

Establishing metrics and standardization for Non-CRF data in EDC

Ashish Indani (Head, Research & Innovation, TCS ADD Platforms), Sharad Sharma (Domain Consultant, Life Sciences & Healthcare, Clinical Data Management, TCS), Shivaji Bote (Domain Consultant, Life Sciences, Clinical Data Management, TCS), Alejandra Guerchicoff (Industry Leader, TCS ADD Safety), Pratibha Potare (Domain SME, TCS ADD Platforms), Susan Korah (Delivery Head, Life Sciences, TCS), Vandita Tripathi (Head, TCS ADD Data Management), Jayathirtha Gopalakrishna (Consultant & Associate VP, Life Sciences, Clinical Data Management, TCS)

Introduction

Clinical Data Management is a pivotal process in clinical research, capable of impacting the success or failure of any study. During clinical research, data is collected on protocol specifications articulated in Case Report Forms (CRFs), however, there is also a significant value addition provided by external data to the CRF, called 'Non-CRF data'. The non-CRF data or third-party vendor data is collected through alternative channels. Thus, in collecting data from external sources, data integrity and quality have a critical influence on clinical trial data management and study success. The non-CRF data includes central and core laboratory data, central imaging (any type of medical images.), subject diaries which includes patient-oriented tools such as questionnaires pertaining to the quality of life, pharmacokinetics and pharmacodynamics data, safety laboratory data, genetic data, biomarkers, devices data and randomization data. A big part of this data is generated from services and components that are either outsourced or automated for direct patient interaction (Figure 1).

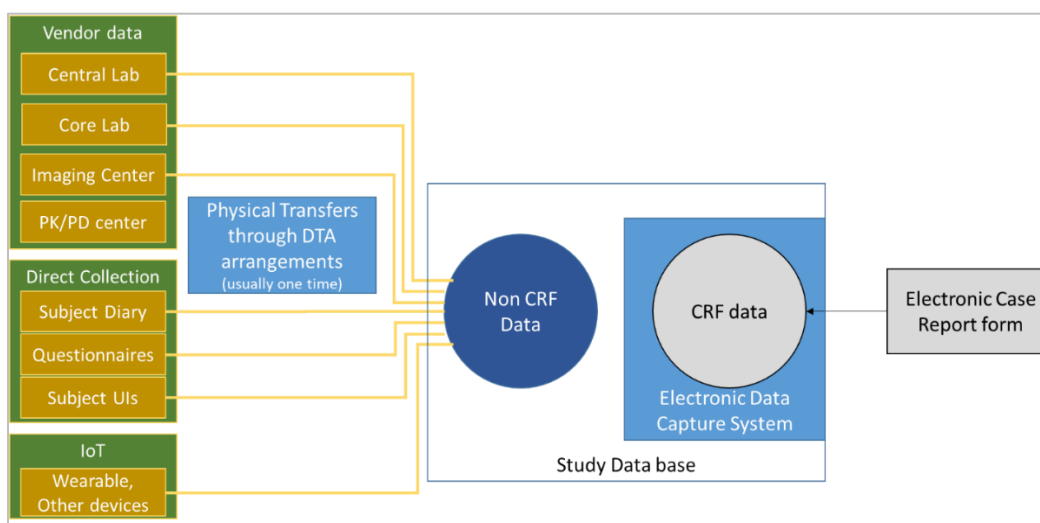


Figure 1: Common sources of Non-CRF data

Process of non-CRF data collection:

While CRF is the major contributor to the collected data, non-CRF data also constitutes a significant portion, thereby contributing to safety and efficacy of the product. The non-CRF data collected during a study is specified in the clinical protocol. To generate this data, Data Transfer Agreements (DTA) are used between sponsor and vendor organizations. However, presently there are no industry-standard formats or procedures to govern this data exchange. For an efficient selection and management of vendors, a critical aspect is to review data transfer agreements for all third-party vendors^[1]. Hence, the DTA process is extremely critical for the quality of a clinical trial data inference. DTA enables receipt of non-CRF data from vendor to the clinical database. It also defines the structure of the database, data exchange timelines, and data definitions. This time consuming and cumbersome process is critically designed (Figure 2) and includes several challenges, limitations and intricacies requiring multiple review cycles.

