Sciences Procedural Outline of Clinical Data Management Under Clinical Research Trials

Pooja Arora*

Department of Pharmacoinformatic, National Institute of Pharmaceutical Education and Research (NIPER), India

Submission: June 8, 2023; **Published:** July 07, 2023

***Corresponding author:** Pooja Arora, Department of Pharmacoinformatic, National Institute of Pharmaceutical Education and Research (NIPER), Sector-67, S. A. S. Nagar (Mohali), India, Email: poojaarora@niper.ac.in

Abstract

Clinical data management is a systematic informative process to generate high quality data that gets converted into information and knowledge essential for decision making [1]. The importance of CDM is highlighted initial role to ensure sufficient integrity in valid results and conclusions with high quality standards that in-turn helps in review and approval of new drugs by regulatory agencies that trust clinical trials. So, rigorous training is induced for all key players involved in clinical trials. Hence, clinical data management is a critical task for collection, integration and validation of clinical trial data [2]. The end result of CDM is to provide a study database that is accurate, secure, reliable and ready for analysis. It accelerates the timeline from data collection to analysis [3]. Multidisciplinary team should be well versed from stage of inception to completion with CDM process under high quality standards. The principle of ICH GUIDELINES, GCP describes that clinical trials should be scientifically sound and should have clear, detailed protocol in the form of data management plan or statistical analysis plan. ICH is a joint venture by both regulators and industry as equal partners in scientific and technical discussions of testing procedures to ensure and assess the safety, quality and efficacy of medicines [1]. The primary objective of CDM is to provide high quality reliable data for randomized controlled trials in line with good clinical practices. These trials are conducted according to GCP standards to ensure internal and external validity with CDM conducted by team members [4]. The procedure under CDM includes development of case report form, CRF designing CRF annotation, data designing, data entry, data validation, discrepancy management, medical data coding, data extraction and database locking [5]. Regulatory compliant data management tools like Open Clinica and QMS PLUS are available for substantial reduction in time and error between the data entry and database locking. It is becoming mandatory for companies to submit electronic data through EDC (Electronic Data Capture) [6]. This review article would focus on the procedure, tools and standards adopted as well as the roles and responsibilities of key players of CDM process.

Keywords: Electronic data capture; ICH; Good clinical practices; Case report form; Ecrf; Data validation

Introduction

Generally, in biomedical and health related studies that involve human subjects with the objective of examining safety and efficacy, a pre-defined protocol is a must. So, clinical data management systems (CDMS) are required to manage the data generated by conducting clinical trials. CDM has consistently been recognized as a primary part of clinical development team and in some instances leads the team. CDM has evolved from data entry process into a diverse process by providing clean data in a usable format in timely manner, providing a database fit for use and ensuring clean data and database ready to lock. Clinical data managers work from entering CRF data, merging of non CRF data, designed to identify bad information, generating and tracking CRF and queries, determining protocol violators to interacting with the site personnel to resolve data issues. At the end, attain high quality data by minimizing errors and missing values as much as possible and gather the data for analysis. For identification and resolution of data deviations, sophisticated software applications are available to handle large trials and ensure data quality in complex trials. High quality data should meet the protocol specified parameters and attain suitable results for statistical analysis as per the set objectives. So overall after development of protocol, it followed by CRF design, database design, data validation /derivation procedures end into database ready to accept the product data [7].

Tools for CDM

For multicentric clinical trials to handle big data, clinical data management systems (CDMS) are available like Oracle Clinical, Clintrial, Macro, Rave and e-clinical suite. For open-source

