Paper ET27
Scaling for the Future: Transitioning from In-House to Commercial Statistical Computing Environments

Majdoub Haloui, Merck & Co., Inc., Rahway, NJ, USA
Hemu Shere, Merck & Co., Inc., Rahway, NJ, USA
Arelys Tarazona, Merck & Co., Inc., Rahway, NJ, USA

ABSTRACT

Defining and implementing a Statistical Computing Environment (SCE) in clinical development has long been a challenge. In the absence of comprehensive commercial solutions, many pharmaceutical companies, including ours, developed in-house systems to meet regulatory requirements and support multiple programming languages. Our internal platform has effectively served the business for more than a decade, enabling scalability and an expanded user base.

With recent advancements in technology, however, commercial SCEs have become increasingly accessible, offering features that were difficult to achieve with custom solutions. This paper outlines our journey in transitioning from an in-house platform to a vendor-based SCE. We describe the drivers for change, the vendor selection process, and the proof-of-concept evaluation criteria. We also summarize early observations and the phased implementation approach we are currently pursuing. Where capabilities are not yet realized, we present them as goals and note current status or dependencies.

INTRODUCTION

Building and maintaining a compliant, flexible, and scalable statistical computing environment has long been a fundamental challenge in clinical development. Statistical programming organizations must balance regulatory compliance, data integrity, traceability, and security with the practical needs of analysts for flexibility, performance, and support for multiple programming languages and tools. Historically, limited availability of robust commercial solutions capable of meeting these combined requirements led many pharmaceutical companies to develop and operate in-house SCE platforms.

This paper describes the company's transition from a legacy in-house SCE to a vendor-based commercial solution. It outlines the key drivers behind the decision, the structured vendor selection process, and the proof-of-concept (PoC) evaluation used to assess candidate platforms. The paper also summarizes the anticipated benefits of the new SCE and outlines a phased implementation approach. Because new tools and concepts, particularly GitHub, along with SAS Studio and the SCE were introduced, significant training, learning, and experimentation were required to identify the best solutions. Several items described herein are planned or in progress rather than fully implemented; we call these out as goals and subject to validation, see Figure 1.

