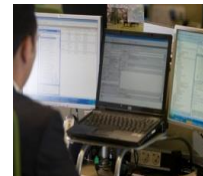


Ensuring Data Quality on Outsourced Studies - The Role of the Data Scientist

Sarah Malbon & Vikki Horton
Early Phase Clinical Data Management



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 are 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 ensures 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 (eCRF, 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

What is Needed During The Study?

- ✓ **Ongoing communication is vital**, and dependant 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 pRED 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

Primary	% missing results from data presently on database
ADCS-Cog 15	Total assessments/visits: scale not completed = 1.4% Missing individual results: 0.3-0.8%
Secondary	% missing results from data presently on database
ADCS-ADL	Total assessments not completed = 1.8% Missing individual results: 0-3.8%
MINI-MRS AD-FIN	Total assessments not completed = 1.4% Missing individual results: 0-0.2%
ADCS	Total assessments not completed = 1.8% Missing individual results: 0.2%
ADCS-COG	Change from baseline: Total assessments not completed = 0.2%
ADCS	Total assessments not completed = 0.2%

A key metric is the percentage of missing primary and secondary endpoint data

Primary	% missing results from data presently on database
ADCS-Cog 15	Total assessments/visits: scale not completed = 1.4% Missing individual results: 0.3-0.8%
Secondary	% missing results from data presently on database
ADCS-ADL	Total assessments not completed = 1.8% Missing individual results: 0-3.8%
MINI-MRS AD-FIN	Total assessments not completed = 1.4% Missing individual results: 0-0.2%
ADCS	Total assessments not completed = 1.8% Missing individual results: 0.2%
ADCS-COG	Change from baseline: Total assessments not completed = 0.2%
ADCS	Total assessments not completed = 0.2%

Standard agreed metrics

Tools

There are many tools that can be used to ensure that quality data for the study is delivered at the appropriate time to key stakeholders. This can take many forms from simple word and excel guidance documents to sophisticated computer programs for data review

- ✓ **Guidance Documents**
 - Examples: CRO Selection; Kick-off Meeting; Set-up/Conduct/Lock; Sample Reconciliation; Thesaurus Standards; SDTMv 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 data required, the delivery type (e.g. from the eCRF or electronic transfer) and blinded requirements for the data review
- ✓ **SDTMv 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.
- ✓ **LBPREF**
 - Metadata tool to allow standardisation of LB test names, and correct use of CT

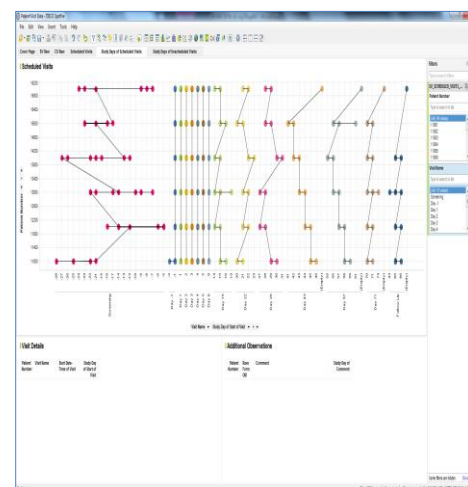
Study	Site	Visit	Parameter	Value	Unit	Source	Comments
121000	10001	1	ADCS-Cog 15	15		ADCS-Cog 15	
121000	10001	1	ADCS-ADL	15		ADCS-ADL	
121000	10001	1	MINI-MRS AD-FIN	15		MINI-MRS AD-FIN	
121000	10001	1	ADCS	15		ADCS	
121000	10001	1	ADCS-COG	15		ADCS-COG	

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

Study	Site	Visit	Parameter	Value	Unit	Source	Comments
121000	10001	1	ADCS-Cog 15	15		ADCS-Cog 15	
121000	10001	1	ADCS-ADL	15		ADCS-ADL	
121000	10001	1	MINI-MRS AD-FIN	15		MINI-MRS AD-FIN	
121000	10001	1	ADCS	15		ADCS	
121000	10001	1	ADCS-COG	15		ADCS-COG	

TIBCO Spotfire



Conclusions

Communication, team work 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