Paper ET07

Legacy to Leading Edge: Reinventing Biostats Programming with Technology, AI, and Agile Innovation

Deepthi BM, GSK, Bengaluru, India

ABSTRACT

The traditional era of clinical trials where a single molecule could take 12 to 15 years to reach the market, is rapidly giving way to a new paradigm. Today, the focus is on accelerating development from bench to market to meet urgent patient needs across multiple regions in parallel. This shift is only possible through transformational changes in technology, tools, processes, and talent capability.

In this session, we share how our GCC Biostatistics Programming team has led and driven an ambitious reinvention journey from legacy, manual approaches to an agile, tech-enabled future. We explore how the team has successfully leveraged open-source tools, AI-powered automation, and Agile methodologies to deliver regulatory submissions across multiple countries in record time.

This transformation has not been without challenges. It has demanded a culture of resilience, where experimentation, learning from failure, and continuous reinvention have been key enablers of progress. Through real-world examples, we will highlight how the team navigated complex landscapes, implemented cutting-edge solutions, and helped redefine what's possible in global clinical development.

INTRODUCTION

Clinical development is undergoing a fundamental shift. The traditional model of sequential, functionally siloed execution, once sufficient for predictable submission cycles, is increasingly misaligned with today's reality of parallel regional submissions, compressed timelines, and continuous regulatory readiness. In this environment, speed is no longer a competitive advantage; it is a baseline expectation.

As this landscape evolves, biostatistics programming can no longer operate as a downstream, output-focused function. Instead, it must serve as a strategic enabler of accelerated, insight-driven decision-making, capable of supporting continuous delivery without compromising quality, compliance, or traceability.

This paper describes a real-world transformation journey in which an Agile delivery model, engineering-oriented design, selective automation, and cultural reinvention were combined to enable predictable, high-quality acceleration at scale. The focus is not on specific tools or proprietary systems, but on the principles, delivery patterns, and ways of working that made the transformation successful.

By emphasizing transferable lessons and practical implementation insights, this paper aims to provide a framework that is broadly applicable across organizations, therapeutic areas, and regulatory contexts navigating similar pressures and expectations.

THE TIPPING POINT: WHEN INCREMENTAL OPTIMIZATION WAS NO LONGER ENOUGH

For years, incremental efficiency improvements, template reuse, standardized macros, and refined handoffs, were sufficient to meet delivery expectations. Over time, however, increasing data volume, analytical complexity, and regulatory scrutiny exposed the structural limitations of this approach.

Key pressures included:

- Parallel submissions across multiple regions rather than sequential filings
- Reduced tolerance for late-stage rework and unstable datasets
- Heightened expectations for traceability, transparency, and audit readiness

Traditional programming models, characterized by monolithic dataset development, manual version tracking, and sequential handoffs, struggled to keep pace. Late-cycle changes frequently cascaded into extensive rework, while critical study knowledge often resided with individuals rather than embedded within systems.

At this stage, it became clear that incremental optimization was no longer sufficient. A fundamental reinvention of how programming work was designed, executed, and integrated was required.

OVERVIEW OF THE REINVENTED DELIVERY MODEL

From Sequential Handoffs to Continuous Readiness

The response to these pressures was the deliberate design of a new delivery model optimized for predictable speed with quality at scale. Rather than pursuing isolated process or technology improvements, the reinvention aligned process, technology, and culture around a single delivery objective and a shared operating rhythm.

At the core of this model sat a clear reinvention framework, built on four mutually reinforcing pillars:

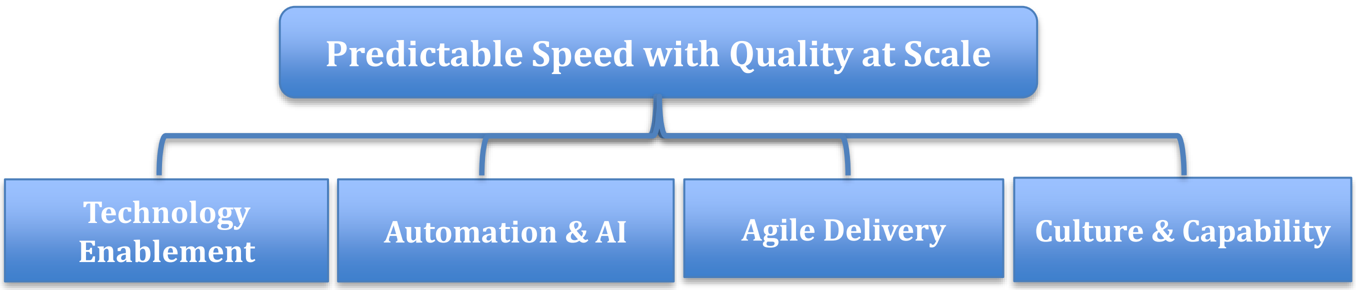

Figure 1: Our re-invention framework

- **Technology Enablement** to enable modular design, interoperability, and transparency
- **Automation & AI** to reduce manual effort, improve consistency, and accelerate insight generation
- **Agile Delivery** to support parallel execution, early integration, and incremental readiness
- **Culture & Capability** to shift accountability from task completion to end-to-end ownership

Together, these pillars provided the foundation for moving from activity-level optimization to the design of an integrated delivery system capable of absorbing change and maintaining continuous readiness.

Work was intentionally structured to progress in parallel, with early integration points replacing late-stage consolidation. Programming activities evolved iteratively alongside protocol development, data maturation, and analysis planning, allowing risks to surface early and reducing downstream rework.

Within this model, programming was repositioned as an end-to-end delivery capability embedded across the study lifecycle. Continuous review replaced episodic validation, and deliverables were maintained in a submission-ready state throughout execution.

As a result, late-stage recovery was replaced by continuous readiness, transforming speed from an exceptional outcome into a predictable characteristic of the delivery system.

TECHNOLOGY ENABLEMENT AND SCALABILITY

Technology enablement played a critical role in operationalizing the reinvention model by providing the foundation for parallel execution, controlled integration, and continuous readiness. Modern, cloud-based platforms replaced fixed, on-premises environments, enabling scalable compute, concurrent access, and faster iteration cycles. This shift reduced operational friction during peak workloads and supported collaborative development across distributed teams.

Version-controlled workflows were central to this enablement, strengthening reproducibility, traceability, and controlled integration, capabilities essential in regulated environments and under compressed timelines.

Within this ecosystem, GitHub served as the backbone for structured, component-based development. It enabled:

- Parallel development through modular, reusable components
- Robust version control supporting controlled change management
- End-to-end traceability from development through delivery
- Efficient search and reuse of existing assets

GitHub repositories were purposefully structured to separate concerns, including study-specific repositories, standard code repositories, and utility repositories, allowing consistency and reuse without constraining study-level flexibility.

In parallel, DOMINO functioned as the scalable data and analytics platform supporting execution. It provided:

- Easy, controlled access to data aligned with evolving data governance policies
- Scalable infrastructure to support intensive computation and concurrent usage
- Seamless integration through multilingual coding and API-based workflows

As a platform for data storage, versioning, and analytics, DOMINO enabled transparent data access while supporting flexible integration with downstream tools and processes.

