Clinical Data Management, 2nd Edition. Richard K. Rondel, Sheila A. Varley, Colin F. Webb

ISBNs: 0471-98329-2 (Hardback); 0470-84636-4 (Electronic)

6 Data Validation

PANKAJ 'PANNI' PATEL

SmithKline Beecham Pharmaceuticals, Harlow, Essex, UK

INTRODUCTION

The primary objective of Clinical Data Management (CDM) is to ensure timely delivery of high-quality data which are necessary to satisfy both good clinical practice (GCP) requirements and the statistical analysis and reporting requirements. CDM data validation activities play a critical role within the drug development programme involving many people, multiple systems and several data transfers. The quality of the data validation process has a direct impact on the quality of data presented as part of an NDA submission.

There is a general misconception that data validation activities commence when clinical trial data are presented to the sponsor's data management department. The author will attempt to dispel this somewhat narrow view and discuss various stages of data validation activities which actually start when the investigator records the data on the case report form (CRF) and when the final medical report is issued as part of the overall clinical trial data handing and reporting process.

CDM REQUIREMENT IN GCP

CDM requirements within the ICH and EU CPMP GCP guidelines are not defined in any great detail, resulting in lack of clarity, or indeed misrepresentation. The FDA Code of Federal Regulations contains no mention of CDM! This should be considered as a major concern, given that CDM plays a vital role in protecting data integrity, and is charged with producing high-quality databases that meet clinical and regulatory requirements. However, the GCP guidelines do devote a chapter to the 'data handling' aspects, including the requirement of quality control/quality assurance mechanisms to ensure reliable data capture and subsequent processing.

DATA VALIDATION PROCESS DURING THE CONDUCT OF A CLINICAL TRIAL

It is the sponsor's responsibility to implement and maintain quality assurance and quality control mechanisms at each stage of the data validation process to ensure data are generated and processed in compliance with the study protocol and GCP requirements.

What is the definition of data validation? It is a defined number of steps needed to turn the original or 'raw' item or items into the finished item, that is to turn CRF data into a clean database. These steps should ensure that the database is accurate, consistent and a true representation of the patient's profile.

Where does the data validation step start? Is it at the investigator site, when the data are first recorded on the CRF or does it begin when the CRF is presented to the sponsor company's CDM department? It starts at the investigator site and stops when the final medical report for the study has been issued by the sponsor company.

Data Validation Steps Performed by the Investigator

The GCP guidelines are quite clear on when the data validation step starts; the ICH guidelines state: 'The investigator should ensure the accuracy, completeness, legibility, and timeliness for the data reported to the sponsor in the CRFs and in all required reports.' The investigator should ensure that any data reported on the CRF are consistent with the patient's medical records and, where applicable, discrepancies should be explained. The CRF should be signed and dated by the investigator and/or the investigator's designate. In addition, all corrections on a CRF should be dated, initialled, and must be made in a way which does not obscure the original value.

Patient diary card data can be an important source of information about drug compliance, drug efficacy and daily activities. However, diary data can be very unreliable and it is imperative that the investigator reviews the diary's completion with the patient for completeness and accuracy of data recorded.

The sponsor should ensure investigator training and education on the need to accurately record data on CRFs and the impact this has on the overall quality of the clinical trial. A perfect data management system can do little to improve sloppy data produced at the investigator site.

Data Validation Steps Performed by the Monitor

GCP states that the 'monitor should check the CRF entries with the source documents and inform the investigator of any errors/omissions' and 'assure that all data are correctly and completely recorded and reported'.

This requirement is achieved through Source Data Verification (SDV), the process by which the information reported on the CRF by the investigator is compared with the original medical records to ensure it is complete and accurate. SDV is a fundamental step in the data validation process to ensure data integrity and maintain quality of data captured at source. Through the SDV process, the monitor should confirm accurate transcription of data from source files to the CRF and that the CRF contains all the relevant information about the patient's participation in the clinical trial.

There are two methods of SDV: Direct Access—the monitor is given direct access to the actual source document, and conducts an independent comparison versus the CRF; Indirect Access—the monitor is not allowed access either to the actual or to the photocopied source document. Key variables are chosen for which the investigator or member of staff reads the source document entry while the sponsor compares it with the CRF entry. This method is the most time-consuming but ensures the highest level of patient confidentiality.