software tools Open Clinica [8], Open CDMS, Trail DB AND PhoSCo are the most prominent software's working in offline mode [5]. Now QMS Plus [9] (online mode) is an innovative approach and is found more efficient because of reduction in workload on main study database and has multiple user capacity. It has an automated system, standardized query management and improves efficiency by reducing the risk of human error. It also has potential to adapt into an open access tool for wider applications tools ensure audit trials and help in management of discrepancies. Each user can access only their user ID and cannot make changes in the database. Where changes are permitted, software will record the change made, user id that made the change, time and date for audit trial. So regulatory auditors, can verify discrepancy management process [5].
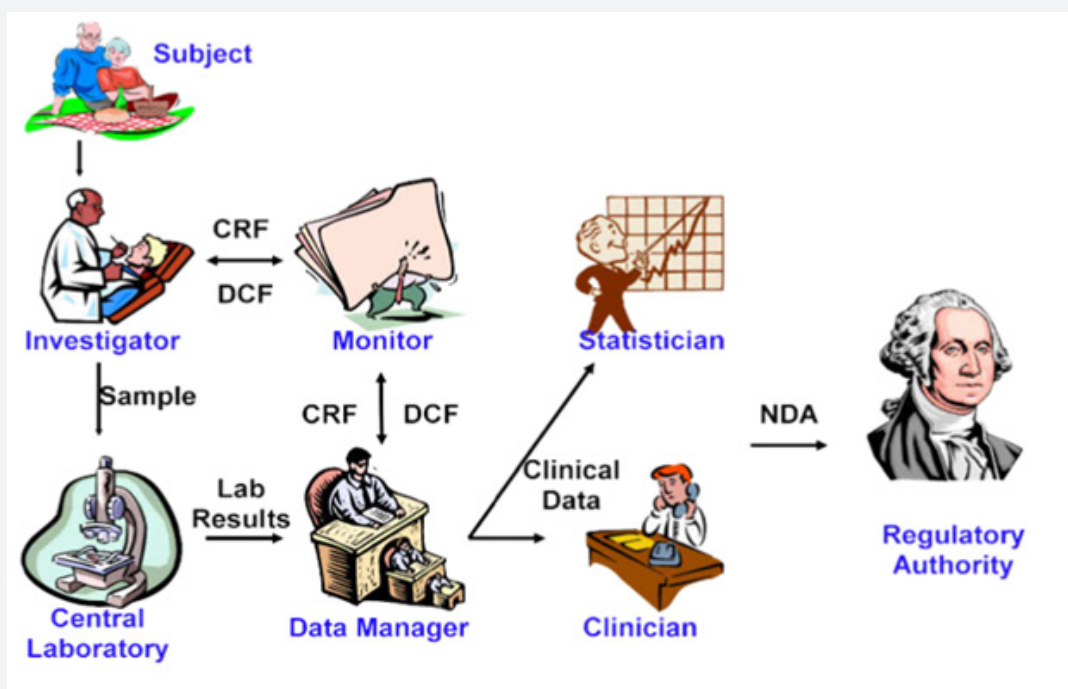

Figure 1.1: Multidisciplinary CDM team.

GCP GUIDELINES: International Conference on Harmonization Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) [10] –a joint initiative involving both regulators and industry as equal partners in the scientific and technical discussions of the testing procedures which are required to ensure and assess the safety, quality, and efficacy of medicines. All clinical research data should be recorded, handled, & stored in a way that allows its accurate reporting, interpretation & verification as per the guidelines. (ICH GCP 2.10, 4.9, 5.5, 5.14 & ICH E9 3.6 & 5.8). Systems with procedures that assure quality of every aspect of research should be implemented. (GCP 2.13) [11]. Quality assurance & quality control systems with written standard operating procedures should be implemented & maintained to ensure that research is conducted & data is generated, documented, recorded& reported in compliance with protocol, GCP & applicable regulatory requirements (GCP 5.1.1). If data is transformed during processing, it should always be possible to compare original data & observations with processed data (ICH GCP 5.5.4). Sponsor should use an unambiguous subject identification number or code that allows identification of all data reported for each subject (ICH GCP 5.5.5) [12]. Protocol amendments that necessitate a change in design of CRF, subject diaries, study worksheets, research database & other key aspects of CDM processes need to be controlled(ICH E9 2.1.2).Common standards should be adopted for a number of features of research such as dictionaries of medical terms, definition & timing of main measurements, handling of protocol deviations(ICH E9 2.1.1). 21 CFR Part 11 applies to all data (residing at the institutional site and the sponsor's site) created in an electronic record that will be submitted to FDA. The scope includes validation of databases, audit trail for corrections in database, accounting for legacy systems/databases, copies of records and record retention [11].

