



REVIEW ARTICLE

Biostatistical Methodologies in Clinical Trials: An Overview of Recent Developments and Pitfalls

Sagar Kumar Patel^{1*}, Rachna Patel² and Sandeep Bolla³

¹LabCorp Drug Development Inc, Somerset, USA

²Catalyst Clinical Research, Wilmington, NC, USA

³Fortrea, North Carolina, USA

*Corresponding author: Sagar Kumar Patel, LabCorp Drug Development Inc, Somerset, USA



Abstract

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

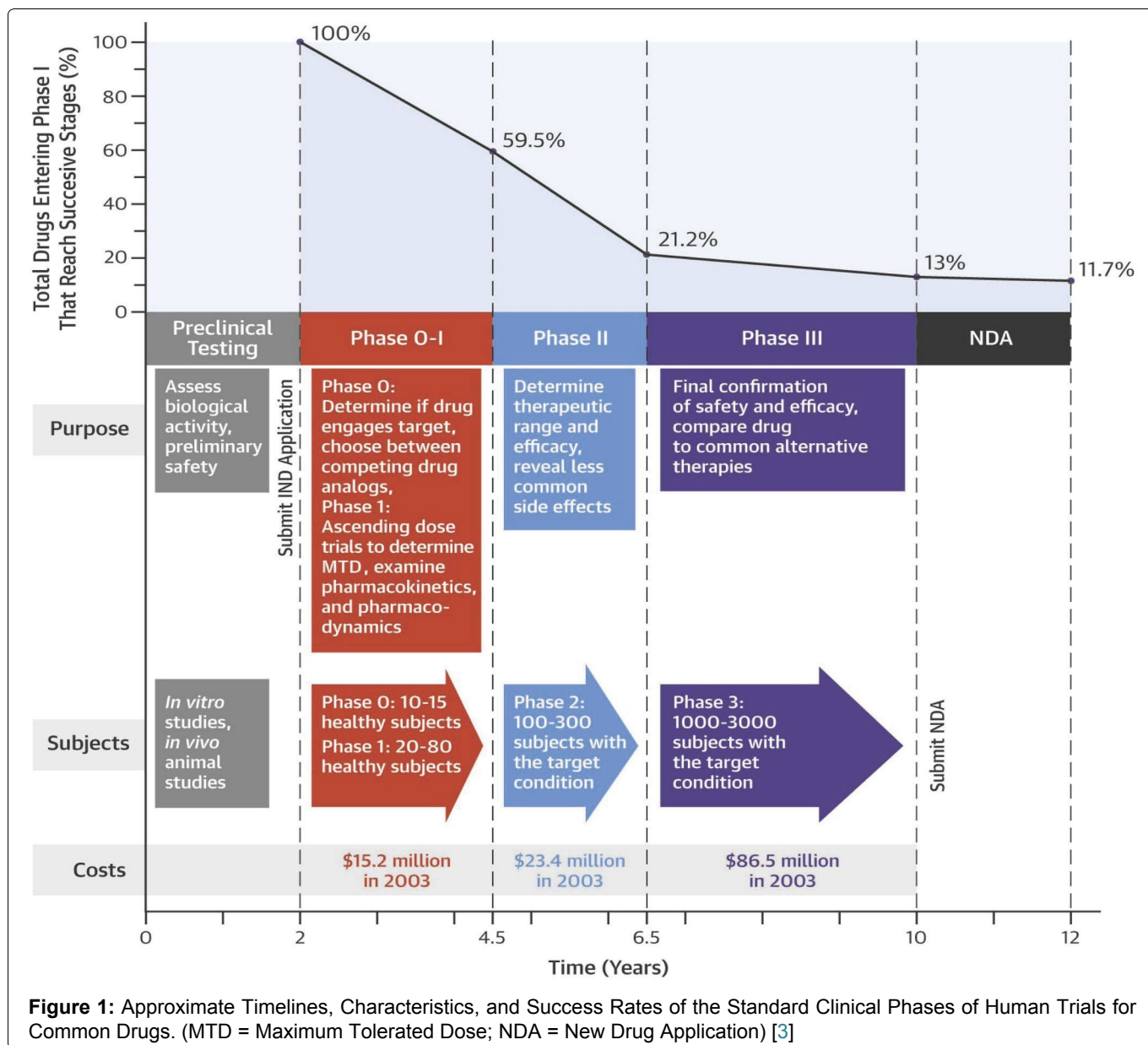
Keywords

USFDA, Clinical trials, Drugs, Biostatistics, SAS, SDTM, ADaM, ARDM, CDISC

Introduction

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

Integration of patient diaries, medical coding, surveillance, adverse event reporting, handling of suppliers, laboratory data, external interfaces, and randomization are all essential components of clinical trial management systems (CDMS). In the CDMS, one will decide when the research will begin and end, what it will focus on, how participants will be selected and removed, and how the design will be commented on and managed [4]. The importance of data analysis inside clinical trial systems has grown in recent years. Errors caused by duplicate data input are reduced by storing clinical trial site data in the CDMS in the form of case record forms. Pre-market approval applications verify operations from laboratory to animal to human clinical investigations for safety and efficacy [5].



