Category: Clinical Trial Operations (Trial Start-up, Regulatory, Data Management, IITs) – Work in Progress

Cross-System Data Validation: Ensuring Consistency and Compliance with SQL Joins

D. Keusch, E. Eells, I. Nzuki, N. Karanja-Meek, M. Willer

The University of Kansas Cancer Center

1. Background

Cancer Centers are required to report clinical trial and accrual data. This information is stored in multiple databases such as the Center's internal Clinical Trial Management System (CTMS), the National Cancer Institute's (NCI) Clinical Trials Reporting Program (CTRP), and the ClinicalTrials.Gov public database. Other entities [e.g. institutions, consortiums, cooperatives, etc.] administrating or participating in trials conducted at the KU Cancer Center also report data into these systems. It can be challenging to keep data across databases synchronized and consistent when it's being reported by multiple sources. A recent change in guidelines for Cancer Center Support Grants require Cancer Centers to pull their Observational and Interventional Trial and Accrual data from CTRP. Furthermore, CTRP uses ClinicalTrials.gov to perform quality checks on data within their database. This has increased the need for a quick way to audit, compare, and update data in internal and external systems.

2. Goals

First, our goal was to create a sustainable method for comparing data from our internal CTMS to external databases. Additionally, we aimed to ensure data was accurate and matched between databases. Finally, we wanted the ability to quickly and easily monitor, validate, and verify data to allow for error correction between the database serving as the source of truth and others, and to maintain compliance with Good Clinical Practice.

3. Solutions and Methods

We created a SQL server database and loaded data tables from ClinicalTrials.gov, our CTMS, and CTRP. We created queries that combined the tables based on a unique identifier (NCI ID) to provide a comprehensive view. SQL JOIN statements were used to return all records from one table and match them to records from another. The queries only selected specific data needed for analysis and ordered columns to enhance crosschecking. This allowed us to spot inconsistencies and inaccuracies through filtering and side-by-side comparisons. For example, studies deemed reportable were checked for identifiers for ClinicalTrials.gov and CTRP. Similarly, we compared accrual data between systems, making sure accrual totals and demographic data were consistent.

4. Outcomes

We discovered that the process we developed through SQL made cross-systems comparisons efficient and effective. Manual comparisons between systems can be error-prone and time-consuming. The new method has reduced 90 percent of the time used for data comparison. It enhanced efficiency and reduced error rates by 98 percent. Finally, we adapted our process to include data audits, which led our team to identifying areas of concern in internal data reporting and allowed us to respond by creating working practice guidelines for clinical and data teams to use to improve data quality. This has yielded a significant improvement in our data governance and reduced more than half of errors in our data.

5. Learned and Future Directions

Consistency between internal and external systems is essential for data integrity and reporting accuracy. Further, our collaboration with external sites has improved as we have begun to realize differences in reporting timelines and priorities. In the future, we hope to acquire deeper insight on external sites'

processes and barriers when completing their reporting. Finally, our team intends to integrate this process with our ClinicalTrials.gov record management.

Figure