PharmaSUG 2025 - Paper SI-253

TRANSITIONING EXTERNAL CLINICAL STUDIES TO INTERNAL: A FRAMEWORK FOR KNOWLEDGE TRANSFER AND OPERATIONAL EXCELLENCE

Chunqiu Xia, Lily Zhang, and Xiaohui Wang, Merck & Co., Inc., Rahway, NJ, USA

ABSTRACT

This paper presents a framework for transitioning ongoing clinical studies acquired from external pharmaceutical companies into internal statistical programming operations. The primary challenge lies in ensuring regulatory compliance with the CDISC (Clinical Data Interchange Standards Consortium) standards, a critical requirement for data integrity and regulatory submissions. To address this, the framework includes extensive efforts to align external data structures to internal formats, standardize deliverables such as Tables, Listings, and Figures (TLFs), and conduct rigorous compliance checks. Additional efforts involve creating centralized knowledge repositories, re-mapping external datasets, and tailoring statistical programming to internal requirements. Operational tools, such as delivery trackers and structured meetings, further enhance the efficiency and coordination of the transition process. By focusing on regulatory compliance while implementing scalable methodologies, this framework ensures a seamless integration of external studies into internal workflows, supporting high-quality and compliant operations.

INTRODUCTION

From the statistical programming perspective, transitioning ongoing clinical studies acquired from external sources into internal operations is a multifaceted process. The primary challenge is ensuring that external study packages meet CDISC standards, which is critical for regulatory submissions and data integrity. Beyond this, teams must align external datasets, methodologies, and deliverables with internal workflows to maintain consistency and efficiency. The process of integration can vary significantly depending on the acquisition scenario. These scenarios include ongoing studies, completed studies, single compound acquisitions, or whole company acquisitions, each with distinct requirements. This paper specifically focuses on transitioning an ongoing study for a compound from an acquired company, where statistical programming teams must ensure compliance with CDISC standards while streamlining internal adoption of external data, tools, and processes. The proposed framework addresses these needs by emphasizing compliance validation, standardization of TLFs, creation of reusable programming tools, and adoption of operational practices like delivery trackers and centralized repositories. This structured approach equips statistical programming teams with the strategies required to manage study transitions efficiently and effectively, ensuring both regulatory alignment and operational excellence.

ENSURING REGULATORY COMPLIANCE WITH CDISC STANDARDS

When acquiring clinical studies from external pharmaceutical companies, ensuring adherence to Clinical Data Interchange Standards Consortium (CDISC) standards, particularly Study Data Tabulation Model (SDTM) and Analysis Data Model (ADaM), is crucial. The following structured workflow details comprehensive steps to systematically achieve compliance, starting from initial collaboration through final dataset validation.

Initial Collaborative Phase and Documentation Review

The transition process typically begins with a collaborative period involving internal statistical programming teams working closely alongside the external programming team to ensure that all necessary documentation, datasets, and programming assets from the external organization are thoroughly understood and effectively transferred. Regular meetings and workshops are conducted between both teams to thoroughly document external data collection methodologies, SDTM and ADaM mapping procedures, dataset derivations, reporting standards, methodologies, systems, vendor selection, data preparation processes, dataset derivation logic, and the logic underpinning the generation of Tables, Listings, and Figures (TLFs). Additionally, documentation verification activities ensure that all relevant

external documentation—including protocol documents, dataset specifications, SAS programming guides, and TLF mock-ups—is fully provided and clear, with early identification of knowledge gaps or missing information and documentation of necessary clarifications or additional information required from external teams.

INTERNAL RE-EXECUTION OF EXTERNAL PROGRAMS AND OUTPUTS

Following successful initial knowledge transfer, the next crucial step involves verifying compatibility by internally executing external programs to exactly reproduce ADaM datasets and TLF outputs as originally delivered by external sources. This step ensures dataset reproducibility, integrity, and seamless operational continuity.

Preparation of the internal environment begins by aligning external programs to internal directories, libraries, and macro configurations. The internal team clearly documents any modifications necessary to successfully execute external scripts within the internal environment, including adjustments to directory structures or macro paths.

Next, the internal team reruns all external ADaM and TLF programs without initial alterations to validate baseline compatibility and confirm reproducibility of outputs. They systematically compare internally generated outputs against externally provided ADaM datasets and TLFs, identifying any differences, discrepancies, or execution issues.

Finally, the internal team thoroughly documents and investigates any discrepancies found, such as programming errors, execution warnings, or output mismatches. They resolve compatibility issues—such as macro reference errors, dataset merging problems, or path-related discrepancies—and clearly document each implemented solution, ensuring transparent tracking and effective resolution of identified issues.

