PharmaSUG China 2025 - Paper DS-134

Implementing CDISC SDTMIG v3.4: Practical Strategies and Lessons Learned from Real-World Programming

Liming Xie, Chunpeng Zhao, BeOne Medicines Co., Ltd.

ABSTRACT

The SDTMIG v3.4, released by CDISC on November 29, 2021, introduces significant enhancements, new domains, and clarified guidance, impacting clinical trial data and regulatory requirements compared to v3.3. This paper addresses the real-world challenges statistical programmers face during this transition, including updating SDTM mapping, integrating new terminology and variables, ensuring ADaM consistency, and adapting validation processes to meet Pinnacle 21 and regulatory expectations. Increased domain complexity, new study structures, and evolving subject visit expectations necessitate careful planning and cross-functional coordination. We present practical solutions from pilot studies, demonstrating how process optimization and automation tools can streamline implementation and maintain traceability. The paper details how SDTM-level changes affect ADaM derivations and TFL programming. Attendees will gain insights to efficiently manage this upgrade, minimize rework, and confidently ensure compliance with the latest CDISC standards.

INTRODUCTION

The release of the Study Data Tabulation Model Implementation Guide (SDTMIG) v3.4 by the Clinical Data Interchange Standards Consortium (CDISC) on November 29, 2021, marks a critical milestone in the evolution of clinical data standards. This latest version introduces structural refinements, additional domains, updated controlled terminology, and clarified implementation guidance designed to meet the increasing complexity of modern clinical trials and the expectations of regulatory agencies.

According to the U.S. FDA Data Standards Catalog, clinical studies that start after March 15, 2025, must comply with SDTMIG v3.4 for regulatory submissions. The Japan PMDA will begin supporting SDTMIG v3.4 starting April 1, 2025.

This paper expands on the structural, technical, and operational implications of the transition, offering granular insights for statistical programmers. Drawing from pilot studies, industry best practices, and CDISC's implementation guides, we provide actionable strategies to mitigate risks and leverage opportunities inherent in this update. This comprehensive review aims to equip stakeholders with the necessary knowledge to navigate the challenges and capitalize on the advancements introduced by SDTMIG v3.4, ultimately enhancing data quality, streamlining regulatory submissions, and fostering greater consistency across clinical trials.

KEY CHANGES IN SDTMIG V3.4

Structural and Domain-Level Enhancements

SDTMIG v3.4 introduces structural and domain-level enhancements to support emerging study designs and accommodate greater granularity and flexibility in data capture.

1. **New Domains**

- **BE (Biospecimen Events), BS (Biospecimen Findings) and RELSPEC (Related Specimens)**: These new domains were copied and minimally updated from the provisional SDTMIG for Pharmacogenomics and Pharmacogenetics (SDTMIG-PGx), published 2015-05-26. BE and BS domains enable precise tracking of biospecimen lifecycle events (e.g., collection, storage, analysis) and genomic findings, which are critical for precision medicine trials; RELSPEC domain represents relationships between specimens.
- **CP (Cell Phenotype Findings)**: This domain contains data related to the characterization of cell phenotype, lineage, and function based on expression of specific markers in single cell or particle suspensions.

- **GF (Genomics Findings)**: This domain facilitates structured reporting of genomic variants (e.g., SNPs, CNVs) with standardized terminology and replaces PF (Pharmacogenomics Findings) domain from SDTMIG-PGx.

2. **Decommissioned Domain**

   The MO (Morphology Findings) domain, which was initially introduced in SDTMIG v3.2 to represent morphology findings separate from physiology findings by body system, has been deprecated in SDTMIG v3.4. The Submissions Data Standards (SDS) team decided to deprecate the MO domain and replace it with body system-based domains covering both morphology and physiology findings due to the difficulty in separating these two types of findings.

3. **Updated Scope of SV (Subject Visits) Domain**

   Following SDTMIG v3.4, the SV domain now allows for the submission of data on the timing of trial visits for a subject, including not only those visits they actually passed through but also those visits that did not occur. This expands upon SDTMIG v3.3, which only included visits that actually occurred.

4. **Updated Definition of IS (Immunogenicity Specimen Assessments) Domain**

   Previously in SDTMIG v3.3, the IS domain holds assessments that describe whether a therapy provoked/caused/induced an immune response.

   Starting with SDTMIG v3.4, the IS domain also holds the following assessments:

   - assessments that describe whether an allergen, microorganism, or endogenous molecule provoked/caused/induced an immune response in a subject, such as a subject's antibody reaction (autoantibodies) against auto/self-antigens for autoimmune studies or antibody production in response to allergens in allergy trials.
   - assessments about all other types of "induced" humoral immune response in a subject.
   - certain types of cellular immune responses using non-flow cytometry techniques.