The non-CRF data is not configured to EDC and is received as a separate electronic file. After the data has been transferred to a clinical database, there is often a need for manual reconciliation with the existing data. Such manual reconciliations take place between visits across CRF data and the data stored in third party datasets using listings.

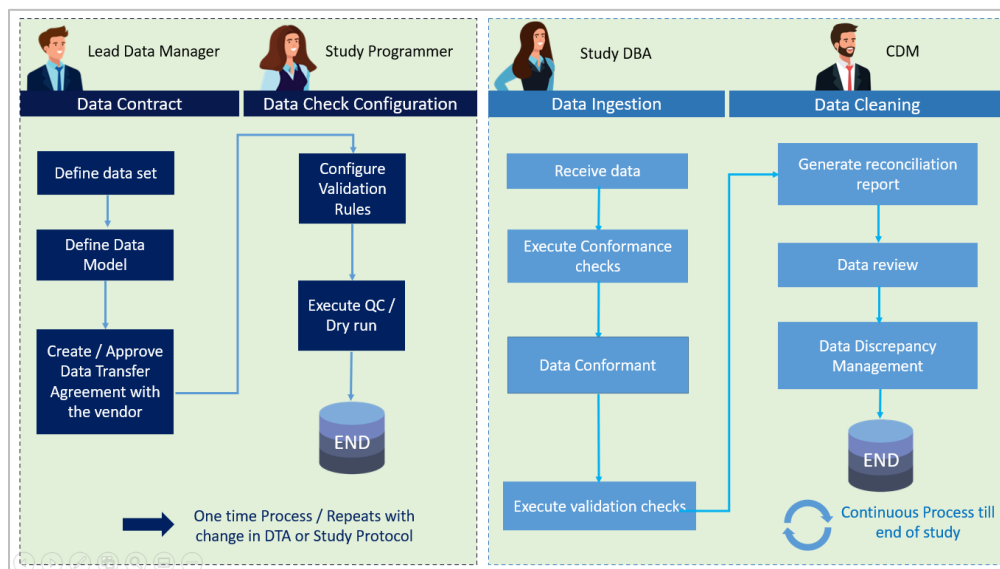


Figure 2 Process of Data Transfer Agreement

Limitations and challenges:

The process of non-CRF data reconciliation is fraught with risks of missing out on errors, missing data across datasets, identifying duplicate records etc. resulting in serious consequences on the safety outputs. These risks are compounded due to several additional challenges such as the use of standards, delivery of expected data file formats, uptake of new technology and adherence to timelines^[2]. For studies that rely heavily on this data, inaccuracies in non-CRF data can be dangerous and provide several underlying risks to patient safety. Despite the largely cumbersome process and careful evaluation of the vendors, there are several challenges (Table 1) in validation and reconciliation of non-CRF data. The challenges result in critical difficulties pertaining to data management, data quality, integrity and confidentiality. In addition,

third party data transfer is required to be handled by the Data Management Group (DMG) during the conduct and closeout phase of the study.

| |
|---|
| Diverse data transfer agreements which are non-reusable and handled manually |
| Variation and errors in vendor provided data needing QC before intake |
| Lack of standards, workflow, oversight and triggers |
| Manual reconciliation review and cumbersome report creation |
| Manual query process in supplier systems |
| Lack of metadata-based structure and standards that lead to delay in data transfer specification document and further activities related to Non-eCRF |
| Frequent changes in specification & checks which leads to Post-Production Changes in the database, resulting in a delay corresponding to the entire study cycle |
| Lack of real time reconciliation |
| Absence of a robust communication and governance structure |

Table 1 Challenges and Limitations in handling non-CRF data

Besides the underlying data issue, there are several reasons for delays in data transfers. Common reasons for the delay include people-oriented challenges (missed data entry, data transfer efforts), technological aspect (e.g. spreadsheets are only limited to include 1 million rows^[3]), process errors (missed metadata, poor set up) etc. resulting in rejection and reiteration, consuming efforts and contributing to delay in database locks in 50% cases^[4].

