

From Listings to Leadership: Navigating 18 Years of Statistical Programming Outsourcing and the Road Ahead.

Girish Havildar, Merck & Co., Inc., Rahway, NJ, U.S.A

Outline

My journey with India-based outsourcing in Statistical Programming began 18 years ago with a small team and a single partner. Today, we collaborate with multiple partners and hundreds of individuals who support our work. Over the past two decades, I have witnessed the evolution from CROs handling simple listings to now taking on protocol-leader responsibilities.

Over time, as individuals' growth aspirations evolve, it becomes challenging for CROs to manage expectations. Furthermore, the introduction of new technologies adds additional challenges in satisfying overall resource needs and meeting sponsor expectations.

With the rapid progress in Artificial Intelligence and Machine Learning, the industry is on the verge of significant change. This prompts important questions: What lies ahead for outsourcing statistical programming? How will CROs respond to these changes? As routine tasks traditionally outsourced by pharmaceutical companies decrease, only a small portion of work may still be outsourced to CROs.

This discussion will explore these challenges and opportunities, highlighting the need for strategic adaptation in the evolving landscape of clinical research outsourcing.

Artificial Intelligence (AI) and Machine Learning (ML) are transforming statistical programming in clinical research by accelerating routine tasks and increasing the demand for advanced analytics, methodological oversight, and value-focused engagement models. Based on industry survey data and partner insights, this paper explores outsourcing trends over the next 3–7 years. We find that although automation will decrease manual efforts in data cleaning, standard analyses, and report creation, the market for outsourced services is likely to expand—driven by the complexity of trials, growing regulatory standards, and the rise of AI-powered analytics.

Introduction

We conducted the survey in collaboration with our partners, and senior management personnel from our partners answered a series of questions, as mentioned below. The partners include 2 big global technology companies, 3 established Statistics/Stat Programming CROs in the U.S, and India. Insights gathered from partner responses in the survey suggest that the outsourcing market will reshape rather than shrink. Respondents expect increased demand for AI-powered analytics and advanced modeling, even as repetitive tasks become faster or completely

automated. There is agreement that human expertise remains essential, especially for ensuring proper methodology, ethical use, and scientific validity when AI is involved.

The questions used to collect feedback from partners contributed to this paper.

- How do you assess the overall impact of AI/ML on the addressable market for outsourced statistical programming over the next 3–7 years?
- Given the likely automation of routine programming tasks, how is your company repositioning its service offering to capture new, higher-value work?
- What proportion of your current revenue do you expect to be most at risk from AI/ML automation, and how are you planning to replace or offset that revenue?
- Are you pursuing a “build, buy, partner” approach to AI capabilities? Why did you choose that route?

Service providers are shifting toward consulting, clinical design support, AI governance, and hybrid human-in-the-loop workflows; revenue risks from automation (estimated at ~20% to 70% by respondents) are offset by new AI-enabled offerings, workforce reskilling, and outcome-based pricing. Build–buy–partner strategies are converging on a hybrid approach that combines proprietary platforms with strategic acquisitions and close ecosystem collaborations. We suggest evolving the operating model from deliverable-based contracts to value- and outcome-based agreements, including transparency clauses for AI tool use, auditability, and data governance. Finally, we categorize tasks that are likely to remain outsourced (complex modeling, regulatory submissions, and QC/validation) versus those migrating in-house or becoming fully automated (data transformation and standard TLFs), along with practical metrics for measuring value in non-routine outsourcing.

Market Impact

AI/ML will not shrink the outsourcing market over the next 3–7 years; it will reshape it as automation reduces cycle times for routine tasks while overall market demand increases due to growing clinical trial complexity, stricter regulatory requirements, and the rise of AI-enhanced analytics. Routine work like data cleaning, standard SDTM/ADaM conversions, and TLF generation are especially vulnerable to automation, with generative AI and agentic systems now able to produce code from specifications, translate code across languages, and perform repetitive QC checks, although human oversight remains crucial to maintain methodological rigor. Partner feedback indicates that 20–70% of revenue linked to routine programming could be at risk as these technologies develop, but this risk is balanced by growth in AI-driven services and the expanding scope and complexity of modern clinical trials.

Strategic Repositioning and Workforce Transformation

From Transactional Programming to Value-Centric Services

Service providers are pivoting toward consulting and design advisory services, oversight and governance of AI tools, and communication/interpretation of complex results.

