eSource Initiative

ISSUES RELATED TO NON-CRF DATA PRACTICES

TransCelerate BIOPHARMA INC.
ACCELERATING THE DEVELOPMENT OF NEW MEDICINES

ISSUES RELATED TO NON-CRF DATA PRACTICES

Introduction

Non-Case Report Form (CRF) data are defined as data which "include collection and transfer of data in electronic format from internal clinical trial sponsor sources (e.g., specialty laboratories) or external vendors (e.g., laboratory results, imaging, ECG, randomization, drug accountability) into clinical research data repositories/warehouses without entering the data on a CRF" in **Optimizing the Use of Electronic Data Sources in Clinical Trials, The Landscape, Part 1**[1]. Although sponsors must collect and retain this data, there is currently no applicable industry guidance.

Well thought-out practices and procedures for the collection and retention of non-CRF data are crucial to maintaining data accuracy and integrity. To guide the development of such practices and procedures, this paper outlines certain high-level best practices related to handling and managing non-CRF data. These are high-level suggestions, as the purpose is to share learnings to enable more efficient, effective use, and/or adoption of eSource. The suggestions can and should be tailored to each company's individual circumstances. These considerations can also help companies to address or reduce challenges that they have encountered and/or to initiate discussions to help improve processes.

1. Considering Availability of Standard Specifications & File Formats

a. Why is this important?

i. Predefines format of incoming data
ii. Facilitates integration with other systems
iii. Can enable automation, which increases the likelihood of discovering errors
iv. Aids interpretation and understanding of expected sponsor requirements
v. Enables development of standard preprocessing procedures needed when receiving files
vi. Creates efficiencies with vendors as they set up their systems in ways that satisfy specifications across sponsors
vii. Can be used across vendors to help with interoperability
viii. Utilizes standards, helps achieve efficiencies by minimizing time and expenses spent on non-CRF activities

b. Potential approaches/options

i. Set up Standard Data Transfer Specifications for non-CRF data by data type (e.g. pharmacokinetic (PK), safety, and biomarker) and/or therapeutic area can generate efficiencies
ii. If possible, harmonize standards company-wide
iii. Consider using existing industry standards as a starting point for non-CRF data to create a template for data specifications and file formats (e.g., CDISC's Study Data Tabulation Model [SDTM], etc.)
iv. Explore whether vendor is capable of providing data according to industry standards – or have a standardized/validated export data format
v. A sponsor might also create alternative processes for those vendors that are not capable of providing data in standard format. Discuss with the vendor what they can do and what could be done by the sponsor/company (e.g., new programming may accommodate vendor-specific variables, but conversion to required controlled terminology is not necessary if the vendor can adhere to industry standards)
vi. Consider including all transfer details in Data Transfer Specifications such as but not limited to:
    1. File naming conventions
    2. File format (e.g., Excel, CSV, text…)
    3. Metadata for content
    4. Variable names/labels
    5. Allowable values and controlled terminology
    6. Transfer frequency
    7. File should be cumulative or incremental
    8. Blinding instructions

2. Considering Standard Processes for Handling Non-CRF Data

a. Why is this important?

i. Consistency reduces potential for error and re-work
ii. Can reduce timelines
iii. Can reduce resources needed
iv. Reduces programming efforts and less chance of errors
v. Reduces training (e.g., especially for new employees)
vi. Meets regulatory and other stakeholders' expectations
vii. Provides clarity for study teams and during formal inspections

b. Potential approaches

i. Consider defining data flow process first if possible (as alluded to in "Defining Data Flow Process", section 3 below) and defining data handling processes that reflect and align with expected data flow
ii. Consider clearly defining the expected use of standard processes so it's understood by appropriate function roles (e.g., provide training to appropriate function roles, etc.)
iii. If a Data Standard Specification template does not exist, consider using existing study specifications and work to create a standard template
iv. Discuss setting data cleaning processes and standards checks during study start up
v. Consider using a software tool that can standardize checks, reports etc., by data type
vi. Think about utilizing preferred vendors that have experience with working with sponsors and skilled personnel
vii. As part of the data transfer specifications, have the vendor specify which checks they are performing on the data and on the data load
viii. Specify which checks will be performed by the sponsor on the data load (e.g., number of observations is higher in following cumulative data transfer)
ix. Suggest archiving non-CRF data specifications in the Trial Master File

3. Defining Data Flow Process

a. Why is this important?

i. Outlines process definition

ii. Identifies sender and receiver as well as the data originator and source data location

iii. Provides data lineage, integrity, and clarity (e.g., to be able to limit access and allow version control)

iv. Ensures audit and inspection readiness

v. Helps ensure IT resources are adequately prepared (e.g., System Owner, Data Owner, server storage, security and compliance)

vi. Enables alignment with stakeholders' process flow

vii. Provides structure and framework for non-CRF data

viii. Helps to identify areas of improvement and innovation

b. Potential approaches/options

i. Consider creating a standard flowchart and a list of roles to notify when data come in by study

Consider having an automated method when files are posted, retrieved, and placed in a company's system platform (e.g., receive an email alert when a file is posted, have scripts that can automatically retrieve and place files in a system platform under a corresponding study directory, and also receive an email alert once completed etc.)

