Ensuring consistently high quality data is delivered to clinical science & statistics for decision making and reporting is the number one priority of a data scientist (DS) in Data Management (DM). Maintaining the right level of sponsor oversight to ensure quality, when DM is outsourced to a CRO, requires a balance of different skills.

What is Data Quality?

Data Quality is a measurement of the state of collected data (validity/accuracy/completeness/conformance) for a specific analysis/business need, including acceptance criteria defined by the stakeholders. If data quality is not sufficient, results may be wrong/unfit for the intended purpose.

The Importance of Planning – Things to Consider

- Relationship building
- Encourage two way discussions
- Create a positive environment
- Fostering strong relationships will help avoid issue resolution
- For strategic partnerships, invest into governance for own function
- Set expectations from sponsor and CRO, clearly documenting roles & responsibilities
- Capture exact scope of work in Work Order

Understanding the study protocol

- It is important that DM is involved early on in the protocol development
- Understand the primary and secondary endpoints, and any composite data
- Understand data collection requirements of different parts of a study, e.g. SAD, MAD, Food Effect components; studies with both healthy volunteer and patient cohorts; studies with both adult and paediatric cohorts etc.
- Understanding what is required ensure collection is optimal and streamlined, i.e. collecting only what’s needed for analysis

Kick off meetings

- Set expectations; review Scope of Work; agree timelines, metrics, DM documentation review and sign-off
- Highlight sponsor specific DM requirements (e.g. coding, sample reconciliation)

Review of DM documents (e.g., Data Management Plan, edit checks, SDTM mapping specs etc.)

- To ensure appropriate data is collected and cleaning is planned to be performed as expected (e.g. listings may be used by a CRO whereas Roche might program a complex edit check or a custom function)
- All cleaning steps should be clearly documented to prevent any misunderstanding

Data Handling & Cleaning Rules

- Important to understand how a CRO’s database flows in order to ensure that appropriate cleaning is in place, and help interpretation of the metrics
- e.g. use of eCRF triggers, derived data
- Understanding data capture within an eForm, e.g. answers on questionnaires dependent on previous answers.

Data Quality is a measurement of the state of collected data (validity/accuracy/completeness/conformance) for a specific analysis/business need, including acceptance criteria defined by the stakeholders. If data quality is not sufficient, results may be wrong/unfit for the intended purpose.

What is Needed During The Study?

- Ongoing communication is vital, and dependent on study type, complexity and duration, DM focused meetings may be needed.
- Consider message to be conveyed; the urgency, sensitivity, timing, and the appropriate method. Although email is convenient, it could lead to misunderstandings. Consider also when a long chain of emails would be better as an ad-hoc meeting. Ensure a study communication plan is in place.

- Metrics
- Standard metrics have been agreed with the pRBD CDM preferred provider CROs, which provide a standardised way to provide metrics to Roche DSs and Study Teams. These can be adaptable depending on study type, duration and size.
- Requirements to be discussed during the DM kick-off meeting and reviewed throughout the study

Tools

- Guidance Documents
  - Examples: CRO Selection; Kick-off Meeting; Set-up; Conduct/Lock; Sample Reconciliation; Thesaurus Standards; SDTM Creation
  - Provide guidance to the Roche DS on level of oversight required whilst ensuring that the CRO remain free to follow their own processes and ensuring that any Roche specific requirements are met

- Data Delivery Expectations Document
  - Used to document data requirements for each milestone, including the type, delivery type (e.g. from the eCRF or electronic transfer) and blinded requirements for the data review

- SDTM Mapping Specification
  - Raw data can vary CRO to CRO, even study to study, with implications on the mapping
  - Need to understand differences, e.g. Safety Lab data transformed from vertical to horizontal format; the use of Controlled Terminology (CT) etc.

- LBREF
  - Metadata tool to allow standardisation of LB test names, and correct use of CT

- TIBCO Spotfire
  - Allows interactive data visualisations, as well as data listing review
  - SDTM domains loaded into specific, access-controlled folders on the Roche server
  - Used by Roche study team members: Science review for dose decisions & science oversight; DS for quality oversight and QC

Sponsor UAT/QC of SDTMv Domains

- Structure
  - UAT performed by Database Programmer
  - Standardised program, called conformance checker, in place for programmer to check SDTMv domains against the mapping structure

- Content
  - QC performed by DS against the eCRF raw data
  - Performed at key stages to check volume of data expected per protocol
  - Independent check to CRO quality checks

Conclusions

Communication, teamwork and having an in-depth understanding of the study protocol are key to delivering quality data to our stakeholders.

It can take a long time to build a good relationship and trust, which could be lost very quickly. Work from both sponsor and CRO is required. Proactivity, communication and transparency regarding study status are key in mitigating issues that might arise.

Simple solutions:
- Capture exact requirements in work order; invest the time in setting expectations
- Check and measure delivery according to expectations
- Build relationships at all levels
- Ask questions, encourage two-way discussion

sarah.malbon@roche.com vikki.horton@roche.com