Taking Data Validation To The Next Level: Automating Data Validation Using CDASH-Standardized Global eCRFs – Phase II

Shreya Rachuri, MSc; Kathe Douglas, BA; Leilani Barry, MSc; Josh Tewell BA; Amaraja Patwardhan, BA; Robin V Johnson, M.Med.Sc; Terry Carroll, BS; Rebeca Mayorga Gattas; Kristin Brock Kennedy, BA; Saianand Balu, PhD; J. Kaitlin Morrison, PhD; Erin Crecelius, MA

[Lineberger Comprehensive Cancer Center]

Background

Ensuring the accuracy and reliability of clinical trial data is crucial for the integrity of research outcomes. Manual data validation processes are time-consuming and susceptible to human error, which can compromise data quality. The Lineberger Comprehensive Cancer Center (LCCC) recognized the need for an automated data validation system to address these challenges. This system aims to enhance efficiency, reduce the risk of errors, and improve the overall quality of clinical trial data. By automating the validation process, LCCC can detect and resolve data issues more quickly and accurately, leading to cleaner and more reliable data. This initiative is essential for streamlining the validation workflow and ensuring robust data management practices in clinical research.

Goals

- Streamline the data validation process by reducing manual efforts, improving accuracy, and increasing efficiency.
- Decrease validation time by quickly identifying and resolving more errors.
- Enhance real-time data monitoring and traceability with automated reports and detailed logs.
- Detecting data discrepancies between cross-platform source systems (EDC & OnCore).

Solutions & Methods

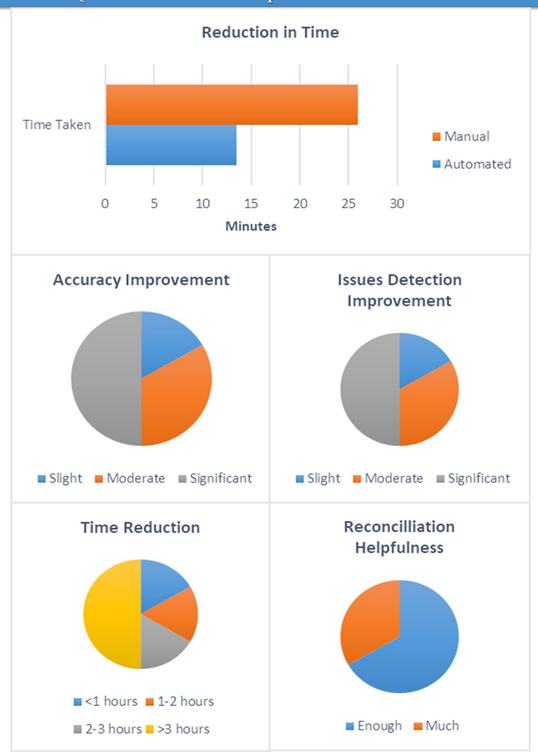
The data management team submits a Data Validation Plan (DVP) in Excel for programming. This spreadsheet contains a data dictionary with variables and metadata for each protocol. The data manager then provides programming logic for automated edit checks. The report output highlights discrepancies, enabling faster issue resolution. Cross-form data validation detects and addresses discrepancies automatically to ensure consistency across datasets. The validation report is integrated into Statistical Analysis Software (SAS) reports, which can be run through a web-based reporting interface for real-time monitoring. The Clinical Data Acquisition Standards Harmonization (CDASH) eCRF Global Library standardizes SAS programs, making them reusable across multiple studies. We compared manual and automated data review times for similar trial subjects to measure success. Data managers were surveyed to gather user feedback on accuracy and efficiency improvements.

Fig 1: Validation Outcome in Excel Sheet Listing The Data Issues And Missing Data is Color Coded for Immediate Identification.

If MHTERM_DIAG not null, MHCAT required
If MHCAT = Other MHACTOTH required
if MHTERM not null, mhstdat and mhendat required

Medical History Event Type	Medical History Term	Cancer Diagnosis	Other specify diagnosis	subtype	other specify, subtype	start date	end date	Number of relapses
Current Diagnosis	a	Colon	a	Abdomen	a	15Oct2021	14Oct2021	0.0000000000
General Medical History Event	Anaemia					04Oct2021	.	.
Episode	abc					26Oct2021	26Oct2021	.
						.	.	.

Fig 2: Impact of Automation on Clinical Data Validation: Quantitative Time Reduction and Qualitative User Perceptions



Outcomes

Feedback from Clinical Data Management Associates (CDMAs, n = 6) demonstrates perceived benefits of automation in the validation process. Regarding accuracy improvement, 50% (3 of 6) reported a significant improvement, 33.3% (2 of 6) a moderate improvement, and 16.7% (1 of 6) a slight improvement. Issue detection showed a similar trend, with 50% (3 of 6) indicating significant improvement and 33.3% (2 of 6) reporting moderate improvement. In terms of time savings, 50% (3 of 6) of CDMAs experienced a reduction of over 3 hours, while the remaining reported smaller savings ranging from under 1 hour to 2–3 hours (each 16.7%, 1 of 6). Additionally, automation improved reconciliation between electronic data capture systems: 66.7% (4 of 6) found it “enough” helpful, and 33.3% (2 of 6) rated it as “much” helpful. These findings reflect not only greater efficiency but also enhanced accuracy and workflow satisfaction associated with automation.

Quantitative findings further support the efficiency of automation in the validation process. On average, tasks completed using automated methods took 13.5 hours, compared to 26 hours when performed manually—representing a 48% reduction in time. This substantial decrease highlights the significant time-saving potential of automation and its role in accelerating clinical data validation workflows. Automation also enhanced communication among team members, with CDMAs noting improved collaboration due to more streamlined data validation processes. This also facilitates clearer and more efficient coordination between departments

Lessons Learned & Future Outcomes

Successful adoption required training to ensure a smooth transition to automated workflows. Continuous performance monitoring was essential for optimization. While automation significantly improved efficiency and accuracy, certain checks still require manual validation due to specific protocol requirements. Overall, automation has strengthened clinical data validation at LCCC by reducing errors, streamlining workflows, and improving data quality. Future efforts will focus on refining automated processes, expanding functionality, and integrating additional validation checks to further enhance efficiency. Collaboration between data management, clinical, and programming teams has also been critical to ensure the automation aligns with real-world workflow needs and regulatory compliance.