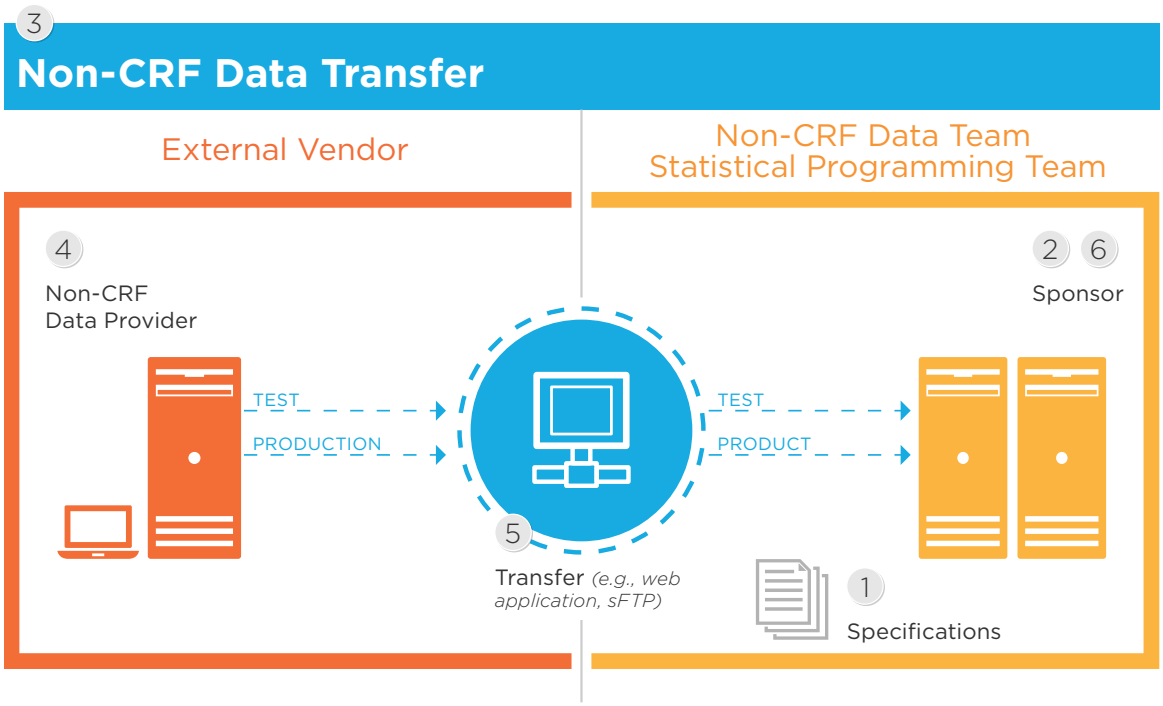
ii. Communicate data flow process to the team, and solicit feedback and comments

iii. Think about revisiting and updating the data flow process on agreed-upon frequency (and if there are processes that fail or improvements that can be made)

iv. Design an agile approach to make the flow consistent but that includes flexibility

c. **Please see example of data flow below. The numbers identify the non-CRF Potential practices listed in this document.**

***Data Flow processes**



*Data Flow Processes

1. Considering Availability of Standard Specifications & File Formats
2. Considering Standard Processes for Handling Non-CRF Data
3. Defining Data Flow Process
4. Using a Test Transfer File
5. Using a Secure Transfer Method
6. Data Review, Cleaning and Reconciliation

4. Using a Test Transfer File

a. Why is this important?

i. Ensures the transfer file process is adhered to and functioning as planned (e.g., access to FTP); End-to-end testing to ensure SYSTEM part is in place

ii. Enables the sponsor to verify that the file format complies with the sponsor specifications (e.g., file naming conventions are in place as expected)

iii. This also enables the team to refine the specifications in case the format of the data collected is different from what anticipated during the design phase

iv. Ensures the provider and sponsor are aligned prior to moving forward with production transfers

v. Ensures study teams and other relevant sponsor functions agree with expectations and what should be collected per protocol

vi. Demonstrates that non-CRF data file passes conformance checks and available for processes

b. Potential approaches/options

If test data are relevant, think about including end users and their requirements:

i. Test data can be created (i.e., 'dummy data') to conform to file specifications provided. However, when possible, it is generally preferable that the test data be comprised of actual study data to ensure compliance with specifications. If using automated transfers including monitoring, testing this part should also be considered

ii. Also, recommend utilizing the same process on test and production files to ensure there are no gaps and/or issues (e.g., same process on receiving, retrieving, storing and checking the file by using the same scripts, programs and/or tool)

iii. In addition, both provider and sponsor should use robust test data that emphasize real production scenarios as much as possible

