IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

RECENT ADVANCES IN BIOSTATISTICS: METHODOLOGIES FOR CLINICAL TRIAL DESIGN AND ANALYSIS

1SURAJ UDHAV SANAP, 2SHRUTI SHAILESH WADEKAR, 3SOHAM EKNATH DAHORE, 4VARSHA KASHINATH CHANDAN

1STUDENT, 2STUDENT, 3STUDENT, 4PROFESSOR

1DBATU UNIVERSITY,

2DBATU UNIVERSITY,

3DBATU UNIVERSITY,

4DBATU

ABSTRACT:

Biostatistical analysis has emerged as one of the most crucial applications for the clinical trial systems. The relevance of statistical analysis in the clinical trials cannot be overstated in advancement of the novel pharmaceuticals and therapeutics. Data volumes, including effectiveness and safety summaries, are produced throughout the process of the developing and approving novel medicines. Data standards developed by Clinical Data Interchange Standards Consortium (CDISC) are well-established, widely accepted, and widely utilized by pharmaceutical sector for submissions to the regulatory bodies. Biostatistical programs like PRM, ODM, CDASH, SDTM, ADaM, SAS, and ARDM have come to long way in the recent years, this article highlights those developments. The present study has made an effort to the report on typical difficulties encountered when using biostatistical methods, with every new discovery, there is always a potential pitfall. Even the main endpoint of experiment might be improperly analysed without adequate validation of the model assumptions. Biostatisticians must collaborate closely with the physicians to choose the most suitable statistical methods for data assessment; otherwise, results may be ambiguous or misleading. To guarantee that a clinical trial satisfies the requirements of the regulatory agencies in the drug approvals, future research should concentrate on the strategies that can enhance compliance and validation processes. Clinical trials play a pivotal role in the advancing medical research and improving patient outcomes. Biostatistics methodologies are essential for the designing robust clinical

trials and analysing the resulting data effectively. This review explores recent advancements in the biostatistics methodologies, focusing on the innovative approaches that enhance the efficiency, reliability, and the ethical considerations of the clinical trials.

KEYWORDS: Clinical trials, Biostatistics, SDTM, ADaM, TLFs, CDASH, SAS, USFDA, Drugs.

INTRODUCTION:

Biostatistics plays a crucial role in the designing, analysing and interpreting clinical trials, which are essential for evaluating the efficacy and safety of the medical interventions. Recent advancements in the biostatistics methodologies have revolutionized the way clinical trials are conducted, leading to more efficient, informative, and the ethical studies. This review article explores some of significant advances in the biostatistics methodologies for the clinical trial design and analysis in the recent years. Clinical trial design and analysis are complex processes that require careful planning and statistical expertise to ensure the validity and reliability of the results. Recent advances in the biostatistics have introduced novel methodologies that address various challenges encountered in the clinical trial research. This review highlights key developments in biostatistics methodologies, including adaptive trial designs, Bayesian approaches and methods for handling missing data and the multiplicity issues.

Clinical trials are designed to collect data on how well a potential treatment or diagnostic tool performs in the real-world situations. Phase 0 (micro-dosing studies), 1, 2, 3, and 4 are all parts of clinical trial process. Phases 0 and 2 are known as exploratory trial phase, phase 1 as non-therapeutic phase, phase 3 as therapeutic confirmatory phase, and phase 4 as post-approval or post-marketing monitoring phase. Phase 0, often known as micro-dosing, is performed in the human volunteers to learn about dosage tolerance (pharmacokinetics) before drug is given to the healthy people in a phase 1 study.

Integration of the patient diaries, medical coding, surveillance, adverse event reporting, handling of suppliers, laboratory data, external interfaces, and the randomization are all essential components of clinical trial management systems (CDMS). In CDMS, one will decide when research will begin and end, what it will focus on, how participants will be selected and removed, and how designs will be commented on and managed. The importance of the data analysis inside clinical trial systems has grown in the recent years. Errors caused by duplicate data input are reduced by storing clinical trial site data in CDMS in the form of case record forms. Premarket approval applications verify operations from the laboratory to animal to human clinical investigations for the safety and efficacy.