The significance of statistics in clinical research cannot be overstated in the era of data-driven clinical healthcare choices. In order to depict the community's health dynamics, the common man is inundated with figures, charts, and evaluations (all statistical data). Modern statistical data convey an observed fact via a combination of text, pictures, voice (acoustics), or any one of these alone. Data originating under uncertainties due to biological variability, sample variations, changes in ambient circumstances, etc. are commonplace in scientific study, especially in biomedical research. These variances have an influence that must be managed effectively. Sample size, sampling strategy, and analytical techniques all affect study results since they are not based on a census of the whole cohort. Clinical trial data may be better managed and analysed with the use of statistics [6].

The relevance of biostatistics in clinical trials cannot be overstated in the advancement of novel pharmaceuticals and therapeutics. To ensure that new treatments are both safe and effective prior to

reaching the general population, clinical trials are essential. Biostatisticians play a crucial role in this field. They have a difficult and lengthy job managing the trial's planning, data processing, and interpretation. The careful preparation of the necessary materials for submission to regulatory authorities is crucial to the overall success of the investigation. The Biostatistician will undertake quality assurance assessments to detect data outliers and verify data correctness after the data management staff has finished collecting all of the data at the conclusion of the trial. Accurately and simply summarising the statistical data is an important part of CSR, and this is where biostatisticians come in. They make sure that the right statistics are used, the research is thorough, and the findings are communicated clearly [7].

Collecting valid, superior, and statistically reliable information from clinical trials is the focus of clinical data management, an integral part of clinical research. As a result, it speeds up the process of bringing new medicines to market. The overflow of data is

a consequence of the increasing complexity of trial data. Hence, adaptive clinical design is preferred by researchers over fixed clinical design because it may evolve in response to individuals' collected data. Clinical Trial Software Applications are databases designed to make the CDM process easier and complete the chores required to conduct several studies. The trials have access to a wealth of information thanks to novel data collection methods (such as wearable devices and sensors), greater data volume and accuracy, advanced quality assurance procedures based on risk assessment, distributed clinical trials, and flexible study layouts. Likewise, new techniques and technology may help clinical data managers carry out their duties and adapt to the evolving field of data management in clinical trials [8,9] (Figure 2).

Types of Statistical Methods in Clinical Research

Research in the biological sciences relies heavily on clinical trials and statistical analysis. Clinical trial statisticians, doctors, investigators, and scientists who

collaborate need a solid understanding of these ideas. By mandating CDISC standards (such as SDTM and ADaM (Analysis Dataset Model), define.xml, and therapeutic area extensions) for submissions to regulatory agencies, the US Food and Drug Administration (FDA) and the Japanese Pharmaceutical and Medical Devices Agency (PMDA) took the crucial first step in harmonizing worldwide standards. Because of the varying requirements across the course of a clinical study's lifespan, CDISC data standards are designed to accommodate the major phases of the clinical data lifecycle [10,11] (Figure 3).

The advancement of clinical medicine and therapies relies on a variety of factors, one of which is the widespread use of data standards. Data standards developed by the Clinical Data Interchange Standards Consortium (CDISC) are well-established, widely accepted, and widely utilized by the pharmaceutical sector for submissions to regulatory bodies. In order to improve the accuracy, productivity, and affordability of clinical research procedures from start to finish, the CDISC Foundational Standards serve as the backbone of