PINNACLE 21 VALIDATION CHECKS FOR SDTM AND ADAM DATASETS

Upon successful internal re-execution and reproduction of outputs, the internal team conducts a thorough CDISC compliance validation using Pinnacle 21 (P21). This validation ensures dataset integrity, regulatory compliance, and submission readiness.

The internal team systematically executes Pinnacle 21 checks on both SDTM and ADaM datasets, categorizing validation findings into Errors, Warnings, and Notices. All validation findings are clearly documented in dedicated compliance trackers, with corrective actions prioritized based on their severity and regulatory implications.

Common SDTM issues identified through Pinnacle 21 validation include missing or inconsistent unique identifiers (USUBJID), non-adherence to controlled terminologies (such as incorrect AE severity scales), and missing or incorrect domain structures and required variables (e.g., required variables absent in AE or DM domains). Typical ADaM-related issues include incorrectly derived analytical variables (AVAL, PARAMCD, BASE), structural deviations impacting compatibility with internal analysis macros, and unclear traceability from ADaM datasets back to source SDTM datasets.

Critical Errors discovered during validation are promptly resolved by restructuring datasets, re-deriving analytical variables, or making necessary metadata adjustments. The internal team thoroughly documents all outcomes and solutions and communicates critical findings clearly to external programmers, explicitly requesting necessary updates to programming scripts and dataset specifications.

Following resolution of critical issues by external teams based on the initial Pinnacle 21 validation, the internal team requests the updated dataset package. They then perform iterative Pinnacle 21 validations on these re-delivered datasets to confirm resolution of previous critical errors, maintaining clear, documented communication regarding validation outcomes and any subsequent actions required for ongoing compliance and submission readiness.

CONTINUOUS COMPLIANCE MONITORING AND FINAL DOCUMENTATION

Continuous compliance monitoring and thorough final documentation require active oversight, rigorous documentation practices, and proactive management of newly identified issues.

The internal team regularly monitors SAS logs using automated checks to promptly identify programming issues, including unresolved macro references, data truncation, or type mismatch warnings. They swiftly address these issues to preserve data integrity throughout the project lifecycle.

The team maintains strict version control for datasets and programs by consistently managing and tracking file timestamps from the Windows operating system. They verify timestamp accuracy and ensure version alignment during each dataset delivery and re-validation cycle, maintaining transparency and consistency throughout the process.

Additionally, the internal team routinely schedules Pinnacle 21 validations, especially at critical milestones like interim data cuts, database locks, and prior to regulatory submissions. They clearly document the history of validation findings, actions taken to resolve them, and maintain comprehensive records for regulatory readiness.

ESTABLISHING KNOWLEDGE REPOSITORIES, DATASET RE-MAPPING, AND INTERNAL PROGRAMMING ALIGNMENT

Beyond regulatory compliance, teams must undertake additional critical efforts when transitioning externally acquired clinical studies into internal operations. These efforts include facilitating effective knowledge transfer, systematically re-mapping external datasets, adapting statistical programming to internal standards, and structuring operational tools that streamline integration and enhance collaboration efficiency.

COMPREHENSIVE KNOWLEDGE TRANSFER

A structured and collaborative knowledge transfer process establishes a solid foundation for effective integration. At the start of the transition, internal and external programming teams actively engage in collaborative sessions to clearly document essential knowledge, including processes, methodologies, and deliverables.

Structured documentation workshops enable teams to identify and capture external data collection platforms, associated protocols, standard operating procedures (SOPs), data sources, and audit trails. Teams thoroughly document data preparation and cleaning methods, clearly defining processes for outlier management, handling missing values, interim data-cut definitions, and maintaining data validation logs.

The internal team explicitly documents vendor-selected methodologies and tools like Pinnacle 21 and SAS, mapping logic, and quality assurance practices for generating compliant SDTM and ADaM datasets. Additionally, comprehensive documentation includes external TLF-generation tools, associated SAS macros, internal logic, validation strategies, and quality control procedures. Raw data listings and continuous data integrity checks, including reconciliation processes, are also clearly standardized and documented.

The internal team consolidates all acquired documentation into a centralized knowledge repository, such as SharePoint or Confluence, structured to facilitate easy navigation and team collaboration. This repository features logical content indexing, intuitive folder structures, and advanced search capabilities. It centralizes critical information such as programming guides, dataset specifications, TLF mock-ups, protocols, macros, and training materials for ongoing reference.

