PhUSE EU Connect 2018

Paper PP11

Robust Test Data

Jennie McGuirk, ICON, Ireland
Ramesh Ayyappath, ICON, US
Debra Anderson, ICON, US

ABSTRACT

Quality and efficiency of clinical trial programming development is reliant on the quality of input data. Database test data is often developed with a sole purpose in mind (e.g. edit checks), ignoring the multiple programming uses of clinical data. ICON has established a process by which robust test data is defined by the cross-functional study team, then entered into the database and made available for the various programming teams shortly after eCRF go- live. This paper/poster will describe the problem statement, the process ICON established for defining, entering and reviewing the test data, the challenges and benefits we have experienced seen since roll out of the process, and plans for building a test data library.

PROBLEM STATEMENT

- Problem Statement
  - Inadequate test cases to fully test program assumptions upfront
  - Inadequate coding / lack of robustness
  - Program re-work / updates during re-delivery cycles
  - Programming on critical path, rather than in advance
- Solution
  - Develop a process to build test cases through cross functional engagement / collaboration and enter into the test database shortly after eCRF go-live.
- Return on Investment
  - Time / Cost / Quality
    - Reduction in re-work time
    - Increase robustness of code
    - Programming can start as earlier as the test data is available

THE PROCESS

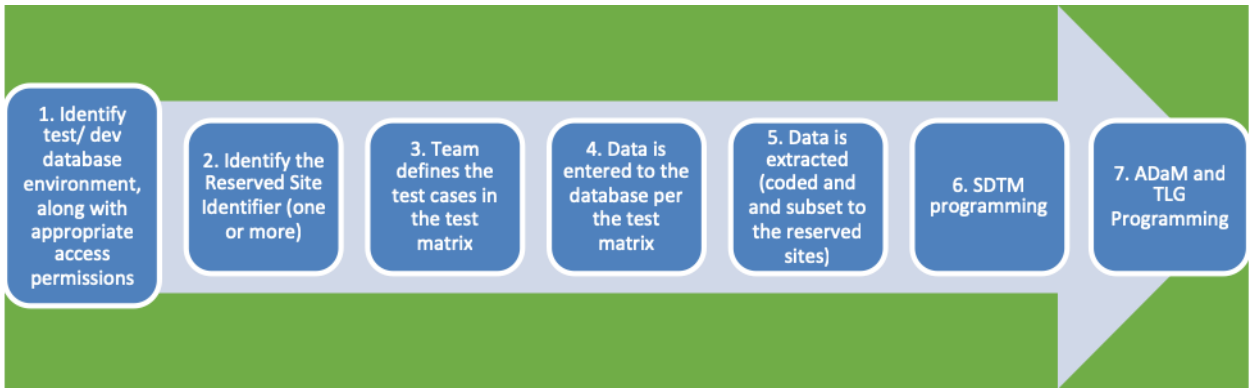
The cross functional team at ICON, includes the following:

Team Member
Data Manager
Database Lead
SDTM Lead
Data Entry and Validation Lead
Statistical Programming Lead
Statistical Lead

The prerequisites for the cross functional team to start the process are listed below.

- Final Study Protocol
- Final approved CRF version
- Final Annotated (database annotations) CRF
- Final Database Specifications
- Other associated guidelines e.g. CRF completion guidelines

As soon as the prerequisites are available, the team will start.



1. Identify a test/development database environment, along with appropriate access permissions
2. Identify Reserved Site Identifier(s) to be used only for robust test data creation and to allow for subset
3. Team meets to define the test cases (subject and associated data) relative to the study in the test matrix
4. Data is entered to the database per the test matrix
5. Data is extracted from the database, terms are coded and data is subset to the reserved sites
6. SDTM programming development, validation and review
7. ADaM and TLG programming.

As an estimate, the ICON process normally take 3 weeks to step 5 above, after which the standard ICON durations for SDTM, ADaM and TLGs applies.

GENERAL PRINCIPLES FOR DEFINING TEST CASES

The following test cases should be defined

- Screen Fail Subject
- Complete Subject (i.e. Completed Study Treatment)
- EOS Subject (i.e. completes the study regardless of treatment completion or discontinuation)
- Early Terminated Subject (i.e. withdrew after Randomisation and prior to completing Study Treatment)
- Subject Termination due to Death
- Additional test cases of interest per protocol design

When designing the additional test cases consider unique aspects of the Study design and Protocol. For example:

- If multiple arms are available a test case should be defined for each arm.
- If re-enrolment can occur a re-enrolled test case should be defined
- If there is a cross over design, test cases for each treatment sequence should be defined.

With a minimum of 8-10 subjects to ensure 2 subjects per test scenario in the matrix, ensuring at least with 2 subjects per treatment arm.

Some general principles are

- Data for each page/domain should be entered for each test subjects
- Complete subjects needs to have all scheduled assessments
- Partial dates and complete dates should be available for test subjects. EX: Jul2010, 2011, 01Jun
- Actual AEs/CMs to permit valid coding
- Date of visits should be sequential
- Unscheduled visits to test visit windowing

Whilst the test cases should be logical and clean, consider developing test cases for protocol non-compliance, for example

- Subjects that have a randomization date but missing treatment records in exposure domain
- Coded and Missing BODSYS/DECOD variables AE/MH/CM
- Multiple records at the same visit window (week/month) with differing result values.
- Multiple records for lab data with same lab test date/time but different results.
- Screen failure subjects having treatment records

- Differing bottle ID numbers for the same drug
- Difference in randomized and actual treatments
- A non-treatment subject having marked as completed in Disposition domain

CHALLENGES AND BENEFITS

There is an art to defining test cases! As this is a cross functional engagement / collaboration, the team members defining the test cases are also coming from different perspectives. The main consideration is to keep focus on the goal i.e. the end game is to have robust programs after first development, and reduce later re-work. When defining test data, the goal is not to introduce test cases so illogical that they break the programs or cause unnecessary 'coding-around' illogical data. Equally, consideration should be given to logical test cases, that are also not per protocol and/or deviate from the eCRF completion guidelines.

This process will not be of benefit to all studies. For example, this process may benefit a study with a first reporting event after a small number of subjects are enrolled. Whereas, it may not be of benefit to a study team, where the time between eCRF and first reporting event is more generous. As this process means additional upfront effort for a study team, the team should assess benefit versus investment.

For studies where this process is deemed beneficial, the next consideration is agree when to start and by when the process should complete. A general recommendation is that this process should start just prior to eCRF live (e.g. in the form of a kick off meeting), with test data available at the same time as the SDTM specification is being finalized. For ICON, this is approximately 3 weeks post eCRF live.

NEXT STEPS

Currently ICON are approaching this as a study by study process, however our plans are to build a library of quality test data overtime.

CONTACT INFORMATION

Jennie McGuirk Jennie.Mcguirk@iconplc.com
Ramesh Ayyappath Ramesh.Ayyappath@iconplc.com
Debra Anderson Debra.Anderson@iconplc.com