Paper PD02

Development of Statistical Scientists for Pharmaceutical Drug Development at India Global Capability Centres

Suresh Chenji, Teva Pharmaceuticals, Bangalore

ABSTRACT

Statistical scientists are indispensable throughout the clinical development lifecycle, overseeing critical tasks from trial design and sample size calculation to protocol development and regulatory submission. Their role extends beyond data analysis to include database locks, integrated analyses, and the final interpretation of study results. Success in this field requires a blend of rigorous technical expertise and strong interpersonal skills, as they collaborate daily with cross-functional teams and regulatory bodies. As the industry evolves, there are expanding opportunities for India-based statisticians to contribute to global drug development. To excel, these professionals must prioritize continuous learning in therapeutic sciences, innovative trial methodologies, and advanced statistical computing.

Growing opportunities for statisticians based in India to contribute to global drug development through virtual collaboration require a sustained focus on professional development. Key areas include a strong understanding of therapeutic and clinical science, regulatory guidance and requirements, innovative clinical trial design methodologies, modern and appropriate statistical methods and applications, proficiency in statistical computing and software, and effective collaborative skills such as clear communication and a genuine intent to provide and receive feedback. This paper discussion will also cover statisticians' roles and responsibilities, accountability, and practical experiences

INTRODUCTION

STATISTICAIN RESPONSIBILITIES

As a Statistical Lead, you serve as a vital member of the core Clinical Trial Team, providing strategic input into trial design and protocol development across diverse therapeutic areas and clinical phases. Your role involves making critical decisions on the choice of control groups, defining primary and key secondary endpoints—whether continuous, binary, or time-to-event—and establishing the appropriate hypothesis framework for superiority, non-inferiority, or equivalence testing. You are responsible for the technical foundation of the study, including rigorous sample size and power calculations, determining the treatment and study duration, and crafting the primary statistical analysis plan.

Your expertise spans the entire development spectrum: from conducting clinical pharmacology and PK/PD studies in Phase 1 to managing dose-finding and optimization trials by assessing safety and tolerability in Oncology. You also lead the design of Phase 2 dose-response trials and navigate the complexities of multiple testing endpoints in Phase 3 confirmatory trials. Beyond interventional studies, you provide essential guidance for non-interventional, observational, and Phase 4 post-marketing requirement or commitment trials, ensuring statistical integrity at every stage of the product development lifecycle.

During trial start-up, the statistician plays a key role by reviewing and providing input to critical activities that ensure data quality and analysis readiness. This includes confirming the design and content of case report forms (CRFs) and electronic clinical outcome assessments (eCOA), reviewing the Data Management Plan, and contributing to the definition of critical data edit checks and data transfer specifications. The statistician also provides input to and reviews the randomization schedule to ensure alignment with the study design. In addition, responsibilities include developing the Statistical Analysis Plan (SAP) and creating mock shells for tables, figures, and listings (TFLs) to support the Clinical Study Report. The role further involves reviewing SDTM and ADaM standards and providing study-specific input to ensure that analysis datasets are appropriately structured to meet the trial's analytical objectives.

Throughout study execution and close-out, the statistician is responsible for ensuring the integrity and readiness of trial data across key analysis milestones. This includes reviewing, discussing, and confirming data quality, as well as formally confirming the trial database lock. The role involves preparing and releasing topline results to internal clinical development leadership, reviewing and finalizing tables, figures, and listings (TFLs) for the Clinical Study Report, and providing critical input to CSR drafts, including interpretation of trial results and conclusions. Statisticians also review and contribute to trial publications and are accountable for reviewing and confirming regulatory submission data packages, such as define.xml, reviewer's guides, and CRT datasets for FDA submissions. In addition, they develop and support integrated (pooled) analysis dataset specifications, analysis plans, and TFL shells, and review analysis outputs and reports to support regulatory submissions and timely responses to health authority questions.