Additionally, the repository incorporates permission-based access controls and audit trails to protect sensitive information, ensure accountability, and maintain regulatory compliance.

OPERATIONAL TOOLS FOR TRANSITION MANAGEMENT

The delivery tracker is a critical tool for monitoring and coordinating the transition of clinical studies. It should include multiple sheets, each tailored to specific aspects of the study transition:

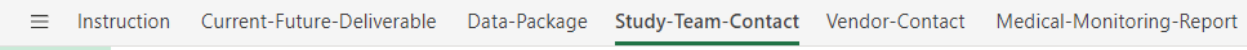


Figure 1.Delivery tracker

Data-Package

The team actively tracks the status of datasets, including raw data, SDTM, ADaM, and TLF deliverables, using a structured tracker. This tracker contains columns for delivery dates, file names, version numbers, and clearly indicates review status, such as pending, in progress, or approved.

	Study	Data Package Name	Data Package Purpose	Data Package Contents e.g., SDTM, ADaM, IB reports, etc.	Data Cut-off	Data Package Transfer Date	File(s) Missing	pacted Deliverables
2	Merck-001	001_Prod_005_20240709-1525.zip	Manuscript	SDTM, ADaM and Programs	4-Jun-24	9-Jul-24		
3	Merck-001	001 Prod 005 20240722-1420.zip	Manuscript	ADaM,TLF/ programs	4-Jun-24	22-Jul-24		

Figure 2.data package tracker

Study-Team-Contact

The tracker lists key contacts from the internal study team, clearly identifying roles such as statistician, programmer, and project manager, along with their contact information. This structure streamlines communication within the team.

Study	Functional Area	Role	Contact	Backup Contact
All	Merck DM	Project Manager	Messi <eail>	
Merck-001	Non-Merck SP	SP Protocol Lead	Jerry <email>	Tom <email>

Figure 3.Study team contact

Vendor-Contact

The tracker clearly documents key contacts from the external vendor or acquired company, specifying their roles such as SDTM programmer or data manager, along with detailed contact information. This setup ensures the team has clear points of contact to efficiently address specific issues or questions.

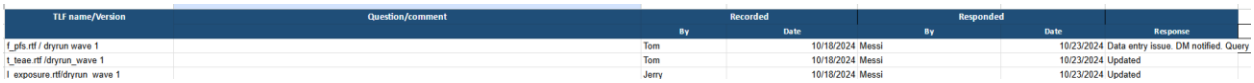
	A	B	C	D	E	F	G
	Vendor (Data)	Scope of Data	Representative	E-mail	Vendor Name	Compound	Current Data Transfer Method
	Company1 (EDC data)	EDC	Tom		Company1	Merck-001, Merck-002	SAS on Demand, then vendor sFTP
	Company2 (Lab data)	Sample Management Vendor Results data for DCL (TM)	Jerry		Company1	Merck-001, Merck-002	On vendor portal

Compound	Current Data Transfer Method	TS Comments	Merck Customer of the Data	Test Data Trans
Merck-001, Merck-002	SAS on Demand, then vendor sFTP			9/26: xxxxx.zip 9/26: xxxx.zip
Merck-001, Merck-002	On vendor portal			

Figure 4.Vendor contact

Comments Tracker

The tracker systematically captures feedback and issues identified during the transition process, clearly documenting, assigning, and resolving them. It includes columns for study name, detailed comments, date identified, responsible parties, resolution status (e.g., open, in progress, resolved), and any follow-up actions. This approach ensures transparency, accountability, and effective monitoring of progress, helping the team address all concerns promptly and efficiently.



TLF name/Version	Question/comment	Recorded		Responded		
		By	Date	By	Date	Response
f_pfs.rtf / dryrun wave 1		Tom	10/18/2024	Messi	10/23/2024	Data entry issue. DM notified. Query
t_teae.rtf /dryrun_wave 1		Tom	10/18/2024	Messi	10/23/2024	Updated
l_exposure.rtf/dryrun_wave 1		Jerry	10/18/2024	Messi	10/23/2024	Updated

Figure 5.Comments tracker

STANDARDIZATION OF INTERNAL KNOWLEDGE AND REPORTING STANDARDS

Establishing consistent internal standards significantly improves operational efficiency, ensures regulatory compliance, and maintains high-quality deliverables. A standardized internal folder structure, including consistent naming conventions and centralized storage on collaboration platforms such as SharePoint or internal servers, enables seamless teamwork and easy retrieval of files. Implementing access controls further safeguards data integrity and confidentiality, while a well-organized repository allows teams to systematically manage and access critical documentation, streamlining workflows and compliance.

