

## **Taking Data Validation to the Next Level: Automating Data Validation Using CDASH-Standardized Global eCRFs**

S. Rachuri, M. O'Dwyer, K. Douglas, L. Logan, J. Tewell, S. Balu, C. Lee, J.K. Morrison, E. Crecelius

*UNC Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill*

### **1. Background**

A high volume of clinical trial data is generated daily. Data management staff then ensure the data's accuracy, reliability, and consistency and prepare high-quality datasets for safety and efficacy analysis by biostatisticians. Hence it is essential to have the data validated from the initial stages of the clinical trial to avoid risks to patient safety and data quality. Data validation is a tedious and time-consuming process highly susceptible to human error, resulting in lower data quality. Therefore, there is a necessity for automating the data validation process, which is time-saving and more efficient than manual data validation. This automation helps in understanding multivariate data relations.

### **2. Goals**

This project aims to implement an automated data validation process using CDASH-standardized global electronic case report forms (eCRFs) to maintain data quality and integrity by automatically detecting non-compliant data. Further goals include:

- Expediting data review by directly reviewing the targeted data
- Increased accuracy of the data validation by eliminating human-prone errors and increasing query volume
- Developing a user interface to execute the data validation programming checks by entering the study reporting parameters that will run the Statistical Analysis System (SAS) program and output the data into the study folder

### **3. Solutions and Methods**

For each study, the clinical data management associate (CDMA) provides the Data Validation Plan (DVP) to the programmer in an Excel sheet. The automation uses the SAS programming language, whereby the checks within and across the CRFs are programmed into separate SAS code files. After executing the programs, data issue reports are generated as Excel workbooks containing multiple sheets corresponding to each CRF. Each Excel sheet contains tables of observations for each issue with titles describing them. The CDASH eCRF global library enables a standardized SAS program file that can be used across studies.

### **4. Outcomes**

The data review process is faster and easier as all the targeted data is in one place. We will collect metrics to assess the number of discrepancies and data review time using the new process. Automated results made a significant difference when there were many records where the issues could be easily missed in manual reviews. This new process also increases frequency of clinical data reviews. The creation of cross-form checks enabled the assessment of multivariate data relationships. Evaluating standard data checks and queries to streamline the eCRF build increased the database build efficiency. Success is contingent on improving the data review time.

### **5. Lessons Learned and Future Directions**

Some checks cannot be automated and require manual intervention. We successfully standardized the checks for forms, wherein we reduced programming time by coding individual reusable forms common

for various studies. However, we still need to write new code exclusively for study-specific forms. Phase 2 of the project will aim to build and host a website where authorized users can choose an Advarra EDC protocol from a list. Upon execution, it connects to the EDC backend and builds a SAS initialization program that generates the validation reports without the involvement of the SAS programmer.

Figure