ESSENTIAL SKILLS

An effective statistician brings a strong understanding of the clinical development life cycle and clinical development plans, supported by solid therapeutic knowledge and awareness of regulatory guidance and requirements. Getting the basics right is essential, including the application of appropriate statistical methods, robust analysis approaches, and clear, accurate data presentation. The role requires the ability to clearly communicate complex data and results, along with providing sound clinical and statistical interpretations. A good understanding of CDISC standards (SDTM and ADaM) and regulatory submission requirements is critical, as is programming proficiency in SAS and R, familiarity with sample size and trial design tools such as PASS, nQuery, and EAST, and experience working in cloud or distributed computing environments. In addition, strong communication, project management, and outsourcing or vendor management skills are important, with cross-functional collaboration being paramount when working closely with medical directors, clinical pharmacology, clinical project managers, regulatory teams, medical writing, data management, and statistical programming colleagues.

INNOVATIONS IN STATISTICS METHODOLOGY AND APPLICATIONS

ESTIMANDS AND SENSITIVITY ANALYSIS

In clinical trials, establishing a robust framework for assessing treatment effects begins with the precise definition of an estimand, which clearly articulates what is being estimated in the presence of potential intercurrent events. To align the trial with specific therapeutic objectives and clinical contexts, researchers must select from five primary strategies: treatment policy, hypothetical, composite, while-on-treatment, or principal stratum. These strategies ensure that the analysis accounts for events like treatment discontinuation or the use of rescue medication in a way that is scientifically meaningful.

Equally critical is the pre-specification of sensitivity analyses, which serve to explore the robustness of the trial's conclusions. By testing the impact of different model assumptions and data limitations—such as varying approaches to handling missing data through a two-way tipping point analysis—investigators can demonstrate that the primary results remain reliable even under alternative scenarios.

MULTIPLE TESTING APPROACHES

In confirmatory clinical trials, the simultaneous testing of multiple hypotheses—such as comparing several dose levels, evaluating a combination of primary and key secondary endpoints for labeling, or conducting interim analyses—necessitates a rigorous approach to control the family-wise error rate. To maintain statistical integrity, researchers must select appropriate multiplicity adjustment strategies tailored to the trial design. These range from single-step procedures like Bonferroni to stepwise methods such as Holm or Hochberg, which offer varying levels of power. For more complex structures, hierarchical approaches like fixed-sequence or gatekeeping procedures ensure that secondary endpoints are only formally tested if primary objectives are met. Furthermore, advanced graphical testing procedures provide a flexible framework for visualizing and navigating the alpha-allocation between multiple interrelated hypotheses, ensuring that the trial's conclusions remain robust and scientifically valid.

CLINICAL TRIAL SIMULATIONS

Simulation-based modeling serves as a vital strategic exercise to navigate drug development uncertainties by facilitating comprehensive "what-if" analyses. This approach allows researchers to optimize trial designs by integrating complex variables, including patient characteristics, drug-disease dynamics through pharmacokinetic and pharmacodynamic (PK/PD) models, and operational execution factors. By leveraging historical data, these models can even support the creation of synthetic control arms, reducing the need for placebo-controlled participants and enhancing efficiency. Furthermore, these simulations are essential for pre-testing and justifying adaptive design elements, such as sample size re-estimation and futility planning, ensuring that the trial remains statistically sound and ethically viable even as it evolves based on accumulating data.

PROBABILITY OF STUDY SUCCESS

While traditional statistical power is calculated based on a fixed, assumed treatment effect, the Probability of Success (PoS) offers a more realistic assessment by utilizing an unconditional, often Bayesian, probability that accounts for real-world uncertainty and a range of potential outcomes. By integrating data from Phase 2 results and historical studies, PoS allows developers to estimate the likelihood that a Phase 3 trial will successfully meet its primary objectives. This metric is a cornerstone of strategic decision-making, providing the quantitative evidence needed to guide Go/No-Go transitions between clinical phases. Ultimately, PoS enables more informed portfolio prioritization and smarter investment allocations by highlighting which drug candidates have the highest statistical and clinical potential for success.

