Paper PD21

Bridging Analysis to Strategy: The Transformative Role of Statistical Analysts in Drug Development

Lingsong Yun, Bayer Inc., Toronto, Canada

ABSTRACT

The role of statistical analysts in pharmaceutical drug development has expanded far beyond traditional programming and reporting. Analysts are now strategic partners in study design, data integration, regulatory interactions, and cross-functional decision-making. Rapid changes in the pharmaceutical industry have heightened expectations, requiring broader expertise in data science, clinical development methodology, and regulatory guidance. Drawing on case studies in oncology and precision medicine, we pioneered a new hybrid statistical analyst role to lead across-function activities such as SAP development, methodology guidance, and CSR review, thereby strengthening study planning, improving interpretability of clinical evidence, and supporting regulatory submissions. We also discuss how the role will continue to evolve with AI-driven tools, data-driven analyses, and integrated evidence generation. Collectively, these shifts highlight statistical analysts not only as data experts, but also as innovators and strategic drivers shaping the future of drug development.

1. INTRODUCTION

Clinical drug development is fundamentally an evidence-generation process, transforming patient data into scientifically valid conclusions that support regulatory approval and clinical decision-making. For decades, randomized clinical trials have served as the gold standard for demonstrating safety and efficacy, with clearly defined roles across clinical, statistical, regulatory, and operational functions. Within this framework, statistical analysts (SAs) have traditionally played a critical but narrowly scoped role, focused primarily on programming, analysis execution, and production of tables, figures, and listings (TFLs) to support clinical study reports (CSRs).

In recent years, the statistical programming and analytics (SPA) function has experienced a transformation driven by technological innovation, including widespread adoption of open-source programming languages such as R and Python, rapid development of automation, data science, and most recently artificial intelligence. Besides, the pharmaceutical industry has undergone profound changes. Advances in precision medicine, oncology targeted therapies with companion diagnostics (CDx), innovative and adaptive trial designs, and the integration of complex and high-dimensional data have dramatically increased the analytical and interpretative demands. These changes have blurred traditional boundaries between analysis execution and analysis strategy, creating both challenges and opportunities for the statistical analyst role. SAs are increasingly expected to contribute beyond code development and TFL reporting, including data exploration, methodological discussions, and cross-functional collaborations with both statistician and other strategic stakeholders such as scientific leads and regulatory.

This article seeks to discuss how SAs position as strategic contributors rather than purely operational resources, specifically:

- Describe key changes in the drug development landscape that drive role evolution
- Examine how the SA role in pharmaceutical is different from SA in real world
- Present a case study of a pioneered hybrid SA role in oncology targeted therapy and precision medicine.
- Illustrate how SA can bridge analysis and strategy through transforming data into evidence-based decisions that support clinical development and regulatory success

2. CHANGES OF DRUG DEVELOPMENT LANDSCAPE

2.1 HISTORICAL DEVELOPMENT OF THE STATISTICAL ANALYST ROLE

The statistical analyst role in the pharmaceutical industry has evolved in parallel with advances in computing technology, data standards, and regulatory expectations. In earlier phases of drug development back to dozens of years ago, analytical support functions were primarily focused on document preparation and basic data tabulation. As electronic data capture, standardized datasets, and regulatory submission requirements became more established, responsibilities increasingly shifted toward SAS programming and standardized reporting.

Over time, the role matured into what is now commonly referred to as statistical programming or statistical analysis, with core responsibilities centered on implementing predefined analysis plans, producing tables, figures, and listings

(TFLs), and supporting regulatory submissions through standardized data structures such as SDTM and ADaM. These activities remain essential to clinical drug development and regulatory compliance. However, they have traditionally been viewed as operational contributions, with limited direct involvement in analytical strategy or decision-making.

In recent years, this traditional characterization has been increasingly questioned. As drug development has become more complex, the role of the statistical analyst has been re-examined within statistical programming and analytics functions and across research and development organizations. Questions such as whether statistical analysts can contribute beyond execution, why analytical impact appears to be expanding in modern development programs, and how statistical analysts can further develop analytical capabilities have become central to ongoing discussions. These questions provide motivation for examining how the statistical analyst role continues to evolve in today's drug development landscape.

2.2 CHANGES IN THE DRUG DEVELOPMENT LANDSCAPE

While the fundamental principles of clinical trials such as randomization, control, and prespecified analysis, remain unchanged, the design, conduct, and interpretation of clinical trials have evolved substantially. Advances in scientific understanding, increasing use of complex and heterogeneous data sources, and heightened regulatory expectations have introduced new analytical challenges and increased the need for closer integration between data generation, analysis, and decision-making across the clinical development lifecycle.

Regulatory and methodological evolution

Health authorities, including the FDA, EMA, PMDA, and CDE, have increasingly emphasized early and continuous engagement with sponsors throughout clinical development, particularly with respect to study design and overall analysis strategy. Regulatory guidance and review practices now place strong emphasis from data collection through analysis and reporting. This includes clear alignment between study objectives, endpoint definitions, estimands, primary statistical analyses, and the selection of sensitivity and subgroup analyses to support robustness and interpretability of results. In addition, increasing attention is given to how analytical results are communicated, both within regulatory submissions and during interactions with health authorities.

This regulatory focus requires analytical considerations to be addressed much earlier in the study lifecycle, rather than being confined to the reporting stage. As a result, statistical analysts are increasingly expected to be familiar with relevant health authority guidance, understand the analytical implications of regulatory expectations, and support analysis planning and implementation in ways that facilitate effective regulatory review and decision-making.

In parallel, innovative trial designs such as adaptive designs, basket trials, umbrella trials, platform studies, and continuous trials have become more common, particularly in oncology and rare diseases. These designs offer increased efficiency and flexibility but also introduce additional analytical and operational complexity.

Effective implementation of these novel designs depends not only on sound statistical methodology but also on precise data structuring, timely data availability, real-time data monitoring, and close collaboration across clinical, statistical, and operational functions. In practice, these requirements increase the need for statistical analysts to develop a working understanding of innovative trial methodologies, enabling them to translate complex design concepts into implementable analytical workflows and to support consistent execution across the study lifecycle.

New data sources

Beyond traditional clinical trial data, drug development increasingly incorporates real-world evidence (RWE), external control arms, digital health technologies, and patient-reported outcomes captured through electronic platforms, imaging data, and high-dimensional laboratory and genomic data. The introduction of these heterogeneous data sources has substantially increased demands on data integration, standardization, and analytical rigor, contributing to greater overall analytical complexity.

Precision medicine, for instance, has become an important development pathway, particularly in oncology targeted therapy. And biomarker, genomic, and companion diagnostic data represent a particularly impactful evolution within this broader data landscape. Patient populations are increasingly defined by molecular, genomic, or biomarker characteristics rather than by clinical features alone. Companion diagnostics are often developed in parallel with therapeutic products, requiring close coordination between drug development and assay development activities. As a result, biomarker data, assay outputs, and diagnostic classifications must be accurately linked with clinical endpoints and maintained with traceable data lineage suitable for regulatory review. From an analytical perspective, high-dimensional biomarker and genomic data introduce additional complexity related to missingness, imputation strategies, longitudinal evaluation, and assessment of associations with clinical endpoints. Addressing these complexities may require detailed exploratory analyses to inform confirmatory analysis strategies and ensure interpretability of results.

Technologies and Artificial Intelligence (AI)

Technologies have reshaped analytical workflows. Increased automation of data processing, widespread adoption of open-source programming environments, and emerging applications of artificial intelligence have enabled more efficient data handling, data visualization, and reporting. More recently, AI-assisted programming tools have begun to influence how code is developed, reviewed, and maintained. These tools can support code generation, debugging, refactoring, and documentation, reducing time spent on repetitive or boilerplate programming tasks.