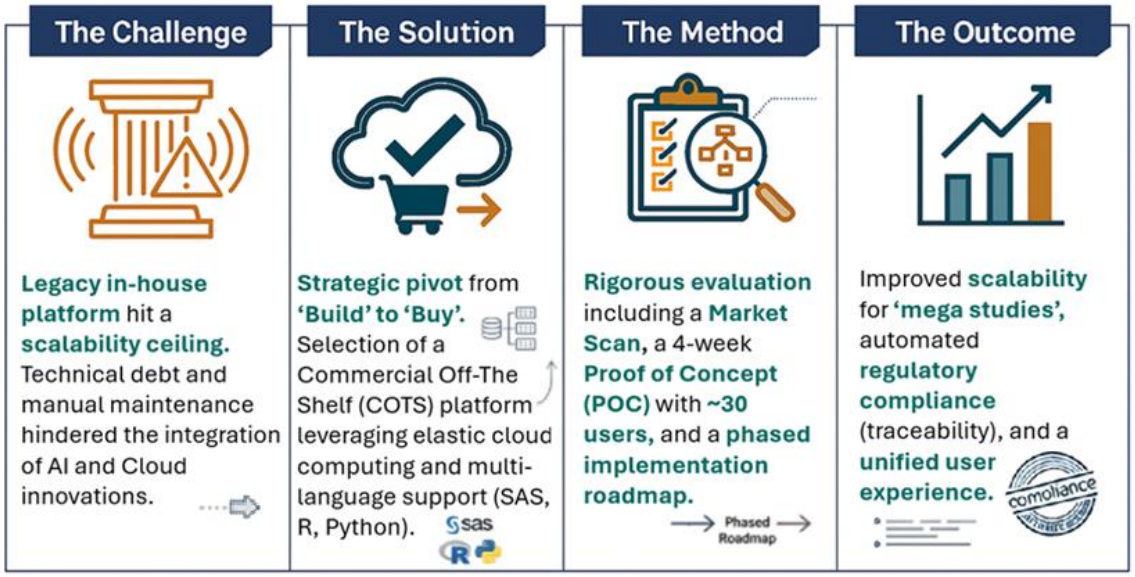

Figure 1: Modernizing the Platform: Challenge, Solution, Method, Outcome

BACKGROUND / REASONS FOR CHANGE

The expectations for an SCE in clinical development are high. It must ensure compliance with regulatory standards, provide traceability and reproducibility, and support multiple programming languages such as SAS, R, and Python. Our in-house platform met these requirements for many years, enabling rapid analysis and relying on manual tasks. However, sustaining this system became increasingly challenging as technology evolved and business needs expanded. Maintenance overhead grew, scalability was limited, and automation capabilities lagged industry standards. Outdated technology resulted in increasing maintenance costs, impacting scalability, challenges to incorporate new technologies such as AI and cloud computing, and prevented the automation of processes for versioning, traceability, and reproducibility, see Figure 2. In addition, manual tasks for workflow management created additional inefficiencies. Moreover, internal stakeholders demanded faster cycle times and improved user experience. Cloud adoption initiatives and cost optimization pressures further highlighted the limitations of our legacy system.

Several strategic and operational factors motivated the transition away from the internally developed SCE:

- **Struggle to keep pace with technological innovation**
  Future of statistical computing became multi-lingual (SAS, R, etc.), cloud-based, and user-friendly. We needed to act now to keep pace with industry-best practices. Our current environment did not provide our scientists with the best solution for doing their jobs within a single platform.

- **Sustainability and Maintenance Overhead**
  Our current environment of custom-built applications with old technology has become difficult to sustain long term. Maintaining an in-house SCE required ongoing investment in infrastructure, validation, security patching, and specialized technical expertise. As regulatory expectations and technology standards continued to evolve, the effort required to sustain the platform increased significantly.

- **Inefficiencies with manual work**
  We have an opportunity to enhance efficient tools and increase system flexibility to fully leverage automation, accelerating timelines, reducing manual work, and strengthening programming workflows through improved traceability, reproducibility, and seamless integration with data-access systems and tools.

- **Evolving Regulatory and Compliance Expectations**
  Regulatory agencies increasingly began to expect strong controls around system access as well as emphasis on audit trails, data lineage, and validation documentation. While these requirements were met in our existing platform, enhancements often required custom development and extensive revalidation efforts.

- **Technology Advancements and Modernization**
  Emerging technologies such as cloud-native infrastructure, containerization, and automation offer improved scalability, resiliency, and efficiency. Retrofitting these capabilities into the legacy SCE posed architectural and operational challenges. For example, company-wide initiative to move all systems to the cloud would require significant time, money, and resources to overhaul the existing architecture. Our current platform faces continued high demand pipeline with growing risk of business disruption due to aging technology. Hence, building new was a better option.

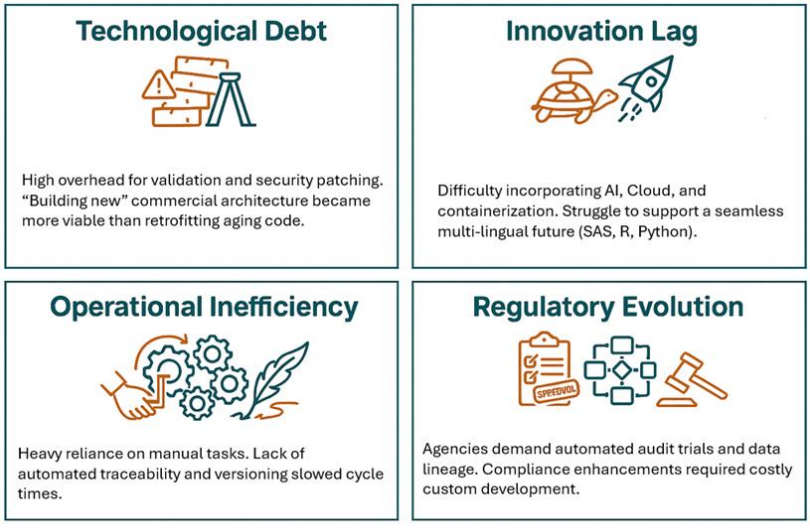

Figure 2: Key Drivers for Platform Transformation

It is worthwhile noting that this was not our first attempt to implement an SCE to replace our existing workspace. Our company has looked at various products over the past years, and these attempts were not successful due to various reasons including lack of a flexible solution. Our previous attempts to transition to a new SCE revealed critical lessons: vendor solutions must align with compliance requirements, integrate seamlessly with existing systems, and offer flexibility for customization.

