International Journal of Clinical Trials
Matkar S et al. Int J Clin Trials. 2017 Feb;4(1):1-6
http://www.ijclinicaltrials.com
pISSN 2349-3240 | eISSN 2349-3259

DOI: http://dx.doi.org/10.18203/2349-3259.ijct20170302

Review Article

An outline of data management in clinical research

Sunmeet Matkar[1]*, Ajit Gangawane[2]

[1]Senior Executive, Arbor Biotech Pvt Ltd, Mumbai, India
[2]Faculty of Applied Sciences, Parul University, Limda, Waghodia Road, Vadodara, Gujarat-391760, India

Received: 24 October 2016
Revised: 18 November 2016
Accepted: 21 November 2016

***Correspondence:**
Dr. Sunmeet Matkar,
E-mail: matkar.sunmeet@gmail.com

ABSTRACT

Clinical data management (CDM) is an indispensable part of clinical research. CDM activities lead to the collection of reliable, high-quality and statistically sound data generating from the clinical trials. Several studies suggest that such data helps in extreme reduction in time from drug development processes to the marketing stage. Several practices in CDM including CRF annotation, case report form (CRF) designing, data extraction, data entry, data validation, database designing, database locking, discrepancy management and medical coding are evaluated for quality checks at regular intervals during clinical trial. In recent times, the huge gap about improvements of the CDM standards for meeting the regulatory requirements remains to be filled. Fulfilling these requirements will help the clinical trial sector to stay ahead of the game. The current article accentuates the practices followed and activities involved in CDM. Therefore, it enables the reader an outline of management of data in the clinical research.

Keywords: Clinical research, Data management, Case report form

INTRODUCTION

Clinical studies are planned to augment the current medical knowledge inter-related with the treatment, diagnosis, and prevention of diseases or conditions. Clinical research involves experiments, also called clinical trials, on human participants designed to answer pertinent questions about biomedical challenges. An enormous number of clinical trials generate huge quantum of clinical data. Quality of this clinical data is of paramount importance for it determines the reliability of data for making decisions. Quality of clinical data can successfully be achieved by the clinical data management system (CDMS).

It has become impossible to perform clinical research without the utilization of CDMS. The use of CDMS has become essential in clinical trials to manage the increasing amount of clinical data.[1] This clinical data must be collected, processed and analyzed for effective outcome of clinical trials in compliance with the regulatory standards.

The principal objective of CDMS is to make sure on-time submission of high-quality data that is essential for compliance with both the requirements of good clinical practice (GCP) and the statistical analysis and reporting requirements. The data validation activities of CDMS perform a pivotal function in the drug development programme, consisting of numerous people, several data transfers and multiple systems.

In fact, the quality of the data validation process has a direct influence on the quality of data presented as part of new drug application (NDA) submission. High quality data is characterized by maintaining the number of errors and missing data to the lowest possible limit and obtaining maximum data for analysis.[2]

In order to achieve this objective, best practices are implemented to make sure that the data is consistent,

absolute and processed precisely. Many software applications enable monitoring and maintenance of audit history and thereby offer uncomplicated access to data inconsistencies. Advanced technological improvements have empowered CDM to conduct large clinical trials and enrich the data quality even in complex trials.[3]

Main aim of the study was to observe different types of Prescription Audit parameters & evaluate the compliance & non-compliance data of audit according to the checklist as per National Accreditations Board of Hospitals Health (NABH).