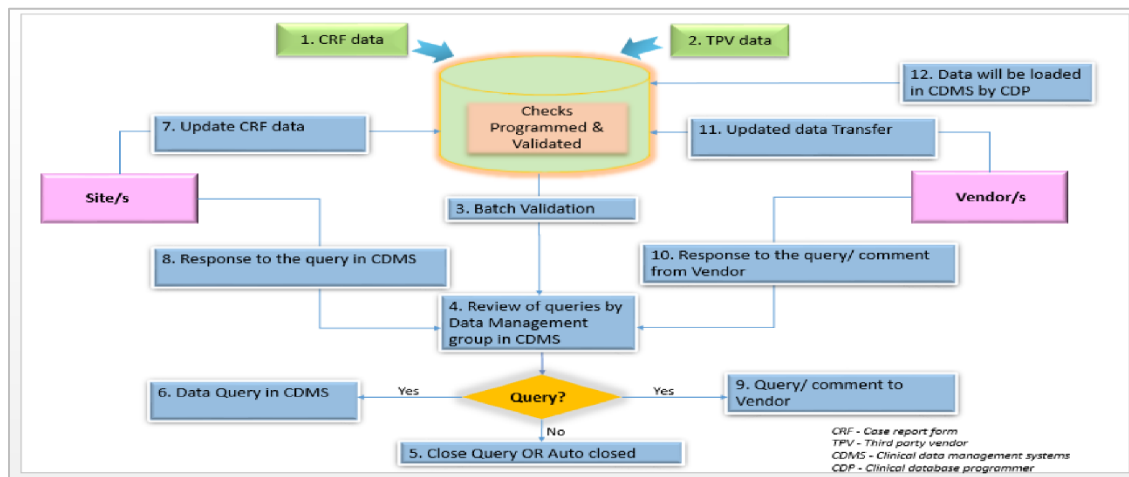


Figure 3: Third Party data reconciliation Process

Non-CRF data capture platform

The challenges surrounding the DTA and non-EDC data capture processes help facilitate the collection of such data within the non-CRF data management platform. Such solutions are critical for enabling consistent performance of activities related to vendor operations without jeopardizing the quality and timeliness during the conduct of the trial. Executing operations such as DTA, set-up, validation,

reconciliation, queries etc. with minimal human intervention is the need of hour to expedite the process, maintain traceability and avoid delays. The platform should support integration with vendor data management systems to upload and manage the data exchanges provided by vendors. Adopting a platform for such processes will not only make the process validated, traceable and recordable, but would also be instrumental in improving the data quality and turnaround time. Additionally, it would also offer data visibility across different data sets collected from various sources.

Non-CRF data transfer process

Due to an absence of common industry-wide standardized DTA and non-CRF data transfer process, the process of data transfers lack parameter-based metrics that would help determine the efficiency of non-CRF data transfers. Hence, amongst the studies, vendors and sponsors there is a large variation in this process as well as its evaluation. Subsequently, it affects the quality, timeliness, and information standards in non-CRF data in comparison to the CRF data that is well managed, monitored and process-driven. In addition, there exists an opportunity for generation of universal or at least sponsor-level standardization of this process solely by determining the standard metadata structures, defining errors and providing methods for error-elimination at the source. Furthermore, to implement a uniform industry-standard data transfer agreement and corresponding data transfers for all types of non-CRF data, a mix of commitment, participation and continuous efforts towards process development, editing and verification procedures based upon workflow driven approach from all the stakeholders is a priority

Role of technology and automation

Another key route is in deployment of automation based processes. The process of DTA generation, reconciliation, query generation, integration of non-CRF data into EDC etc. present immediate opportunities for technology to streamline the process for non-CRF data utilization. The principle idea behind this opportunity is to manage all clinical data in a single data repository, providing visibility across all datasets, facilitating automated reconciliations between the current CRF and non-CRF data, generating reconciliation report, error reports and reject erroneous data before entry into the clinical database. Automation provides a low-risk and efficient technique for reconciliation, without any data loss and saves significant amount of time and cost.

