Creation of Test Data for SDTM QC Process

Stefan Bordasch, 18th FEB 2014

for the 18th German CDISC User Group Meeting

PAREXEL.
Right where you need us™

SDTM QC process

- SDTM is defined for cleaned, finalized studies, having all relevant data in place without data issues as missing or incorrect information
- SDTM mapping means to create programs to map all relevant information from the original clinical database into the SDTM structure
- Every existing program has to be validated, QCed, before it might be used for mapping the production environment and create the SDTM datasets
- QC process contains two parts:
  - Functional QC – prove that the data are mapped correctly as specified before – and document it
  - Technical QC – check that the programs run without producing errors, warning and special notes (e.g. format conversion), but also without hardcoding information (with some possible exception)

Test Data – why do we need them

- To be able to do the functional QC you need test data to run the programs and create a meaningful output

- In difference to the On-Line Check testing where it's enough to work with single data points, for SDTM you need complete subjects
  - In all domains exist required information that have to be in place
  - All domains are cross-linked, e.g. Finding domains to SV and that one with the TV domain
  - Subjects have to have all important decision point in place, especially for the DS domain (and so as input for the SE domain)
  - eDC forms have to be filled at best completely to check all parameter and information, as well as Supplemental qualifiers

Test Data – possible Scenarios

- In addition you have to cover all possible scenarios during a trial:
  - subject screened but not randomized
  - subject randomized, but did not receive study treatment
  - subject randomized, received study treatment, but discontinued
  - subject died
  - Subjects being randomized, treated, completed (at best 2 per treatment group having slightly different entries)
  - for open-label studies, test data should include at least one subject randomized / enrolled into each possible treatment arm
  - for blinded studies you have define dummy information for DM (arm information), EX (blinded treatment), SE (blinded elements) and if needed for DS (entries like “Randomized to Group XYZ”)

Test Data - documentation

- As for every step there should be a documentation about the test subjects
  - At least the information per subject containing the general scenario
  - Quality Management (QM) tells you to document all data points for tracking purpose

- Exact documentation means:
  - Having a very simple study: open-label, 5 visits, 10 forms containing 10 data points as average, using the previous scenarios (6 subjects needed) => we talk about 3000 data points
  - Complex Oncology study with dynamic visits and cross-over treatment => there might be several 100.000 data points

- Data entry could take several weeks !!! Documentation even longer…

Test Data – Limits…

- QM tells you to QC your programs once – and you're done for the whole study…
- In general: real live data is much more complicated than ever expected:
  - You're not able to know about all possible scenarios, especially about possible data issues that will happen
  - For the use of Controlled Terminologies – how will you know beforehand which entries you'll find in "Other, specify" fields, e.g. for Routes, Frequencies and Units
  - …and especially for Local Lab Data: how will you know what units will be entered by the sites, as you have to convert all results per lab-parameter to the same standardized unit
- Reprogramming for every data transfer might happen !!!

Data Transfer Impacts

- For each SDTM data transfer has to be a review of the datasets
  - Check for issues due new data
  - Re-run the OpenCDISC checker to identify issues
  - You’re no longer able to send out the SDTM data on the same day as the extract has been done – the review needs time

- If there is a new issue due new data, there has to be a reprogramming
  - Every re-programming means also a re-QC process
  - Timelines have to be extended
  - In worst case the mapping specification has to be updated and new sign-offs has to be requested before the SDTM outputs could be send to the ADaM team…

Test Data – Summary

- You’re not able to create perfect test data for a study
- You’re not able to copy test data between studies (no functionality in eDC systems to copy such data, in addition there is normally a different visit schedule)
- Every data transfer needs a review of the data quality
- There is always the chance of reprogramming has to be done per data transfer
- You need some time between data extract and delivery for SDTM mapping – so the timelines need to be accepted by the study team

Thank you