As a result, analytical work has increasingly shifted away from primarily manual and repetitive programming activities toward higher-level analytical problem solving and greater involvement in analysis planning and analytical strategy. At the same time, the use of AI-assisted programming introduces new requirements for analytical judgment, including verification of code correctness, assessment of methodological appropriateness, validation of outputs, and consistency with prespecified analysis plans and regulatory expectations. In regulated settings, analysts must ensure that AI-generated or AI-assisted code is transparent, reproducible, and subject to appropriate review and governance.

Consequently, statistical analysts are increasingly required to combine strong analytical reasoning with technological literacy: understanding not only how to apply analytical tools, but also when and how to use AI-assisted capabilities responsibly. This evolution further reinforces the transition of the SA role from execution-focused programming toward integrated analytical contribution, where domain knowledge, methodological understanding, and interpretive skill are essential for translating analytical results into clear and consistent narratives for cross-functional and regulatory stakeholders.

The Statistician role and redistribution of analytical responsibilities

The focus areas of study statisticians have changed substantially over time. Whereas statisticians were historically involved across the full analytical spectrum, their role has increasingly concentrated on the development and application of advanced statistical methodologies, statistical innovations, and theoretical efforts. Examples include, but are not limited to, Bayesian approaches to sample size determination, sophisticated statistical modeling, adaptive and innovative trial designs, real-world evidence (RWE) methodologies, and predictive modeling to support trial execution and decision-making. More recently, engagement with artificial intelligence (AI) and advanced computational approaches has further expanded the methodological scope of the statistician role.

At the same time, statisticians' direct involvement in certain downstream in data-driven and analytical exploratory activities has decreased in many development settings. These activities include detailed data exploration, hands-on investigation of data anomalies, integration of new or non-traditional data sources into clinical trial databases, and analytical support for medical affairs or exploratory analyses. In practice, many of these tasks have been partially or fully absorbed by statistical analysts as part of routine analytical workflows.

This redistribution of responsibilities has occurred largely in response to practical needs. By virtue of their continuous proximity to clinical trial data throughout the study lifecycle, statistical analysts are often well positioned to support data-driven exploration and analytical implementation. Their responsibilities increasingly include assuring completeness of data collection, assessing data quality and consistency, integrating diverse data sources, conducting exploratory analyses, and supporting elements of analysis planning from an implementation perspective. Consequently, analysis execution by statistical analysts now frequently requires analytical judgment, scientific understanding, and familiarity with statistical concepts that extend beyond traditional TFL reporting.

Statistical training and capability of SA workforce

The expanding analytical responsibilities observed among statistical analysts are further supported by changes in workforce composition and hiring practices within the pharmaceutical industry. Unlike earlier periods in which statistical analyst roles were primarily filled by candidates with computer science or programming-focused backgrounds, many organizations now actively recruit statistical analysts with formal training in statistics or related quantitative disciplines.

In the Toronto, Ontario, Canada region, for example, industry hiring patterns across several large pharmaceutical companies and contract research organizations indicate that more than one third of SA new hires in the past decade possess formal education in statistics or closely related fields. Job descriptions increasingly emphasize statistical knowledge, understanding of clinical research methodology, and the ability to engage in analytical reasoning, rather than focusing exclusively on programming proficiency.

This shift in hiring reflects changing expectations for the SA role. Statistical analysts with formal statistical training are better equipped to understand the rationale behind analysis plans, engage meaningfully in discussions of methodology, and assess the appropriateness of analytical approaches from a feasibility and implementation perspective. Their statistical foundation enables them to contribute constructively to analysis planning and designing, exploratory investigations, and interpretation of results.

2.3 COMPARISON OF SA ROLE ACROSS SETTINGS (ACADEMIA VS. INDUSTRY)

Statistical analyst roles are widely established in academic real-world clinical research settings, including health administrative organizations, hospitals, clinical and epidemiological research institutions, and universities. To better understand differences in analytical responsibilities across environments, we evaluated statistical analysts in real-world research organizations in Canada, including Institute of Clinical Evaluative Sciences (ICES), Cancer Care Ontario (CO), Canadian Institute for Health Information (CIHI), and University Health Network (UHN). We compared activity profiles for statistical analysts in the pharmaceutical industry, statistical analysts in real-world research settings, and statisticians in the pharmaceutical industry.

The comparison spans activities across study design, data-related tasks, analysis strategy and execution, reporting, decision-making, programming and analytical tools, health authority (HA) interaction and submission, and collaboration with strategic stakeholders. The left-hand side of Figure 1 lists specific activities within each category. For comparison purposes, each activity was scored on a relative scale, with a maximum value of 10 indicating full responsibility for the activity and a minimum value of zero indicating no responsibility.

As illustrated in Figure 1, statistical analysts in real-world settings and statisticians in the pharmaceutical industry show high involvement in protocol development with study design, analysis strategy planning, data collection and exploration, and decision-making activities. These include result communication, interpretation, clinical study report (CSR) review, and publication-related work, as well as close collaboration with scientific stakeholders. In contrast, statistical analysts in the pharmaceutical industry demonstrate comparatively lower involvement in certain upstream design and decision-oriented activities, reflecting the structured and role-specialized nature of regulated development environments.

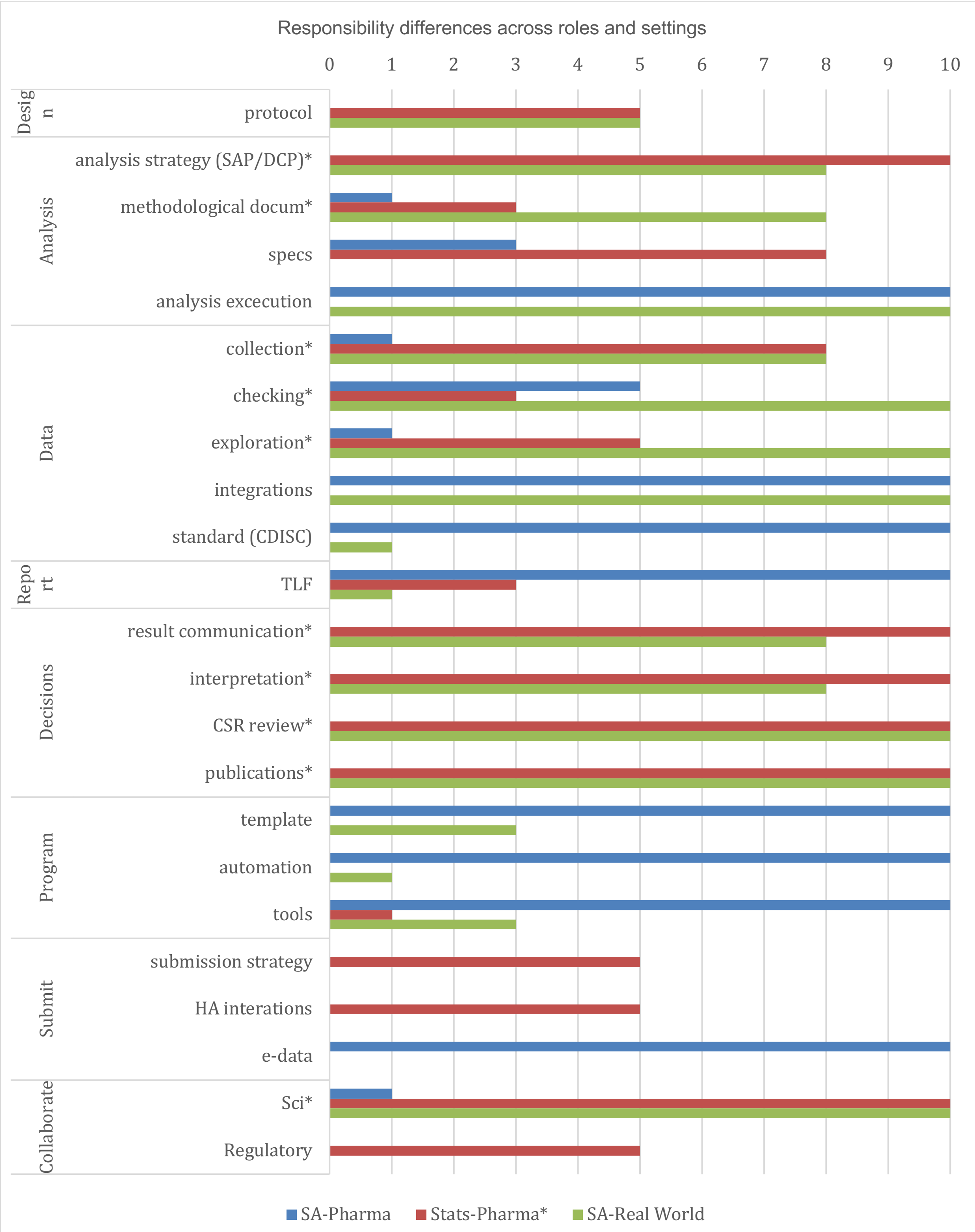
Across analysis execution and data integration activities, both statistical analysts in real-world settings and statistical analysts in the pharmaceutical industry assume full responsibility. Statistical analysts in real-world settings also exhibit substantially greater involvement in data checking and validation activities compared with statisticians in the pharmaceutical industry, consistent with the less standardized and more heterogeneous nature of real-world data sources.