The significance of the statistics in clinical research cannot be overstated in era of data-driven clinical healthcare choices. In order to depict the community's health dynamics, common man is inundated with figures, charts, and evaluations (all statistical data). Modern statistical data convey an observed fact via the combination of text, pictures, voice (acoustics), or any one of these alone. Data originating under uncertainties due to the biological variability, sample variations, changes in the ambient circumstances, etc. are common place in scientific study, especially in the biomedical research. These variances have an influence that it must be managed effectively. Sample size, sampling strategy, and the analytical techniques

all affect study results since they are not based on a census of whole cohort. Clinical trial data may be better managed and analysed with use of the statistics. The relevance of biostatistics in the clinical trials

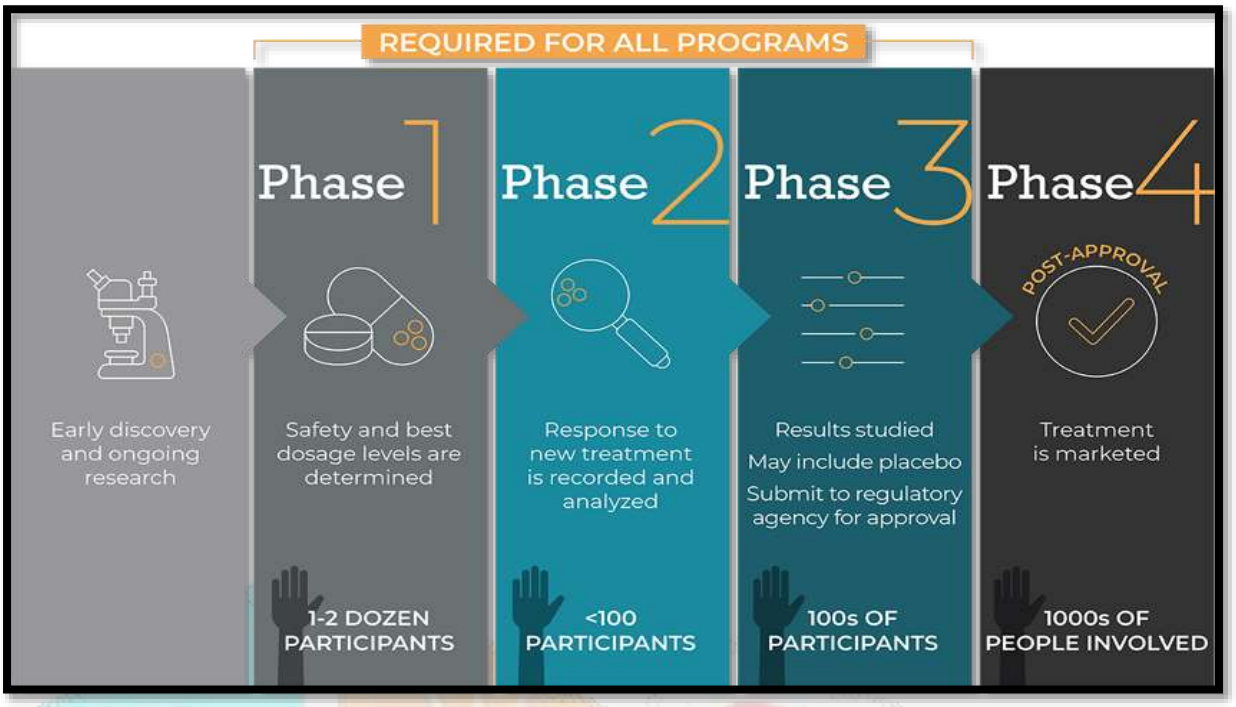

Fig no - 1

cannot be overstated in advancement of the novel pharmaceuticals and therapeutics. To ensure that new treatments are both safe and effective prior to research