Direct access to source documents must be the preferred choice in order to maintain data integrity and improve quality of data at source (i.e. at the investigator site). Sponsors should exclude investigators who do not allow direct access by sponsor and regulatory personnel to source documents. The USA FDA have recognised the importance of reviewing source documents and as such demand direct access to these documents. The responsibilities of both the sponsor and the investigator in SDV must be finalised at the outset of the clinical trial with a view to ensuring there are no misunderstandings of the requirements of SDV.

SDV is an integral part of data validation procedures, as required by GCP, and one could argue that if it is not possible to verify data in CRF as part of SDV due to unavailability of source documents, serious consideration should be given to excluding the data from the final study report.

Once the CRFs have gone through the SDV process, they are sent to the sponsor's CDM site for subsequent processing.

Data Validation Steps Performed by CDM

CDM data validation activities are an integral part of GCP and fundamental to the delivery of high-quality data for statistical analyses and reporting. Attention should be focused on ensuring that the data are a reasonable representation of what actually happened at the investigator site. The aim is to transform data recorded on CRFs into information that can be used in the final clinical report from which the right conclusions about the new drug can be made. Figure 6.1 represents a generic model for processing CRFs through CDM's data validation activities. Data clarification queries are issued to the investigator at various stages in the process, in particular, as a result of pre-entry review, data entry, and the running of edit checks.

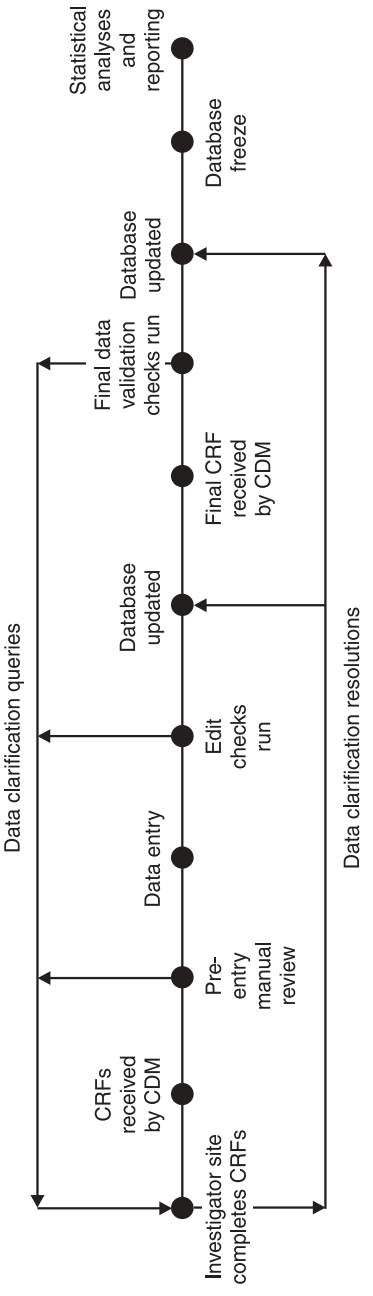

Figure 6.1 CDM data validation activities

CDM data validation guidelines should be developed to ensure data are processed in such a way as to maximise data integrity and to deliver high-quality data for analyses and reporting.

Process for defining and implementing edit checks

Edit checks consisting of manual and computer checks need to be performed on the data to ensure the database is accurate and consistent. The definition stage consists of producing an Edit Check Specifications (ECS) document and the implementation stage involves the programming and testing of the checks.

Figure 6.2 represents a generic model for defining and implementing ECS checks for which the data management is to be conducted by the sponsor company's own CDM group. The finalisation of the ECS document is the responsibility of the ECS Team, consisting of all functional groups who have a vested interest in the data generated from the clinical trial. In particular, the clinical and statistical groups are key players of the ECS Team, whose input in the development of the ECS document is critical to ensuring adequate checks are defined and implemented in the cleaning effort to deliver as high-quality database.

The first step in the process is for the clinical data manager to prepare and circulate a draft ECS document to the ECS Team, subsequent to which a document review meeting is held to finalise the document. It is essential all members of the ECS Team attend the meeting so the implications of the checks can be clearly understood. However, there may be a need for a further meeting if approval by all team members is not obtained. At this meeting all outstanding issues are resolved and the document signed-off.