Clear role distinctions are observed in regulatory-facing activities. Only statisticians in the pharmaceutical industry are consistently involved in health authority interactions, submission strategy development, and collaboration with regulatory functions. Statistical analysts in both industry and real-world settings typically do not assume primary responsibility for these activities. In addition, statistical analysts in real-world settings generally do not develop detailed analysis specifications, which are routinely produced by statisticians and statistical analysts in the pharmaceutical industry. Instead, real-world SAs often create and maintain methodological analysis documents that support exploratory and hypothesis-generating research.

Conversely, statistical analysts in the pharmaceutical industry demonstrate unique and substantial responsibility for standardized analytical deliverables and programming-intensive activities. These include data standardization, TFL reporting, electronic data submission, and extensive development of programming templates, automation processes, and analytical tools. These responsibilities reflect the operational scale, regulatory requirements, and efficiency demands of pharmaceutical development programs.

In summary, statistical analysts in real-world settings combine responsibilities typically associated with both statistical analysts and statisticians in the pharmaceutical industry. Their roles encompass study design, analysis strategy development and execution, data-related activities, and decision-support functions, while involving comparatively less emphasis on standardized programming and reporting. In this respect, the real-world SA role reflects emerging trends in the evolution of statistical analyst responsibilities within the pharmaceutical industry, where expanded analytical scope and deeper engagement with data and decision-making are increasingly observed.

Figure 1: Responsibility differences in SA-Pharma, Statistician-Pharma, SA-Real World.



2.4 ENABLING STRATEGIC ANALYTICAL IMPACT IN THE STATISTICAL ANALYST ROLE

As described in earlier sections, the drug development landscape has evolved substantially due to regulatory and methodological advances, increasingly complex data sources, rapid technological and AI-enabled innovations, redistribution of analytical responsibilities, and changes in the training and experience of the statistical analyst workforce. These changes have created new expectations for statistical analysts that extend beyond traditional execution-focused responsibilities.

At the same time, the statistical analyst role in real-world research settings reflects emerging trends relevant to the pharmaceutical industry, particularly the emphasis on the methodology–data–analysis–decision continuum. These trends highlight the importance of analytical reasoning, data-driven insight, and interpretation alongside technical execution.

Therefore, strategic analytical impact can be understood as an outcome of capability development rather than formal role definition. This section outlines key capability areas that enable statistical analysts to contribute meaningfully to strategic analytical impact within clinical drug development.

Proximity to data as an analytical enabler

Statistical analysts occupy a unique position at the intersection of data collection, standardization, transformation, and analysis. Continuous proximity to raw data, derived datasets, and data lineage provides data insights with how data are generated, structured, and transformed across the study lifecycle. This proximity also enables early identification of data quality or completeness, inconsistencies across data domains or data sources, and limitations in data capture.

Close interaction with data allows statistical analysts to conduct systematic data exploration, assess distributions and relationships across data domains, and evaluate patterns of missingness and sparsity. These activities support identification of data limitations, feasibility of planned analyses, and appropriateness of analytical assumptions. In addition, examination of intermediate outputs, alternative summaries, and graphical displays during analysis execution enables assessment of result plausibility, detection of unexpected patterns, and identification of questions that need further investigation or clarification.

These data-driven activities inform analysis planning and analytical strategy. Proximity to data thus serves as a critical enabler of analytical reasoning and supports alignment between data, analysis, and interpretation.

Expanding analytical scope beyond execution

Traditional statistical analyst responsibilities have focused on implementing prespecified analyses, generating standardized datasets, and producing tables, figures, and listings according to approved statistical analysis plans. These activities remain essential but are no longer sufficient on their own to support the analytical demands. To enable broader analytical impact, statistical analysts increasingly engage in activities such as:

- Developing a clear understanding of analytical intent and scientific rationale, beyond technical specifications
- Contributing to analysis planning discussions by assessing feasibility, data readiness, and implementation considerations
- Conducting exploratory and descriptive analyses that inform endpoint definitions, subgroup evaluation, and interpretation
- Reviewing analytical outputs critically to assess consistency, plausibility, and clinical relevance

These activities ensure that analytical execution remains closely aligned with scientific objectives and methodological intent, thereby strengthening the contribution of analytics to strategic decision-making.

Cross-functional collaboration with strategic stakeholders

Strategic impact in drug development is inherently cross-functional. For statistical analysts, expanding influence requires sustained collaboration with key stakeholders, including:

- Clinical scientists, to align analytical outputs with clinical questions and interpretation
- Biostatisticians, to translate methodological concepts into implementable analyses
- Regulatory affairs, to understand evidence expectations and submission requirements
- Clinical operations, to understand operational constraints that affect data completeness and timing

Effective collaboration requires more than participation in meetings. It depends on the ability to communicate analytical concepts clearly, ask scientifically informed questions, and translate between technical detail and strategic intent. While these skills have not traditionally been emphasized in statistical analyst training, they are essential for extending analytical contributions into strategic domains.

In summary, the ability of statistical analysts to expand analytical impact depends on individual capability development, organizational support, and opportunities for structured learning and collaboration. The case study presented in the following section describes a pioneering hybrid statistical analyst project in oncology targeted therapy and precision medicine. This initiative was intentionally designed to integrate expanded analytical responsibilities, deep data engagement, and cross-functional collaboration, illustrating how strategic analytical impact can be enabled when statistical analysts are empowered to operate beyond traditional execution and when appropriate conditions and governance are in place.

3. CASE STUDY: A PIONEERED HYBRID SA PROJECT IN TARGET THERAPY AND PRECISION MEDICINE ONCOLOGY

To explore how the statistical analyst (SA) role can contribute to more strategic analytical impact in clinical drug development, the Statistical Programming and Analytics (SPA) function initiated an experimental project in 2020. The project was co-designed in collaboration with two other core functions, Statistics and Translational Science Oncology (TSO), to pilot a hybrid SA project with expanded analytical responsibilities under defined governance and oversight.