The importance of the statistical data in clinical research needs no emphasis in these days of the data-driven clinical health-care decisions. The common man is flooded with numbers, charts and the reviews (all statistical data) to describe the health dynamics of community. Statistical data, as seen today, are not mere numbers but it includes text, images and voice (acoustics) or a mixture of all, which reflects an observed fact. In scientific research, particularly biomedical research, one has to deal with the data arising under uncertainties attributable to the biological variability, sampling fluctuations, changes in environmental conditions and so on. The effect of these variations has to be properly handled in the drawing conclusions. Since most research is based on the samples rather than a census of cohort, the sample size, method of drawing sample and analytical methods used for analysis will impact the findings. Recognising the importance of need for competence in the biostatistics and research methodology, the Medical Council of India has made it mandatory for the postgraduate medical student to learn research methodology by taking up an online course; and has also made it necessary for faculty in the medical colleges to complete the online basic course in biomedical research to get promoted. The results from a sample study serve as estimates for target group but prone to carry errors which can be evaluated and controlled. The researcher has to report all sources of variability (confounders) and account for them in summarising the findings or making the comparisons. It is the science of the statistics that plays a key role in conduct of a clinical study. Statistics-related errors in the biomedical research have been a concern among the research community for quite some time.

TYPES OF STATISTICAL METHODS IN THE CLINICAL RESEARCH

- THE OPERATIONAL DATA MODEL

The operational data model (ODM) is conceptual model that defines the structure and relationships of the data within an organization's operational environment. It serves as blueprint for organizing and managing the data to support the organization's day-to-day operations. The ODM typically includes entities (such as customers, products, orders), attributes (characteristics of entities), and the relationships (connections between entities).

Key components of an operational data model include following:

1. Entities: These are the objects or the things that are of the interest to the organization, such as customers, products, employees, and the orders.
2. Attributes: These are properties or characteristics of the entities. For example, a customer entity may have attributes such as name, address, and the phone number.
3. Relationships: These define how the entities are related to each other. For instance, a customer may place many orders, establishing one-to-many relationship between the customers and orders.
4. Keys: These are unique identifiers for the entities, such as customer ID or product code, which ensure each entity instance can be uniquely identified within model.
5. Constraints: These define rules or conditions that must be satisfied by data, such as mandatory attributes or referential integrity constraints.

Overall, operational data model provides a structured framework for organizing and understanding the data that drives an organization's day-to-day operations, facilitating effective data management and the decision-making.

- THE PROTOCOL RERESENTATION MODEL

The Protocol Representation Model (PRM) is conceptual framework used in the field of the computer networking to represent and analyse communication protocols. It provides a structured way to describe the components, behaviours, and interactions of protocols in standardized manner. Here are the key aspects of Protocol Representation Model:

1. Components: PRM identifies the various elements that make up protocol, such as messages, data structures, states, events, and the actions. These components are defined in the hierarchical manner to represent the protocol's structure.
2. Behaviours: PRM describes the behaviours exhibited by protocol under different conditions or scenarios. This includes specifying how the protocol responds to the incoming messages, changes in state, timeouts, errors, etc.
3. Interactions: PRM models interact between entities participating in the protocol, including message exchanges, handshakes, acknowledgments, and the error handling mechanisms. It defines the sequence and format of the messages exchanged during communication.

4. State Transitions: PRM represents the transitions between different states of protocol entities in the response to events or actions. State transition diagrams or tables are often used to visualize these transitions and specify a conditions under which they occur.
5. Protocol Hierarchies: PRM supports the modelling of the complex protocols by defining hierarchies of protocols, where higher-level protocols may use lower-level protocols to achieve their functionality. This enables modular design and the analysis of protocols.
6. Formal Specification: PRM facilitates formal specification of the protocols using mathematical notation or formal languages, which allows for rigorous analysis, verification, and validation of the protocol designs.

- THE MODEL FOR DATA COLLECTION (CDASH)

CDASH, which stands for the Clinical Data Acquisition Standards Harmonization, is a set of standards developed by Clinical Data Interchange Standards Consortium (CDISC) for clinical research data collection. It aims to standardize how data is captured in the clinical trials, ensuring consistency and interoperability across different studies and systems. CDASH provides guidelines for structuring data collection forms, known as the Case Report Forms (CRFs), to facilitate the capture of high-quality data that can be easily analysed and shared. By adhering to the CDASH standards, researchers can streamline data collection processes, enhance data quality, and improve efficiency of the clinical research.