ADAPTIVE TRIAL DESIGNS

Adaptive trial designs provide a flexible framework that allows for prospectively planned modifications to one or more aspects of an ongoing study based on accumulating participant data. This "learn-and-confirm" process typically involves an interim analysis conducted by an independent Data Monitoring Committee (DMC), ensuring that the trial can be adjusted without compromising its integrity. Key modifications may include sample size re-estimation to ensure adequate power, the addition or removal of treatment arms, or narrowing the study population to those most likely to benefit. Throughout these changes, the primary focus remains on maintaining statistical validity by strictly controlling the Type I error rate and ensuring robust data blinding, thereby protecting the trial's ability to provide definitive regulatory evidence.

MASTER PROTOCOLS

Master protocols represent an overarching clinical trial framework designed to increase efficiency by evaluating multiple drugs, multiple diseases, or multiple patient subgroups simultaneously within a single infrastructure. These designs are categorized into three primary types: basket trials, which test a single targeted therapy across multiple diseases sharing a common molecular origin; umbrella trials, which evaluate multiple different drugs for a single disease based on specific biomarkers; and platform trials, which function as perpetual studies where treatment arms can enter or leave the trial as evidence accumulates. While highly efficient, these complex designs require sophisticated statistical management to address multiplicity and alpha control, manage subgroup heterogeneity, and mitigate the risk of population drift over time, ensuring that the results for each sub-study remain rigorous and reliable.

BAYESIAN STATISTICAL METHODOLOGY

Modern oncology dose-finding trials have evolved beyond traditional designs toward more sophisticated methods to identify the maximum tolerated dose (MTD). By utilizing advanced frameworks such as the Continual Reassessment Method (CRM), Bayesian Logistic Regression Models (BLRM), and the Bayesian Optimal Interval (BOIN) design, researchers can more accurately model the dose-toxicity relationship and protect patient safety. These Bayesian approaches allow for the dynamic borrowing of information from historical trials or the augmentation of concurrent control arms with external data, significantly increasing statistical efficiency. Furthermore, these methods facilitate pediatric extrapolation through hierarchical models and the use of predictive probability to support early stopping for either success or futility, ensuring that drug development remains both agile and ethically sound

COVARIATE ADJUSTMENT AND SUBGROUPS IDENTIFICATION

To ensure the reliability of clinical findings, researchers must deeply understand and adjust for diverse patient characteristics that can influence the magnitude and direction of a treatment effect. By employing advanced statistical methods such as G-computation, investigators can estimate unconditional (marginal) treatment effects, providing a clearer picture of how a therapy performs across the entire target population rather than just a specific subset

Furthermore, it is essential to identify and evaluate treatment effects within specific patient subgroups to determine who benefits most from an intervention. While these analyses must be pre-specified in confirmatory trials to maintain regulatory rigor, they also serve a vital role in exploratory research to generate new hypotheses for future clinical development

PERSONALIZED MEDICINES DEVELOPMENT

Precision medicine utilizes biomarkers to tailor treatments to specific patient subgroups, ensuring that observed treatment effects are genuine rather than the result of natural disease progression or random chance. A critical distinction in this framework is the role of predictive biomarkers, which identify patients likely to have a favorable response to a specific therapy—statistically represented by a treatment-by-biomarker interaction. In contrast, prognostic biomarkers predict the overall course of a disease, such as survival or recurrence, regardless of the treatment administered, and are modeled as a main effect. Implementing this approach requires navigating complex challenges, including managing small sample sizes through the use of master protocols and maintaining rigorous Type I error control amidst increased multiplicity. Furthermore, the clinical utility of these findings depends heavily on the technical validation of assays and the precise selection of cut-points to accurately categorize patients.

OTHER DEVELOPMENTS

Causal inference has become part of the core DNA of modern statistics and is increasingly moving into the mainstream, enabling the principled integration of observational and real-world evidence (RWE) data alongside randomized controlled trial (RCT) data.

cardiovascular outcome trials, which are typically time-to-event studies focused on major adverse cardiovascular events (MACE), often enrolling tens of thousands of patients at high cost and facing challenges such as equal weighting of endpoints, dilution and delayed treatment effects due to non-proportional hazards, and the presence of competing risks that must be carefully addressed both statistically and clinically.