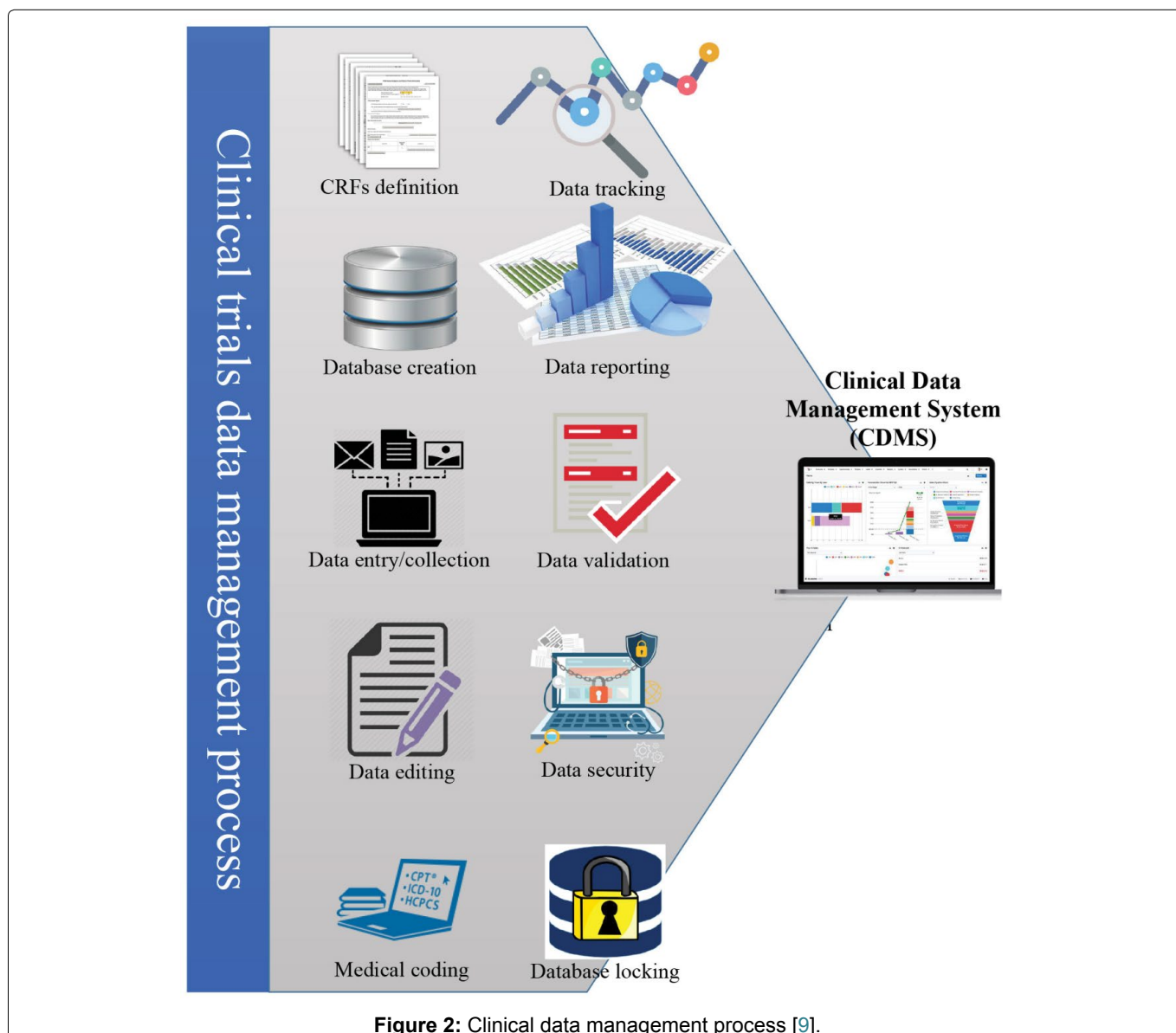


Figure 2: Clinical data management process [9].

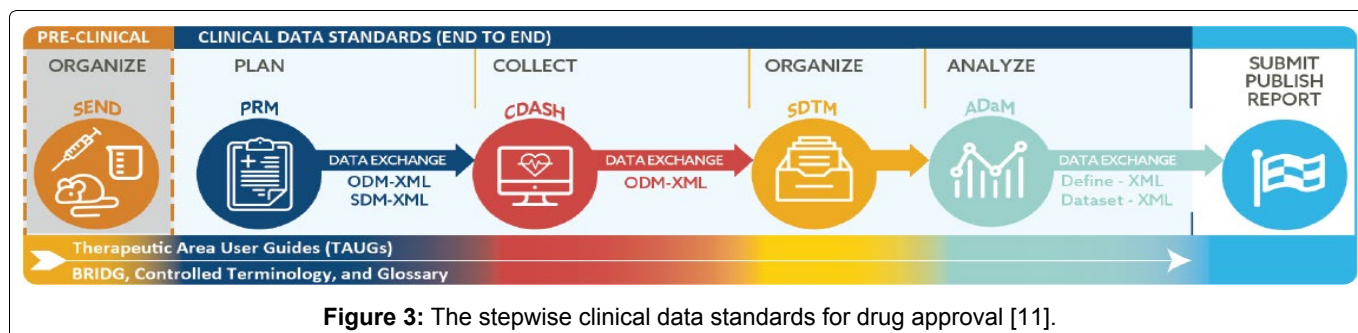


Figure 3: The stepwise clinical data standards for drug approval [11].

a comprehensive suite of data standards. Core principles for developing data standards are the emphasis of Foundational Standards, which also define data models, domains, and specifications. When submitting a new drug application (NDA), one must include clinical data that follows the guidelines established by the Clinical Data Interchange Standards Consortium (CDISC). There is widespread adherence to CDISC's requirements in the pharmaceutical industry [12].

