

CLINICAL DATA MANAGEMENT IMPORTANCE IN CLINICAL RESEARCH

DEEPA MURUGESAN^{1*}, RANGANATH BANERJEE², GOPAL RAMESH KUMAR²¹Sri Koraka Technologies (P) Ltd., Coimbatore, Tamil Nadu, India. ²AU-KBC Research Centre, Anna University Chennai, MIT Campus, Chennai, Tamil Nadu, India. Email: deepabioinfo75@gmail.com

Received: 14 May 2016, Revised and Accepted: 25 May 2016

ABSTRACT

Over the last few decades, most of the pharmaceutical companies and research sponsors are facing a lot of challenges in clinical research for their new drug approval. The sponsor research needs a high-quality data report for getting new drug approval from Food and Drug Administration for their medical products. Clinical trial data are important for the drug and medical device development processing pharmaceutical companies to examine and evaluate the efficacy and safety of the new medical product in human volunteers. The results of the clinical trial studies generate the most valuable data and in recent years; there has been massive development in the field of clinical trials. A good clinical data management system reduces the duration of the study and cost of drug development. Further a well-designed case report form (CRF) assists data collection and makes data management and statistical analysis. Nowadays, the electronic data capture (EDC) is very beneficial in data collection. EDC helps to speed up the clinical trial process and reduces the duration, errors and makes the work easy in the data management system. This article highlights the importance of data management processes involved in the clinical trial and provides an overview of the clinical trial data management tools. The study concluded that data management tools play a key role in the clinical trial and well-designed CRFs reduce the errors and save the time of the clinical trials and facilitate the drug discovery and development.

Keywords: Pharmaceutical, Clinical trial, Clinical data management, Data capture.

© 2016 The Authors. Published by Innovare Academic Sciences Pvt Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>) DOI: <http://dx.doi.org/10.22159/ajpcr.2016.v9s2.13940>

INTRODUCTION

Drug discovery is a too lengthy, expensive, and most difficult process. Drug discovery requires identification of drug compound, screening, and assays for therapeutic safety and efficacy. Clinical trials are research studies that explore whether a medical drug treatment or device is safe and effective for humans. Before the clinical trial, the sponsor research may include pharmacodynamics, pharmacokinetics, absorption, distribution, metabolism and excretion studies, and toxicity testing of the test article in animals (pre-clinical studies). During pre-clinical studies, *in vitro* and *in vivo* testing is performed and based on the result of pre-clinical studies; the sponsor research starts their clinical trials in human volunteers. Hence, the studies follow strict scientific standards and these standards protect patients and help produce reliable study results [1].

PHASES IN CLINICAL TRIAL

The main aim of the clinical trial is to investigate whether the investigational new drug (IND) has efficacy and safety in human subjects and bring out IND to the market. More than thousands of molecules screened after the clinical trial and one or two IND reaches to the market after the trials. A clinical trial has four main phases: Phase I, Phase II, Phase III, and Phase IV; the main purpose of the study is to investigate the efficacy and safety (Fig. 1). Phase 0 or micro-dosing studies are also included in the clinical trial recently to minimize the cost and time duration. These studies can be a very effective phase to determine very early in the drug development process whether the IND has potential expected biologic effect or not. Phase 0 trials serve as a good tool for clinical researchers in testing the safety and efficacy of drugs at micro level before the onset of Phase I trial [2] using a small number of participants. The clinical trial phases, purpose and length of the study are given in Table 1.

In Phase I trial, 20-100 volunteers will be involved in the study. The primary goal of this study is to investigate safety and dosage of the drug. 70% of the drugs will move to Phase II trials [3]. The maximum tolerated dose and side effects, tolerability, pharmacokinetics, and pharmacodynamics are evaluated in this phase [4] and the investigator

find the dose that works best without causing a severe side effect. The main aim of the Phase II study test the efficacy and side effects of IND by recruiting 100-300 people with the disease/condition are participated. This second phase of testing can last from several months to 2 years. Most Phase II studies are randomized clinical trial and only 33% of the drugs move to the Phase III trial [3].