Team of CDM

The team of CDM is multidisciplinary in nature [2]. It comprises of clinical investigator, site coordinator, pharmacologist, trialist/methodologist, biostatistician, lab coordinator, project manager, clinical research manager/associate, monitor, regulatory affairs, clinical data managers, clinical safety surveillance associate (SSA),IT,IT/IS personnel, trial pharmacist, clinical supply and

auditors all these are involved in CDM process as shown in figure 1.1 After the review of the protocol, the data managers will identify the data items to be collected and frequency of collection. Every clinical study should have a data management plan to ensure and document the adherence to good clinical data management practices for all phases of the study. Although a study protocol contains the overall clinical plan for a study, separate plans such as a data management plan (DMP) [13] or statistical analysis plan, should be created for other key areas of emphasis within a study. A well-designed DMP provides a road map of how to handle data under any foreseeable circumstances and establishes processes for how to deal with unforeseen issues. DMP is a document which describes and defines all data management activities and helps an organization develop and standardize data management procedures along with DMP data validation plan containing all edit checks to be examined and calculations of derived variables [1]. These edit checks help in data cleaning and identifying the discrepancies. The elements of DMP are preparation of CRF design, study setup, data capture, tracking CRF data (CRF workflow), cleaning data, managing lab data, SAE handling, coding reported terms, creating reports and transferring data and closing studies. The most important step in ensuring data quality is appropriate form design in terms of CRF content and CRF layout. CDM process begins with development of CRF.A case report form (CRF) that is a printed or electronic form used in a trial to record information about the participant as identified by the study protocol record data in a manner that is both efficient and accurate. Data is recorded in a manner that is suitable for processing, analysis and reporting. It must be designed to capture all the data required as per the study protocol and to collect data elements in standardized format. Data elements should be captured on the CRF in a fashion that ensures that data is suitable for summarization and analysis. Data elements planned to be transcribed to the CRF from source documents should be organized and formatted on the CRF to reduce the possibility of transcription error and to facilitate subsequent comparison to source documents [5]. Ease of completion for the investigator and study coordinator is the key to accurate and timely CRF completion. Redundant data elements within the CRF should be avoided and unnecessary data should not be collected. Typical approach for the development of CRF is the so-called backwards approach. It starts with assessing the type and format in which information will be presented in the study report. Information is then tracked backwards, and the CRF is designed to capture data in a manner that allows it to be easily converted or directly placed in the designed tables or figures.

In terms of CRF content, the following things should be taken care of:

a) CRF questions (prompts and instructions to be clear and concise)

b) Avoiding open-ended questions.

c) Phrasing questions in the positive in order to avoid confusion.

d) Using appropriate, mutually exclusive responses.

e) Including suitable units of measurement (e.g., ml/min/mmHg)

f) Explicitly identifying data (e.g., first name, middle name, last name versus name)

g) Avoiding referential and redundant data points.

h) Including an identifier for the protocol version.

i) Keeping subject identifiers to a minimum.

While the points during CRF layout is concerned, key data used in the analysis must be prominently placed on the page in a well-ordered, structured, easy to follow CRF format with a consistent style for all the CRFs in the study. CRFs must be designed to follow the data flow from the perspective of the person completing the CRF. Prior to study initiation, the study investigators should be clear with completion guidelines for error free data acquisition annotation is done by naming the variables as per the data coded schemes according to SDTMIG. The sample CRF has been depicted in figure 1.2. When entry is performed from paper or images, companies need to track case reports forms (CRFs) through the data management process to assure that all the data makes it into the storage database.