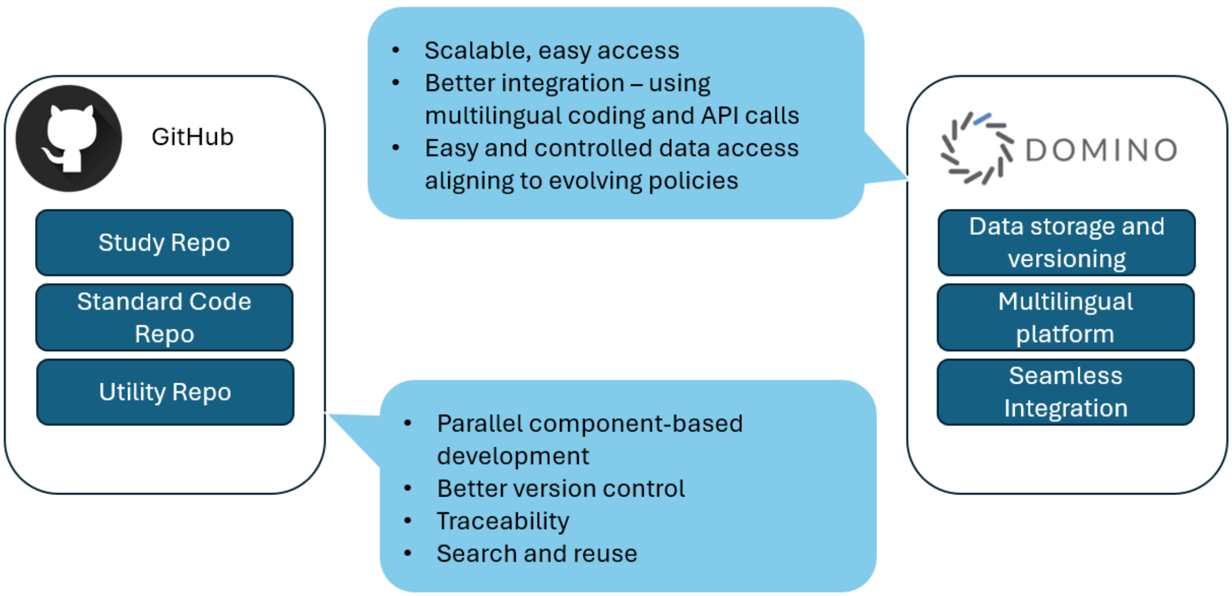

Figure 2: Technology usage to achieve acceleration

Together, these technology capabilities formed an integrated enablement layer that supported the reinvention framework, allowing programming teams to work in parallel, integrate early, and maintain traceability throughout the study lifecycle. Rather than accelerating delivery through late-stage effort, technology enabled predictable speed by design, even under increasingly compressed timelines.

FROM PROGRAMMING OUTPUTS TO ENGINEERING SYSTEMS

Modular, Component-Based Development

A critical enabler of acceleration was the transition from dataset-centric programming to modular, component-based development. Rather than building entire datasets as single units, logic was decomposed into reusable components that could be developed and validated independently.

For example, when multiple outputs required different subsets of variables derived from the same source domain, parallel development was enabled by separating variable derivation into independent modules. These components were developed concurrently and integrated through version-controlled workflows.

This engineering-based approach enabled:

- Parallel development without overwriting work
- Earlier integration and validation

- Reduced dependency on individual programmers
- Improved reuse across studies and indications

As a result, programming teams began to operate more like engineering teams than traditional report generators.

AGILE DELIVERY AS THE OPERATING MODEL

The transformation of biostatistics programming was guided by Agile delivery principles, adopted both at the study team level and within the programming function itself. Agile implementation went beyond simple sprint mechanics, extending to how work was planned, executed, tracked, and reviewed throughout the study lifecycle.

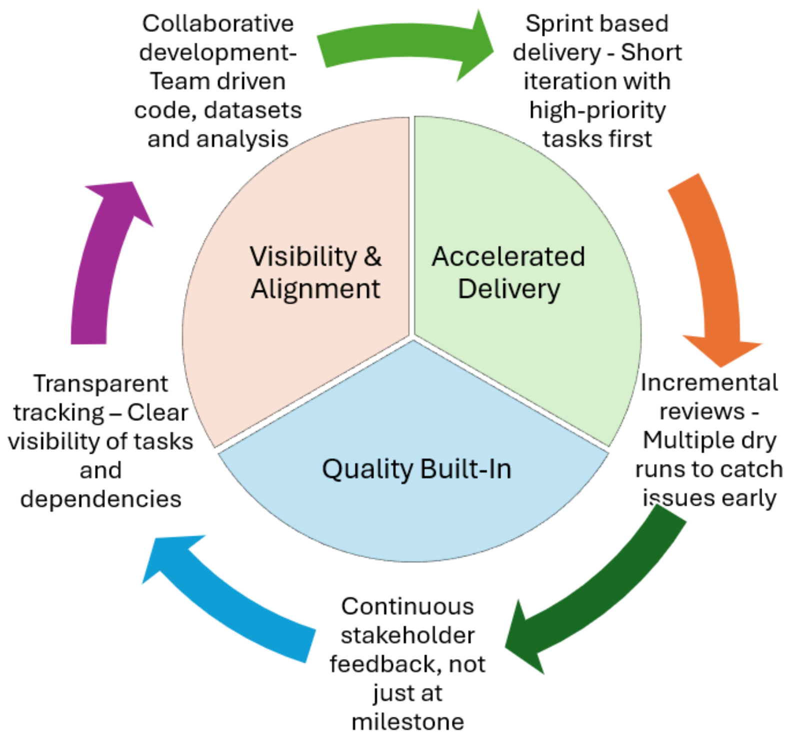

Figure 3: Agile methodology implementation

Core principles and practices included:

- **Sprint-Based Delivery and Incremental Reviews**
  Work was organized into short, iterative sprints, with high-priority tasks defined and executed first. Incremental reviews and multiple "dry runs" of outputs enabled early detection of issues and continuous refinement of datasets, analysis outputs, and code. This iterative approach reduced the risk of late-stage rework and supported predictable timelines.