Phase III trials are a very important phase and also long study phase because the study purpose of this Phase III is to evaluate the efficacy and monitoring the adverse event in the patients. It is a pre-marketing phase of clinical trials and the main goal is to find out the new drug is better than standard drug. In Phase III, 300-3000 patients are involved in the study. Phase II and III clinical trials are usually randomized, where one group of patients receives the experimental drug, while a second "control" group receives a standard treatment drug or placebo. The investigator does not choose which person gets the new drug or the current standard treatment drug. Randomization ensures that each patient has an equal chance of receiving any of the treatments under study, generate comparable intervention groups, which are alike in all the important aspects except for the intervention each group receives. It also provides a basis for the statistical methods used in analyzing the data [5]. Phase III study is more expensive and time-consuming trials. 25-30% of the drug moved to the Phase IV study. In the Phase IV study is a post-marketing surveillance study and this studies several thousands of volunteers who have the disease/condition. The focus of the trials is on how drugs work in the real world [3,6].

DATA MANAGEMENT IN CLINICAL TRIALS

Clinical data management (CDM) is a process of handling the data during clinical trials. Clinical data are one of most valuable assets for sponsor research and pharmaceutical product manufacturers. CDM is a critical phase in clinical research, which leads to generation of high-quality, reliable, and statistically sound data from the clinical trials [7]. The clinical trial can be conducted in one or multi-center research trial, and during the clinical trial, the investigator collects the data in the patients. Clinical research has become impossible without the use of CDMS to handle the increasing amount of data that must be collected, processed

and analyzed for clinical trials [8]. Traditionally, the data are recorded on paper case report forms (pCRFs), and the data management process is tedious with more data errors in pCRFs. Research sponsors and clinical research organizations (CROs) are transitioning from paper-based data collection to electronic data capture (EDC) system [9]. Data entry and data assessment process are moving toward with greater automation software solutions which increase the reliability of clinical trial research. EDC systems are the method of choice for collecting data in single/multisite clinical and non-clinical studies [10].

EDC studies are popular and more because it saves the time with real-time access to data and less time spent on query management and user-friendly navigation. EDC process may bring from 49% to 62% of savings when compared to paper data capture (PDC) [11]. EDC is reducing the investigator's time to work with the subjects in patient care. The use of EDC and modified site management practices were associated with significant reductions in total trial costs [12]. CDM

Table 1: Clinical trial research phases and its purposes

Study	Patients	Purpose	Length of study
Phase 0	<20	Clue on safety and efficacy	Few months
Phase I	20-100	Safety and dosage	Several months
Phase II	100-300	Efficacy and side effects	Several months to 2 year
Phase III	300-3000	Efficacy and monitoring the adverse event	1-4 years
Phase IV	Several thousand	Safety and efficacy	Post-marketing

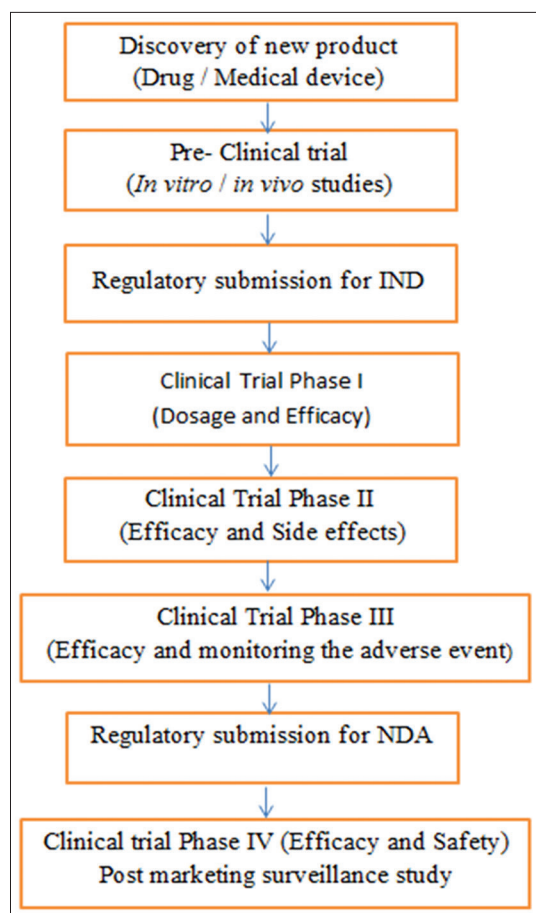


Fig. 1: Drug development process

includes the CRF designing and annotation, data entry, data validation, medical coding, database locking, data analysis, and clinical report generation (Fig. 2).