The pharmaceutical business now makes significantly more use of the CDASH and TA (Therapeutic area) guidance than does non-commercial sector. The Food and Drug Administration (FDA) in the United States and Pharmaceuticals and Medical Devices Agency of Japan (PMDA) (but not the European Medicines Agency; EMA) have both mandated that data submitted in the pursuit of marketing authorization adopt CDISC's Study Data Tabulation Model (SDTM), standard meant to offer the uniform structure to submission datasets. If original data was gathered using CDASH, which is designed to support and map over to submission standard, then creating SDTM structured data is much simpler. Although the CDASH system is straightforward in the principle, it contains a vast amount of the data that must be explored and understood before it can be put to the good use

The model metadata in the CDASH Model v1.2 lists the standard variables in model and gives a broad structure for creating fields to gather the data on CRFs. Root-naming standards for CDASHIG variables are provided in the CDASH Model v1.2 to ease mapping to the SDTMIG variables. The CDASH Model can work with SDTM v1.7 and has comparable structure. To better correlate with the SDTMIG and SDTM.

- THE MODEL FOR DATA ANALYSIS (ADaM)

ADaM (Analysis Data Model) is a standardized format used in the clinical research for organizing and analysing data from the clinical trials. It provides a consistent structure for datasets, making it easier to analyse and report results. ADaM datasets typically include variables relevant to the statistical analysis, such as demographics, efficacy measures, safety data, and other endpoints defined in study protocol.

Key components of ADaM include:

1. Dataset Structure: ADaM datasets are organized into the specific domains, such as Demographics (DM), Efficacy (EG), and the Safety (SA), each containing predefined variables.
2. Variable Definitions: Variables in ADaM datasets are defined according to the CDISC (Clinical Data Interchange Standards Consortium) standards, ensuring consistency across studies and facilitating data sharing and the integration.
3. Mapping: ADaM datasets are mapped back to original raw data collected during the clinical trial. This mapping ensures traceability and reproducibility of analysis.
4. Analysis-ready Data: ADaM datasets are designed to be analysis-ready, meaning they are formatted and annotated to facilitate the statistical analysis, including efficacy and safety assessments, subgroup analyses, and the other exploratory analyses.
5. Documentation: Comprehensive documentation accompanies ADaM datasets, detailing variable definitions, dataset structure, analysis methods, and any assumptions made during data processing and the analysis.

- THE MODEL FOR TABULATION OF STUDY DATA (CDISC/STDM)

Standardized tabulations of data from clinical trials are defined by CDISC/SDTM. Domains are used to describe each of these tables used. As part of data management process, clinical trial data is often obtained and mapped to SDTM variables. The Food and Drug Administration suggests using most well-known CDISC standard, the Study Data Tabulation Model (SDTM), for submitting data from clinical trials for the regulatory approval. An expansion of SDTM, the systematic Data Task Method Implementation Guide for the Medical Devices (SDTMIG-MD) provides a framework for systematic collection of data related to the device attributes and events. Data collection, administration, analysis, and the reporting may all be simplified with the help of the SDTM's standardized organization and formatting of the data. Due diligence and other critical data review processes are aided by use of the SDTM, and quality of the regulatory review and approval process is enhanced overall. In addition to the clinical research, SDTM has found applications in fields of medical device research, pharmacogenomics, and genetics. Submission of data to FDA (United States) and Pharmaceuticals and Medical Devices Agency, Japan (PMDA) must adhere to the certain criteria, one of which is the SDTM. SDTM mapping to newest edition of the SDTM standards is considered essential for the regulatory submission by any pharmaceutical business or the Clinical Research Organisation (CRO) if data is not in the SDTM standards. Clinical programming relies heavily on analysis of clinical trial data to demonstrate the effectiveness and safety of any Investigational New Drug Application (INDA), and the SDTM can facilitate data gathering, warehousing, mining, and reuse. It is also a preferred standard for European Medicines Agency (EMA) and Health Canada. When it comes to the data reuse, however, SDTM has two key drawbacks: (1) It is complicated and difficult for the people with no expertise to comprehend, and (2) The format does not lend itself well to the standard analytical tools and links with other data resources. To encourage the reuse and incorporation of the trial data with other data types, such as registry data and electronic health records, stakeholders must establish effective and uncompressed