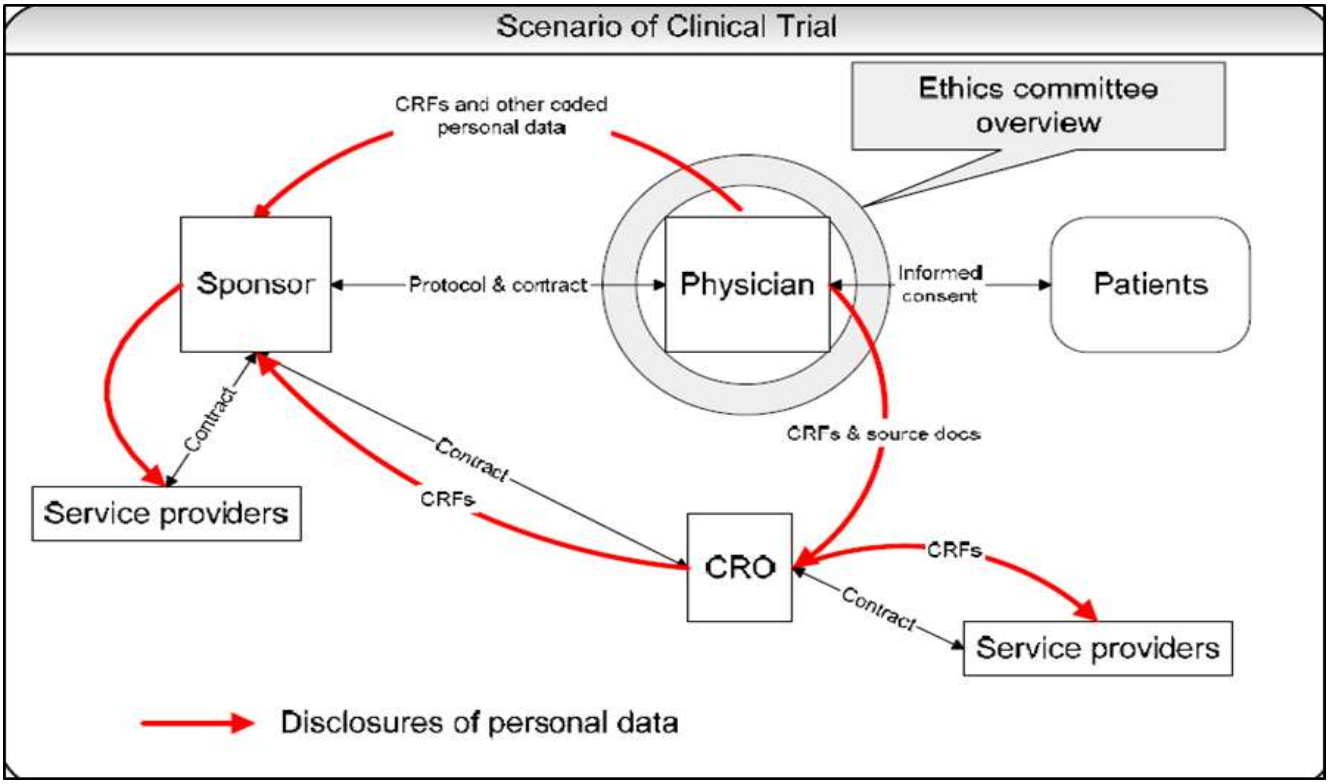

Figure 1: A typical scenario of clinical trails.

FACTORS INFLUENCING THE QUALITY OF DATA

There are several factors that have a serious impact on the overall quality of data collected.

1. ***Case report form (CRF) design***

   CRFs needs to be thoroughly prepared in order to collect complete and accurate data. Both the protocol and the CRF need to be designed in parallel to establish consistency between the two.

2. ***Field monitoring guidelines***

   The quality of field monitoring guidelines has a direct association with the quality of data presented to the CDM.

3. ***Source data verification (SDV)***

   Source data verification is one of the significant phases of the data validation process, without which the integrity and quality of data may get affected.

4. ***Data conventions***

In the multicenter clinical trials, varying date conventions being followed by several investigators may pose difficulties and problems when it comes to entering the data on the database.

MISSING DATA

Missing data is a serious issue having drastically compromised inferences in clinical trials. Missing data are defined as values that are not available and that would be meaningful for analysis if they were observed.[4]

TOOLS FOR CDM

Most of the software tools for CDMS employed in pharmaceutical institutions are commercial, albeit a limited number of open source tools are available. Commonly used CDM tools are RAVE, ORACLE CLINICAL, CLINTRIAL, MACRO and eClinical Suite. RAVE is a cloud-based clinical data management system used to electronically capture, manage and report clinical research data. Like RAVE, ORACLE CLINICAL is also

a data management system designed by Oracle to provide data management, data entry and data validation functions to support clinical trial operations. Similarly, CLINTRIAL is also multilingual, regulatory-compliant clinical data management system. MACRO and eClinical Suite are the other software tools for CDMS.