- **Continuous Stakeholder Feedback**
  Rather than relying solely on milestone-based approvals, Agile enabled continuous engagement with study teams, statisticians, and other stakeholders. Ongoing feedback informed incremental adjustments, improved alignment, and ensured outputs met scientific, operational, and regulatory expectations.

- **Visibility and Alignment**
  Transparent planning, task tracking, and dependency mapping enabled all team members to see work in progress in real time, identify bottlenecks early, and coordinate efficiently across functions. By eliminating siloed working and fostering cross-team collaboration, shared backlogs and dashboards provided continuous visibility into progress, dependencies, and priorities. This approach enhanced alignment between study teams and programming groups, improved accountability, and supported informed, timely decision-making without disrupting ongoing work.

- **Accelerated Delivery**
  Agile practices enabled sprint-based execution and parallel work across programming, protocol

development, dataset creation, and analysis planning. By prioritizing high-value tasks and applying risk-based quality control, outputs were integrated early, allowing timely identification and resolution of issues. This approach transformed speed from an exceptional outcome into a predictable, reliable feature of the delivery system, while maintaining quality and reducing downstream rework.

- **Quality Built-In**
  Continuous review cycles, incremental testing, and automated checks embedded quality into the workflow. Version-controlled repositories and reusable code components ensured traceability, reproducibility, and submission-ready outputs at each stage of the study.

- **Collaborative Development**
  Cross-functional teams contributed collectively to programming, dataset preparation, and analysis logic. This approach leveraged domain expertise, standardized methods, and shared learning, improving efficiency and reducing errors.

Outcomes of Agile Implementation

Implementing Agile in biostatistics programming resulted in:

- **Predictable speed** through iterative delivery and parallel execution
- **Embedded quality** via continuous validation and automated checks
- **Enhanced collaboration** across distributed teams and study functions
- **Early risk detection** through transparent tracking and continuous stakeholder feedback
- **Scalability and reuse** through modular coding, version-controlled workflows, and integrated technology platforms

By embedding Agile principles across planning, execution, and review, the programming function evolved from a downstream, output-focused activity to an end-to-end delivery capability, supporting high-quality, submission-ready outputs under compressed timelines and increasing regulatory expectations.

AI AND AUTOMATION

Automation and artificial intelligence (AI) were selectively and responsibly introduced to enhance efficiency, accelerate delivery, and reduce repetitive manual work, while maintaining strict regulatory compliance. Human oversight and accountability remained central to all processes, ensuring that AI acted as a decision-support mechanism rather than a replacement for expert judgment.

Three primary use cases illustrate the impact of AI and automation in clinical programming:

- **AI-Assisted Programming and Logic Design**
  AI engines, such as GPT-based tools, were leveraged to generate complex programming code. Carefully crafted prompts enabled scenario-driven, dynamic, and defensive coding, accelerating logic design while retaining human review and ownership. This approach reduced repetitive effort, allowed rapid iteration of program logic, and maintained regulatory traceability.

- **AI-Enabled Data Review and Visualization**
  Interactive applications, developed using Shiny and SDTM datasets as the source, empowered stakeholders to explore data early and independently. By providing live, scenario-based visualization, the team reduced reliance on static outputs, accelerated decision cycles, and enhanced transparency. Users could test hypotheses, examine subgroups, and interact with exploratory datasets without waiting for sequential programming outputs.

- **AI-Supported Medical Writing Acceleration**
  Analysis-ready datasets were integrated with AI-enabled medical writing tools to prototype CSRs more efficiently. Post-analysis exploration with visualization tools allowed rapid identification of trends and insights, reducing the number of outputs while increasing interpretability. This enabled medical writers to iterate faster, ensuring early alignment with the statistical narrative and sponsor expectations

By implementing AI in this manner, teams-maintained control, traceability, and compliance while accelerating timelines and enhancing the quality and usability of outputs. The integration of AI and automation transformed clinical programming from a manual, output-focused function into a strategic enabler of high-quality, accelerated decision-making.

CULTURE AS THE TRUE MULTIPLIER

Growth Mindset, Resilience, and Reinvention

Reinvention did not follow a linear path. Early adoption of new tools and delivery models required experimentation, acceptance of imperfection, and continuous course correction. A growth mindset openness to feedback, willingness to learn, and smart risk-taking within clear guardrails proved essential to enable progress without compromising quality.