This experimental initiative was designed to build strategic analytical impact in established statistical, methodological, and regulatory frameworks through an evidence-based approach spanning the methodology-data-analysis-decision cycle. The project created a developmental environment in which selected statistical analysts could progressively develop analytical and methodological capabilities beyond traditional execution responsibilities. Specifically, it explored whether statistical analysts could expand their analytical scope by learning and performing selected study statistician activities within biomarker and companion diagnostic oncology programs. The project focused on integrating analytical execution with deeper engagement in analysis planning, methodological learning, scientific interpretation, and decision support alongside clinical drug development.

3.1 PROJECT DESIGN AND GOVERNANCE FRAMEWORK

This hybrid statistical analyst project was designed as a learning and mentorship-based experiment. A study-based design allowed statistical analysts to engage in higher-level analytical activities while maintaining scientific, methodological, and regulatory expectations in alignment with biomarker and CDx alongside drug development strategy. Key elements of the project design are summarized below:

Cross-functional co-design and oversight

- Primary mentorship by senior statisticians, who maintained accountability for statistical methodology, inferential strategy, and regulatory-facing statistical decisions across studies and multiple compounds.
- Secondary mentorship by clinical scientists and regulatory affairs partners, providing guidance on biomarker strategy, companion diagnostic development, and health authority expectations.
- Defined scope of responsibility, ensuring that expanded activities were focused on analytical, scientific, and regulatory learning objectives.

Candidate selection criteria

Candidate selection was restrictive to ensure readiness for expanded analytical responsibilities.

- Formal education in statistics or a closely related training in quantitative approach (Master's degree or equivalent)
- A minimum of five years of experience in late-phase pharmaceutical drug development
- One to two years of experience supporting biomarker and/or companion diagnostic studies

Study-based design rather than time-based rotation

The project was study-based rather than duration-based. Each candidate was assigned to one or more biomarker or CDx studies, with learning objectives tied to specific analytical, scientific, and regulatory challenges rather than a fixed training timeline. This approach ensured that learning was contextual, outcome-driven, and completeness of development strategy cycle.

3.2 LEARNING OBJECTIVES AND CAPABILITY DEVELOPMENT

The hybrid statistical analyst project defined explicit learning objectives across three dimensions: data-driven operations, statistical and analysis strategy operations, and scientific and regulatory knowledge development. These objectives were aligned with the gaps and opportunities identified between SA and strategic stakeholders as discussed in earlier sections.

Data-driven analytical operations

Learning objectives in data-driven operations focused on expanding analytical engagement beyond standard reporting activities. These activities emphasized analytical reasoning and data understanding rather than programming efficiency alone and included.

- Sample selection strategies for biomarker-defined populations
- Exploratory data analysis of complex and high-dimensional biomarker and genomic data
- Integration of heterogeneous data sources, including clinical, genomic and genomic annotation, and companion diagnostic data
- Development and optimization of analysis-ready datasets to support exploratory and confirmatory objectives

Statistical operations and analysis planning

In the statistical operations domain, learning objectives included:

- Participation in analysis strategy planning at the study level
- Contribution to the statistics section in protocols
- Development of statistical analysis plans (SAPs) and analysis specifications
- Contribution to statistical planning and approaches across studies and compounds
- Engagement in result interpretation and communication within clinical study reports (CSR), including CSR review
- Involvement of submission development and understanding of submission strategy from regulatory perspective

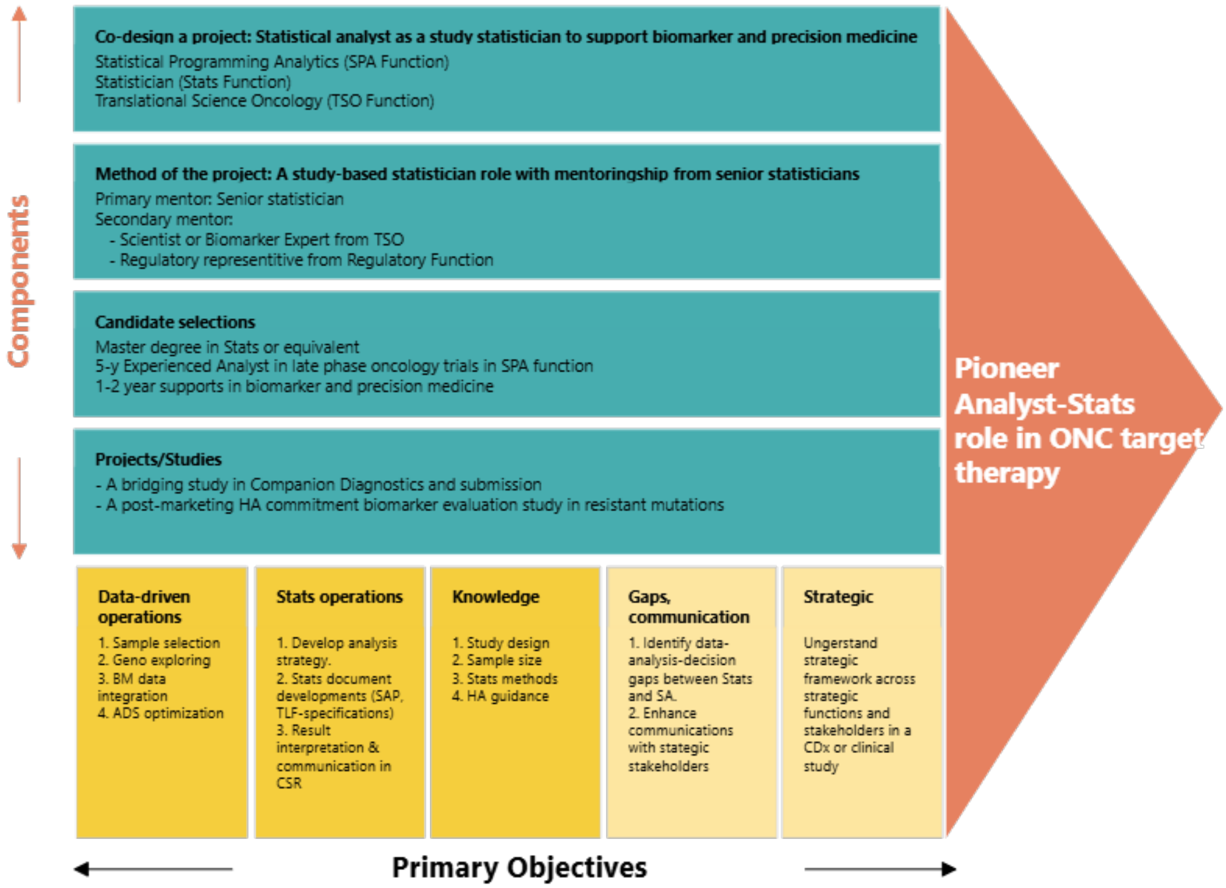
Scientific, regulatory, and methodological knowledge development

Knowledge learning objectives addressed gaps between analysis execution and scientific decision-making and included:

- Sample size considerations, including binomial-based approaches for biomarker studies
- Bridging study designs and population comparability
- Statistical methods for confounding adjustment, covariate analysis, modeling, and multi-layer variance components
- Handling of missing data and imputation strategies
- Sensitivity analysis planning in biomarker and CDx contexts
- Familiarity with health authority guidance relevant to CDx development

In parallel, the project emphasized identification of gaps across the methodology-data–analysis–decision continuum, as well as development of communication skills with strategic stakeholders. Particularly, attention was given to articulating analytical insights using narrative and descriptive language, rather than purely statistical, programming, or data focused terminology. Through these activities, candidates explored how statistical analysts could contribute to strategic discussions by acting as a strategic analytical partner in specific drug development domains, while leveraging existing strengths in analytical programming, tools, technology, and reporting.

Figure 2: Design of a hybrid statistical analyst project in target therapy and precision medicine oncology



3.3 PROJECT IMPLEMENTATION AND RESULTS

The hybrid statistical analyst project was implemented progressively between 2021 and 2025 through a series of biomarker and companion diagnostic assignments in two oncology target therapy compounds. During this period, selected statistical analysts participated in precision medicine oncology programs under senior statistician mentorship, with responsibilities aligned to the learning objectives described previously.

