

Innovative Data Cleaning to Prevent Unblinding Errors in Double-Blind Trials

Ling Zhang, Hoffmann-La Roche Limited, Mississauga, Canada

Nan Zhou, Genentech, A Member of the Roche Group, South San Francisco, USA

ABSTRACT

In double-blind clinical trials, accurate handling of data, such as administered kit IDs, is crucial for maintaining the integrity of the trial outcomes. Investigational sites manually enter kit IDs into the clinical database, which are later merged with the Kit Label Database following the release of Treatment Assignment Information (TAI). This process unblinds the actual treatment assignments for analysis. However, errors of kit IDs in the clinical database can lead to incomplete unblinding, thereby compromising the integrity of analyses.

Given the importance of data accuracy and challenges associated with cleaning kit ID data prior to TAI release, a Shiny App is introduced to enhance data handling processes. The shiny supports the proactive identification and resolution of potential unblinding issues. This tool is a key component of the unblinding process, aimed at improving data quality and minimizing the risk of incorrect unblinding, thereby safeguarding the integrity of clinical trial outcomes.

INTRODUCTION

Double-blind clinical trials are the gold standard for evaluating the efficacy and safety of interventions, as they minimize bias in data collection and analysis. The integrity of such trials hinges on robust data management practices, especially in the handling of administered kit IDs, which link subjects to treatment assignments. Errors in kit ID can significantly disrupt the unblinding process, leading to incomplete or inaccurate analyses and potential regulatory concerns.

This paper discusses the development and implementation of a Shiny App designed to enhance the accuracy and reliability of kit ID data handling prior to the release of Treatment Assignment Information (TAI). By identifying and addressing discrepancies proactively, the application ensures a streamlined and error-free unblinding process.

CHALLENGES IN KIT ID DATA MANAGEMENT

The process of managing kit ID in double-blind clinical trials is intricate and prone to errors. Investigational sites manually enter kit IDs into the clinical database, which introduces opportunities for human error, such as:

1. **Data Entry Errors:** Mistyped or incomplete kit IDs.

2. **Mismatches with the Kit Label Database (KLD):** Discrepancies between the clinical database and the reference dummy Kit Label Database.
3. **Timing Constraints:** The necessity to clean data before TAI release, which can be labor-intensive and error-prone under time pressure.

Errors in kit IDs can lead to several significant issues, including incomplete unblinding, where treatment assignments are not accurately linked to subjects. This can compromise data integrity, resulting in flawed analyses that may lead to misinterpretations of efficacy and safety results. Additionally, such errors pose regulatory risks by potentially violating compliance requirements that emphasize the importance of data accuracy.

The Kit ID Shiny App plays a crucial role in addressing challenges within the data quality check workflow. It is designed to identify discrepancies by cross-checking kit IDs entered into the clinical database with the Kit Label Database (KLD), helping to detect mismatches. The app also flags potential issues, such as errors or missing entries, that could impact the unblinding process. To facilitate resolution, it provides user-friendly tool for flagging discrepancies, including suggestions for fixes and interactive visualizations for review.

OVERVIEW OF DATA TYPES

The table provides a clear summary of the different data types mentioned in this paper, their TAI status, and their specific purposes within the clinical data workflow.

Data Type	TAI Status	Purpose
Kit Label Database (KLD)	TAI	Used for unblinding
Test KLD	Not TAI	Used for pre-unblinding data quality checks
Interactive Voice Kits Assignments (IVK)	Not TAI	ongoing check against RAVE clinical database
Interactive Voice Treatment (IVT)*	TAI	Used for unblinding

*Note: IVT data is not described in this paper

DESIGN AND FEATURES OF SHINY APP

The Kit ID Shiny App, developed using the R Shiny framework, provides an intuitive web-based interface to streamline kit ID quality check in clinical trials. Its core functionality centers around Kit ID check to enhance the accuracy and reliability of trial data. The application performs automated checks to ensure the completeness and correctness of kit IDs, with real-time comparisons against the Kit Label Database. An interactive dashboard offers a visual representation of data discrepancies, enabling users to identify and address issues efficiently. Features like filters and search

functionality allow quick access to specific study or patient level data. Additionally, the app offers suggestions for correcting mismatched entries.

To maintain data quality, the study data science team is encouraged to perform regular kit ID checks using this Shiny App. These checks include verifying that all kit IDs recorded in the clinical database are present in the test KLD and flagging any mismatched kits for further investigation. Exposure data should be examined for missing, incomplete, or improperly formatted kit IDs, with any discrepancies promptly queried. As shown in Table 1, patient 30453 was dosed on 2024-07-11 and 2024-08-09, Kit ID XXX6189535 contains an extra X between letters and numbers, which does not match the corresponding Kit ID in the test KLD file. Kit ID XX3246592 is missing from the test KLD file, indicating potential error.

Table 1 Scenario 1: Example illustration of Kit ID Errors

Patient	Treatment	Treatment Administration Date	Kit ID
30453	Tecentriq/Placebo	2024-07-11	XXX6189535
30453	Tecentriq/Placebo	2024-08-09	XX3246592

The app also facilitates the comparison of kit IDs between treatment exposure data and Interactive Voice Kits Assignments (IVK) by patient ID. Although mismatches may occur due to delays in data entry or transfer, all kit IDs in Rave should align with IVK after CCOD if no errors or accidents have occurred.

Furthermore, the app helps identify instances where the same kit ID is improperly entered to multiple patients or used by a single patient at different treatment visits. In Table 2, patient 30448 was dosed on 2024-05-29 and 2024-06-19 using the same Kit ID (XX3282403). This indicates that Kit IDs may not have been entered correctly.

Table 2 Scenario 2: Example illustration of Kit ID Errors

Patient	Treatment	Treatment Administration Date	Kit ID
30448	Tecentriq/Placebo	2024-05-29	XX3282403
30448	Tecentriq/Placebo	2024-06-19	XX3282403

By integrating these checks into the app’s automated processes, the Shiny App ensures a robust workflow for kit ID data management, enabling the proactive resolution of discrepancies and safeguarding the integrity of clinical trial outcomes.

IMPLEMENTATION AND IMPACT

The Shiny App was successfully integrated into the data quality check workflow of multiple double-blind clinical trials, delivering significant improvements across data cleaning processes. The implementation notably enhanced data quality by reducing the number of errors identified during the TAI unblinding process, ensuring more accurate and reliable results. Additionally, the app facilitated proactive issue resolution by enabling the early detection of discrepancies, allowing for timely corrections and minimizing delays in data analysis.

Operational efficiency was another key benefit, as the app's automated checks and interactive interface significantly reduced the manual effort required for data cleaning, streamlining the overall workflow. Together, these improvements underscored the app's value in ensuring data integrity and supporting the successful execution of unblinding

CONCLUSION

The Shiny App for kit ID check has proven to be a valuable tool for enhancing data integrity in double-blind clinical trials. By proactively identifying and resolving potential issues, the application minimizes the risk of incomplete unblinding and ensures the reliability of trial outcomes. Its user-friendly design and robust features make it a key component of modern clinical trial data management workflows.

Future developments may include extending the app's functionality to other critical data elements and integrating machine learning algorithms for predictive error detection. These advancements would further enhance the app's utility in safeguarding the integrity of clinical trial data.

ACKNOWLEDGEMENT

REFERENCE

- ICH E6(R2) Guidelines for Good Clinical Practice.
- Wickham, H., & Grolemund, G. (2017). *R for Data Science*. O'Reilly Media.
- RStudio, PBC. (n.d.). Shiny: Web Application Framework for R. Retrieved from <https://shiny.rstudio.com>.