translators to other formats if the SDTM or a similar paradigm is supported for the clinical trial data. Nearly two decades' worth of the data has been gathered using CDISC SDTM Questionnaires domain. A more open and standardized approach to data collecting, tabulation, and analysis is possible with the further development and use of SDTM standard, including the questionnaire domains.

- THE MODEL FOR DATA VALIDATION(SAS)

SAS is widely used in fields of survey information analysis and the data management because of its flexibility, dependability, instructional assistance, and comprehensive documentation. SAS, or Statistical Analysis System, is the gold standard for handling and analysing huge datasets. When it comes to executing statistical tests and getting summary statistics, SAS is a top contender for best software. Data processing, statistical description, inference, and data display are only some of numerous uses for SAS's various functions and processes. Descriptive tables may be generated with the help of SAS functions PROC TABULATE and PROC REPORT. Statistical analysis and data visualization are only two of the many uses for the SAS, a command-driven software tool. Unfortunately, it's only compatible with the Microsoft Windows. In both business world and the academic world, it is one of a most used statistical programs. SAS is utilized in many different industries, including business and the health, and has a large, active online community. Its main selling points are its huge variety of the statistical techniques and algorithms, particularly for advanced statistics, its highly configurable analysis choices and the output options, and its publication-quality graphics with an output delivery system (ODS). By harmonizing the data storage, programs like SAS and R, which produce outputs in the incompatible formats, may be used together for the analysis

Advanced analytics, business intelligence, data management, and predictive analytics are just a few of many functions that SAS can do. SAS software is accessible through a graphical user interface in addition to SAS programming language, also known as Base SAS. The following tasks are made easier using the SAS software:

• Retrieve information stored in wide variety of a media, including SAS tables, Excel tables, and database files.

• Manage and transform data that is already there and gets information that is needed. One may add new columns, merge parts of data collection, and create subsets of data.

• Statistical approaches are as simple as the correlations and logistic regression and as complex as the model selection and Bayesian hierarchical models which may be used to examine the data.

• Share the findings of the research with others by writing a comprehensive report on the findings. HTML, RTF , and PDF are just formats that support custom report exports.

All data, including data from clinical trials is standardized, there will be more of it to analyse in the healthcare business of world. SAS is ideal option since it simplifies both data collection and analysis. SAS is not the only program available, but is among the finest at analysing and manipulating data and data sets.

The present engagement of SAS programmers in the clinical study means that their responsibilities will grow day by day. By providing a wide variety of the analytics capabilities, SAS help pharmaceutical manufacturers increase their output and ensures that the patients get safe and effective medications with other treatments. Better demand forecasting is possible due to SAS's ability to enhance the quality through output and equipment performance. SAS provides GMP-compliant environment for integrated analytics making it an ideal for the pharmaceutical business.

Fig no - 2

- **ADVANTAGES OF BIOSTATISTICAL METHODS IN CLINICAL RESEARCH**

Clinical research studies relies heavily on the statistical methods as they provide a basis for concluding a target population. Clinical researchers should have a firm grasp of statistical methods to properly evaluate the study datasets. Clinical research studies relies on statistical analysis to establish norms for the patient characteristics, healthcare delivery and the outcome. Clinicians in today's evidence based practice environment would do well to familiarise themselves with the basic statistical ideas and procedures in order to better assess and implement clinical research findings. To draw the reliable findings from clinical studies, researchers must have a well-defined process by which they can recruit individuals without bias and use appropriate analytical methods. The effectiveness of the clinical trials depends on the use of data analytics, which is made feasible by data analysis tools backed by the software. Preparing regulatory filings and gaining the market clearance in the clinical arena requires effective data management systems. CDISC data standards represent the essential components of the clinical data lifecycle which can accommodate the demands at each stage of clinical research lifecycle.