Database Design

Database needs to include querying and reporting tools. Data needs to be coded into numbers to facilitate statistical analysis. Database allows for adequate storage of study data and for accurate reporting, interpretation analysis. These tools are designed in lieu of compliance with regulatory requirements and are easy to use. To test or validate their need, test data entry screens are used to ensure data is mapped to the correct fields and data field definitions in terms of length and type. It ensures that out-of-range data is flagged and error messages trigger properly. It identifies that primary key fields are assigned correctly without any duplication. It even validates, edit, range and logic checks. It not only ensures the validation of clinical data but also Metadata. Metadata consistently and effectively describes data and reduces the probability of the introduction of errors in the data framework by defining the content and structure of the target data [5]. Meta data is data of data entered (data about an individual who made the entry or change in clinical data). Database systems tend to co-exist alongside one another. First is study management database and other is clinical database. Study management database is linked to personal information, recruitment, data completeness (CRFs) and Clinical database: clinical information (study outcomes).

Figure 1.2: CRF Layout1.

Database Manager Responsibilities

Retrieving and comparing data, analyzing discrepancies, resolving conflicts, and making the appropriate revisions in the clinical or safety database are the four - simple - tasks that make up the SAE reconciliation process. The procedure, however, is more difficult than it first seems because it necessitates properly managing the many activities involved, particularly because all choices and actions must be completely documented for GxP-compliance (Figure 1.3).
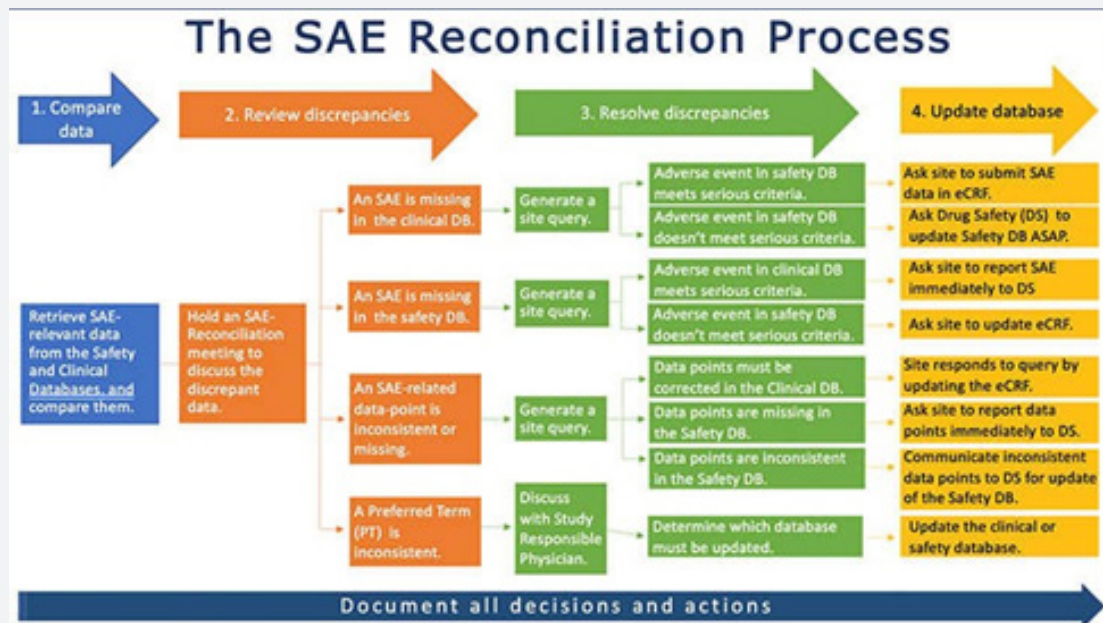

Figure 1.3: SAE Reconcilation process map.

A Data Manager controls the SAE reconciliation process by effectively monitoring clinical database data and comparing defined variables with corresponding records in the safety database. The clinical and safety databases frequently have different verbatim descriptions, coding terms, dates, and other information. Some SAEs may also be missing from one or both databases. Drug Safety works with Data Managers or CRO-responsible Data Managers to find a solution [14].