Significance of Metric Driven Approach

Each study and data type has varied demands. Hence, the success of standard data transfer ecosystem depends upon an effective monitoring of quality, compliance and efficiencies through metrics. Metrics also contribute in driving delivery governance and enables effective activity management. Periodic analysis of metrics not only reflects partnership maturity but also provides a “*true picture of delivery*” by representing gaps and quality of deliverables for all processes under purview. Key metrics suggested include:

(1) Key Performance Indicators

Key performance indicators for the non-CRF data integration are listed below:

| Area | Metric Name | Description | KPI level |
|--------------------------|---|---|--|
| Quality | Data Accuracy : Number of data errors | Measure the quality of the data received from external lab, devices and any other sources | Study, site, country, source (vendor, device etc.) |
| | Data Accuracy : Category wise errors | Measure the data accuracy by identifying the trend for 'reasons / error categories' for not able to process the received data | |
| | Data Accuracy : Error frequency | Measurement of error categories which are most frequently reported | |
| | Data Accuracy : % Reduction of errors / cycle | Measure the improvement in data accuracy post each cycle | |
| | Data Contract : Number of DTA changes | Measure how frequently the DTA was amended/updated | |
| | Data Contract : Reason for DTA changes | Identify the reasons/factors leading to DTA amendments/updates | |
| | Data Quality : Number of queries raised | Measure the number of queries raised for each data transfer | |
| | Data Quality : % Reduction in queries / cycle | Identify the reasons/types for queries for each data transfer | |
| Performance & Efficiency | Data Latency : Total time to process the data | Measure the total time taken from receipt of data to query generation and closure | Study, site, country, source (vendor, device etc.) |
| | Data Latency : Time to generate reconciliation report | Measure the time taken after data is successfully received and reconciliation report is generated | |
| | Data Latency : % Improvement / cycle | Measure and compare the time taken for data processing / cycle | |

(2) Key Quality Indicators

Quality of deliverables should be continuously monitored using key performance indicators from study start up to closeout. The below two quality parameters impacts study timelines, resulting in late submissions of clinical study report to the respective regulatory authorities.

1. Poor quality in interpretation of results or missing data
2. Poor turnaround time

It is also important to maintain an efficient and effective communication with vendors to plan schedules and ensure timeliness of data receipt. Hence it is vital to arrive at an agreement prior to the study start up. In addition, vendor data should be reviewed at regular intervals during the study conduct phase to identify generic issues and address them immediately.

Best practices to handle situations such as training, improvement areas or issues with the vendor, include leveraging reference documents, reconciliation checks and transfer of file according to the DTA, standard test names, quality assurance and quality control of vendor data at a nascent stage to avoid any delays.

(3) Key Risk Indicators

Clinical laboratory tests influence approximately 70% of the medical decisions. Hence, it is vital that trial results which are critical for safety and efficacy of the treatment, are reported with utmost accuracy. Risk management involves anticipation of errors that might happen, the assessment of errors frequency and the consequences or the severity that might come into play as a result of it, and finally the decision taken to mitigate the risks to an acceptable threshold.

(4) Service Level Agreement (SLA)

Service Level Agreement (SLA) as a metric driven approach encompasses the DTA between the sponsor and corresponding vendors. The SLA comprises of a standard flowchart and portrays some of the key roles that are instrumental in delivery of services pertaining to the collection and transfer of Non-CRF data.

Metrics as a Strategic CDM Instrument

Metrics are critical for the reduction of time and effort. Governing SOPs, timely stakeholder involvement, adequate QC procedures and well-defined guidelines are catalysts in ensuring elevated efficiencies.

In addition, unique data identifiers, a structured process for management & reconciliation of non-CRF data, redundancy checks & leveraging the already established standards (CDISC, HL7, etc.) are instrumental for ensuring a high data quality and integrity.

Conclusion

Non-CRF data is critical to various aspects of a clinical study. Despite all the importance that have been identified and laid out, there are no standard processes, quality parameters or automations that can be leveraged to capture the non-CRF data for a clinical study. Standardization and quality metric-based automation of the data transfer agreements will not only enable a seamless validation of the quality pertaining to the available data but also will be instrumental in reducing the overall cycle time. Artificial intelligence as a technology can be seen as the future of this process, that will help unlock capabilities to improvise and transform the three core pillars of the process i.e., data transfer agreement, data capture and data quality along with monitoring by automation-based technologies.

References

- [1] <https://scdm.org/wp-content/uploads/2020/01/GCDMP-Vendor-Selection-and-Management-2020-Edition.pdf>
- [2] <https://journals.sagepub.com/doi/full/10.1177/2168479016670689>
- [3] <https://support.microsoft.com/en-us/office/excel-specifications-and-limits-1672b34d-7043-467e-8e27-269d656771c3>
- [4] <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3915369/>