VENDOR SELECTION PROCESS

Our high-level goal with a new SCE aimed to streamline and update our technology architecture and simultaneously enhance operational efficiency, improve user experience, and capitalize on industry accelerators.

To aid and achieve higher likelihood of success with a new SCE, in 2023, our company conducted an assessment in collaboration with a consulting firm which included vendor assessments, options between vendors, a hybrid approach with multiple vendors, and keeping building on our in-house solution. The result of the assessment identified a few potential vendor options with solutions that could act as accelerators of our future platform and prioritized two options for further diligence, one due to its incumbency for other tools used by the vendor within our clinical space and another due to the maturity of their SCE solution and proven track record among other pharmaceutical peers.

A structured and cross-functional vendor selection process was established to ensure that the chosen SCE met our key business needs as well as all technical and regulatory requirements. Key steps in the process included:

- **Requirements Definition**
  Functional, technical, and compliance requirements were defined with input from statistical programming, statisticians, IT and quality assurance stakeholders. These requirements covered areas such as language support, business validation approach, access controls, data security, performance, and integration with existing systems.

- **Market Scan and Vendor Shortlisting**
  A landscape review of available commercial SCE solutions was conducted to identify vendors with proven experience in regulated clinical development environments. Vendors were shortlisted based on alignment with core requirements and maturity of their offerings.

- **Request for Information**
  Shortlisted vendors were asked to provide detailed responses describing their platform architecture, compliance model, system validation deliverables, support model, and roadmap.

- **Stakeholder Review and Scoring**
  Vendor responses were evaluated using a scoring model to objectively compare solutions across predefined criteria. Feedback from technical, operational, and quality perspectives was incorporated into the final assessment.

PROOF-OF-CONCEPT (POC) EVALUATION CRITERIA

To complement the paper-based evaluation, a hands-on proof-of-concept was conducted with the two selected vendors from the assessment. Project team members involved felt it was necessary to test the SCE tool and not base acceptance solely on vendor demos or discussions. It was important to perform testing for key capabilities and ensure the SCE delivered on these points. A 4-week POC in 2024 was conducted where ~30 programmers and statisticians (across 6 groups such as late development, standards, esubmission, R and HTA) evaluated the SCE in our company hosted environment. We identified 70 requirements for the new SCE of which 33 were deemed critical. These critical requirements were assessed and categorized to various outcomes such as Met, Partially Met, Not Met, Not Tested, On Vendor Roadmap, or IT-relevant, see Figure 3.

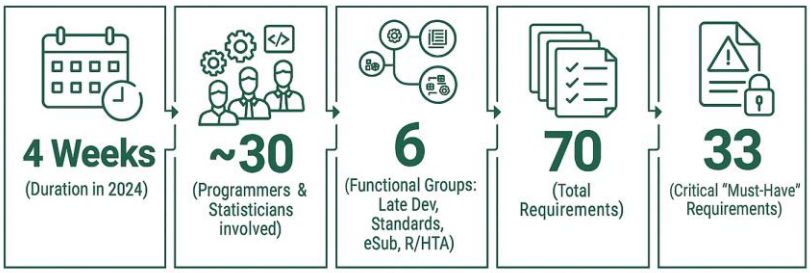

Figure 3: Key Metrics From the 4-Week SCE Proof-of-Concept

The PoC aimed to confirm the key requirements, covering 10 categories from the evaluation criteria:

- Backup/Restore/Archive
- Dashboarding
- Data ingestion and handling of multiple sources/formats including large data
- Integration with upstream/downstream platforms
- Intelligent automation
- Job scheduling and prioritization
- Language-agnostic support
- Security and access controls (including catalogs and inspection support)
- User experience, performance and scalability
- Workflow/process management (data lineage, versioning, reproducibility)

Important note on interpreting requirements:

- Some items were tested and met or partially met during the PoC
- Others were not tested or are dependent on vendor roadmaps or internal integration work
- A subset remains aspirational at this time. For example, “browse files and map study folders to a Windows Explorer drive letter” is under exploration and faces setup challenges; it is not yet a proven or available capability in our environment
- Integration with Pinnacle 21 was tested primarily because it is a critical esubmission system and also as a basis to confirm integration concepts

OVERVIEW OF THE NEW SCE PLATFORM

The solution our company selected is a commercial off-the-shelf platform that is designed to meet the complex needs of clinical development within our organization. It leverages elastic cloud computing to provide scalability and cost efficiency, automatically adjusting resources based on workload demands. The platform supports multiple programming languages, including SAS, R, and Python, enabling statisticians and programmers to work in their preferred environments, all this within a single platform. Integration with GitHub facilitates version control and collaborative development. which will be validated against our processes and regulatory expectations during pilots. Security and governance are embedded within the new solution, providing automated audit trails and capabilities to reliably reproduce results. Demonstrating this end-to-end remains part of our upcoming validation, see Figure 4.