Clarified Guidance and Updated Controlled Terminology

The update also includes clarified implementation guidance, helping to reduce ambiguity and improve standardization across studies. With the integration of new controlled terminology and additional variables, teams must carefully review mapping rules and data derivation logic to ensure alignment.

1. **Added Variables**

   68 variables have been added to domains like AE, CE, DA, EG, IS, LB, PP, QS, RS, SC, VS, EC, DM, SE, and SV. Some variables, such as --CLSIG, --PTFL, and --PDUR, were moved from the supplemental domains in SDTMIG v3.3 to the main domains in SDTMIG v3.4.

2. **Updated Variable Labels**

   The labels of 10 variables have been updated. Table 1 provides a comprehensive list of these changes.

Dataset Name	Variable Name	Variable Label (Updated)	Variable Label (Previous)
DV	DVENDY	Study Day of End of **Deviation Event**	Study Day of End of **Observation**
FT	FTMETHOD	Method of Test **or Examination**	Method of Test
LB	LBTESTCD	Lab Test or Examination Short Name	Lab Test or Examination Short Name.
SM	DOMAIN	Domain **Abbreviation**	Domain
SV	SVENDTC	End Date/Time of **Observation**	End Date/Time of **Visit**
SV	SVENDY	Study Day of End of **Observation**	Study Day of End of **Visit**
SV	SVSTDTC	Start Date/Time of **Observation**	Start Date/Time of **Visit**

SV	SVSTDY	Study Day of Start of **Observation**	Study Day of Start of **Visit**
TM	DOMAIN	Domain **Abbreviation**	Domai**n**
TS	TSVALNF	Parameter **Value** Null Flavor	Parameter **Null Flavor**

Table 1. The list of variable labels updated in SDTMIG v3.4 comparing with SDTMIG v3.3

REAL-WORLD CHALLENGES

The changes introduced in SDTMIG v3.4 have a cascading effect across the full spectrum of deliverables handled by statistical programmers. Understanding these challenges is crucial for effective mitigation.

IMPACT ON PROGRAMMING WORKFLOWS

The transition to SDTMIG v3.4 necessitates significant adjustments to existing programming workflows:

1. **SDTM Mapping Specifications**

   Revisions are required to reflect new domains, variables, and terminologies. These changes often demand extensive review of current mapping documents and rework of existing programming code.

2. **ADaM Consistency**

   Since many ADaM derivations rely on SDTM inputs, even minor changes at the SDTM level can introduce inconsistencies or errors in downstream datasets. This necessitates thorough impact analysis and careful re-validation of ADaM datasets.

3. **TFL Outputs**

   As ADaM datasets are used to produce Tables, Figures, and Listings (TFLs), any instability or inconsistency in upstream data can compromise the accuracy and reproducibility of statistical outputs. Programmers must ensure TFLs accurately reflect the updated data structures and content.

4. **Validation Processes**

   Pinnacle 21 validation rules must be re-evaluated in the context of SDTMIG v3.4, as updates may trigger new warnings or errors. Adjustments to validation scripts and checklists are necessary to maintain compliance and ensure data quality.

CHALLENGES IN IMPLEMENTING SDTMIG V3.4

1. **Integration of New Domains**

   The introduction of new domains presents unique challenges:

   - **Complexity Increase**: The addition of five new domains (BE, BS, RELSPEC, GF, CP) requires programmers to adapt to new structures and understand their content, which can be complex due to unfamiliar data types and structures.
   - **Mapping Challenges**: Establishing clear mapping rules from these new SDTM domains to ADaM datasets is necessary. This involves thorough documentation and careful consideration of how these new data points will be utilized in statistical analysis.

2. **Updated SV (Subject Visits) Domain**

The expanded scope of the SV domain introduces new complexities:

- **Handling Missing Visits**: The inclusion of non-occurring visits in the SV domain may complicate the interpretation and derivation of visit-based data in ADaM datasets. Programmers will need to implement specific logic to differentiate between actual and missed visits, impacting calculations like visit compliance or exposure windows.
- **Impact on TFLs**: TFLs may need adjustments to display or analyze visit data accurately, including scenarios where visits were planned but did not occur.

3. **Expanded IS (Immunogenicity Specimen Assessments) Domain**

The broadened definition of the IS domain impacts data handling and analysis:

- **Diverse Data Types**: The wider scope now encapsulates various types of immune responses, necessitating versatile mapping and derivation strategies in ADaM datasets. This requires a deeper understanding of the biological context of these assessments.
- **Complex Analysis Requirements**: Analysis of immunogenicity might become more complex, requiring comprehensive datasets that accurately reflect the new types of assessments. This could involve developing new analysis dataset or variables.