It is believed that the benefits of using CDISC standards for purposes outside of regulatory submission are many, such as advancements in the exchange of data, cross-study analysis, and meta-analysis of data for all clinical researchers, in addition to streamlining the regulatory submission, review, and approval processes. The CDISC standards cover all aspects of the clinical research process, from the standard for exchange of nonclinical data (SEND) to the collection of data using case report forms (CRFs); clinical data acquisition standards harmonization [CDASH]), from the gathering of data and tabulation (study data tabulation model [SDTM]) to the logical model for biomedical research (BRIDG) and the operational data model (ODM) for transport. The biopharmaceutical sector and several academic institutions worldwide extensively employ these standards, which are given in data models, installation instructions, and user manuals [11]. Electronic submissions of research data are now mandated by the US Food and Drug Administration and the Japanese Pharmaceutical and Medical Devices Agency and suggested by Chinese and European agencies where raw data is needed [13]. The following are examples of some of the most frequent statistical approaches used over the course of a clinical study.

Model for Planning

Protocol Representation Model (PRM)

Protocols are the foundational documents in clinical research, outlining the methodology and rationale for the study's execution. The Protocol Representation Model (PRM) stores protocol information as numbers rather than words. Case report forms (CRFs) and research plans may also be produced with its aid. PRM provides a structured description of the most frequently used components of a protocol that can be understood by computers. The model's classes and their connections

are described using UML, or the Unified Modelling Language. Focusing on study parameters including study design, eligibility criteria, and standards from the ClinicalTrials.gov, World Health Organisation (WHO), and EudraCT registries, the Protocol Representation Model (PRM) offers a standard for planning and constructing a research protocol. To facilitate clinical research and data exchange, PRM facilitates the automation of CRF development and EHR setup [11,14].

The Operational Data Model (ODM)

ODM, or the Operational Data Model, is the foundational standard for information interchange in clinical trials. The ODM architecture uses Extensible Markup Language (XML) for interoperability across Electronic Data Capture (EDC), electronic health record (EHR), data collecting, tabulation, analysis, and archiving processes in the healthcare ecosystem. Clinical data contains datasets in Standard Data Tabulation Model (SDTM) format and clinical report formats [15]. To simplify the process of deriving analysis datasets for regulatory submissions, the ODM allowed for the development of standards like the SDTM and the analysis data model (ADaM). CDISC data standards represent the essential phases of the clinical data lifecycle in light of the requirements at every phase of the clinical research lifecycle. While regulatory processes have always revolved around document submission, there has been a growing interest in evaluating the underlying facts [16].

Technical issues, such as the statistical software used, and unrecorded analytical decisions at any point in the pipeline, from input data to output, might make for a particularly challenging study. Even the simplest routine analyses have the potential for inaccuracy and are not always easily reproduced. Overall, it's a drain on both time and energy. Automating the analysis has been presented as a possible fix. In light of this, Brix, et al. [16] created the ODM Data Analysis, a tool for automatically validating, monitoring, and generating descriptive statistics from clinical data saved in the CDISC Operational Data Model format, with the goal of facilitating clinical development.

This program's greatest strength is its simplicity of use. In seconds to under half an hour (depending on file size and hardware), anybody may upload an ODM file

and obtain a fast overview of the included data and its distribution. Because ODM-DA is free and easy to use, statistical questions may be addressed without incurring any further expenses or requiring any specialized knowledge. Doctors reported having a better grasp of the anatomy and contents of their research data after engaging in the analytic procedure. Physicians and research staff both found ODM-DA's ability to identify inaccurate data inputs helpful in improving data quality as a whole. ODM Data Analysis was created so that just generalized statistics on each item would be generated, with no cross-item comparisons. The next major advancement might be the implementation of automated bivariate statistics creation [17].

Model for Data Collection

CDASH

Case report forms (CRFs) are used to document data gathered from clinical studies; these forms are typically created for single research. The suggestions for creating CRFs provided by the Clinical Data Acquisition Standards Harmonisation (CDASH) Model include fields that are standard across all CRFs. It offers a standard approach to gathering data consistently across trials and sponsors, which will increase openness for regulators and those doing data reviews. This guideline guarantees that all data submissions into the Study Data Tabulation Model (SDTM) may be traced back to their original sources. Interventions, events, results, and unique purposes are the four main sections into which CDASH divides up clinical data [14].