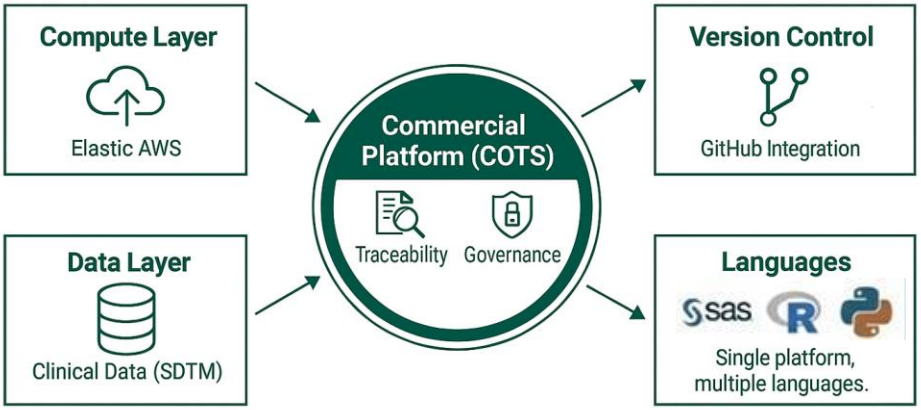

Figure 4: Foundation for a Scalable, Multi-Language Platform

IMPLEMENTATION STRATEGY OF NEW SCE

The strategy emphasized a phased implementation approach, system integration, mock testing, controlled migration, and structured change management starting with 2 pilots and further user adoption. This approach helps maintain business continuity and supports long-term scalability and continuous improvement.

- **Phased implementation approach**
  To minimize risk and ensure business continuity, the platform was introduced using a phased rollout strategy. The first phase (Phase 0) involved informal business testing, allowing subject matter experts to

explore the platform's capabilities and validate workflows. The second phase (Phase 1a) included a technical release focused on infrastructure setup, integration with Clinical Data, design a GitHub flow, and performance tuning. The third phase (Phase 1b – current phase) is targeting an end-to-end analysis and reporting (A&R) execution with SDTM integration, finalization of business processes, and training materials to facilitate adoption. The last phase (Phase 2) will focus on early adopter studies and iterative improvements. Success criteria are being tested and refined as we progress, see figure 5.

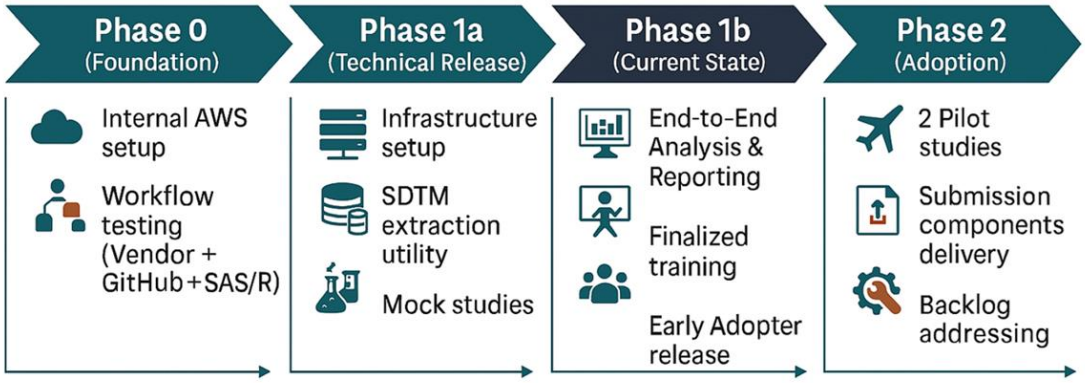

Figure 5: Implementation Strategy: The Phased Roadmap

The table below summarizes each rollout phase, its objectives, the success criteria achieved, and status:

Phase	Objective	Success Criteria Achieved	Status
Phase 0	• Implement SCE Platform on our internal AWS • Experimentation of the SCE platform including its 3 core components: Vendor product, GitHub, SAS Studio/RStudio • Identify/develop key aspects of the business process in the new environment	• Installed & Tested Basic workflow with SCE, GitHub, SAS Studio, and R • Drafted 1st version of Business Process Best Practices • Completed in-depth testing against crucial business requirements and business process • Evaluated existing systems for capabilities and eventual integration with the SCE	Achieved
Phase 1a	• Initial release of working SCE platform with analysis and reporting business processes to facilitate execution on mock studies • Further elaborate/define key aspects of the business process	• Successfully created ADaM and TLFs • Successfully delivered apps to complement SCE e.g. apps for project setup, provisioning • Refined business process for key activities such as working with blinded/unblinded data, folder naming conventions, roles, etc. • Successfully implemented SDTM data extraction utility • Conducted Learning Lab to obtain feedback on the SCE and training material	Achieved
Phase 1b	• E2E A&R execution with integration of clinical data layer (SDTM data) • Finalize business processes and training to facilitate adoption and rollout.	• Execute an E2E A&R project with data integration capabilities • Conduct impact assessment & buildout of standard macros and R packages • Finalize business process for SCE, GitHub, and SAS Studio/R • Establish a training Strategy, Training Material & Business Processes including	In-Progress

Phase	Objective	Success Criteria Achieved	Status
		a 2nd round of Learning Lab • Implement a collaboration solution to allow sharing and editing of documents • Conduct Performance Testing ***This is our Official release of the SCE for Early Adopters.***	
Phase 2	• Execution of early adopter studies (2 pilots) with end-to-end CSR deliverables • Address backlog list and prioritize based on early adopter study needs	• Success will be determined by delivery of submission components and closure of backlog items	Planned

Table 1: Objectives and Success Criteria

- **Guidance from SCE Vendor**
  Vendor consultants provided guidance on technical and business processes, complementing our strong internal resources and enabling cost-effective implementation. Final process designs and validations remain our responsibility and are ongoing.

- **Integration**
  Integration was a key component of the strategy. The platform is being connected to the clinical data layer for study data ingestion and is under development to close gaps (e.g., automated study setup, access provisioning, data extraction interfaces). While integration was tested during the PoC (e.g. Pinnacle 21), the production integration with systems in the A&R space is planned and subject to further design, testing, and validation; it is not yet in place.

- **Testing**
  Mock study testing played a vital role in proving the newly developed processes, Initial performance testing has begun and will continue during pilot studies to validate scalability and reliability under peak workloads.

- **Legacy coexistence and transition planning**
  We plan to support parallel operation of legacy and new platforms during transition. Criteria and timelines for decommissioning legacy systems will be defined after pilot outcomes.

- **Migration**
  We intend to prioritize the workloads for initial migration, with clear criteria for moving studies fully into the new environment. This plan is in development.

- **Change management and training**
  Role-based training and onboarding materials are being developed. We are conducting periodic sessions to share updates and gather feedback. Full training rollout and adoption support activities (e.g., hyper-care) are planned but not yet executed; effectiveness will be evaluated during pilots.

DRIVING ADOPTION

Adoption will be supported by structured management across business and technical dimensions. Stakeholder engagement, targeted communications, and feedback mechanisms are ongoing as we prepare for early adopter pilots, see Figure 6.

Current adoption plan (in development):

- A multi-tiered training strategy will be implemented to accommodate varying levels of experience and roles within the organization. This will include foundational end-user training focused on core workflows, considerations for trainings methods such as train-the-trainer model to establish internal subject matter experts and promote self-sufficiency or direct user training. During adoption, hyper-care support will be provided through dedicated office hours, real-time issue resolution, and direct access to platform and

process experts. This approach will reduce barriers to entry, accelerate proficiency, and minimize productivity impacts during early use.

