Report Generation

After setting up edit checks according to the study requirements, users can access the Generate Report tab to facilitate the generation and downloading of the validation report (Figure 8). Within this tab, the established study edit checks are executed on both the study data and intermediate data. Discrepant observations identified by these edit checks are aggregated and stored as an Excel file. The generated validation report is then stored in the study area, with users having the ability to download it to their local machines using the provided Download option.

Additionally, the tab offers users the functionality in the form of a dashboard with a list of available checks, enabling users to easily manage their preferences and selectively enable or disable specific checks as needed. This feature enhances flexibility, allowing users to tailor the validation process to suit the specific requirements of their study.

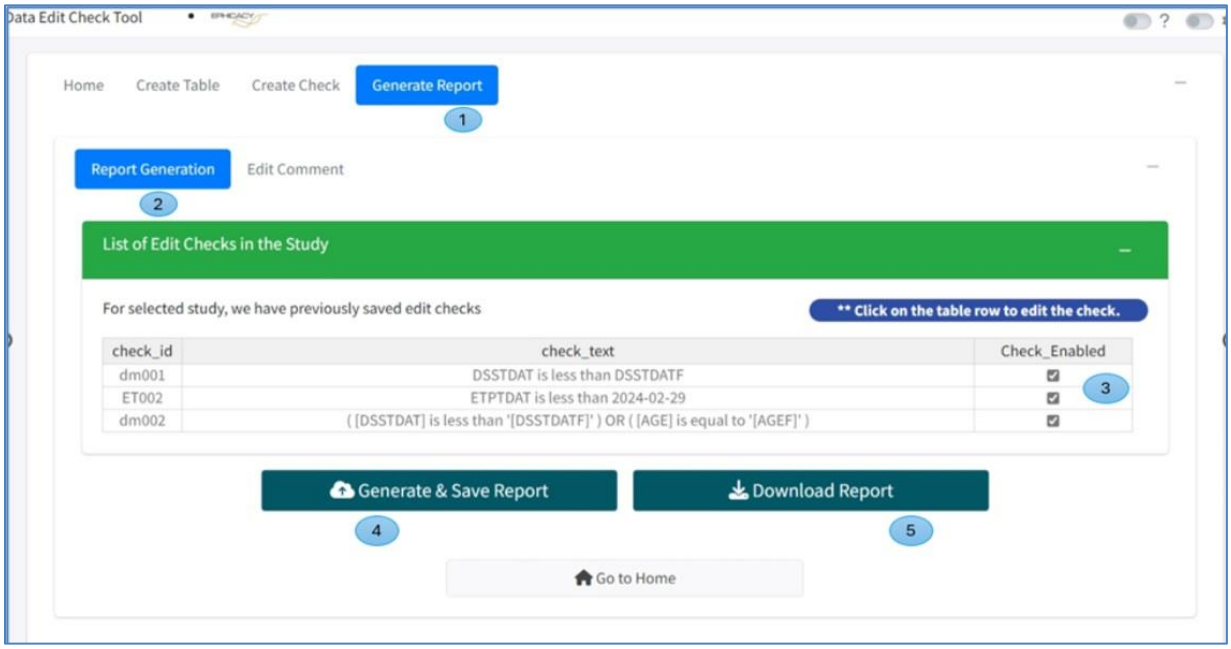

Figure 8. Screen capture of Report Generation

Sample Report

Figure 9 depicts the screen capture of sample report.

CheckNo	STUDY	PT	INVSITE	CPEVEN	REP	Q_TEXT	Review Date	DM Comments
EG001	XXXXX	01001	331013_01	DAY 7	1	ECG performed is Yes however Date of ECG collection and Interpretation are missing, please check and query.		
EG001	XXXXX	01002	331013_01	DAY 7	1	ECG performed is Yes however Date of ECG collection and Interpretation are missing, please check and query.		
EG001	XXXXX	01003	331013_01	DAY 7	1	ECG performed is Yes however Date of ECG collection and Interpretation are missing, please check and query.		
EG001	XXXXX	02001	331013_02	DAY 7	1	ECG performed is Yes however Date of ECG collection and Interpretation are missing, please check and query.		
EG001	XXXXX	02002	331013_02	DAY 7	1	ECG performed is Yes however Date of ECG collection and Interpretation are missing, please check and query.		
EG001	XXXXX	03001	331013_03	DAY 7	1	ECG performed is Yes however Date of ECG collection and Interpretation are missing, please check and query.		
EG001	XXXXX	04002	331013_04	DAY 7	1	ECG performed is Yes however Date of ECG collection and Interpretation are missing, please check and query.		
EG001	XXXXX	05001	331013_05	DAY 7	1	ECG performed is Yes however Date of ECG collection and Interpretation are missing, please check and query.		
EG001	XXXXX	05002	331013_05	DAY 7	1	ECG performed is Yes however Date of ECG collection and Interpretation are missing, please check and query.		
EG001	XXXXX	05003	331013_05	DAY 7	1	ECG performed is Yes however Date of ECG collection and Interpretation are missing, please check and query.		
EG001	XXXXX	05004	331013_05	DAY 7	1	ECG performed is Yes however Date of ECG collection and Interpretation are missing, please check and query.		
EG001	XXXXX	05005	331013_05	DAY 7	1	ECG performed is Yes however Date of ECG collection and Interpretation are missing, please check and query.		
EG001	XXXXX	05007	331013_05	DAY 7	1	ECG performed is Yes however Date of ECG collection and Interpretation are missing, please check and query.		
EG001	XXXXX	05008	331013_05	DAY 7	1	ECG performed is Yes however Date of ECG collection and Interpretation are missing, please check and query.		
EG001	XXXXX	06001	331013_06	DAY 7	1	ECG performed is Yes however Date of ECG collection and Interpretation are missing, please check and query.		