The pharmaceutical business now makes significantly more use of CDASH and the TA (Therapeutic area) guidance than does the non-commercial sector. The Food and Drug Administration (FDA) in the United States and the Pharmaceuticals and Medical Devices Agency of Japan (PMDA) (but not the European Medicines Agency; EMA) have both mandated that data submitted in pursuit of a marketing authorization adopt CDISC's Study Data Tabulation Model (SDTM), a standard meant to offer a uniform structure to submission datasets. If the original data was gathered using CDASH, which is designed to support and map over to the submission standard, then creating SDTM structured data is much simpler. Although the CDASH system is straightforward in principle, it contains a vast amount of data that must be explored and understood before it can be put to good use [18].

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

and CDASHIG variable labels have been revised where necessary to reflect the new SDTMIG v3.3 domains (i.e., AG, ML, OE, and RE) [19]. The design of CDASH and SDTM is guided by the underlying philosophy of the respective protocols, which are optimized for particular cases. Clean, final CRF data is represented in SDTM, which is then organized in a standard way that makes data transfer, review, and reuse more manageable. CDASH's user-friendly, EDC/CRF-compatible data collection method optimizes data quality and ensures a seamless data flow into SDTM. There is a lot of overlap between the two standards in terms of data, but they serve distinct objectives and have different underlying philosophies [20].

Model for the Tabulation of Study Data

CDISC/SDTM (Study Data Tabulation Model)

Standardized tabulations of data from clinical trials are defined by the CDISC/SDTM. Domains are used to describe each of these tables. As part of the data management process, clinical trial data is often obtained and mapped to the SDTM variables. The Food and Drug Administration suggests using the most well-known CDISC standard, the Study Data Tabulation Model (SDTM), for submitting data from clinical trials for regulatory approval. An expansion of the SDTM, the systematic Data Task Method Implementation Guide for Medical Devices (SDTMIG-MD) provides a framework for the systematic collection of data related to device attributes and events [21].

Data collection, administration, analysis, and reporting may all be simplified with the help of SDTM's standardized organization and formatting of data. Due diligence and other critical data review processes are aided by the use of SDTM, and the quality of the regulatory review and approval process is enhanced overall. In addition to clinical research, SDTM has found applications in the fields of medical device research, pharmacogenomics, and genetics. Submission of data to the FDA (United States) and Pharmaceuticals and Medical Devices Agency, Japan (PMDA) must adhere to certain criteria, one of which is SDTM [22]. SDTM mapping to the newest edition of SDTM standards is considered essential for regulatory submission by any pharmaceutical business or Clinical Research Organisation (CRO) if the data is not in SDTM standards. Clinical programming relies heavily on the analysis of clinical trial data to demonstrate the effectiveness and safety of any Investigational New Drug Application (INDA), and SDTM can facilitate data gathering, warehousing, mining, and reuse.

Submission of clinical trial data in SDTM is required by the Food and Drug Administration (FDA), the Pharmaceuticals and Medical Devices Agency (PMDA), and the National Medical Products Administration (NMPA). It is also a preferred standard for the European

Medicines Agency (EMA) and Health Canada. When it comes to data reuse, however, SDTM has two key drawbacks: (1) It is complicated and difficult for people with no expertise to comprehend, and (2) The format does not lend itself well to standard analytical tools and links with other data resources. To encourage the reuse and incorporation of 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 SDTM or a similar paradigm is supported for clinical trial data [23]. Nearly two decades' worth of data has been gathered using the 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 the SDTM standard, including the questionnaire domain [24].

Model for Data Analysis

ADaM (Analysis Data Model)

The Clinical Data Interchange Standards Consortium (CDISC) mapping approach was used to convert datasets from separate trials into a common format prior to integration. Data analysis in clinical trials has been standardized since the publication of the Analysis Data Model (ADaM) in 2006. Patients were assigned to either the Safety Analysis Set or the Full Analysis Set in ADaM's analysis sets. Validated ADaM datasets for medical history, effectiveness, and adverse events (AEs) are now available, and their creation gave substantial insight into individual trials and the derivation choices taken in each investigation. ADaM establishes dataset and metadata standards that enable traceability across analysis findings, analysis data, and data contained in the Study Data Tabulation Model (SDTM) and facilitate the rapid development, replication, and evaluation of clinical trial statistical analyses. The FDA (United States) and the PMDA (Japan) both recognize ADaM as one of the mandatory requirements for submitting data for approval [25].

The Analysis Data Model (ADaM) and the Analysis Data Model Implementation Guide (ADaMIG) offer guidelines for the construction of analysis datasets used in clinical trials of human medicinal products. Clinical study reports (CSRs), Integrated Summaries of Safety (ISS), Integrated Summaries of Efficacy (ISE), and other analyses needed for a full regulatory evaluation, should all make use of ADaM datasets to develop and support their respective outcomes. Data in ADaM models may be obtained from SDTM models or be imputed. When sponsors and biostatisticians are interested in how the timing of one event or discovery at a one-time point in the research relates to the timing of a subsequent occurrence or findings throughout an additional point in the study, they might ask that programmers use the ADaM variable choices [26].