To debate the disparities, decide whether they are acceptable, or decide on the best course of action, an SAE-Reconciliation meeting is frequently required. In most situations, sites are contacted for clarification, or if a Preferred Term (PT) is inconsistent, the PT is reviewed with the Study Responsible Physician or Safety Surveillance Physician. Clinical monitors may participate in inquiry resolution and corrections and are in constant communication with the investigational sites throughout the SAE reconciliation process. A sponsor specific GxP compliant process is used for the update when the query response contains information that necessitates updating the safety database.

Documentation of all choices and actions is required. The reason for the mismatch must also be noted if any data points do not match exactly. As a study progresses towards the end, the number of SAEs increases, and the reconciliation activity becomes more intense. This is when the use of a GxP-compliant automated tool like Ethical Reconciliation offers competitive advantage by helping to save time while guaranteeing the quality and compliance of the SAE reconciliation process.

Capture

The term data capture refers to the accumulation of clinical data onto a database in a consistent, logical fashion so that it can be retrieved and searched. Data collection is done using CRF that may exist in form of paper or electronic system. But proper handling is done to avoid no loss of quality when an electronic system is in place of a paper system. Validity of the data collection must be ensured. Source data should be identified, and data should be transcribed correctly onto the data collection system. Monitoring of CRF and declaring clean file comes under the process of data collection/transcription and is audited throughout the process. After setup, test or pilot study is done before the actual data collection. Data is electronically or manually captured under standardized operating procedures (SOP). After piloting, the next step is to train all users of the system. A detailed diagram and description of how data will be collected should be provided at the training. There is a need for maintenance of audit trail of data changes made in the system. The paper CRF is filled by investigator as per the guidelines. But for e-CRF, the investigator will be logging into CDM system and entering the data directly at the site. The chances of error and discrepancies are resolved at a faster pace so, most of the pharma companies are relying on e-CRF for reduction in time and human error and enhancement of drug discovery and development process.

Tracking CRF

The entries will be monitored by clinical research associates for completeness and filled up CRF's are retrieved and handed over to CDM team. When entry is performed from paper or images, companies need to track case reports forms (CRFs) through the data management process to assure that all the data makes it into the storage database. The benefits of tracking are that it increases the quality of data by reducing time and assuring that there is no loss of data. It also allows in removing outstanding discrepancies like repeating pages with the same page number, pages with no data, duplicate pages and pages with no page. In case of any missing data, the investigator clarifies, and the issue is resolved.

Data Entry

The design of data entry screens is an important factor in determining the speed at which data can be entered. Data entry procedures should be tested at the earlier design stage, and testing adequately should be documented. Adequate training in these procedures should be provided. Appropriate quality control procedures must be set up [5]. There are three entry systems first is single entry manual entry of data by trained specialist; data entry operators who input data from a paper case record form (CRF) onto a central database [15]. Data entry may occur only once (single entry) or successively (double entry) and second pass entry (entry made by second person) [16].

Cleaning Data

There are certain errors or inconsistencies, or missing data spotted at different time points depending on the study and methods used [5]. Errors should be corrected wherever possible, but no changes should be made without proper justification. Appropriate audit trails should be kept to document changes in the data (queries form, SPSS syntax). Discrepancies that are sent to the sites for resolution are often called queries. The kinds of checks found in an edit check specification can be categorized as follows:

a)   Missing values (e.g., dose is missing) [1].

b)   Simple range checks (e.g., dose must be between 100 and 500 mg).

c)   Logical inconsistencies (e.g.no hospitalization is checked but a date of hospitalization is given). Checks across modules (e.g., adverse event (AE) action taken says study termination, but the study termination pages do not list AE as a cause of termination).