All the systems stated have the necessary features for monitoring and processing quality clinical trials data. With respect to functionality, there appears to be no significant advantage of one system over the other. Nonetheless, to procure entire benefits of sophisticated systems such as these, a good support network and comprehensive training programmes are essential since their day-to-day use demands a high level of technical competence.[5]

Furthermore, a handful of multinational pharmaceutical giants spend on customized CDMS tools to suit their operational requirements and procedures. The open source tools like TrialDB, openCDMS, PhOSCo and OpenClinica are equally utilized in the clinical trial processes. They can be downloaded from their corresponding websites.

Overall, these software's are of immense help in the regulatory submission studies where they help to maintain an audit trail and in the management of discrepancies of data management activities.

REGULATIONS, GUIDELINES AND STANDARDS IN CDM

Like the other areas of clinical research, CDM also has guidelines and standards whose compliance is a must. The clinical research and pharmaceutical industry heavily depends on the electronically captured data for the assessment of drugs and medicines of clinical use. Therefore there is a stringent requirement for CDM to follow good practices and maintain the standards in electronic data capture. Post capturing, these electronic records need to comply with a code of federal regulations (CFR), Title 21 CFR Part 11. Title 21 CFR Part 11, a constituent of the code of federal regulations is defined as the criteria under which electronic records and electronic signatures (ERES) are considered trustworthy, reliable and equivalent to paper records.

Clinical data interchange standards consortium (CDISC) is global, open, multidisciplinary, non-profit organization that has established standards to support the acquisition, exchange, submission and archival of clinical research data. The two most important standards established by CDISC are study data tabulation model implementation guide for human clinical trials (SDTMIG) and the clinical data acquisition standards harmonization (CDASH) standards.

The SDTMIG standard functions as a guide to the organization and summarizes the details of model and standard terminologies for the data.[6] Whereas, CDASH enlists the basic data information required from a scientific, regulatory clinical and perspective. It also outlines the standards for the collection of data in a clinical trial.[7]

DATABASE DESIGNING

Database designing and design of CRF show positive correlation. The eCRF facilitates the entry of data into an underlying relational database. For a clinical trial employing a paper CRF, the relational database is built separately. In both cases, the relational database allows entry of all data captured on the CRF. Computer system validation involves that all the computer systems used in the processing and management of clinical trial data must go through sufficient validation testing to make sure that they perform as intended and that results are reproducible.

DATA COLLECTION

Data collection, also called as data acquisition can be accomplished by the use of paper or electronic medical records, interactive voice response systems, local electronic data capture systems or central web based systems. The ICH guidelines on GCP utilize the term 'case report form' or 'CRF'. A CRF is a data collection tool used in clinical trials to support investigators and coordinators in capturing all protocol-required information. A well-designed CRF facilitates data collection and entry, and directly benefits other facets of data management and statistical analysis. An informative and structured CRF simplifies database design and data validation processes as well as manipulation of data during statistical analysis. Quite recently, there is a changing trend to perform clinical studies using an electronic case report form (eCRF), instead of paper CRF. eCRF offer advantages of being faster and efficient, environment friendly and with the best security features.

CRF TRACKING

A CRF tracking system is essential to ensure that all the required CRFs are collected. Also, it becomes useful in determining which monitors or sites are having trouble completing or collecting the CRFs in a timely manner. To add, CRFs can also be tracked for missing pages and illegible data manually to assure that the data are not lost. The entries documented in the CRF will be examined by the clinical research associate (CRA) for completeness. These CRFs are then handed over to the CDM team. The CDM team will track the retrieved CRFs and maintain their record.

DATA ENTRY

The process of data entry is undertaken as per the guidelines prepared along with the data management process (DMP). This is valid only when the paper CRF is

retrieved from the study sites. When an eCRF is in use, data entry is carried out at the investigative site where the clinical trial is conducted. On the other hand, when using a paper CRF the pages are entered by data entry operators. Best practice is for a first pass data entry to be completed followed by a second pass or verification step by an independent operator. Commonly, double data entry is executed wherein the clinical data is entered by two different operators on separate system.[8] The second pass entry (entry made by the second person) aids in the authentication and resolution by locating the transcription errors and discrepancies caused by illegible data. Also, double data entry helps in preparation of a cleaner database as compared to a single data entry. Studies carried out by Reynolds note that the double data entry ensures better consistency with paper CRF as denoted by a lesser error rate.[9]

DATA VALIDATION

Data validation helps to develop clinical trial databases that are comprehensive and superior in quality to meet the study's objectives and comply with regulatory standards. One of the common procedures adopted to validate a clinical trial database is, batch validation. In batch validation, one executes a series of checks that are developed to validate the clinical trial database. Discrepancy in the data is one of the major issues in data validation. It is defined as a data point which does not pass a validation check. This may be because of range checks, deviations from the protocol inconsistent and missing data. Edit check programs are embedded in the database to identify the discrepancies in the entered data, thus ensuring validity of the data. In e-CRF based clinical research studies, data validation process will be running on frequent mode to identify discrepancies in the data.