<u>*Candidate participation*</u>

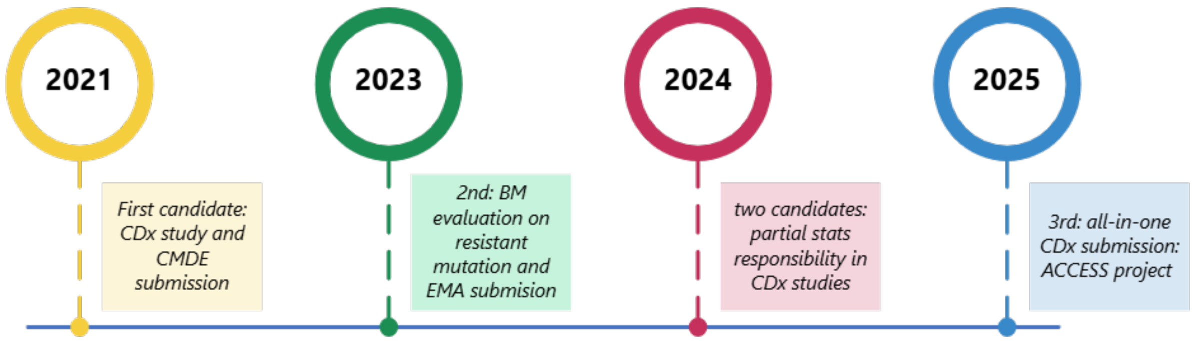
The first candidate started in 2021 and was assigned to a companion diagnostic study supporting a China Medical Device Evaluation (CMDE) submission. This assignment focused on analytical support for CDx-related data integration, sample definition, and analysis planning in parallel with clinical development activities.

The second candidate enrolled in the project in 2023 and supported biomarker evaluation for resistance mutation analysis in in precision medicine, contributing to an EMA regulatory commitment submission. This work emphasized exploratory data analysis, sensitivity assessment, and alignment of biomarker findings with clinical outcomes using sophisticated statistical approach.

The third candidate joined the project in 2025 and contributed to the development of a companion diagnostic submission platform, with a focus on scalable analytical workflows, data integration, and consistency across studies and regulatory regions.

In addition to these three primary assignments, hybrid statistical analysts contributed to total 12 studies in two oncology targeted therapies between 2021 and 2025, spanning different development phases.

Figure 3: Candidate participation diagram



Analytical and statistical deliverables

Across the implementation period, hybrid statistical analysts contributed to a broad range of analytical and statistical deliverables:

- Statistical analysis plan (SAP): Hybrid SAs supported development of 12 study-level SAPs related to biomarker and companion diagnostic evaluations. Contributions included analysis strategy input, alignment of analytical objectives with protocol and health authority expectations, development of analysis specification, and supports for analysis standardization across compounds under senior statistician mentorship.
- Data-driven analytical support: SAs supported bio-sample identification and selection for retrospective genomic sequencing testing, explored longitudinal analyses, interpreted complex genomic and diagnostic data. These activities included harmonization and integration of clinical, biomarker, and CDx data sources, as well as optimization of analysis-ready datasets for downstream use.
- Regulatory submissions: Hybrid SAs contributed regulatory submissions across multiple regions, including FDA, CMDE, EMA, and PMDA. In addition to analytical preparation supporting CSRs, contributions included participation in submission strategy discussions, review of submission documents beyond statistical and data sections, and collaborations with partner diagnostic companies during submission preparation.

Methodological and scientific learning outcomes

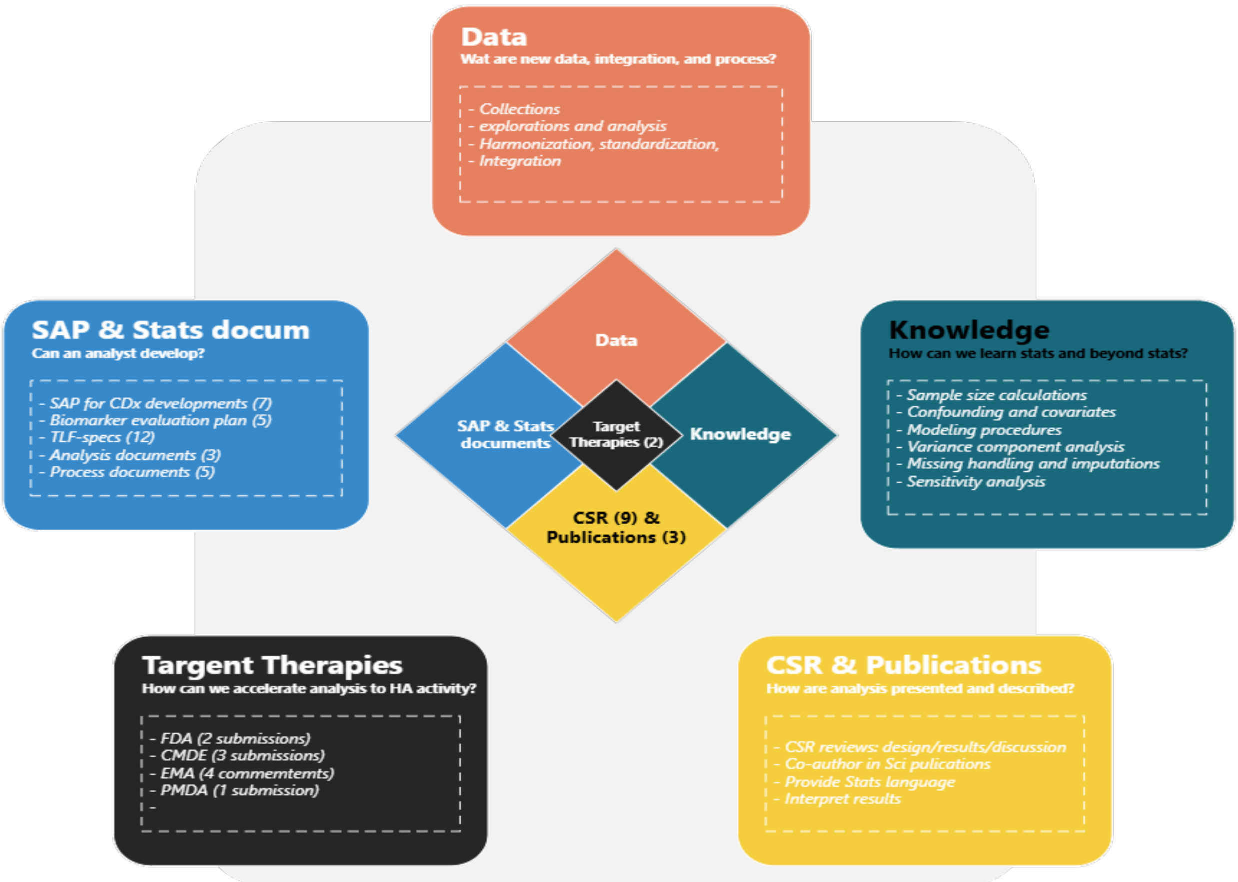
Participation in the project enabled hybrid statistical analysts to develop practical methodological knowledge across several key areas:

- Sample size calculations for biomarker-driven studies, including binomial-based approaches
- Evaluation of confounding and covariate effects
- Application of statistical modeling for longitudinal data and variance component analysis
- Handling of missing data and implementation of appropriate imputation strategies
- Planning and execution of sensitivity analyses to support robustness of primary analysis results
- Increased familiarity with biomarker-driven study designs, companion diagnostic development pathways, and health authority guidance relevant to CDx evaluation and submission

Result interpretation, and scientific communication / publications

Hybrid statistical analysts extended their involvements into result interpretation and publication activities. Contributions included participation in clinical study report (CSR) review, drafting and refinement of statistical language in study protocols and CSRs, and support for interpretation of analytical results in collaboration with senior statisticians and clinical scientists.
In addition, hybrid SAs served as co-authors on three scientific publications, contributing to analytical content, interpretation of results, and alignment of analytical findings with scientific narratives.