d)   Protocol violations (e.g., time when blood sample is drawn is after study drug taken and should be before, per protocol)

e)   Some of the checking methods are. Automatic check, Manual review, External check (e.g., SAS) (Figure 2).
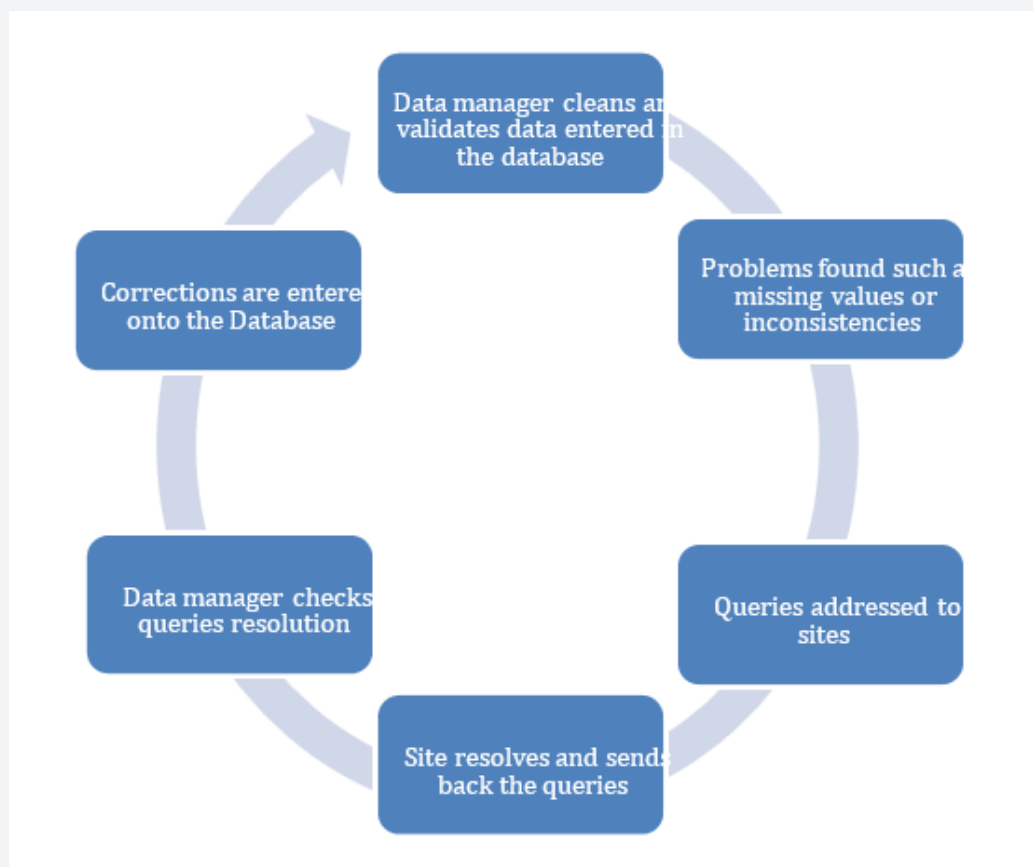
Figure 2: Data cleaning process.

Data Review and Coding

The use of medical coding dictionaries for medical terms data such as advertisements, medical history, medications and treatments/procedures are valuable from the standpoint of minimizing variability in the way data is reported and analyzed. To minimize frequently encountered errors, medical coding properly classifies the medical terminologies. For manual review of CRF, manual coding using common coding dictionaries (icd-10, ICPM, WHO drug dictionary, medical dictionary for regulatory activities "MedDRA") [3]. This activity needs knowledge of medical terminologies, understanding of disease entities, drugs used and basic knowledge of pathological process. Commonly used medical dictionaries available are cost art – coding symbols for thesaurus of adverse reaction terms, ICD9CM - International Classification of Diseases [8] Revision Clinical Modification, MedDRA - Medical Dictionary for Regulatory Activities.15WHO-ART – World Health Organization Adverse Reactions Terminology and WHO-DDE - World Health Organization Drug Dictionary Enhanced [4]. Medical coding helps in classifying reported medical terms on the CRF to standard dictionary terms to achieve consistency and avoid duplication. In case of same adverse event investigator may use different terms at multiple sites, but by coding all the events can be standardized, and uniformity can be maintained in the dataset. Incorrect coding will lead to wrong safety concerns about the drug.