DISCREPANCY MANAGEMENT

Discrepancy management is a process of cleaning subject data in the clinical data management system (CDMS). It includes manual checks and programmed checks. Trivial discrepancies are closed as per self-evident correction method or internal rulings. Some discrepancies that require response from the study site are queried by raising data clarification forms (DCF). In some cases, it also includes reviewing the discrepancies, scrutinizing the reason, and concluding them along with the documentary evidence or stating them as unworkable.

Majority of the CDMS contain a discrepancy database including all the discrepancies that will be recorded and stored with audit trail.

CDM team revaluates every discrepancy in database at the regular intervals to ascertain that they have been resolved. After revaluation, the resolved data discrepancies are recorded as 'closed'. However, closure of the resolved discrepancies is seldom possible. Like for some cases, the investigator in clinical studies may not be able to resolve certain discrepancies. Such discrepancies will be considered as 'irresolvable' and therefore will be updated in the discrepancy database. Among all the activities, discrepancy management is the most critical activity in the CDM process.

MEDICAL CODING

In multi-centric clinical trials, several investigator or medically qualified experts are from different sites/centers and therefore recording the medical term(s) in a uniform manner is a big challenge. Therefore, medical coders from CDM team process these terms and perform medical coding.

Medical coding demands basic knowledge of the pathological processes involved in the disease, different types of drugs used, sound knowledge of medical terminology and understanding of disease entities. More so, it also requires knowledge about the structure of electronic medical dictionaries and the hierarchy of classifications available in them.

Medical coding is performed to categorize the medical terms reported appropriately so that they can be analyzed/reviewed. For the classification of events, medical dictionaries available online are used. There are several standardized medical coding dictionaries; however the following five dictionaries are frequently used for medical coding: COSTART- coding symbols for thesaurus of adverse reaction terms, MedDRA- medical dictionary for regulatory activities, WHO-DDE- world health organization drug dictionary enhanced, ICD9CM- International classification of diseases 9 revision clinical modification and WHO-ART- World health organization adverse reactions terminology.[10]

The available medical dictionaries assist in coding the adverse events that may occur during the study, prior to and concomitantly administered medications and pre-or co-existing illnesses. Medical coding helps to achieve the data consistency and avoid unnecessary duplication.

DATABASE LOCKING

Database locking is done to prevent further changes to a clinical trial database. Database locking is done after review, query resolution, and a determination has been made that the database is ready for analysis. All the data management activities must be concluded preceding the database lock. To ensure this condition, a pre-lock checklist is prepared and successful accomplishment of all the activities is confirmed. After obtaining the consent from all the stakeholders, database is locked and clean data can be extracted for statistical analysis. Post locking; the database cannot be altered in any way. However, only in case of a critical issue or any crucial operational reasons, privileged clients can access and amend the data even after the database is locked. However, this entails absolute documentation and an audit trail to be

maintained with satisfactory explanation for updating the previously locked database. Database locking is followed by data archival.

DATA ARCHIVING

It is the long-term storage of all essential documents which individually and collectively permit the evaluation of the conduct of a clinical trial and the quality of the data produced. Archiving of clinical trial data must be carried out in compliance with the EU clinical trials directive (2001/20/EC), Volume 10 of Eudralex- the rules governing medicinal products in the European Union, International conference for harmonization-good clinical practice (ICH-GCP) guidelines (CPMP/ICH/135/95) and GCP directive. Essential documents must be archived for sufficient periods to allow for audit and inspection by regulatory authorities.