Once the ECS document has been signed-off, the next phase is to complete the Edit Check programs. Sufficient time should also be allocated for the testing of the programs through the use of robust data prior to running the programs on live data. Test data should be created for all ECS checks specified, comprising both good and bad data to ensure only bad data are located in the output.

Figure 6.3 represents a generic model for defining and implementing Edit Checks for studies which are outsourced to Contract Research Organisations (CROs). The main differences to the in-house model are:

- The clinical data manager at the CRO is responsible for preparing and circulating the ECS document to the ECS Team
- The CRO is responsible for the programming and testing of the ECS checks

It should be noted that the need to ensure an ECS Team is set up applies equally to outsourced studies as for in-house studies. In addition, the timelines are the same as for in-house studies.

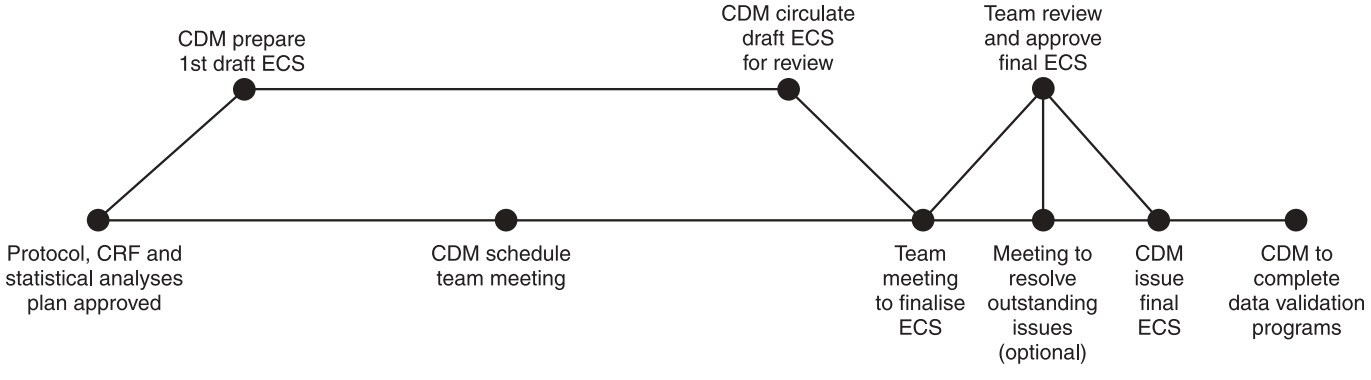

Figure 6.2 ECS finalisation process for in-house studies

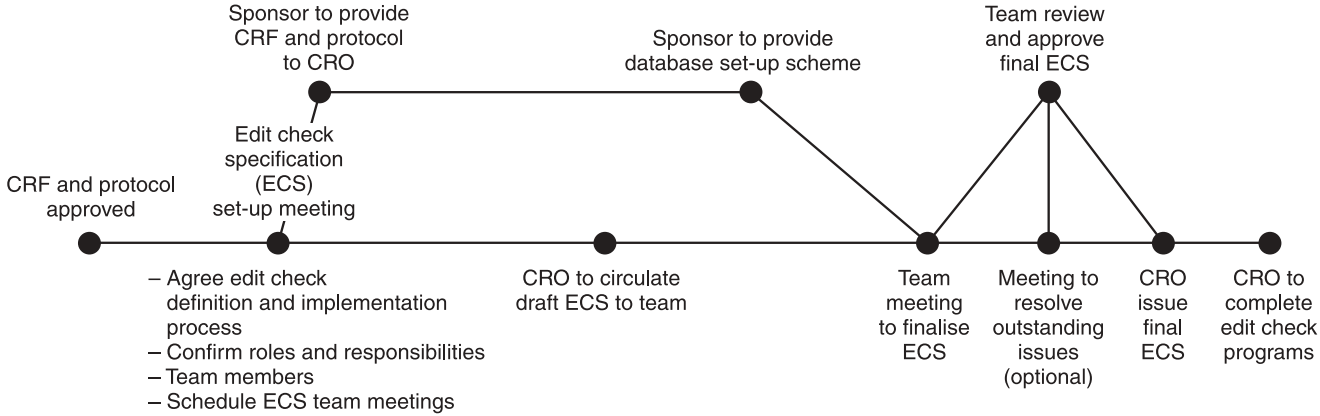

Figure 6.3 ECS finalisation process for studies contracted to CRSs

Any changes to the ECS document post finalisation need to be reviewed and approved by the ECS Team prior to implementation, irrespective of whether a study is in-house or contracted to a CRO. The clinical data manager should complete a 'request for amendment to ECS' form (see Figure 6.4), outlining the impact of the proposed change. It is essential that