5. Using a Secure Transfer Method

a. Why is this important?

i. Required as part of a validated system

ii. Maintains consistency on file exchange process for all vendors

iii. Ensures the security and data integrity of proprietary data

iv. Complies with laws and regulations

v. Helps ensure sensitive or un-blinded data are only accessed by the appropriate personnel via access controls/permissions in the secure transfer method

b. Potential approaches/options

i. Using a secure data transfer method that works for all vendors

1. Consider encryption methods during data transfer or when data are at rest if data are ultimately stored in cloud-based services
2. Consider having and maintaining read-only copy of file with read-only access
3. It may be necessary to have a back-up option for vendors that cannot use standard process (i.e., CD flash drive). If possible, think about using one method to exchange files for all vendors. Using many different vendor web portals, FTP servers, etc., may increase confusion and errors

6. Data Review, Cleaning and Reconciliation

a. Why is this important?

i. To ensure data integrity and completeness

b. Potential approaches/options

i. Consider creating non-CRF Data Handling Plan Template and a section in regards to non-CRF datatypes: this template can be used at study level and modified as needed. Therefore, this plan can define what to expect and how to handle unexpected scenarios for a specific study

ii. Set up standard checks to be performed on the data transfer files for non-CRF data (e.g., creating checks by therapeutic area and/or data types e.g., PK, safety, and biomarker)

iii. Set up standards reports for therapeutic area and/or data type. Medical monitors and/or bioanalysis groups may need to review the actual data and provide feedback/assessments. Please keep in mind that some function roles may be blinded or un-blinded depending on company's organization structure, so think carefully about this approach

iv. Think about adding conditional and data integrity checks along with formatting checks (e.g., check that date of visit is within expected window from protocol and/or against visit date in electronic data capture (EDC) system; have checks to define what file records may be missing or overdue based on pre-defined schedule of events, etc.)

v. Consider requesting inventory file from vendors, which confirms what has been transferred (e.g., inventory notification such as size of file, number of records, etc.)

vi. Think about defining a process with vendor on how to handle missing records, how to manage changes if a file is incremental or define if the file will be cumulative

vii. Perform ongoing non-CRF discrepancy management. This may include executing the check file and conducting data content checks every time a file is received (i.e., test and production) to check for consistency and identify existing and/or new file and data discrepancies. Here are some examples of points to consider when reviewing and reconciling data:

1. Received data compliant to specifications
2. Received complete set of data (all patients and their visits expected per protocol where possible, etc.) should be performed for each data transfer by study group (DM responsible)
3. Check for duplicate and obsolete records
4. Ensure data values match if more than one source is collecting the same information (e.g., subject identifiers, date of visit, subject initials, birth date, date of collections, PK records and samples received in data transfer file compared to data in EDC system)

viii. Perform ongoing non-CRF discrepancy management. This may consist of outlining how queries get resolved (who processes the queries and what happens during the process of getting them resolved)

ix. Think about clearly defining the roles and responsibilities on the tasks of data review, cleaning and reconciliation. This ensures all stakeholders are clear on their duties and that nothing is left out or duplicated

x. Identify timelines with each vendor. Clearly discuss and document frequency of when transfers are expected, how long reconciliation process will take after each transfer, when vendor should expect data reconciliation feedback after each transfer, how long vendor has to make requested corrections/ updates before next transfer delivered

Acknowledgements

Thank you to the following non-CRF team members from the eSource Initiative, which contributed to this document:

Aleny Caban (Medical & Development - Data Science, Astellas Pharma, Northbrook, IL, USA), Mari Clovis (Clinical Data Integration and Technology Bristol-Myers Squibb Hopewell NJ, USA), William Kesil (Clinical Data Management, Roche, New York, NY, USA), Michelle M Lewis (Clinical Trial Solutions – Center of Excellence, Pfizer Inc., Groton, Connecticut, USA),, Patricia McIntire (Clinical Development & Operations, Pfizer Inc., New York, New York, USA), Gerardo Ortiz (Safety and Biomarkers Laboratory, Pfizer Clinical Research Unit, New Haven CT), Suma Suresh (Clinical Trial Solutions – Center of Excellence, Pfizer Inc., Groton, Connecticut, USA).

In addition, the non-CRF team members gratefully thank all of the TransCelerate BioPharma Member Companies, which provided helpful input to this document.

References

1. TransCelerate BioPharma. Optimizing the Use of Electronic Data Sources in Clinical Trials, The Landscape, Part 1
http://journals.sagepub.com/doi/pdf/10.1177/2168479016670689, Accessed June 2017