CRF DESIGN

CRF is a printed, optical, or electronic document designed to record all of the protocol-required information to be reported to the sponsor on each trial [13]. CRF design is the initial step in translating the protocol into standard questionnaires and is paramount to a successful clinical trial [14]. For a well-designed clinical trial requires a protocol and the main purpose of the CRF is to capture all protocol required information. The study protocol can be viewed as a written agreement between the investigator, the participant, and the scientific community. The contents provide the background, specify the objectives, and describe the design and organization of the trial. Every detail explaining how the trial is carried out does not need to be included, provided that a comprehensive manual of procedures contains such information [15]. A well-designed CRF should be capture the data in only one place; it reduces errors in data query management and increases the efficiency of data analysis and report generation.

Clinical data interchange standard consortium (CDISC) is incorporates some standards for clinical data acquisition standard harmonization (CDASH), study data tabulation model, analysis data model and laboratory data model, standard for exchange of non-clinical data, study design model (SDM-XML), operational data model (ODM-XML) and pharmacogenomics and genetics, etc. CDISC standards are vendor-neutral, platform-independent and freely available via the CDISC [16].

Most of the pharmaceutical companies and sponsor research used the commercial CDM, due to their high-security data. Some pharmaceutical companies and sponsor research prefer custom-made CDM. A few tools are available such as Oracle Clinical, Clintrial, Macro, InForm, Rave, DATATRAK, eTrials, and openClinica. These software tools are expensive and need sophisticated information technology infrastructure to function. The Oracle Clinical provides both EDC and PDC. Clintrial provides only PDC. The Macro, InForm, Rave, DATATRAK, eTrials, and openClinica have EDC system. Among them, openClinica is the only open source tool.

DATA ENTRY

Data entry is the process of entering or transferring the data from CRF. There are two types of data entry: single data entry and double data entry. In earlier days, the data entry is in double data entry due to the usage of pCRF. Nowadays, electronic CRFs are increasingly chosen by investigators and sponsors of clinical research instead of the traditional pCRFs [17] due to the data entry process is easy and less duration for data entry. Most of the answers given in choice basis, which will very useful for the investigators, make their work simple and comfortable.

Data entry is carried out at the investigative site where the clinical trial is conducted. The data are entered by either investigator or data entry operator. When using, a pCRF the pages are entered by the investigator. The scanned copy of the pCRF sent to CDM team, the data transferred from pCRF to EDC. Best practice is for transferring pCRF into EDC is double pass entry, the first pass data entry to be completed followed by a second pass entry. This second pass entry is for verification whether the first pass entered data are entered correctly or not. Double pass entry method verification is to reduce data entry error but it is time-consuming process. EDC is web based data entry process. Most of the CRO's prefer EDC due to its fast data entry, rapid data cleaning with edit checks and errors also less. Queries are fixed in real time during data entry. The data are three types, and they are non-time dependent data, time-dependent data and cumulative data. In clinical trials, laboratory results play a key role in evaluating new drugs determining the safety and efficacy of the drug in patients [18]. All samples shipped to the central lab from site for analysis and reporting. Usually, data can be electronically transferred and uploaded into the database. The non-

time dependent data is also called as snapshot data, in which the data are collected at one time. The best example of non-time dependent data is demographic, medical history, exclusion and inclusion criteria CRF data. In the time-dependent data, the data are collected every time either from patients, investigator or from lab time dependent data is vital sign CRF, cumulative data collected over time but not linked to a specific visit example for cumulative data is adverse event CRFs [19].

DATA VALIDATION

The main aim of the CDM is to produce high-quality data. The statistical report is an essential document for getting approval for new drug application (NDA) from Food and Drug Administration (FDA). EDC system provides streamlined communication between monitors, data managers and coordinators for getting high-quality data. Data validation is the process to validate and rectify the errors and missing data that ensures the quality of data. Data validation tools are edit check, double entry, manual check, and SAS program. Most of the CDM edit check tools are auto-generated queries and the ability to manually add queries.