Quality Assurance

All those planned and systematic actions that are established to ensure that the trial is performed and the data is generated, documented (recorded), and reported in compliance with Good Clinical Practice (GCP) and the applicable regulatory requirement(s) are subjected to audit which is a systematic and independent examination of trial related activities and documents to determine whether the evaluated trial related activities were conducted and the data was recorded, analyzed and accurately reported according to the protocol, sponsor's standard operating procedures (SOPs), Good Clinical Practice (GCP), and the applicable regulatory requirement(s) [17].

Database Locking

After the last patient's data has been collected from the sites, the race is on to close and lock the study. Once a study has been locked in, the final analysis can be made, and conclusions can be drawn. Proper closeout activities for these studies are crucial, especially for clinical research presented to regulatory agencies such as the Food and Drug Administration (FDA). The time, energy, and resources employed to collect clinical data is wasted if data cannot be properly verified, validated, transferred, and stored [18]. the term —lock may refer to not only locking a study or database but may also refer to locking specific forms or casebooks.

Statistical Analysis Plan

As per ICH E9, it is a document that contains a more technical and detailed elaboration of the principal features of the analysis described in the protocol and includes detailed procedures for executing the statistical analysis. Statisticians could be described as make-up artists for data. It is a lot easier to dress up a woman as a woman than dressing up a man as a woman. So how much effort do we need to put into a statistical analysis is dependent on the raw data. If the data are appropriate to the trial, clean and complete, then very little effort is needed to make them look good. If the data is not appropriate, incorrect or missing, it will be a lot harder to dress them up nicely and there will be a limit to what can be done. A close examination might reveal that the underlying problems have not been resolved. This can lead to problems when trial data is later submitted to a journal or the authorities for registration. The SAP should not be an afterthought, but an essential part of a clinical trial. It is recommended by the authorities and required by some journals now to ensure that data analysis has been performed in the right way. It includes detailed procedures for executing statistical analysis. The protocol defines how a trial should be performed and if the trial goes according to plan, the SAP will add further detail to the analysis. If there are any changes made to methods or data occurring during the trial, these will be addressed in the SAP and the data analysis will be adapted to match those changes. If the SAP should take into account changes to trial conduct, it cannot be written and finalized prior to the trial.

Therefore, the description of statistical methods is relatively brief in the protocol. The SAP itself should be a document which is written during the trial and adapted as it becomes necessary, e.g., after protocol amendments or when data problems become apparent (imbalanced recruitment). It should always be finalized before the trial ends and before the data collection is final to avoid bias. Where a study is blind, the SAP must be finalized prior to unblinding. Interim analysis poses a problem in the SAP as it means that analysis must be final before the first analysis is performed. This is another reason to avoid interim analysis, if possible, as later changes will need amendments and a lot of administrative work. Usually, we expect the trial to run as the protocol defined

it. However, there are several problems which have an impact on SAP. A common problem is protocol amendments during the trial, which might have an impact on how data is collected or what data is collected. Changes in the treatment can also lead to a change in statistical analysis: e.g., if a treatment arm is discontinued due to safety concerns or inefficacy. A trial that was planned for [4] groups might suddenly be down to [2]. There can be instances where recruitment is not going to be planned, additional sites will be added, patients will be recruited from more active sites, or the patient numbers will be reduced overall. This impacts the statistical analysis as well. It usually started with a short description of the trial, and this links back to the original protocol. A lot of the information is taken straight from this document. It then describes the study rationale and which variables/ data describe the primary outcomes.