Figure 4.1: Results in analytical and statistical deliverables, methodological and scientific learning, and result interpretation and scientific communication.

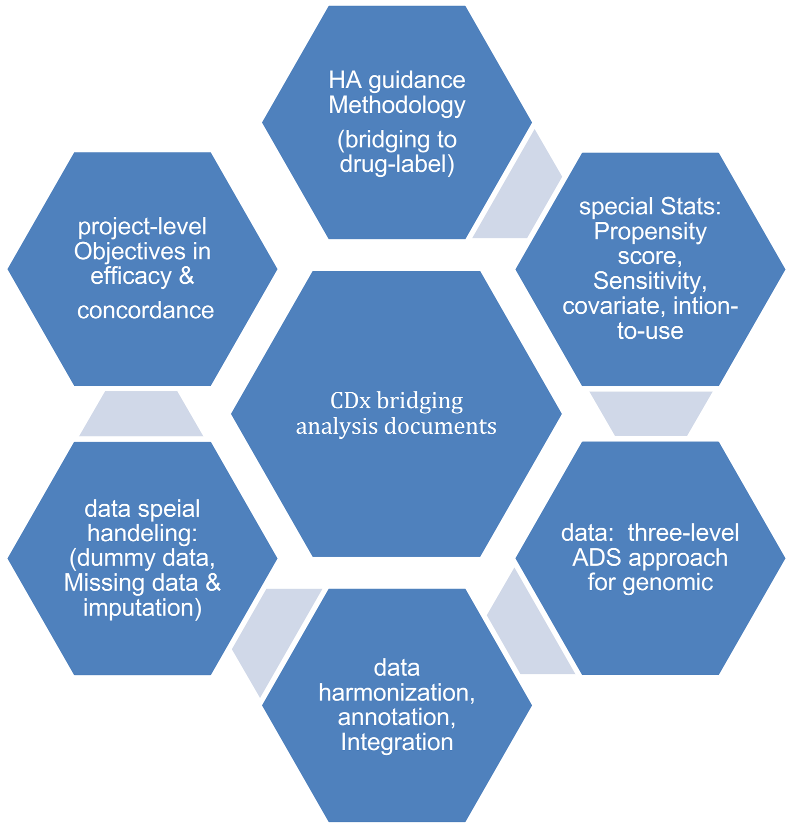


Analytical documentation creation

In addition to SAPs, hybrid SAs developed and maintained a comprehensive set of analytical documentation to support biomarker and CDx evaluations. These materials included program/project-level analytical and statistical frameworks for CDx development; guidance for creation of dummy CDx analysis datasets; structured approaches for identification of confounding factors and categorization of covariates; analytical summary for biomarker change-from-baseline evaluation; implementation guidance of propensity score analysis; harmonization approaches for genomic data; and methodological guidance for handling missing data using multiple imputation and bootstrap methods. Additional documentation summarized key principles for CDx bridging study design and relevant health authority guidance for CDx submissions.

Figure 4.2 shows analytical documents developed to support CDx bridging study analysis. These documents include six categories as genomic data standardization, three-level ADS approach, special data handling for dummy data creation and missing imputation, analysis to support objectives, and health authority guidance and methodology summary. These documents facilitated across function alignment, provided consistent statistical and methodological reference across compounds and studies, and supported training, knowledge transfer, and continuity within the hybrid SA project.

Figure 4.2 Analytical documents supporting CDx bridging study analysis



4. OUTCOMES: BRIDGING ANALYSIS TO STRATEGY

The primary objectives of the project were to assess the feasibility of a pilot hybrid SA role expansion, characterize learning outcomes for participating statistical analysts, and support the development of practical knowledge and analytical skills. The hybrid statistical analyst project demonstrated that statistical analysts can extend their analytical contributions in ways that meaningfully support strategic decision-making, particularly in biomarkers and companion diagnostic to support precision medicine within oncology drug developments. Observed outcomes illustrate how statistical analysts can function as analytical partners through:

- Enhanced data-dive analytical contribution
- Expanded involvement in analysis planning, analysis strategy,
- Improved interpretation and communication of analytical results
- Strengthened cross-functional analytical collaboration with strategic stakeholders

Developing extended strategic analytical mindsets through analytics

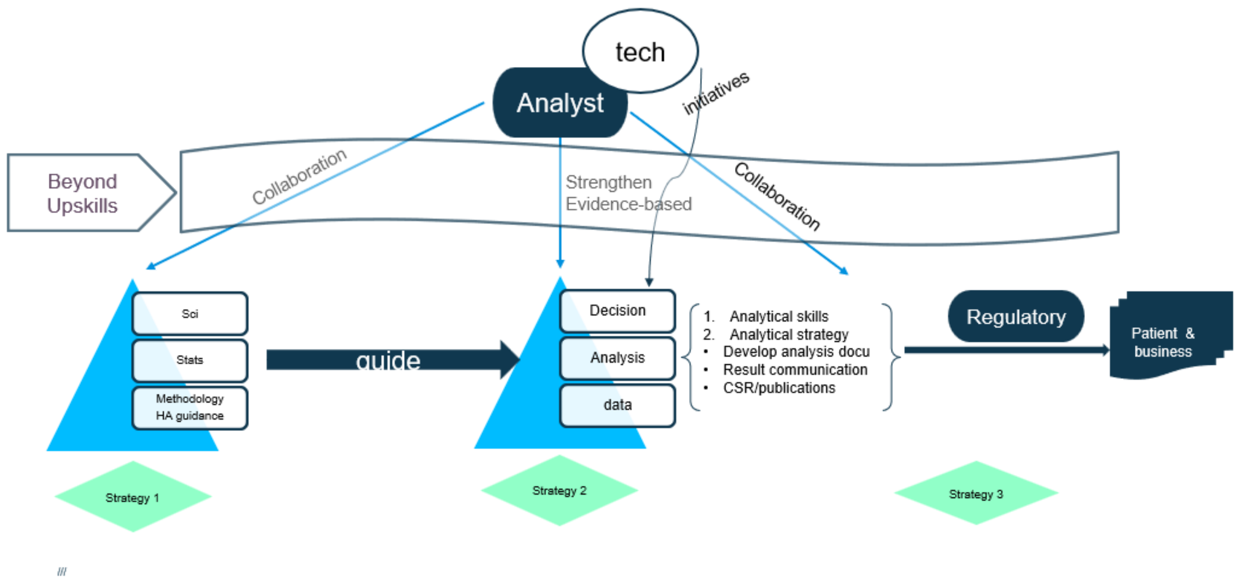
Traditionally, statistical analysts have considered strategy primarily in data processing, analysis execution, and TFL reporting. In regulatory and submission activities, involvement has typically been limited to e-data packages and data related documentation. Figure 5 presents an integrated view of three interconnected strategy domains in Research and Development of clinical drug development. Statistical analysts in the future should consider developing extended strategic analytical mindsets by using analytics as a bridge across all three domains.

The first strategy domain represents scientific, statistical, methodological, and health authority guidance. This domain defines the guiding context for development programs and shapes what evidence is required and how it should be generated. Historically, statistical analysts have had limited direct engagement with this upstream strategy space beyond implementation of prespecified analyses.