Figure 9. Screen capture of Validation Report

Edit Comment

This section allows the Data Manager to edit the validation report to add the DM comments based on the issue identified and has a date picker to collect the Date time stamp.

Figure 10 illustrates the steps involved in editing the validation report to add comments.

The tool provides Data Managers with the capability to input review comments directly within its interface. Users access the Edit Comment section, where they can browse to choose a validation report of their choice. Upon selection, the report is loaded into the interface, enabling editing of columns Review_Date and DM_Comments along with additional review column which are added in Home tab (Figure 4).

Once the comments are finalized, User can save the changes to the existing Excel file within the study area using Save Report option or can download new Excel file locally using Download Report option.

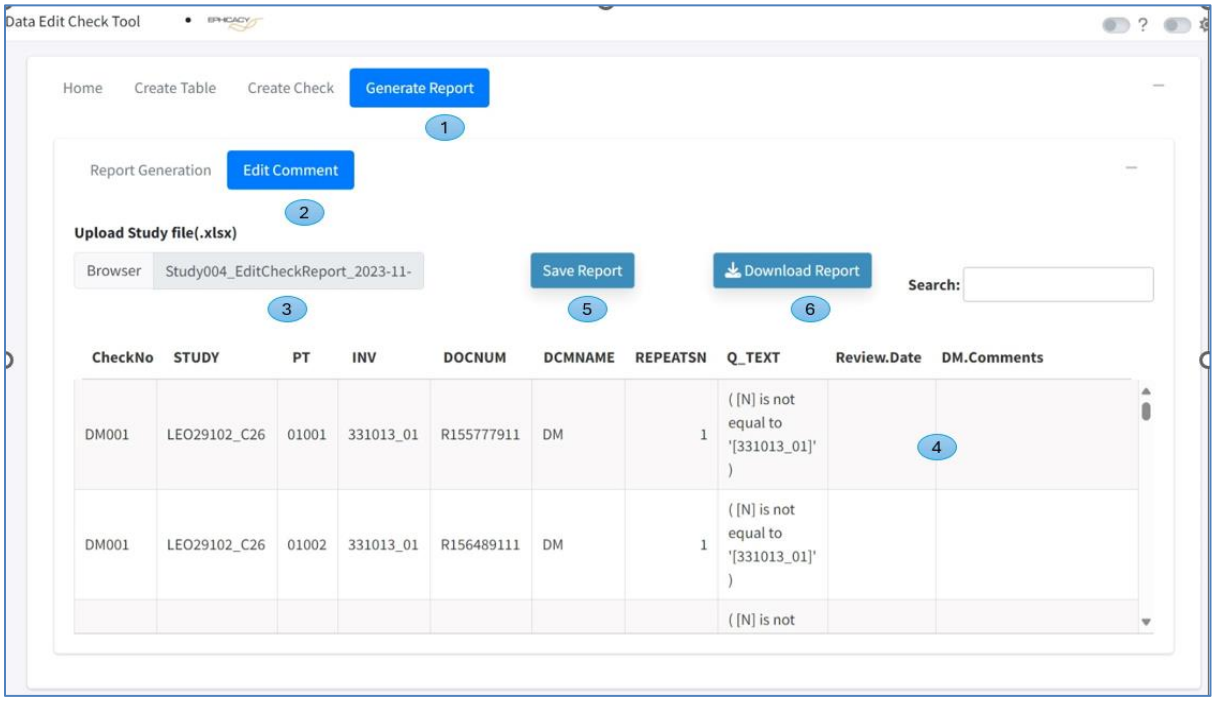

Figure 10. Screen capture of Edit Comment

CONCLUSION

The "Edit Check Tool" described in the paper is a user-case solution developed using R programming language. It serves the purpose of setting up data validation checks and generating reports with minimal training required for Data manager and other stake holders to use the tool.

In addition to features above the tool facilitate users in editing reports by allowing them to add comments within the tool interface and the library feature of the tool serves as a centralized repository for storing and reusing checks across different studies are valuable enhancements.

We are working on enhancements which enable table creation via multi-table joins and implement functionalities to derive new columns using distinctive character, numeric, and date operations. These enhancements aim to provide advanced data manipulation features for the tool.

REFERENCES

https://shiny.rstudio.com/

ACKNOWLEDGMENTS

Sincere thanks to Tyagrajan Swaminathan, Senior Director at Ephicacy Lifescience Analytics., for the continued support and encouragement throughout.

Our gratitude to Akshata J Salian and Rajan Thind, Senior Statistical Programmers at Ephicacy Lifescience Analytics., for their contribution and collaborative work towards tool development.

CONTACT INFORMATION

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

Madhavi Gundu
Ephicacy Lifescience Analytics
madhavi.gundu@ephicacy.com

Vivek Jayesh Mandaliya
Ephicacy Lifescience Analytics
vivek.mandaliya@ephicacy.com