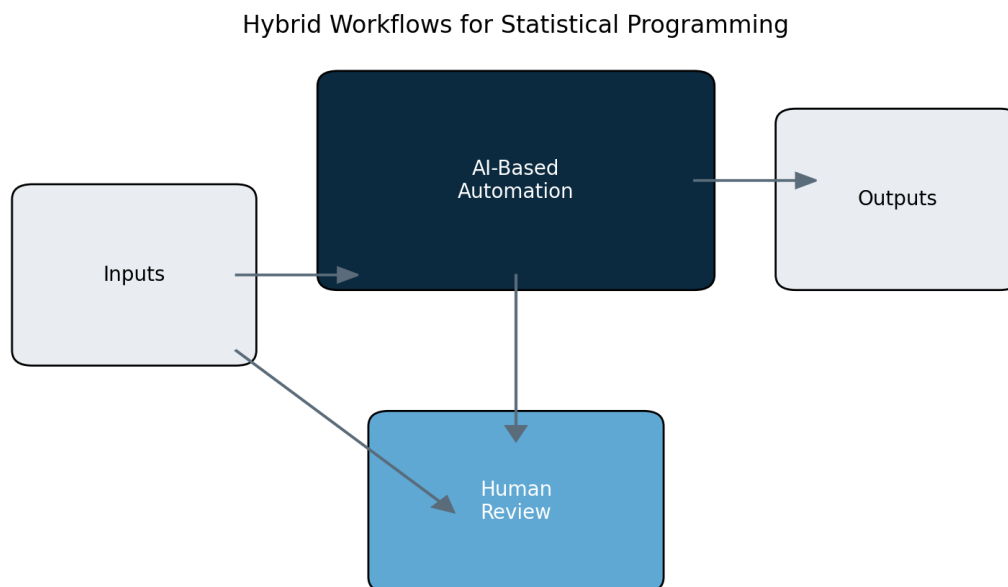
Workforce Transformation: Skills for the AI Era

Upskilling in AI literacy, ML tools, data science integration, and regulatory/ethical expertise is essential. Organizations are investing in centers of excellence, hybrid teams, and crowdsourcing models.

Business Model Evolution

Shift from volume-based deliverables to value-based engagements with tiered pricing for AI-assisted versus expert-reviewed work, and subscriptions for continuous AI-supported analytics.

Figure: Evolution of Stat programming roles and hybrid workflow.



Revenue Risk and Mitigation Strategies

Revenue at Risk from Automation

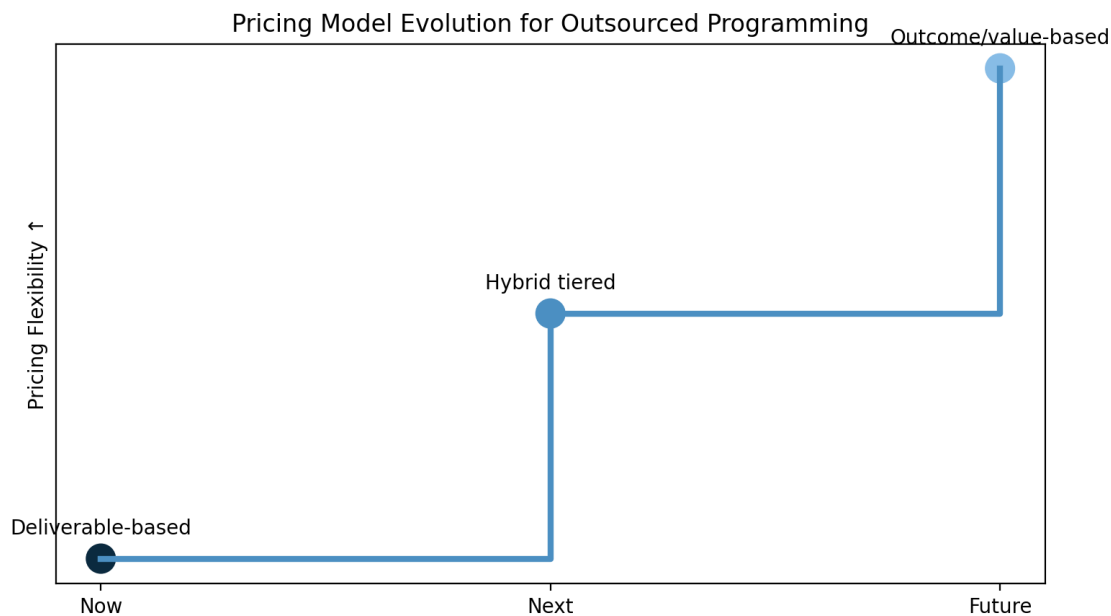
Estimates range from ~20% (conservative) to ~70% (aggressive), reflecting differences in service mix and AI readiness.

Drivers of Revenue Risk

Automation of routine deliverables, price pressure from procurement, and shifts toward outcome-based expectations drive revenue risk.

Mitigation Strategies

Diversifying into AI-driven services (predictive analytics, RWE), transforming the workforce (new oversight roles), introducing new pricing models (value-based, tiered), and pursuing hybrid tasks like interim analysis and RWE analysis will result in partial automation but will still require human oversight for interpretation and decision-making.