While statistical analysts are most deeply involved in the second strategy domain focusing on the data–analysis–decision foundation, we used to emphasize analytical execution and TLF deliverables. Through proximity to data and continuous analytical engagement such as data exploration, analytical planning and strategy, and result communications, statistical analysts are uniquely positioned to strengthen this domain and to use analytics as a mechanism for connecting upstream guidance with further decision-making.

The third strategy domain encompasses regulatory and submission strategy, including pivotal trial submissions, cross–health authority considerations, indication and lifecycle decisions, and patient and business impact. Direct involvement of statistical analysts in this domain has historically been limited. The hybrid statistical analyst project demonstrated that familiarity with the full set of regulatory submission materials that are beyond e-data packages alone, and understanding overall submission strategies across health authorities and lifecycle stages can support more efficient analytical execution and better aligned reporting planning.

Figure 5: Three interconnected strategy domains in R&D of clinical drug development.



Methodology–Data–Analysis–Decision Foundation

With an extended strategic analytical mindset, statistical analysts are better positioned to contribute across methodology-data-analysis-decision evidence-based continuum when they develop deeper analytical skills, extend beyond knowledge in methodology, and gain stretched familiarity with regulatory and scientific frameworks.

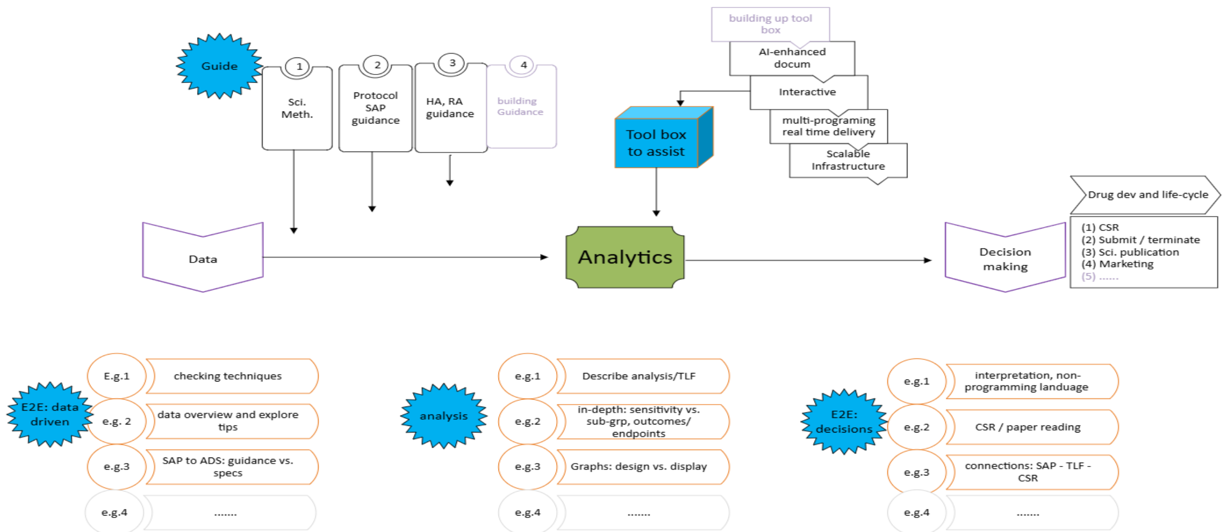
Figure 6 illustrates a methodology–data–analysis–decision foundation that emerged from the hybrid statistical analyst project and provides a framework for understanding how analytical capability can support strategic decision-making. The framework comprises five interrelated components: Guide, Data, Analysis, Decision, and Toolbox/Technology. These components describe how statistical analysts can integrate analytical execution, methodological reasoning, and decision support in complex development settings.

- Guide: The Guide component represents the scientific, methodological, and health authority guidance that frames analytical work. Exposure to this guiding context enables statistical analysts to better understand why specific analyses were required, how analytical objectives aligned with scientific hypotheses, and how regulatory expectations influenced analytical approaches.
- Data: The Data component reflects the central role of clinical trials as the foundation of evidence generation. Proximity to the data enables statistical analysts to translate data-driven insights into analysis strategy, planning, and execution.
- Analysis: The Analysis component represents the transformation of data into interpretable analytical outputs. Expanding analytical scope from execution toward analytical reasoning, contributing to analysis planning, evaluation of alternative analytical approaches, and interpretation of results enable statistical analysts to engage more effectively in further decision-making.
- Decision: The Decision component captures how analytical outputs support decisions. Through experience in result interpretation, CSR and publication review, and scientific communication, statistical analysts gain insight into how analytical results are synthesized into evidence that informs clinical and regulatory decision-making.

- Technology: The Toolbox or Technology component includes programming environments, automation, visualization tools, and emerging AI-based capabilities that support analytical work. By leveraging strengths across the other four components, statistical analysts can identify use cases by themselves and deploy technology in a targeted manner to support the full methodology–data–analysis–decision continuum.

These five components form an integrated foundation that connects methodology, data, analysis, and decision-making through enabling technology to support statistical analysts to contribute to strategic discussions.

Figure 6: Components of methodology–data–analysis–decision foundation enables SA being a strategic partner in drug development



<u>Statistical Analyst Capability Development in Pharmaceutical Research</u>

As clinical drug development continues to evolve toward greater complexity, statistical analysts in the pharmaceutical industry are increasingly expected to develop expanded analytical skills and capabilities. These expectations align closely with the analytical profiles commonly observed for statistical analysts in real-world research settings, particularly with respect to engagement across the methodology–data–analysis–decision foundation described in the previous section. Developing such capabilities enables statistical analysts to more effectively bridge data and analysis with strategy and decision-making in advanced drug development contexts.

Instead of prescribing specific responsibilities, the five-component framework particularly methodology–data–analysis–decision foundation illustrated in Figure 6 can be used as a reflective tool to assess analytical readiness and guide capability development. When statistical analysts seek to expand their analytical contributions, the following questions may be asked as examples but are not intended to cover all considerations.

Data-driven analytical foundation:

- How do we examine relationships within and across clinical trial data domains?
- Do we systematically assess missing data patterns before deriving analysis variables?
- What data checking, overviewing, and exploratory techniques are most informative at different stages of development?
- How do we assess and support completeness of planned data collection?
- Can we explore new or external data sources and develop optimized approaches to integrate them into existing clinical trial databases?

Analysis and analytical strategy:

- Can we clearly describe the study SAP, including its analytical objectives and rationale?
- Can we articulate the analyses that are implemented and explain how they address study questions?
- Can we describe the purpose and interpretation of key tables and figures?
- Can we share in-depth analytical knowledge related to specific endpoints, health outcomes, novel study designs, or uncommon analytical approaches?
- Can we define and negotiate top-line result TFLs based on analytical objectives rather than template defaults?
- Can we design meaningful graphical data displays rather than focusing solely on their technical creation?

Methodological guidance and analytical continuity:

- Do we develop analytical guidance documents to support consistent analysis data development across

studies, compounds, or therapeutic areas?
- Can we develop analysis datasets based on SAP and analytical guidance and intent, rather than relying solely on prescriptive specifications?
- Can we provide analytical overviews that support data cutoff planning, CSR development, and submission preparation?

Decision support and regulatory strategies:
- What are the methodological expectations of health authorities in specific therapeutic or scientific contexts?
- Which TFLs are included in clinical study reports and scientific publications, and what analytical questions do they address?
- What discussions, interpretations, and conclusions are ultimately presented in CSRs and publications, and how are analytical results translated into decision-relevant narratives?

Profound, decision-oriented questions:

❖ Health authority and evidence strategy

  - What are the key health authority considerations for an indication-agnostic development strategy (e.g., defining the target population, strength of evidence across tumor types, and expectations for confirmatory evidence)?
  - What are typical regulatory expectations for tumor-agnostic claims regarding consistency of efficacy effect across tumors?
  - Under what conditions might a successful Phase II study support multiple settings (e.g., first-line and second-line) or support regulatory decisions, and what additional evidence is typically needed?
  - When could ORR be considered sufficient for an accelerated pathway, and what factors would drive the need for Phase III confirmatory evidence?