Adopting early created space for learning through practice. Multiple dry runs, incremental reviews, and continuous stakeholder engagement formed a learning flywheel, where each iteration informed the next. Learning became an embedded capability rather than a one-time activity.

System integration across functions was more complex than anticipated. Progress accelerated only when open communication, shared ownership, and close collaboration with technology teams became the norm, reducing friction and enabling faster resolution of dependencies.

Resilience emerged as the ability to absorb constant change, stay engaged through uncertainty, and continue delivering predictably. Teams operated within a continuous learn–unlearn–relearn cycle, where iterative learning compounded into deeper expertise and stronger delivery judgment.

A culture of clear ownership, rational decision-making, and strategic thinking enabled faster decisions with fewer downstream corrections. This clarity empowered teams and reduced rework while supporting long-term sustainability.

Ultimately, culture acted as a multiplier. A growth-oriented culture enabled rapid adoption of new capabilities while maintaining compliance and quality. In environments of constant change, culture determines whether transformation creates fatigue, or lasting advantage.

Change is inevitable. Culture determines whether it breaks organizations or builds them. A growth mindset creates resilience. Resilience sustains change.

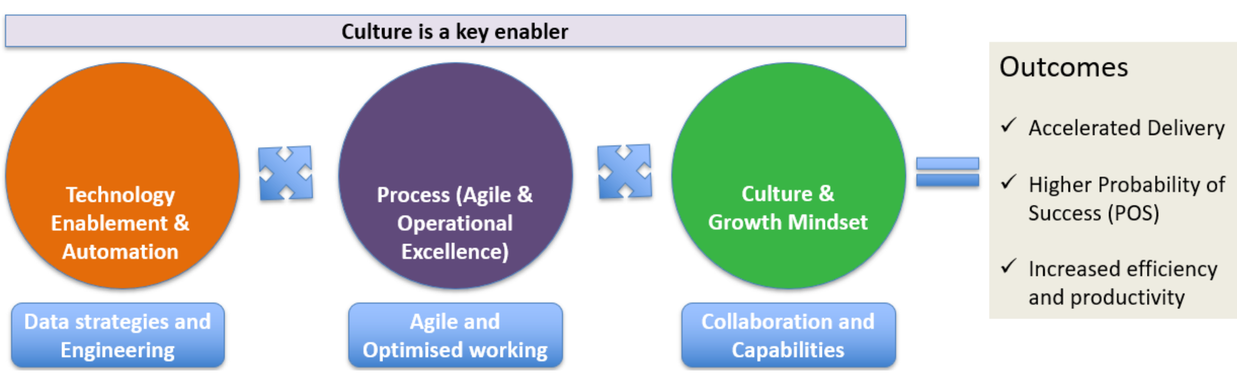

Figure 4: Growth Mindset

REAL WORLD IMPACT AND OUTCOMES

The integrated Agile and technology-enabled delivery model delivered clear, measurable improvements across clinical execution, statistical delivery, and regulatory outcomes, while maintaining quality, traceability, and regulatory compliance.

- **Lifecycle Acceleration and Execution Efficiency**
  Early integration of planning, data review, and execution accelerated delivery across the clinical lifecycle. Recruitment for the studies was advanced by approximately six months, enabling parallel indication submissions. This momentum was sustained downstream, with database lock activities for pivotal studies consistently completed within ≤2.5 weeks. Proactive resource planning and early risk mitigation, defined up

to eight months in advance, enabled efficient handling of variable data transfers, while strict data cut-offs and structured stakeholder communication minimized late-cycle rework.

- **Predictable and Accelerated Statistical Delivery**
  Statistical analysis and reporting timelines were materially reduced through lean analysis plan alignment, risk-based quality control, early pharmacokinetic access, and in-stream data review. Exploratory visualization reduced reliance on large volumes of static outputs, enabling fewer deliverables with greater insight. Six analysis packages across 2 indications for a same molecule were delivered within 1.5 months, including three analysis packages (~1,000 outputs) completed in a single week, with individual analysis packages cycles completed in ≤7 working days.

- **Earlier and More Efficient Regulatory Submissions**
  A single, harmonized data package strategy enabled efficient multi-regional submissions and reduced duplication. Study reports and clinical overviews were delivered ahead of plan, seven and six weeks post-analysis complete, supported by early study report prototyping and streamlined review cycles. Reliance and priority pathways enabled filings up to 270 days earlier in selected regions, with approvals achieved months sooner across multiple markets. An Accelerated Innovation Route pilot supported Group 1 market submissions within seven months of EU filing.