Model-Informed Drug Development (MIDD) has emerged as a strategic framework that leverages mathematical and statistical models to integrate data across the entire drug development lifecycle, improving efficiency by optimizing dosing, trial design, and regulatory decision-making, and ultimately accelerating the delivery of safe and effective therapies to patients.

These developments are not just statistical methods to plan and implement but address modern drug development strategies & challenges and drive optimum decision-making, which require statistical leadership, *close partnership and ongoing dialogue with core clinical team (physicians, regulatory, commercial etc.) and regulatory agency*

Statistician role is not only limited to implement complex statistical methods and analyze trial data but also getting involved in the decision-making process to drive strategic business decisions

EVOLVING SPECIALIZED ROLES

RWE STATISTICS

Clinical research increasingly relies on analyzing data collected outside the highly controlled environment of randomized controlled trials, including observational studies, electronic health records, disease registries, and insurance claims, particularly for rare diseases or new indications where traditional trials may be infeasible. In these settings, eliminating bias and confounding is the primary statistical challenge, addressed through methods such as propensity score matching, inverse probability of treatment weighting (IPTW), and propensity score stratification. When important confounders are unmeasured, instrumental variable analysis provides an alternative approach to support causal interpretation. More advanced causal inference techniques, including g-computation and marginal structural models, are especially valuable when treatments change over time, while the target trial emulation framework offers a principled way to design and analyze observational studies to estimate causal treatment effects. These approaches have enabled the use of real-world evidence as external control arms or for confirming long-term outcomes, as illustrated by regulatory approvals of drugs such as Zolgensma, Vimizim, and Voxzogo that incorporated RWE into their evidence packages

HTA STATISTICS

Health Technology Assessment (HTA) bodies evaluate the value of health technologies by balancing clinical benefit against cost and real-world effectiveness to inform reimbursement and pricing decisions that ultimately determine patient access. In Europe, the new HTA Regulation seeks to harmonize these assessments across member states through Joint Clinical Assessments (JCAs) for innovative medicines and devices, streamlining processes, reducing duplication, and aligning regulatory and HTA evidence requirements to support timely access for patients. Economic modeling plays a central role in HTA, using decision trees, Markov models, or simulation approaches to estimate outcomes such as the Incremental Cost-Effectiveness Ratio (ICER) or cost per Quality-Adjusted Life Year (QALY). To synthesize evidence or conduct indirect comparisons, methods including meta-analysis, network meta-analysis (NMA), and individual patient data approaches such as Matching-Adjusted Indirect Comparison (MAIC) or Simulated Treatment Comparison (STC) are commonly applied. Advanced survival analysis techniques—such as parametric curve fitting, restricted mean survival time (RMST), spline models, and methods to adjust for treatment switching—are often required when clinical trials are too short to fully capture life expectancy. In India, HTA initiatives (HTAIn) aim to generate robust, transparent evidence to support informed decision-making on medicines, medical devices, and public health programs.

NON-CLINICAL STATISTICS

Statistical methods play a critical role before human testing during discovery, toxicology, and chemistry, manufacturing, and controls (CMC) activities by supporting safety assessment, dose determination, and stabilization of the manufacturing process. Guided by regulatory frameworks such as ICH Q1, Q2, and Q10, Good Laboratory Practice (GLP), and the 3Rs principle—Reduction, Refinement, and Replacement of animal models—optimized statistical designs help improve efficiency while minimizing animal use. Dose–response modeling, often using the four-parameter logistic (4PL) model, is applied to estimate key metrics such as IC_{50} and ED_{50}, while statistical ranking and selection methods support lead compound identification by prioritizing candidates with the most favorable therapeutic window. Design of Experiments (DoE) enables systematic evaluation of critical factors such as pH and temperature to define the manufacturing design space, ensuring product robustness. In addition, stability and shelf-life are assessed using linear regression to determine when the lower 95% confidence interval crosses predefined acceptance limits, typically 90% potency, and assay validation demonstrates accuracy, precision, and linearity in compliance with ICH Q2 requirements.