❖ Endpoints, efficacy evaluation, and interpretation

  - For a composite endpoint (e.g., PFS; or composite cardiovascular events such as MI and stroke), what analyses are required to ensure interpretability (e.g., component-wise analyses, time-to-event methodology, competing risks considerations, and sensitivity analyses)?
  - What additional analyses may be necessary to demonstrate robustness when composite endpoint components differ in clinical importance or frequency?
  - What is an appropriate minimum follow-up duration to evaluate efficacy for a given oncology setting and endpoint, and how should maturity of data be assessed and communicated?
  - How do follow-up requirements differ by endpoint type (e.g., ORR vs PFS vs OS) and treatment line?
  - What constitutes a clinically meaningful and statistically interpretable objective response rate (ORR) in a specific disease context, and what thresholds (if any) are typically considered meaningful for decision-making?
  - How should uncertainty (e.g., confidence intervals), durability of response, and response assessment criteria be incorporated into interpretation?

Answers for these questions may not be obtained uniformly across programs and projects. However, these questions illustrate the types of methodology, endpoint, and regulatory considerations that statistical analysts may engage with as their analytical capabilities expand.

5. ANALYTICS IMPLICATIONS

To illustrate how analytical capability can be strengthened in practice, this section introduces brief examples in three areas.

Analytical Skills: data exploration mindset beyond programming execution

Data exploration often differs between analysts focused primarily on programming execution and those with a stronger analytical or statistical mindset, such as statisticians or statistical analysts working in real-world research settings. For example, programmers commonly explore data using filters or subsets within programming languages such as SAS, R, or Python, focusing on specific records or conditions to debug or validate code. While this approach is effective for targeted checks, it may limit visibility of the full data structure. In contrast, statisticians and SA-real-world analysts more frequently adopt multi-way cross-tabulation to obtain a global view of data structure, relationships, and sparsity. This approach displays one row per observed combination of variables along with frequency counts, enabling rapid identification of structural issues and data patterns.

A simple SAS example code of multi-way cross-tabulation is shown below and the Table 1 summarizes what statistical analysts can explore data by using multi-way cross-tabulation.

```
Proc freq data=dsn;
    Tables var1*var2*var3/list missing;
Run;
```

Table 1: Examples of multi-way cross-tabulation.

Exploring/checking purpose	Examples	What to look?
Structural data validation, data missing diagnosis	Treatment*visit*parameter	• Impossible combinations appear • Expected combinations are missing
Data error identifications	VISIT*ANALYSIS_FLAG*AVAL	Error: Analysis flag = Y but AVAL missing
Data sparsity & feasibility	TRT*composite-event*event-component	• Cell with small count • Combinations never occur
Data balance and distribution	TRT*SITE*RANDOMIZED	• Severe imbalances • Randomization issues

Recognizing and learning and applying exploratory techniques such as multi-way cross-tabulation helps statistical analysts build stronger analytical skills beyond programming execution, drawing on analytical practices commonly used by statisticians and statistical analysts in real-world research settings.

Analysis-Focused Documentation: Building Analytical Strategy Through Documentation

A second analytical implication is analysis-focused documentation creation which is a mechanism for analytical learning, continuity, and strategy development.

Traditionally, analytical documentation in clinical trials has focused on specifications for dataset derivation and TLF production, with an emphasis on implementation and reproducibility. While these documents remain essential, they often provide limited visibility into analytical intent, methodological rationale, or decision context. In contrast,
As illustrated in the previous section for biomarker and CDx programs, statistical analysts developed and maintained analysis-focused documentation such as project level analytical frameworks, guidance for biomarker and genomic data harmonization, structured approaches to confounding and covariate categorization, and principles for bridging study design. These documents outlined how analytical questions were framed, why specific analytical approaches were selected, and how analytical consistency was maintained across studies and compounds.

Creating analytical documentation may function as a learning process. Translating scientific and methodological concepts into written analytical guidance required statistical analysts to engage deeply with analysis strategy. Documentation thus becomes a tool for developing analytical strategy thinking, rather than a downstream record of completed work. Analysis-focused documentation can also support analytical continuity and knowledge transfer. these documents enable consistent application of analytical approaches across studies, facilitate onboarding and collaboration, and reduce implicit.

Result interpretation and communication: learning through CSR and publication engagement

Learning result interpretation and communication through clinical study reports and scientific publications is strongly recommended. Through structured reading, review, or contribution to CSRs, statistical analysts can gain insight into how analytical findings are framed within study objectives, how analytical results are contextualized and translated into narratives, how uncertainty and limitations are communicated, and how analytical results support primary and secondary conclusions. This learning experience may improve analytical capability beyond generating tables and figures by understanding which results are emphasized, how consistency across analyses is demonstrated, and how analytical evidence is aligned with scientific hypotheses and regulatory expectations. Importantly, this learning experience strengthens communication skills by encouraging statistical analysts to describe analytical findings using descriptive and interpretive language rather than technical or programming terminology alone. This shift supports more effective collaboration with cross-functional stakeholders and enhances the ability to communicate analytical insight in a decision-relevant manner.

6. CONCLUSION

The evolving complexity of clinical drug development driven by regulatory and methodological advances, novel data sources, redistribution of analytical responsibilities, and rapid technological innovation, has reshaped how analytical evidence is generated and used to support decision-making. In this environment, sstatistical analysts are increasingly expected to extend beyond traditional execution-focused responsibilities.

This paper described an experimental hybrid statistical analyst project designed to explore whether, under structured mentorship and defined governance, statistical analysts could expand their analytical contributions by learning selected study statistician activities in biomarker and companion diagnostic developments. Through a multi-year case study spanning two oncology targeted therapies, the project demonstrated the feasibility of such capability expansion and documented concrete outcomes across analysis planning, data-driven analytical support, analytical documentation creation, regulatory submission support, and scientific communication.

The methodology–data–analysis–decision framework presented in this paper provides a conceptual foundation for statistical analyst to contribute beyond reporting. Through expanded analytical reasoning, systematic data exploration, analysis-focused documentation, and active participation in result interpretation and communication, statistical analysts can extend their analytical impacts to support strategic decision-making.

The reflective questions proposed for statistical analysts offer a practical lens for assessing analytical readiness and identifying areas for growth, particularly as expectations for analytical roles in pharmaceutical development increasingly resemble those observed in real-world research settings. The analytical practices highlighted in this paper including data-driven exploration, analytical documentation, and narrative-focused interpretation are represent transferable analytical capabilities that can be adapted across development programs and organizations as analytical complexity continues to increase.

In conclusion, analytics serves as a critical bridge between data and strategy in modern drug development. By investing in analytical capability development, fostering cross-functional collaboration, and leveraging emerging technologies, organizations can enable statistical analysts to contribute analytical insight that supports decision-making across the clinical development lifecycle.

REFERENCES

[1] Sascha Ahrweiler, Training Statistical Programmers on SAP Review Skills, IS01, PHUSE 2011.
[2] Forbes H. Statistical Programming in the Pharmaceutical Industry: Advancing and Accelerating Drug Development. SAS Gloal Forum 2020; Paper 4648-2020
[3] Lingsong Yun, Hong Zheng, Challenges of Genomic Data in Clinical Trials. DS12, PHUSE US Connect 2025

CONTACT INFORMATION

Author Name: Lingsong Yun
Company: Bayer Inc.
Address: 2920 Matheson Blvd E, Mississauga, ON L4W 5J4, Canada

Work Phone: 1 (416) 571-6495
Email: lingsong.yun@bayer.com
Website: http://www.bayer.ca