ADaM often helps throughout the FDA review process. However, the data it provides is not always organized in a manner that allows for all of the analyses that should be presented for review. For instance, multiple dependent variable analyses or correlation analyses spanning multiple response variables are both outside the scope of ADaM structures. Therefore, after consulting with the relevant review division, sponsors should augment their ADaM datasets as necessary. Conforming analysis datasets to ADaM is intended to facilitate analysis by reducing the complexity of the necessary programming. The information in the SDTM datasets should be used to create the ADaM datasets, as previously mentioned. The ADaM standard includes components that facilitate the transfer of information between analysis results and ADaM datasets, as well as between ADaM datasets and SDTM datasets. All SDTM variables used for variables derivatives in ADaM should be included in the ADaM datasets if at all possible to guarantee traceability. The define.xml file must provide a thorough description of the contents of each given ADaM dataset. To facilitate effectiveness and safety evaluations, sponsors are encouraged to contribute ADaM records. Data description files should point to at least one dataset that contains the major efficacy variables [12].

Model for Data Validation

SAS (Statistical Analysis System)

SAS is widely used in the fields of survey information analysis and data management because of its flexibility, dependability, instructional assistance, and comprehensive documentation. SAS, or the 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 the best software. Data processing, statistical description and inference, and data display are only some of the numerous uses for SAS's various functions and processes. Descriptive tables may be generated with the help of the SAS functions PROC TABULATE and PROC REPORT.

Statistical analysis and data visualization are only two of the many uses for SAS, a command-driven software tool. Unfortunately, it's only compatible with Microsoft Windows. In both the business world and the academic world, it is one of the most used statistical software programs. SAS is utilized in many different industries, including business and health, and has a large, active online community. Its main selling points are its huge variety of statistical techniques and algorithms, particularly for advanced statistics, its highly configurable analysis choices and output options, and its publication-quality graphics with an output delivery system (ODS). By harmonizing data storage, programs like SAS and R, which produce outputs in incompatible formats, may be used together for analysis [27].

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

- Retrieve information stored in a wide variety of media, including SAS tables, Excel tables, and database files.
- Manage and transform data that is already there and get the information that is needed. One may add new columns, merge parts of the data collection, and create subsets of the data.
- Statistical approaches are as simple as correlations and logistic regression and as complex as model selection and Bayesian hierarchical models may be used to examine the data.
- Share the findings of the research with others by writing a comprehensive report on the findings. HTML, PDF, and RTF are just some of the many formats that support custom report exports [28].

As all data, including data from clinical trials, is standardized, there will be more of it to analyse in the healthcare business. SAS is the ideal option since it simplifies both data collection and analysis. SAS won't be the only program available, but it will be among the finest at analysing and manipulating data. The present engagement of SAS programmers in the clinical study means that their responsibilities will grow. By providing a wide variety of analytics capabilities, SAS helps pharmaceutical manufacturers increase their output and ensure that patients get safe and effective medications and other treatments. Better demand forecasting is possible because of SAS's ability to enhance quality, throughput, and equipment performance. Only SAS provides a GMP-compliant environment for integrated analytics, making it ideal for the pharmaceutical business. With over 40 years of experience in the life sciences business and extensive knowledge of manufacturing, SAS is a recognized leader in analytics. It allows for a more natural and efficient shift from batch processing to real-time analysis of data [29].

Analysis Results Data Model (ARDM)

The ARDM may evolve and expand as required. For instance, the structure may be modified to include additional columns for each analytical standard. The findings may be inspected and visualized in a number of different ways, and this is not a component of the analysis standard. The focus of the analysis is shifted from the final products that will be used (such as tables and figures) to the underlying data model. Applying ARDM streamlines what was previously a laborious

process of information extraction and verification which is meta-analysis. Instead of redoing the study, the tables, lists, and figures might be prepared automatically. In order to demonstrate the feasibility of the ARDM, it has been implemented in the R programming language with a relational SQLite database; however, these decisions are flexible. This implementation is not intended to be comprehensive, but rather a springboard for further exploration. Using the CDISC Pilot Project ADaM dataset, appropriate steps can be followed to build the ARDM, which incorporates three analytical standards (descriptive statistics, safety analysis, and survival analysis) [30].

Having an ARDM that supports query-based searches makes it easier to keep track of, search for, and get results. For each particular trial in the database, one may do a search based on secondary endpoints such as point estimates, p-values, and the incidence of adverse events. Data selection for cohorts may be automated using query-based searches, and the use of analytic standards can be used to speed up the generation of findings from the gathered information. Decision-making and improvements are aided by this as well. To extrapolate to cohorts that need particular considerations, such as paediatric patients, it is helpful to have access to entire trial findings beyond the main objective. The analytical findings may be treated as data with the help of the ARDM, which offers a centralized repository for all relevant information. Outputs like tables, images, and listings may be readily created from the findings, making the ARDM useful for regulatory bodies as well [31] (Table 1).