REQUEST FOR AMENDMENT TO EDIT CHECK SPECIFICATION (ECS)

Sponsor study number:

CRF page numbers plus section:

Change: ☐ New specification – specific new edit check number

☐ Change/amendment to existing edit check specification (specify edit check number)

Reason/Impact

Approved: Yes/No If 'Yes':

Planned date of implementation:..... Actual date of implementation:.....

Approved by representative member of:

Sponsor study team .....

CRO data manager .....

Figure 6.4 Request for amendment to ECS form

the impact on the statistical analyses plan and database be assessed, together with whatever back validation may be required as a result of the change in ECS.

The ECS document should itemise all manual and computer checks which will be performed on the data at either pre-entry review or post-entry. General assumptions as to how to handle dates, timefields, text strings, partial dates, units, continental decimals/commas and so on all need to be specified. Any derived data points should also be included in the document, that would wherever possible impact on computer checks. An example of an ECS format and content is represented in Figure 6.5.

EDIT CHECK SPECIFICATION DOCUMENT

DEMOGRAPHY: EDIT CHECKS

CRF module: DEMOG
Page: 1

1(a) List if date of birth is missing

(b) If DOB is missing then output:

(c) 'Please provide the patient's date of birth'

2(a) List if study date – date of birth is ≤ 18 and ≥ 65.
If age is not within range then query. If query confirms that the date of birth is correct then record this as a protocol violator in the data handling file

(b) If DAT – DOB is ≤ 18 or ≥ 65 then output:

(c) 'Please confirm patient's date of birth'

(a) English terminology, (b) Technical terminology, (c) Investigator query text

Page 1 of XXX

Figure 6.5 Example of demography ECS page

Timelines for defining and implementing edit checks

The ECS finalisation process should commence after the protocol CRF have been finalised. The objective is to ensure that all checks are defined and implemented upfront, within a matter of weeks, and prior to the first CRF in-house. This helps to promote in-stream validation of CRF data and timely issue of any data clarification queries to the investigator sites.

Factors affecting quality of data

There are a number of factors which have an underlying impact on the overall quality of the data collected. These considerations warrant further discussion.

1. *CRF design.* CRFs need to be carefully prepared to collect data completely and accurately. Both the protocol and the CRF need to be designed in parallel to ensure consistency between the two. The CRF should allow collection of the data as requested in the protocol and the format should follow the protocol's treatment procedure. Adequate quality control procedures need to be implemented to ensure timings of visits and examinations match and that duplicated data are not being captured in different places. If the CRF does not allow data capture as requested by the protocol then errors are built into the study instead of quality, which would inevitably result in a high number of queries being generated as the CRF is processed.
2. *Field monitoring guidelines.* The quality of field monitoring guidelines has a direct correlation to the quality of data presented to the sponsor's CDM department and, subsequently, the volume of queries that need to be generated. Field monitoring guidelines should be developed in parallel with the CDM's data validation guidelines to ensure consistency of data monitoring and cleaning between the monitor and CDM. Field monitoring guidelines should be developed to ensure data integrity, and to check that the transcription of data from source documents to CRF is correct, complete and reliable.
3. *Source Data Verification (SDV).* As previously discussed, SDV is a critical phase of the data validation process, without which the integrity and quality of data would suffer. SDV is an effective way to ensure that the data reported by the investigator to the sponsor are accurate and valid.
4. *Missing data/CRF pages.* GCP guidelines clearly state that 'appropriate measures should be taken by the monitor to avoid overlooking missing data . . .'. However, large boluses of queries often get generated by CDM to retrieve missing data.

5. *Date conventions.* For multicentre clinical trials, differing date conventions being used by various investigators can present problems when it comes to entering the data on the database. It is vital that this issue is recognised at the outset of the clinical trial during the CRF development phase.
6. *Electronic laboratory data.* The main considerations are reconciling electronic laboratory data to the database:
   - How do you match a patient's screening lab sample when the patient's unique identifier is yet to be generated?
   - What course of action should be taken if the patient's demography details on the electronic lab data do not match the demography details recorded on the patient's CRF?
   - What units are going to be used? will these be familiar to the investigator?