Conclusion

With the recent developments in pharmaceutical industries for reducing timelines in drug discovery or drug development processes, maintaining data quality and integrity is an important concern. Clinical data management is the most important aspect for data quality. Various innovative approaches like electronic data capturing and QMS PLUS: an automated standardized tool which can handle complex synchronization process to provide reliable, high-quality data with good clinical practice are currently in trend. The biggest challenge for high quality is the requirement of sufficient resources, expertise and multidiscipline team efforts for multicentric study [19], setups for data standardization and procedures. As old technology is becoming obsolete, proper planning, execution and implementation is required amongst regulators and industrialists. The data should be well integrated and be transferred to clinical data management system (CDMS) that is monitored from the inception to completion, maintained and quantified to ensure a reliable and effective base for operational optimization [20].

References

1. Binny K, Shantala B, Naveen BR, Latha S (2012) Data management in clinical research: An overview. Indian J Pharmacol 44(2): 168-172.

2. Stephen BJ, Frank JF, Kevin P, Leon R (2016) Data management in clinical research: Synthesizing stakeholder perspectives. Journal of Biomedical Informatics 60: 286-293.

3. Nguyen Thi My Huong (2013) Clinical Data Management. CDM.

4. Omollo R, Ochieng M, Mutinda B, Omollo T, Owiti R, et al. (2014) Innovative approaches to clinical data management in resource limited settings using open-source technologies. PLoS Negl Trop Dis 8(9): e3134.

5. Krishnankutty B, Bellary S, Kumar NBR, Moodahadu LS (2012) Data management in clinical research: An overview. Indian J Pharmacol 44(2): 168-172.

6. Walther B, Hossin S, Townend J, Abernethy N, Parker D, et. Al. (2011) Comparison of Electronic Data Capture (EDC) with the Standard Data Capture Method for Clinical Trial Data. PLoS ONE 6(9): e25348.

7. https://issuu.com/cliniversity/docs.

8. LLC, collaborators (2011) OpenClinica version 3.1. Waltham (Massachusetts): OpenClinica.

9. Miralles R, Gicqueau A, Clinovo, Sunnyvale, CA Ale Gicqueau (2010) Improve your Clinical Data Management with Online Query Management System. In: Proceedings of Pharma SUG conference; Orlando, Florida, United States p. 2-18.

10. International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (1996) ICH Harmonized Tripartate Guideline: Guideline for Good Clinical Practice E6(R1).

11. https://ccrod.cancer.gov/confluence/display/CCRCRO/ Clinical+Data+Management

12. Lu Z, Su J (2010) Clinical data management: Current status, challenges, and future directions from industry perspectives. Open Access J Clin Trials 2: 93-105.

13. World Health Organization (2012) Accelerating work to overcome the global impact of neglected tropical diseases: a roadmap for implementation.

14. Ethical GmbH the SAE Recociliation Process: A MULTIFACET challenge.

15. Cummings J, Masten J (1994) Customized dual data entry for computerized data analysis. Quality Assurance 3(3): 300-303.

16. Reynolds-Haertle RA, McBride R (1992) Single vs. double data entry in CAST. Controlled clinical trials 13(6): 487-494.

17. Ottevanger PB, Therasse P, Van de Velde C, Bernier J, Van Krieken H et al. (2003) Quality assurance in clinical trials. Crit Rev Oncol Hematol 47(3): 213-235.

18. Haux R, Knaup P, Leiner F (2007) On educating about medical data management - the other side of the electronic health record. Methods of information in medicin 46(01): 74-79.

19. Gerritsen MG, Sartorius OE, Vd Veen FM, Meester GT (1993) Data management in multicenter clinical trials and the role of a nation-wide computer network. A 5-year evaluation. Proc Annu Symp Comput Appl Med Care 659-662.

20. (2011) Study Data Tabulation Model [Internet] Texas: Clinical Data Interchange Standards Consortium.

Your next submission with Juniper Publishers will reach you the below assets

- Quality Editorial service
- Swift Peer Review
- Reprints availability
- E-prints Service
- Manuscript Podcast for convenient understanding
- Global attainment for your research
- Manuscript accessibility in different formats
  (Pdf, E-pub, Full Text, Audio)
- Unceasing customer service

Track the below URL for one-step submission
https://juniperpublishers.com/online-submission.php