Advantages of Biostatistical Methods in Clinical Research

Clinical research relies heavily on statistical methods because they provide a basis for concluding a target population. Clinical researchers should have a firm grasp of statistical methods to properly evaluate study datasets. Clinical research relies on statistical analysis to establish norms for patient characteristics, healthcare delivery, and the outcome. Clinicians in today's evidence-based practice environment would do well to familiarise themselves with basic statistical ideas and procedures in order to better assess and implement clinical study findings [32].

To draw reliable findings from clinical studies, researchers must have a well-defined process, recruit individuals without bias, and use appropriate analytic methods. The effectiveness of a clinical trial depends on the use of data analytics, which is made feasible by data analysis tools backed by software. Preparing regulatory filings and gaining market clearance in the clinical arena requires effective data management. CDISC data standards represent the essential components of the clinical data lifecycle to accommodate the demands

Table 1: Biostatistical methodologies and their key functions.

Model for Planning	
PRM (Protocol Representation Model)	The PRM captures information as data instead of text in a protocol and helps create case report forms (CRFs) and study outlines
ODM (Operational Data Model)	The ODM standard plays a key role in clinical research informatics, including areas such as data exchange, archival, U.S. Food and Drug Administration (FDA) submission, and inter-operability with healthcare data.
Data Collection	
CDASH (Clinical Data Acquisition Standards Harmonization)	CDASH organizes 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 data to streamline processes in collection, management, analysis, and reporting.
Data Analysis	
ADaM (Analysis Data Model)	ADaM datasets provide significant information about individual studies and the derivation decisions made in each study.
Data Validation	
SAS (Statistical Analysis Software)	SAS is a command-driven software for data manipulation, statistical description and inference, and data presentation.
ARDM	ARDM facilitates the tracking, searching, and retrieving of outputs as it enables query-based searches.
Submission for approval	

at each stage of the clinical research lifecycle. While regulatory processes have historically revolved around document submission, there has been a growing interest in evaluating the data behind document creation [17].

Challenges

Researchers confront a difficult task in the reporting phase in creating an understandable account of the simulation study's methodology, findings, and analysis. At this point, a wide variety of questionable research practices (QRPs) are possible. For instance, the findings with the highest ICH E6 (R2) for the technique of interest may be highlighted in the report. Overly positive impressions are created, and readers may be persuaded to believe that the strategy consistently surpasses rivals if negative results are not mentioned. Reporting the degree of uncertainty associated with simulation findings is also important. Failure to report Monte Carlo uncertainty (R3) by providing error bars or confidence intervals that reflect uncertainty in the simulation makes it difficult for readers to evaluate the reliability of the simulation study's findings and it permits the presentation of random variations between results as if they were systematic [33].

The potential for flawed statistical analysis in biological studies has long been a source of worry for scientists. Numerous authors have detailed the problems of statistical reporting, and guidelines have been recommended for both researchers and journal reviewers. Even the main endpoint of the experiment might be improperly analysed without adequate validation of model assumptions. If biostatisticians

don't collaborate closely with doctors to choose the best statistical methods for analysing data, the resulting findings might be problematic or deceptive [32].

When dealing with massive amounts of personally identifiable health information, data management, and security pose a significant obstacle to data analytics. In this context, the difficulty for biostatisticians is in selecting suitable statistical models for the examination of data with a complicated structure. Failure to adhere to SDTM's conformance requirements and pass validation tests is a typical complaint about adopting CDISC standards in observational research. Rules for conformity may be applied to a dataset, a variable, or a restricted vocabulary. These and other regulations relevant to regulatory submissions are checked for validity throughout the validation process [34]. In contrast to clinical trials, the definition of a visit in observational research is far more flexible. Similarly, a common obstacle with ADaM is figuring out how to define visits and how to set up windows for them. This may be especially true in retrospective research when incomplete or missing dates are more likely to occur and may be impossible to complete. Staff without an in-depth understanding of CDISC standards and datasets may find it difficult to complete SDTM mapping, hence it is important for them to acquire this information [35].

Conclusion

There is a rich and varied multi-disciplinary atmosphere in all areas of biostatistical work. When submitting an NDA, one must include clinical data that follows the guidelines established by the Clinical Data

Interchange Standards Consortium (CDISC). In recent years, CDISC standards have received much attention from the scientific community. There is still a high need for standard implementation assistance in this area. There are certain inconsistencies that must be fixed, in spite of the fact that the vast majority of the current CDISC standards and accompanying guidelines are suitable for use with research data. Therefore, this research has made an effort to present the many statistical approaches used in the FDA's drug approval process. The strengths and weaknesses of biostatistical software and analysis have also been investigated. The merits and drawbacks of any available statistical study are different. To guarantee that a clinical trial satisfies the requirements of regulatory bodies in drug approvals, more work has to be done to improvise the compliance and validation methods.