Edit check programs are written to identify the discrepancies in the entered data, which are embedded in the database, to ensure data validity. Edit check programs are initially tested with fake data containing discrepancies. Edit checks are consisting manual and computer checks. Edit checks detect the data entry issues and errors. The main errors are typography errors, copying errors, coding errors, and range errors. The discrepancy may be due to inconsistent data, missing data, range checks, and deviations from the protocol. Data cleaning is the process to fix invalid data and missing data and send queries to the investigator. The main work of the discrepancy management team is to identify the errors and cleaning data. Data cleaning as a three-stage process, involving repeated cycles of screening, diagnosing, and editing of suspected data abnormalities [20]. This edit check ensures the database evaluate and consistent. Data cleaning process will be run frequently for identifying discrepancies during the course of CDM.

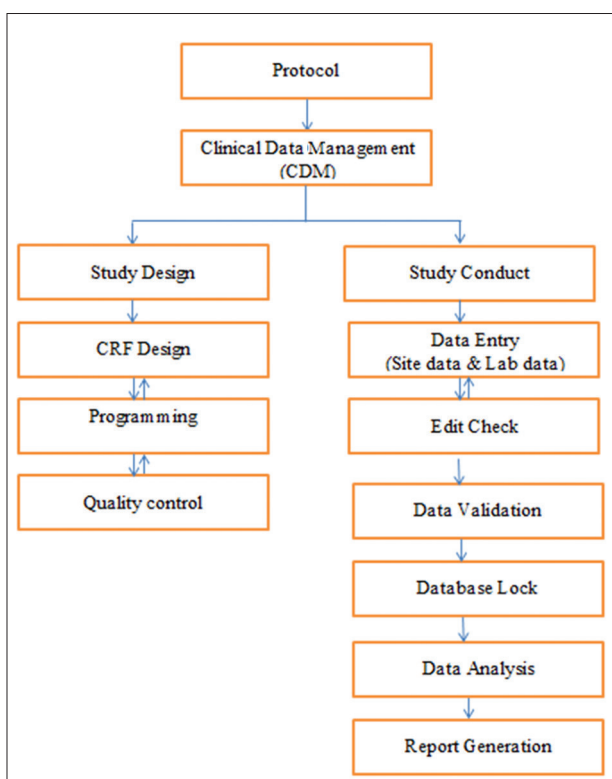


Fig. 2: Process flow of clinical database management in clinical trials

MEDICAL CODING

Medical coding helps in identifying and properly classifying the medical terminologies associated with the clinical trial. The concomitant medications, medical history adverse event, serious adverse event data are generally coded using medical dictionaries. These medical dictionaries are available in online and used in all the phases of clinical trial. There are several standardized medical coding dictionaries are available for coding. Coding symbols for a thesaurus of adverse reaction terms, International Classification of Diseases 9 Revision Clinical Modification, medical dictionary for regulatory activities (MedDRA), World Health Organization (WHO) Adverse Reactions Terminology, WHO Drug Dictionary Enhanced (WHO-DDE) [21].

The medical coding process will take place at the time of query management and data validation. The two most commonly used medical dictionaries in CDM are MedDRA and WHO-DDE used in clinical trial. MedDRA is used for the coding of adverse events, and WHO-DDE is used for coding the medications.

DATABASE LOCKING

Database lock is the process that ensures no manipulation of data during the analysis. After a proper edit check, no discrepancies the database and cleaning data, the database will be locked. Database lock has two types, Interim lock and final lock. Interim lock is generally done at any time in between the study, and the final lock is done when all data entry is completed. To ensure this, a pre-lock checklist is used and completion of all activities is confirmed. Database lock is done because data cannot be changed at any time. Once the database is locked no modification, corrections, updates in the database. Data extraction is done from the database after locking. The main goal of the sponsor research and pharmaceutical companies is to get the high-quality reports for NDA approval. The result analysis and report generation done by the statistician, the statistician extracts the needed data for report generation. Statistical analysis is takes place by using statistical software like SAS, SUDAAN, and SPSS.