4. **Clarified Guidance and Updated Controlled Terminology**

Updates to guidance and terminology bring their own set of challenges:

- **Controlled Terminology Updates**: Introducing new or updated terminology often leads to mismatches with legacy data, requiring harmonization across historical and current studies. Managing terminology version control becomes critical for ensuring consistency and traceability.
- **Integration of Additional Variables**: With 68 new variables and updates to existing ones, there is a need for extensive review and potential modification of ADaM datasets and their associated analysis logic to thoughtfully incorporate these variables. This includes assessing their impact on existing derivations and analysis plans.

5. **Cross-Functional Dependencies**

Successfully transitioning to SDTMIG v3.4 is a collaborative task. It requires close coordination between data managers, statisticians, and statistical programmers to ensure all aspects of the study data lifecycle are aligned with the new standard. Lack of inter-departmental collaboration can lead to misinterpretations, inconsistencies, and delays in regulatory submissions.

MITIGATION STRATEGIES AND BEST PRACTICE

Navigating the complexities of SDTMIG v3.4 requires a proactive and multi-faceted approach. Based on the identified real-world challenges, the following strategies offer actionable solutions to mitigate risks and leverage opportunities.

SPECIFICATION MANAGEMENT AND CHANGE CONTROL

Effective management of documentation and change is paramount to a smooth transition:

1. **Robust Version Control**

   Implement a comprehensive version control system for all specifications, with detailed change logs that clearly track updates related to SDTMIG v3.4. This promotes standardization, ensures changes are accessible to relevant stakeholders, and supports traceability and consistency across projects.

2. **Standard Operating Procedures (SOPs)**

   Develop or revise SOPs to include clear guidance on managing and implementing specification changes. This should cover protocols for introducing new domains, updating existing mappings, and aligning study-level documentation with evolving standards.

3. **Transparent Communication**

   Establish routine communication channels, such as regular update meetings or a centralized document repository, to keep all stakeholders informed about specification changes and their implications. This is vital for addressing the cross-functional dependencies and ensuring a shared understanding of the evolving requirements.

4. **Proactive Adoption of New SDTMIG Standards**

   Whenever possible, it is considered best practice to adopt new domains and variables from newer versions of the SDTM Implementation Guide (SDTMIG), even if your company is primarily following

an earlier version. Doing so helps ensure that your data structure stays aligned with the most up-to-date CDISC standards, supports future regulatory expectations, and facilitates data integration across studies.

For example, while implementing SDTMIG v3.3, you may choose to adopt newly introduced variables from SDTMIG v3.4 in the Supplemental Qualifiers (SUPP--) domain. This proactive approach minimizes the need for retrospective updates, promotes consistency with future studies, and can help streamline the submission process by reducing discrepancies between study datasets and the evolving standards.

Before adopting elements from a newer version, be sure to assess their compatibility with your current data and analysis plan, and document any deviations from the official version in your Study Data Reviewer's Guide (SDRG) or relevant submission documentation.

SDTM AUTOMATION TOOLS

Leveraging automation can significantly improve efficiency, reduce manual errors, and support consistent compliance with evolving SDTMIG standards:

1. **Develop Automated Mapping Tools**

   Design and implement tools that can automatically map raw data to updated SDTM domains and variables. Automation reduces manual effort and human error while allowing for rapid scalability as new domains or variables are introduced. These tools are especially valuable for managing the increased complexity of expanded domains such as IS and the integration of numerous new variables.

2. **Deploy Automated Validation Scripts**

   Utilize automated validation scripts to verify the accuracy and compliance of datasets against SDTMIG v3.4 requirements. Automated checks help maintain data integrity, ensure alignment with updated Pinnacle 21 validation rules, and proactively identify any issues before submission.

3. **Leverage Machine Learning Capabilities**

   Incorporate machine learning algorithms to predict potential mapping challenges or integration issues based on historical study data and known patterns. This can further enhance efficiency, strengthen data quality, and support complex transformations, particularly for updated domains like SV and IS.

PILOT STUDIES

Conducting pilot programs provides critical insights and opportunities to refine processes before full-scale adoption of SDTMIG v3.4:

1. **Implement Pilot Programs**

   Execute small-scale pilot projects that apply SDTMIG v3.4 standards to representative sample datasets. Pilots help uncover unforeseen challenges in a controlled setting — for example, addressing the complexities of managing non-occurring visits in the SV domain or handling diverse data structures in the IS domain — before broader rollout.

2. **Establish Iterative Feedback Loops**

   Use pilot study results to fine-tune processes, specifications, and mapping strategies. Set up clear feedback channels with key stakeholders to address issues early, incorporate improvements iteratively, and ensure that valuable lessons inform future implementations.