CLINICAL TRIAL CENTRALIZED MONITORING

Risk-Based Quality Management (RBQM), also known as Centralized Statistical Monitoring (CSM), uses statistical tools to proactively identify data anomalies, potential fraud, and performance issues across clinical trial sites and is a key component of quality by design as encouraged by the ICH E6 (R2) guidelines, making statistical monitoring a regulatory expectation. The primary objective is to detect performance outliers, including fabricated or fraudulent data where values appear unrealistically consistent, under-reporting of adverse events by unusually “quiet” sites, procedural non-compliance such as systematic protocol deviations or missed visit windows, and biased results caused by equipment malfunction or poor calibration. This is typically implemented through a structured, waterfall-style statistical process that begins with data ingestion from multiple sources such as electronic data capture (EDC), laboratory systems, and interactive response technologies (IRT). Automated univariate screening is then applied to assess summary metrics like means, standard deviations, and skewness for each variable at each site, followed by multivariate analyses using methods such as principal component analysis (PCA) or Mahalanobis distance to identify sites that are outliers across multiple dimensions. Finally, signal prioritization ensures efficient oversight, with tools such as heat maps helping statisticians and monitors focus attention on the highest-risk sites rather than reacting to isolated statistical signals.

ROLE IN AI/ML EVOLUTION IN DRUG DEVELOPMENT

As referenced in review article 'Artificial intelligence in clinical trials: A comprehensive review of opportunities, challenges, and future directions' AI applications in drug development life cycle are,

AI Driven	Key Inputs	AI Outcomes
Protocol Optimisation	Trial history, protocols, regulations	Feasible design, eligibility criteria
Site Selection	Investigator & population data	Ranked sites, enrollment forecasts

AI Application	Key Data / Inputs	Primary Function
Digital Biomarkers & Monitoring	Wearables, voice, sleep, HR sensors	Detects changes in patient status
Risk-Based Monitoring	Site & trial data	AI risk scoring, deviation & issue detection
Automated Data Cleaning	Clinical trial datasets, free text	NLP standardization, missing-data imputation

Validation of AI systems in regulated industries is critical, as failures or algorithmic bias can have direct consequences measured in patient lives. While machine learning algorithms excel at uncovering complex patterns in high-dimensional data that are difficult or impossible to detect with traditional statistical methods, statisticians play a key role in validating these predictive AI tools by ensuring robust uncertainty quantification, appropriate performance assessment, and interpretability aligned with regulatory and clinical expectations. By combining a deep understanding of the mathematical foundations of machine learning with domain knowledge of clinical research, statisticians provide essential context for evaluating model assumptions, limitations, and real-world applicability. As a result, statisticians are emerging as indispensable partners in the development and validation of AI systems, helping ensure they are reliable, transparent, and compliant with regulatory standards. Statisticians provide input, review and confirm every automated tools or workflows developed by statistical programming and data sciences team.

Interconnected barriers to AI implementation in clinical trials

Area	Key Challenges
Technical Barriers	Data silos & legacy systems; algorithm complexity; integration difficulties; slow deployment; poor scalability
Stakeholder Trust & Acceptance	Black-box models; lack of transparency/training; clinician resistance; delayed adoption; underutilization
Regulatory Uncertainty	No harmonized standards; unclear validation pathways; prolonged approval; inconsistent expectations; innovation bottlenecks

UNIQUE OPPORTUNITIES

Over the past two decades, global pharmaceutical companies have established Global Capability Centers (GCCs) in India, and through CRO FSP teams, statisticians have increasingly become integral members of global clinical trial teams, contributing directly to clinical development portfolio deliverables. As their contributions and capabilities gain greater visibility, expectations for statisticians have grown, creating a need for continuous capacity building. This includes staying current with the latest statistical methodologies, software tools, and evolving regulatory and drug development requirements. Equally important is the development of soft skills to support career progression, including effective communication, negotiation and influencing, cross-functional collaboration and leadership, conflict resolution, adaptability and resilience, strategic and critical thinking, problem-solving, data storytelling and visualization, as well as cultural awareness and the ability to align insights with audience and business priorities. Together, these technical and interpersonal competencies position statisticians to deliver high-impact contributions in a complex, global environment.