Consistent and standardized internal Tables, Listings, and Figures (TLFs) are also essential to maintain submission quality and internal efficiency. Uniform footnotes and clearly defined formatting guidelines—including standardized fonts, pagination, and layout—enhance clarity and consistency across deliverables. Internally developed SAS macros automate and consistently append source data references in TLF footnotes, while periodic review and updates of TLF mock-ups ensure alignment with internal standards and evolving regulatory requirements.

PROGRAMMING EXPLORATION AND ADAPTATION

Adapting and customizing statistical programming helps align external programming assets with internal workflows and analytical standards, addressing existing limitations and inefficiencies. The internal team develops tailored statistical programs explicitly designed to close identified gaps, streamline internal analysis, and improve reporting processes. These programs utilize modular, reusable, and efficient SAS code that integrates smoothly with internal data structures and workflows. Examples include programming logic for preprocessing ADAE datasets by collapsing overlapping adverse events during AESI analyses, creating custom scripts to remap external variable names (e.g., converting START_DATE to TRTSDTM), and developing tailored programs that produce internally compliant tables, plots, and annotations consistent with organizational standards.

Systematic SAS data re-mapping and the creation of robust internal data structures are critical for seamless integration. The internal team clearly documents mappings from external variables to internal conventions, optimizing downstream processing and analysis by establishing standardized dataset formats, uniform variable naming conventions, and reusable dataset templates to minimize redundant rework in future studies. For instance, if external ADAE datasets differ structurally from internal requirements, the team proactively transforms or aggregates adverse events by SMQ to ensure compatibility with internal analytical macros and enhance overall analytical efficiency. Additionally, programmers should consult with statisticians to understand the rules behind data derivation, ensuring that any changes or combinations of different data structures produce outputs that accurately reflect the correct results.

Conclusion

In summary, successful integration of externally acquired clinical studies into internal operations requires a structured, two-fold approach. Initially, internal teams collaborate closely with external programming teams to completely transfer and accurately document datasets, programs, and associated metadata. Next, internal verification involves reproducing external outputs (ADaM datasets and TLFs) to confirm dataset compatibility, followed by rigorous CDISC compliance validation using Pinnacle 21 for both SDTM and ADaM datasets. Concurrently, the internal team establishes comprehensive, centralized knowledge repositories, employs structured operational tools such as delivery trackers and regular meetings, standardizes internal documentation and reporting conventions, systematically re-maps external datasets, and customizes statistical programming to align with internal analytical standards. Collectively, these

structured steps guarantee regulatory compliance, improve internal efficiency, and enable seamless, high-quality integration of externally sourced clinical studies into internal workflows.

Looking ahead, we will need to continuously manage ongoing complexities and implement proactive solutions, even after successful integration. Maintaining consistency and compliance becomes critical as regulatory requirements evolve and internal standards change. Ensuring sustained accuracy and integrity of integrated datasets during subsequent data updates, protocol amendments, or interim analyses presents additional challenges. Furthermore, retaining and effectively transferring knowledge among internal team members becomes increasingly important during staffing changes or organizational restructuring. Future acquisitions of additional studies may further test existing infrastructure and resources, underscoring the importance of scalable solutions. Proactive strategies—including regular training, thorough documentation practices, robust version control, and periodic reassessment of data and programming standards—will enable the organization to successfully address these challenges and sustain operational excellence.

ACKNOWLEDGMENTS

The authors would like to thank Hong Qi and Qi Liu for their guidance.

CONTACT INFORMATION

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

Name: Chunqiu Xia
Enterprise: Merck & Co., Inc., Rahway, NJ, USA
Address: 126 East Lincoln Avenue 7 Public City, State ZIP: Rahway, NJ 07065-4607
E-mail: chunqiu.xia@merck.com
Web: www.merck.com

Name: Lily Zhang
Enterprise: Merck & Co., Inc., Rahway, NJ, USA
Address: 126 East Lincoln Avenue 7 Public City, State ZIP: Rahway, NJ 07065-4607
E-mail: lily_zhang2@merck.com
Web: www.merck.com

Name: Xiaohui Wang
Enterprise: Merck & Co., Inc., Rahway, NJ, USA
Address: 126 East Lincoln Avenue 7 Public City, State ZIP: Rahway, NJ 07065-4607
E-mail: xiaohui_wang@merck.com
Web: www.merck.com