3. **Document and Share Lessons Learned**

   Maintain thorough documentation of all lessons learned during pilot activities, including effective strategies, challenges encountered, and mitigation steps. Sharing this knowledge across teams helps build institutional best practices and supports smoother adoption in future studies.

IMPLICATIONS FOR ADAM AND TFL PROGRAMMING

Adapting ADaM datasets and TFL programming is essential to ensure accurate analysis and reporting aligned with SDTMIG v3.4:

1. **Align ADaM with SDTMIG v3.4 Updates**
    - **Refine ADaM Specifications**: Develop or update ADaM specifications to incorporate new domains (e.g., BE, BS, CP, GF) and variables introduced in SDTMIG v3.4. Address the decommissioning of the MO domain by remapping relevant data to appropriate body system-based domains and integrating updated immunogenicity (IS) assessments.
    - **Enhance Derivation Logic for Visit Data**: Modify ADaM datasets to distinguish between actual and non-occurring visits, as required by the revised SV domain structure. This supports accurate derivations for visit-based analyses, such as defining analysis windows or calculating exposure, and mitigates challenges associated with missing visits.
2. **Ensure TFL Outputs Accurately Reflect New Data**
    - **Content Validation**: Confirm that TFL outputs incorporate new variables (e.g., --CLSIG, --PTFL) and relevant domain-specific data (e.g., CP, GF) sourced from updated ADaM datasets, even when using existing templates. Outputs should meet regulatory expectations for evolving study designs, including those focused on immunogenicity, autoimmune, or allergy studies.
    - **Address Missing Visits in TFLs**: Update TFL logic to account for planned but non-occurring visits, ensuring that patient disposition, visit compliance, and related summaries correctly reflect SV domain updates.
3. **Implement Iterative Validation and Feedback Loops**
    - **Cross-Check ADaM and TFL Consistency**: Use automated validation checks to verify that ADaM datasets and TFL outputs consistently align with SDTMIG v3.4 updates, including new terminology and variables.
    - **Leverage TFL Outputs Insights**: Use discrepancies or gaps identified during TFL generation to refine ADaM derivations, particularly for complex immunogenicity assessments or re-mapped morphology/physiology data.

CROSS-FUNCTIONAL REVIEW FORUMS

Establishing cross-functional review forums is a key element of an effective SDTMIG v3.4 implementation strategy. These forums bring together representatives from data management, statistical programming, biostatistics, and other relevant functions to ensure a consistent and aligned interpretation and application of the updated standards across studies.

Regular, structured discussions create a platform to proactively identify and resolve implementation challenges, share best practices, and document collaborative decisions for transparency and future reference. These forums help maintain consistency across studies, support impact assessments for related process changes, and identify emerging training needs as standards evolve.

By fostering shared accountability and strengthening interdepartmental collaboration, cross-functional review forums play a vital role in delivering high-quality, submission-ready datasets that fully meet regulatory expectations under SDTMIG v3.4.

CONCLUSION

The adoption of SDTMIG v3.4 represents a pivotal evolution in clinical data standards, demanding meticulous planning, coordinated execution, and a proactive mindset from all stakeholders. While this transition presents tangible real-world challenges across programming workflows, data integration, and validation processes, it simultaneously offers invaluable opportunities to modernize data management practices, significantly enhance data quality, and bolster regulatory readiness.

By strategically leveraging pilot implementations, optimizing existing processes, and deploying advanced automation tools, statistical programmers are well-equipped to navigate this transition with confidence.

Crucially, by anticipating downstream impacts on ADaM datasets and TFLs, and by fostering robust cross-functional collaboration through established review forums, teams can transcend mere compliance. This holistic approach will not only ensure adherence to the new standard but will also drive unparalleled efficiency, consistency, and analytical rigor throughout the entire clinical data lifecycle, ultimately accelerating the delivery of safe and effective therapies.

REFERENCES

Certara. 2025. "Exploring Changes in SDTMIG 3.4 & ADaMIG 1.3." Accessed June 23, 2025. https://www.certara.com/on-demand-webinar/exploring-changes-in-sdtmig-3-4-adamig-1-3/.

ACKNOWLEDGMENTS

Thanks to our managers (Leo Li, Cindy Song and Tony Guo) for their great support on the conference and programming colleagues for their help on preparing the paper.

RECOMMENDED READING

- *CDISC SDTMIG v3.4*

CONTACT INFORMATION

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

Liming Xie
BeOne Medicines (Beijing) Co., Ltd. (formerly BeiGene (Beijing) Co., Ltd.)
liming.xie@beigene.com

Chunpeng Zhao
BeOne Medicines (Shanghai) Co., Ltd. (formerly BeiGene (Shanghai) Co., Ltd.)
chunpeng.zhao@beigene.com