CONCLUSION

CDM is essential for evaluating one or more interventions aimed at identifying or diagnosing a particular disease or condition. In the drug development process, the value report and data ensures the accurate drug evaluation and full fill the regulatory authorities' expectations for getting NDA approval from FDA. Due to the development of the information technology, data management assessment and evaluation became easy with quality data. Pharmaceutical companies and sponsor research are encouraging EDC tools for high-speed data generation and high-quality reports. Our study concluded that the pharmaceutical companies and sponsor research need a sound and effective CDM. Well-designed CRF offers the opportunity to minimize data processing. EDC system is one of the best tools for clinical trial data collection and data management. It enhances in monitoring and data management, saving time duration, minimal error and query rates, quick assessment, and reliable results.

ACKNOWLEDGMENT

The authors would like to thank AU-KBC Research Centre, Anna University, Chennai, MIT Campus, Chennai, for providing facilities and comments that greatly improved the manuscript.

REFERENCES

1. Available from: <https://www.nlm.nih.gov/studies/clinicaltrials>. [Last accessed on 2016 Jun 07].
2. Vijayaraghavan R, Kumar GR. Impact of phase zero trials (micro-dosing) in clinical trial research. *Int J Appl Biol PharmTech* 2010;2:486-90.
3. Available from: http://www.fda.gov/ForPatients/Approvals/Drugs/ucm405622.htm#Clinical_Research_Phase_Studies. [Last accessed on 2016 Jun 09].
4. Pocock SJ. *Clinical Trials: A Practical Approach*. Chichester: John Wiley & Sons; 2013.

5. Suresh K. An overview of randomization techniques: An unbiased assessment of outcome in clinical research. *J Hum Reprod Sci* 2011;4(1):8-11.
6. Mahan VL. Clinical trial phases. *Int J Clin Med* 2014;5(21):1374.
7. Krishnankutty B, Bellary S, Kumar NB, Moodahadu LS. Data management in clinical research: An overview. *Indian J Pharmacol* 2012;44(2):168-72.
8. 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 centres. *Trials* 2010;11(1):79.
9. Nahm ML, Pieper CF, Cunningham MM. Quantifying data quality for clinical trials using electronic data capture. *PLoS One* 2008;3(8):e3049.
10. Shah J, Rajgor D, Pradhan S, McCreedy M, Zaveri A, Pietrobon R. Electronic data capture for registries and clinical trials in orthopaedic surgery: Open source versus commercial systems. *Clin Orthop Relat Res* 2010;468(10):2664-71.
11. Pavlovic I, Kern T, Miklavcic D. Comparison of paper-based and electronic data collection process in clinical trials: Costs simulation study. *Contemp Clin Trials* 2009;30(4):300-16.
12. Eisenstein EL, Collins R, Cracknell BS, Podesta O, Reid ED, Sandercock P, et al. Sensible approaches for reducing clinical trial costs. *Clin Trials* 2008;5(1):75-84.
13. ICH Guidance E6: Good Clinical Practice: Consolidated Guideline. Available from: <http://www.fda.gov/downloads/Drugs/./Guidances/ucm073122.pdf>. [Last accessed on 2016 Jun 05].
14. Bellary S, Krishnankutty B, Latha MS. Basics of case report form designing in clinical research. *Perspect Clin Res* 2014;5:159-66.
15. Friedman LM, Furberg CD, De Mets DL, Reboussin DM, Granger CB. Issues in data analysis. In: *Fundamentals of Clinical Trials*. Switzerland: Springer International Publishing; 201. p. 403-62.
16. Available from: <http://www.cdsc.org/about/mission>. [Last accessed on 2016 Jun 12].
17. Le Jeannic A, Quelen C, Alberti C, Durand-Zaleski I; CompaRec Investigators. Comparison of two data collection processes in clinical studies: Electronic and paper case report forms. *BMC Med Res Methodol* 2014;14(1):7.
18. Bassion S. Toward a laboratory data interchange standard for clinical trials. *Clin Chem* 2002;48(12):2290-2.
19. Moon KK. Techniques for designing case report forms in clinical trials consideration for efficient data management and statistical analysis. *CRF Design. Scian News* 2006;9:1.
20. Van den Broeck J, Cunningham SA, Eeckels R, Herbst K. Data cleaning: Detecting, diagnosing, and editing data abnormalities. *PLoS Med* 2005;2(10):e267.
21. Babre D. Medical coding in clinical trials. *Perspect Clin Res* 2010;1(1):29.