- **Scientific Impact and Capability Transformation**
  Advanced statistical methods, including Bayesian dynamic borrowing, increased power (81% to 89%) and adaptive approaches improved sensitivity on novel imaging endpoints. A distributed leadership model with Biostatistics leads and India-based teams supported delivery across multiple indications, including four pivotal studies with positive readouts. Critical path activities were increasingly in-housed, strengthening ownership, efficiency, and inspection readiness.

Overall, these outcomes demonstrate that sustainable acceleration is achievable when Agile delivery, responsible automation, and a resilient operating culture are deliberately aligned, enabling faster decisions, earlier submissions, and improved scientific insight without compromising compliance.

INDUSTRY IMPLICATION

This experience demonstrates how integrated Agile delivery, responsible automation, and cultural transformation can reshape biostatistics delivery for global clinical trials. While the approach was developed to address real-world program pressures, the resulting lessons have broader relevance for organizations navigating increasing complexity, parallel development, and compressed timelines.

- **Scalable Delivery Models for Global Trials**
  A key industry implication is the need to design delivery models explicitly for scale and parallel execution. Traditional sequential handoffs limit responsiveness and amplify late-stage risk. In contrast, platform-based and modular approaches, supported by reusable assets and standardized data packages enable parallel handoffs across functions, studies, and regions. This model supports consistent execution while allowing flexibility to accommodate indication-specific and regional requirements, making it particularly relevant for global, multi-indication programs.

- **Sustainable Speed with Compliance**
  This experience reinforces that sustainable speed is not achieved through late-cycle acceleration but through early readiness and continuous control. Early data availability, in-stream review, and risk-based quality management reduced rework while maintaining regulatory rigor. Continuous review cycles replaced episodic validation, improving predictability without compromising compliance. The case study highlights that quality and speed are complementary when embedded into process design rather than treated as downstream checkpoints.

- **Future-Ready Biostatistics Talent**
  The evolving delivery model reshaped the role of biostatistics teams. Statisticians increasingly acted as strategic partners in decision-making, supported by engineering-oriented thinking and technology fluency. Teams adopted new tools responsibly, led integrated delivery across functions, and assumed ownership of platforms and processes. This shift illustrates an emerging industry expectation: future-ready biostatistics talent must combine statistical expertise with systems thinking and the ability to operate within technology-enabled delivery models.

- **A Learning Journey, Not a Blueprint**
  Importantly, this transformation did not follow a linear or prescriptive path. Progress was driven by experimentation under real-world constraints, with deliberate reflection on what worked, what did not, and

what required course correction. Some initiatives scaled quickly; others required iteration or were deprioritized. This learning-based approach enabled continuous improvement while maintaining delivery commitments.

Broader Lessons for the Industry

The lessons from this case study extend beyond a single organization:

- Scalable delivery models must be designed for parallel execution from the outset
- Sustainable speed is enabled by early readiness and risk-based quality, not late-stage heroics
- Future-ready biostatistics talent integrates statistical rigor, engineering mindset, and technology fluency

Organizations that invest in people, platforms, and adaptive ways of working, while fostering a growth-oriented and resilient culture, will be better positioned to meet the evolving demands of global clinical development. This case study illustrates that reinvention is not a one-time transformation, but a continuous capability built through aligned process, technology, and culture.

CONCLUSION

Accelerating clinical development is not about working faster, it is about working smarter, together. This case study demonstrates that predictable, high-quality speed is achievable when Agile delivery, modular engineering, AI-supported decision-making, and a resilient learning culture are intentionally aligned.

While specific implementations may vary, the principles and practices described are broadly transferable. By embracing reinvention as a continuous capability rather than a one-time change, organizations can build delivery systems that remain resilient in the face of ongoing complexity and change.

REFERENCES

1. Ching, E. (2022). *AS01: SPACE – Launching GSK Biostatistics Towards New Horizons.* https://phuse.s3.eu-central-1.amazonaws.com/Archive/2022/Connect/US/Atlanta/PAP_AS01.pdf

2. Scaled Agile Framework: https://www.scaledagileframework.com/

3. Ching, E. (2024). *AS09: SPACE – GSK Biostatistics continues its journey to a new frontier*. https://phuse.s3.eu-central-1.amazonaws.com/Archive/2024/Connect/US/Bethesda/PAP_AS09.pdf

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Author Name: Deepthi BM
Company: GSK
Email: deepthi.x.bm@gsk.com