Build–Buy–Partner Strategies

Organizations are increasingly adopting a hybrid Build–Buy–Partner strategy to incorporate AI into clinical trial data processes, developing internal capabilities for control and quality, using specialized tools to enhance efficiency, and collaborating externally to advance methods. When integrated into clinical trial data workflows, AI-enabled components must operate within strict governance frameworks that ensure transparency, privacy, and auditability. To meet regulatory

expectations, SOPs should require that AI-supported processes maintain data reliability, patient safety, and decision-making integrity, supported by clear validation standards, lifecycle controls, and complete traceability within statistical outputs.

Operating Model Evolution (Scoping, Pricing, Contracting)

Scope emphasizes value-added tasks, including design consulting, advanced modeling, and interpretation. Pricing shifts from deliverable-based to hybrid, tiered, and outcome- or value-based models. Contracts should include transparency regarding AI usage, data privacy, and model training boundaries, as well as audit trails for reproducibility and compliance.

Task Landscape—Outsourced vs. In-House/Automated

Likely outsourced: complex modeling, regulatory submissions, QC/validation, niche studies.
Likely automated or handled in-house: data cleaning, standard analyses, report generation.
Hybrid: interim analyses and RWE where automation enhances throughput with expert oversight for interpretation.

Value Proposition and Metrics for Non-Routine Outsourcing

Quantify value through ROI, cycle-time reduction, error rates, audit findings, and submission success. Use case studies, POCs, and continuous improvement metrics to demonstrate quality and efficiency improvements.

Changes for Sponsors

- Bring routine analytics closer to core teams; deploy AI for speed, with experts ensuring accuracy.
 - Use AI to pre-screen SDTM datasets for structural issues, validate flags
 - Allow stat programmers to generate initial TLF shells or mock outputs using AI
 - Develop an AI model to detect issues with dates, mismatched visits, deviations, etc.
- Contract for outcomes: define KPIs (time-to-lock, QC defect rate, on-time TLF delivery), adopt automation-tier pricing, and require transparency on AI tool usage and validation evidence.
 - Metrics such as QC defect rates, on-time delivery of TLF batches
 - Instead of paying hourly, negotiate pricing based on the partner's automation maturity

- Make sure the partners disclose that AI components were used in programming or QC, how models were validated, and how outputs were reviewed
- Invest in governance: establish oversight committees, standardize metric definitions, and implement reproducibility and lineage tools across portfolios.
 - A cross-functional group to review AI usage across studies
 - Standardize KPI definitions and performance metrics across all studies and vendors.
 - Create dashboards visualizing end-to-end data lineage (raw → SDTM → ADaM → TLFs).

Changes for CROs

- Rebalance portfolios toward complex, high-trust services; adjust rate cards to reflect advanced expertise.
 - Shift resources towards interpretation-heavy analysis, offer regulatory-ready packages, and create advanced consulting tiers that justify higher rate cards
- Build validation platforms and reusable libraries; purchase targeted QA and metadata tools; collaborate with cloud/GenAI providers and PHUSE working groups to co-develop standards
 - Maintain validated libraries for TLF generation, build a standardized system for validating AI models used in programming
 - Build a standardized system for validating AI Models used in Programming
 - Contribute to working groups on AI validation, dataset standards, and metadata models
 - Collaborate with cloud/GenAI vendors to develop a standardized framework for safe, auditable AI use
- Upskill teams in ML, validation science, and regulated delivery; deploy agile teams and Centers of Excellence
 - Create modules covering basic AI/ML basics, risk-based testing, and regulatory expectations
 - Build an AI COE of a stat programming group responsible for validated AI tools, managing model updates and change control, and training delivery teams

Conclusion

Revenue from routine programming will face challenges, but the market for outsourced statistical programming is expected to expand as organizations shift toward AI-enabled, value-focused services. Investing in governance, talent, and hybrid operating models enables CROs and sponsors to effectively leverage AI while preserving scientific integrity and ensuring regulatory compliance.

Acknowledgments

The author would like to acknowledge the colleagues who provided valuable feedback during the preparation of this paper. Special thanks are extended to **Amy Gillespie, Saigovind Chenna, and Sujay Bothisathuvan** for their careful review and constructive comments.

The author also expresses sincere appreciation to **Pratibha Jalui** for her comprehensive support throughout the submission process.

Finally, the author gratefully acknowledges the participants from our preferred partners, who provided essential input to this work by answering all the survey questions.

Contact Information:

Name: Girish Havildar

Enterprise: Merck & Co., Inc.

Address: 351 N Sumneytown Pike

City, State & Zip: North Wales, PA 19454, USA

E-mail: girish_havildar@merck.com

Web: www.merck.com