Data privacy collection, processing and transfer of personal data, in particular individuals' patient records, in compliance with applicable laws are vital to the success of clinical studies. According to Article 3(2)c of the EC clinical trials directive3, a trial may be undertaken only if "the rights of the subject to physical and mental integrity, to privacy and to the protection of the data concerning him in accordance with directive 95/46/EC are safeguarded.[11]

DATA QUALITY CONTROL AND QUALITY ASSURANCE

Both the ICH GCP and EU GCP guidelines state that the "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. Audits help to ensure that the CDM process operate effectively and conducted to GCP.[12]

DATABASE LOCKING

Database locking is done to prevent further changes to a clinical trial database. Database locking is done after review, query resolution, and a determination has been made that the database is ready for analysis. All the data management activities must be concluded preceding the database lock. To ensure this condition, a pre-lock checklist is prepared and successful accomplishment of all the activities is confirmed. After obtaining the consent from all the stakeholders, database is locked and clean data can be extracted for statistical analysis. Post locking; the database cannot be altered in any way. However, only in case of a critical issue or any crucial operational reasons, privileged clients can access and amend the data even after the database is locked. However, this entails absolute documentation and an audit trail to be maintained with satisfactory explanation for updating the previously locked database. Database locking is followed by data archival.

CONCLUSION

CDM is a multidisciplinary endeavor that involves the handling of data or information. CDM has developed in response to the ever-increasing exigencies from the pharmaceutical companies to expedite the drug development process. It also suffices the demands of the regulatory authorities to put the quality systems in order to ensure generation of high-quality data for precise drug evaluation. Also, there is a gradual up gradation from the conventional paper-based data management to the electronic systems of data management. Technological advancements have had profound influence on the outcome of CDM activities thereby generating rapid and high-quality of clinical data. Regulatory dominion poses major challenge in terms of the standardization of data management process across clinical research organizations. Another issue from regulatory aspect includes the development of regulations to define the procedures to be followed and the data standards. Clinical trials industry also face the mounting pressure of planning, monitoring and implementing data management systems in a vibrant clinical trials research arena where the swift strides towards state-of-the-art technology makes the existing infrastructure redundant. Regardless of the challenges, data management in clinical research is rapidly developing into a standardized system.

Funding: No funding sources
Conflict of interest: None declared
Ethical approval: Not required

REFERENCES

1. Kuchinke W, Ohmann C, Yang Q, Salas N, Lauritsen J, Gueyffier F, et al. Heterogeneity prevails: The State of Clinical Trial Data Management in Europe - results of a survey of ECRIN. Trials. 2010;11:79.
2. Gerritsen MG, Sartorius OE, vd Veen FM, Meester GT. Data management in multi-center clinical trials and the role of a nation-wide computer network. A 5 year evaluation. Proc Annu Symp Comput Appl Med Care.1993:659-62.
3. Lu Z, Su J. Clinical data management: Current status, challenges, and future directions from industry perspectives. Open Access J Clin Trials. 2010;2:93-105.
4. Roderick J, D'Agostino R, Cohen ML, Dickersin K, Emerson SS, John T. The Prevention and Treatment of Missing Data in Clinical Trials. New Eng J Med. 2012;367:1355-60.
5. Tai BC, Seldrup J. A review of software for data management, design and analysis of clinical trials. Ann Acad Med Singapore. 2000;29(5):576-81.
6. Study Data Tabulation Model. Texas: Clinical Data Interchange Standards Consortium.; c2011.

Available at http://www.cdisc.org/sdtm. Accessed on 1 October 2016.
7. CDASH. Texas: Clinical Data Interchange Standards Consortium. c2011. Available at http://www.cdisc.org/cdash. Accessed on 8 October 2016.
8. Cummings J, Masten J. Customized dual data entry for computerized data analysis. Qual Assur. 1994;3:300-3.
9. Reynolds-Haertle RA, McBride R. Single vs. double data entry in CAST. Control Clin Trials. 1992;13:487-94.
10. Babre D. Medical Coding in Clinical Trials. Perspect Clin Res. 2010;1(1):29–32.
11. Elizabeeth Ness, Clinical data management Centre for Cnacer Research National Cancer Institute. Available at https://ccrod.cancer.gov/confluence/display/CCRCRO/CCR+Orientation. Accessed on 15 October 2016.
12. Rondel RK, Varley SA, Webbs CF. Clinical Data Management. 2nd edition. Wiley Publications; 1993.

Cite this article as: Matkar S, Gangawane A. An outline of data management in clinical research. Int J Clin Trials 2017;4(1):1-6.