References

- 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.
- 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.
- 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*.
- Deglise-Hawkinson J, Kaufman DL, Roessler B, Oyen MPV (2020) Access planning and resource coordination for clinical research operations. *IISE Trans* 52: 832-849.
- Sharma H (2021) Statistical significance or clinical significance? A researcher's dilemma for appropriate interpretation of research results. *Saudi J Anaesth* 15: 431-434.
- Beaconcure (2023) The essential role of biostatistics in clinical trials. *Beaconcure Technologies*.
- Dhudasia MB, Grundmeier RW, Mukhopadhyay S (2023) Essentials of data management: An overview. *Pediatr Res* 93: 2-3.
- Nourani A, Ayatollahi H, Dodaran MS (2019) A review of clinical data management systems used in clinical trials. *Rev Recent Clin Trials* 14: 10-23.
- Pugh SL, Torres-Saavedra PA (2021) Fundamental statistical concepts in clinical trials and diagnostic testing. *J Nucl Med* 62: 757-764.
- Facile R, Muhlbradt EE, Gong M, Li Q, Popat V, et al. (2022) Use of clinical data interchange standards consortium (CDISC) standards for real-world data: Expert perspectives from a qualitative delphi survey. *JMIR Med Inform* 10: e30363.
- FDA/CDER (2020) Study data technical conformance guide - technical specifications document, US Food & Drug Administration.
- USFDA (2020) Study data technical conformance guide: Guidance for industry providing regulatory submissions in electronic format - standardized study data.
- Clinical Data Interchange Standards Consortium (CDISC) (2023) PRM Protocol Representation Model.
- Turner MA, Hildebrand H, Fernandes RM, de Wildt SN, Mahler F, et al. (2021) The conect4children (c4c) Consortium: Potential for improving European clinical research into medicines for children. *Pharmaceut Med* 35: 71-79.
- Brix TJ, Bruland P, Sarfraz S, Ernsting J, Neuhaus P, et al. (2018) ODM data analysis - A tool for the automatic validation, monitoring, and generation of generic descriptive statistics of patient data. *PLoS One* 13: e0199242.
- European Industrial Pharmacists Group (EIPG) (2021) European medicines regulatory network data standardisation strategy.
- Clinical Data Interchange Standards Consortium (CDISC) (2023) CDISC standards in the clinical research process.
- Clinical Data Interchange Standards Consortium (CDISC) (2023) CDASH Model v1.2.
- Clinical Data Interchange Standards Consortium (CDISC) (2023) SDTM and CDASH: Why you need both.
- (2023) HCL Tech, What is CDISC and SDTM?
- (2023) Clinical Data Interchange Standards Consortium (CDISC), SDTM.
- Observational Health Data Sciences and Informatics (OHDSI) (2023) OMOP CDM to SDTM.
- Dugas M, Neuhaus P, Meidt A, Doods J, Storck M, et al. (2016) Portal of medical data models: Information infrastructure for medical research and healthcare. *Database (Oxford)* bav121.
- Clinical Data Interchange Standards Consortium (CDISC) (2023) ADaM.
- Fulton J (2022) CDISC ADaM Phases, Periods, and Subperiods: A Case Study. *Journal of the Society for Clinical Data Management* 2.
- Libguides (2023) Statistical & qualitative data analysis software: About SAS and SAS on demand for academics.
- SAS (2023) What can I do with SAS?
- SAS (2023) Pharmaceutical manufacturing software.
- Barros JM, Widmer LA, Baillie M, Wandel S, et al. (2022) Rethinking clinical study data: Why we should respect analysis results as data. *Scientific Data* 9: 686.
- R Core Team (2021) R: A language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria.
- Sarma KVS, Mohan A, Vedururu SS (2022) Statistical methods in clinical studies: An overview. *Journal of Clinical and Scientific Research* 11: 34-39.
- Pawel S, Kook L, Reeve K (2023) Pitfalls and potentials in simulation studies: Questionable research practices in comparative simulation studies allow for spurious claims of superiority of any method. *Biometrical Journal*.
- Sakir A, Aadhaar D (2020) Pitfalls in statistical analysis - A Reviewers' perspective. *Indian Journal of Rheumatology* 15: 39-45.
- Yamamoto M, Nagai K, Saito TI (2022) Implementation of Clinical Data Interchange Standards Consortium (CDISC) Study Data Tabulation Model (SDTM) in investigator-initiated studies on hematologic malignancies. *SSRN*.