- **CONCLUSION**

There is a rich and varied multi-disciplinary atmosphere in all areas of the biostatistical work. When submitting an NDA(New Drug Application) one must include clinical data that follows the guidelines which are established by the Clinical Data. Model for Planning PRM (Protocol Representation Model). The PRM captures information of data instead of text in the protocol and helps to create case report forms (CRFs) and to study outlines ODM (Operational Data Model). The ODM standards play a key role in clinical research, informatics including areas such as data exchange, archival, U.S. Food and Drug Administration (FDA) submissions, and inter-operability with the healthcare data. Data Collection CDASH (Clinical Data

Acquisition Standards Harmonization) CDASH organizes the clinical data into interventions, events, findings, and special purposes. Data Tabulation SDTM (The Study Data Tabulation Model) SDTM provides a standard for organizing and formatting the data to streamline processes in the collection, management, analysis, and reporting of data. Data Analysis ADaM (Analysis Data Model) ADaM datasets provide significant information about individual studies and the derivation decisions made in the each study. Data Validation SAS (Statistical Analysis Software) SAS is the command-driven software for data manipulation, statistical description and inference with data presentation. ARDM facilitates the tracking, searching and retrieving of outputs as it enables the query based searches. CDISC standards have received much attention from scientific community. There is still a high need for the standard implementation assistance in this arena. There are certain inconsistencies that should be fixed in spite of the fact that the vast majority of current CDISC standards and accompanying guidelines are suitable for the use with research data. Therefore this research has made efforts to present many statistical approaches used in FDA's drug approval process. The strengths and weaknesses of the biostatistical software and analysis have also been investigated. The merits and drawbacks of any available statistical study are different from them. To guarantee that a clinical trial satisfies the guidelines statistical methods are very useful.

- **REFERENCES**

1. K V S Sarma, Alladi Mohan, Sai Sarada Vedururu, An overview of Statistical Methods in clinical studies,(2021)
2. Singh AP, Shahapur PR, Vadakedath S, Bharadwaj VG, Kumar P, et al. (2022) Research question, objectives, and endpoints in clinical and oncological research: A comprehensive review. Cureus 14: e29575.
3. Kandi V, Vadakedath S (2021) Clinical research: An overview of study types, designs, and their implications in the public health perspective. AJCMR 9: 36-42.
4. Norman GAV (2020) Update to drugs, devices, and the FDA. JACC Basic Transl Sci 5: 831-839.Marker A (2023) Clinical trial data management and professionals. Smartsheet.
5. Deglise-Hawkinson J, Kaufman DL, Roessler B, Oyen MPV (2020) Access planning and resource coordination for clinical research operations. IISE Trans 52: 832-849.
6. Sharma H (2021) Statistical significance or clinical significance? A researcher's dilemma for appropriate interpretation of research results. Saudi J Anaesth 15: 431- 434.
7. Board of Governors in Supersession of Medical Council of India. Notification. New Delhi, the 11th December, 2019. No. MCI-18(1)/2019-Med./171700.
8. Board of Governors in Supersession of Medical Council Of India. Notification New Delhi, the 12th February, 2020 No. MCI-12(2)/2019-Med. Misc./189334.
9. Lang TA, Altman DG. Basic statistical reporting for articles published in biomedical journals: The "Statistical Analyses and Methods in the Published Literature" or the SAMPL Guidelines. Int J Nurs Stud 2015;52:5-9. 4

10. Indrayan A. Basic statistical reporting for articles published in biomedical journals: The "Statistical Analyses and Methods in the Published Literature" or the SAMPL guidelines. Int J Nurs Stud 2015;52:5-9.
11. Ahmed S, Dhooria A. Pitfalls in statistical analysis– A reviewers' perspective. Indian J Rheumatol 2020;15:39-45.