Statisticians today need to develop strategic competencies that go beyond technical expertise, including a deep understanding of regulatory requirements, the commercial impact of statistical decisions, innovative methodology and implementation, market access considerations, and competitive dynamics. Equally important is the ability to translate complex statistical findings into clear, actionable insights for key stakeholders. The increasing complexity of drug development, coupled with regulatory emphasis on evidence-based decision-making and the integration of emerging technologies, presents significant opportunities for statisticians who proactively embrace this evolution and cultivate the skills required for the future. With continued scientific learning, cross-functional collaboration across regions, and a growth mindset, pharmaceutical statistics holds a promising and dynamic future.

ROLE OF EMERGING STATISTICAL LEADERS

To advance the field of pharmaceutical statistics, it is essential to promote innovative clinical trial methodologies and encourage their adoption through knowledge-sharing sessions and workshops. Early career development should be prioritized, giving junior statisticians diverse experiences and supporting their growth through structured coaching and mentorship programs. Bridging academia and industry is equally important: statistical leaders can collaborate with universities to shape education programs, strengthen partnerships, attract future talent, and provide opportunities such as guest lectures, internships, and guidance for PhD or research projects relevant to pharmaceutical applications. Promoting innovation in drug development, methodologies, processes, and technologies requires closing the gap between creative ideas and practical implementation. Development opportunities can be further enhanced through collaboration across companies, professional working groups, and inter-regional organizations. Leaders must remain aware of the evolving nature of skills and capabilities, staying both relevant and authentic, while fostering strategies that maintain curiosity, encourage continuous learning, and enable statisticians to thrive in a dynamic and rapidly advancing field.

CONCLUSION

As the complexity of drug development increases, statisticians are expected to develop both technical expertise and strategic competencies, including regulatory understanding, commercial insight, cross-functional collaboration, and advanced soft skills such as communication, leadership, and problem-solving. Early-career mentoring, industry–academia partnerships, and continuous scientific learning are key to nurturing the next generation of statistical leaders. By embracing innovation, applying rigorous methodology, and bridging science with business strategy, statisticians are emerging not only as data analysts but as strategic influencers driving decisions in global pharmaceutical development.

REFERENCES

Few relevant journals with focus on biostatistics, clinical trials, and industry applications:
Statistics in Medicine
Pharmaceutical Statistics
Journal of Biopharmaceutical Statistics
DIA Therapeutic Innovation & Regulatory Science
Clinical Pharmacology & Therapeutics
other therapeutic medical journals.

ICH, FDA, EMA and other regions regulatory guidance documents

Collaborative open-source initiatives Openstatsware and Pharmaverse foster innovation and efficiency in statistical methodologies and clinical reporting

Professional Organizations and Industry working groups, workshops and conferences:
American Statistical Association (ASA) Biopharmaceutical Section
European Federation of Statisticians in the Pharmaceutical Industry (EFSPI)
Statisticians in the Pharmaceutical Industry (PSI)
Indian Association for Statistics in Clinical Trials (IASCT)
Pharmaceutical Users Software Exchange (PHUSE)

David B. Olawade et al., Artificial intelligence in clinical trials: A comprehensive review of opportunities, challenges, and future directions, International Journal of Medical Informatics, Volume 206, 2026

This paper describes personal views of the author based on his experience as a statistical leader based out of India. This paper details are based on publicly available information

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:
Author Name: Suresh Chenji
Designation: Director of Global Statistics and Data Sciences
Company: Teva Pharmaceuticals
Address: Bangalore, India
Email: suresh.chenji@teva.co.in
Website: www.tevapharm.com/