- Initial pilot studies will be selected to engage early adopters and validate the SCE within controlled, lower-risk contexts. Insights from these pilots informed refinements to governance, validation, and support models before broader deployment.
- Adoption will be further strengthened through a phased rollout strategy designed to balance current project work with goals to move to the new SCE. The rollout then scaled progressively to enterprise-wide use, supported by clear readiness criteria and success metrics at each stage.
- A structured migration plan will ensure continuity and compliance throughout the transition. High-value and high-risk studies will be prioritized based on business impact and regulatory considerations, with all new studies initiated in the SCE once baseline capabilities were established. Legacy content will be migrated incrementally using defined standards, risk assessments, and validation checkpoints, minimizing disruption to ongoing deliverables.

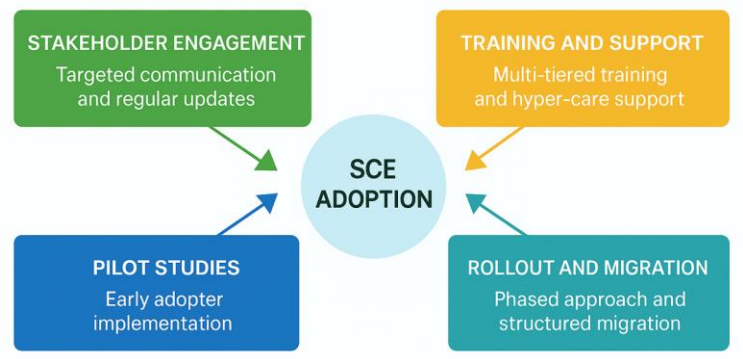

Figure 6: Driving Adoption

Collectively, these strategies will enable a sustained adoption of the SCE, embedding it into standard statistical workflows and positioning the organization for scalable, future-ready analytical capabilities.

ANTICIPATED BENEFITS

The implementation of the new SCE is expected to deliver substantial benefits across compliance, scalability, operational efficiency, and user experience, while establishing a future-ready analytical foundation, see Figure 7.

From a compliance and regulatory perspective, the SCE introduces automated traceability and reproducibility capabilities that significantly enhance audit readiness. These capabilities will be validated through system testing and then again with pilot studies.

The cloud-based architecture of the SCE provides inherent scalability and elasticity. Cloud-based architecture is intended to enable dynamic scaling of compute and storage resources, supporting peak activities (e.g., database locks, interim analyses, submission preparation). We plan to validate cost and performance characteristics during pilots.

Efficiency gains are anticipated through automation. Streamlined workflows should reduce manual handoffs and variability. Integration with tools such as Pinnacle 21 and direct access to SDTM datasets are planned to accelerate validation and improve oversight; these integrations have not yet been completed.

User experience represents another key area of improvement. The platform offers a modern interface intended to support day-to-day analytical workflows. User experience improvements will be assessed and refined based on pilot feedback.

Adopting a Commercial Off-The-Shelf (COTS) solution offers additional strategic advantages. A commercial solution can reduce development time and maintenance burden compared to custom builds, with ongoing vendor enhancements aligned to industry trends. Risks include dependencies on vendor roadmaps and potential change fatigue; we plan to mitigate these through governance, vendor engagement, and continuous feedback loops.

Figure 7: Anticipated Benefits

CONCLUSION

The transition from a legacy, in-house statistical computing environment to a vendor-based solution represents a strategic modernization initiative that delivers enterprise-level value. Early phases have established foundational capabilities and informed process design. As we proceed into pilots, our focus is on validating end-to-end workflows, compliance features (including automated traceability and reproducibility), integrations, performance, and user experience.

Looking ahead, the focus will shift toward expanded automation, strengthen analytics (including AI-enablement where appropriate), and continuously optimize workflows. The phased approach, change management emphasis, and vendor partnership are intended to support a smooth transition. We will adjust plans based on pilot outcomes and evidence gathered, and we will update this paper with proven results as they become available.

Learning Labs have supported experimentation and feedback during early phases. We plan to continue scenario-based sessions aligned with study deliverables during pilots, using feedback to refine training materials, workflows, and guidance. Their effectiveness will be assessed as part of the adoption plan.

RECOMMENDED READING

PHUSE SCE Whitepaper (2019-2021):
https://phuse.s3.eu-central-1.amazonaws.com/Events/2021/US+Connect+2021/SCE_White_Paper_Final_15Dec2021.pdf

CONTACT INFORMATION

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

Merck & Co., Inc

Majdoub Haloui: Majdoub.haloui@merck.com
Hemu Shere: hemu.shere@merck.com
Arelys Tarazona: arelys_tarazona@merck.com