There are further considerations for multicentre clinical trials:

- How do you deal with variations of tests used by different labs?
- Should conversion factors be used to make multiple ranges compatible or for a central laboratory's appropriateness of reference ranges for a patient population spanning across different countries?

QUALITY CONTROL AND QUALITY ASSURANCE OF CDM PROCESSES

Both the ICH GCP and EU GCP guidelines state 'Quality control must be applied to each stage of data handling'. The CDM process is quite complicated and can involve many people and multiple systems. It is important, therefore, to have an effective, quality-controlled system so that the process runs smoothly and efficiently. One possible way of ensuring that the CDM process is operating effectively and conducted to GCP requirements is through audits. It is important to have written policy that describes the auditing process and has been agreed by senior managers in data management. The policy should describe the range of audits to be performed and whether they are study-specific or system audits. It may specify that audits be performed by sampling across all clinical projects and all phases of a clinical trial program. Depending on the type of audit, sampling could also focus on other criteria, such as the critical data collected during the clinical trial. It is important to note that data management audits should be no different from audits conducted in any other area working to GCP standards.

In the mid to late 1980s, the FDA in the US, followed by France and Germany, began asking sponsor companies to have written SOPs as part of

the GCP. The impetus for this came from several drug withdrawals, such as benoxaprofen, from the market and the media publicity about side effects of recently introduced products such as non-steroidal anti-inflammatory drugs.

In the early 1990s, several countries belonging to the EU followed suit with similar requirements. Those involved include the UK, Japan, the Nordic countries; Canada, EU member states, Spain and also the World Health Organisation. The FDA Code of Federal Regulations (CFR) and CPMP GCP guidelines stipulate the establishment of operational SOPs for conducting and monitoring clinical trials. CDM plays a vital role in protecting data integrity and the need to ensure that standard operating procedures (SOPs) are defined that encompass all aspects of the clinical trials' CDM process, which helps assure adherence to the FDA CFR and CPMP GCP guidelines and regulatory requirements.

These SOPs should not state all the details of these guidelines but should highlight the key points and present systematic ways of performing CDM activities to ensure compliance to the guidelines. SOPs are a tool which ensures the generation of quality data to support drug development. Drug development that does not conform the internationally accepted standard of GCP cannot be justified on ethical, moral, or economic grounds.

SOPs concerning the preparation of documents such as protocols, study reports, safety summaries and Investigational New Drug applications must encompass all GCP and regulatory requirements. Some typical CDM SOPs which ensure compliance to the above mentioned regulations include:

- Generation and Maintenance of Study File documentation
- CDM QC/Audit procedures
- Database design
- Query generation and resolution
- Data entry
- Dictionary coding
- Document Management including archiving
- Data Validation activities
- CRO selection and monitoring
- Database Freeze
- Systems validation and maintenance

CDM DATA VALIDATION IN THE FUTURE

In the quest to reduce development times of new drugs, new technologies and working practices are being tried and tested within CDM; for example,

pen-based systems, optical imaging, voice recognition and, last but not least, remote data entry (RDE). These new systems have a direct impact on the data validation process. If we were to look at RDE, the investigator would enter data directly into the RDE system via electronic CRFs. The core of the edit checks could be implemented within the RDE software. Thus, the majority of the validation checks would be performed in 'real-time' at the investigator site. RDE would streamline processes and make data capture more efficient by displacing activities which are a bottleneck or by removing those which do not provide significant added value.

In the future, the success of new systems such as those mentioned above will be measured in terms of:

- Time savings (data flow from investigator site to sponsor, processing time)
- Reduction in resource requirements (with sponsor's clinical and CDM groups)
- Improvement in Data Quality
- Endorsement by regulatory authorities

SUMMARY

CDMs are charged with producing high-quality databases that meet clinical and regulatory requirements. The quality of a clinical trial determines the acceptability of the results and care must be taken to ensure that high standards of quality are present both in the clinical trial design and in the integrity and interpretation of data. To this end, all participants in the clinical trial have a role to play in safeguarding data integrity. As discussed, data validation activities start at the investigator site and end with a statement in clinical or expert reports to indicate that the clinical trial was conducted in accordance with GCP and that the report provides a complete and accurate account of the data collected during the trial.