Paper DS11

Evolving with SDTM-IG v3.4: What Clinical Data Professionals Need to Know

Prashanth Jatothu, Ephicacy Life Science Analytics, Bangalore, India
Nagaraju Salugu, Ephicacy Life Science Analytics, Bangalore, India

ABSTRACT

CDISC continues to refine its data standards to meet evolving regulatory requirements and the increasing complexity of clinical trials. With SDTM-IG v3.4 recommended for implementation starting March 2025, organizations are encouraged to align with updated expectations and best practices. This abstract highlight key updates in SDTM-IG v3.4 compared to SDTM-IG v3.3, offering insights for clinical data professionals, programmers, and regulatory teams. Enhancements include new domains like Genomics Findings (GF) and Cell Phenotype Findings (CP), replacement of SUPPQUAL with standardized variables, improved clarity in inter-domain relationships, and stronger metadata guidance.
These changes impact SDTM mapping and downstream processes such as ADaM Specification and programming, QC validation, and submission packages. Understanding and applying these updates is essential for regulatory compliance, data traceability, and submission readiness. This presentation provides practical guidance to help teams adapt to SDTM-IG v3.4 and future-proof their clinical data standards strategy.

INTRODUCTION

The clinical data landscape is rapidly evolving, driven by advances in precision medicine, increasingly complex trial designs, and heightened regulatory expectations for data traceability and standardization. In response, CDISC has released **SDTM-IG v3.4**, which introduces significant enhancements beyond **SDTM-IG v3.3** that directly impact how clinical trial data are structured, reviewed, and submitted.

SDTM-IG v3.4 includes several important updates to better support modern clinical trials. Specimen-based laboratory domains such as **Cell Phenotype Findings (CP)**, **Genomics Findings (GF)**, and **Laboratory Tests (LB)** are now grouped with a generic specification for greater consistency. These include the introduction of new domains such as **Genomics Findings (GF) and Cell Phenotype Findings (CP)** to better support molecular and cellular data that were previously handled using non-standard or workaround approaches.

The **IS domain** has been enhanced with new variables to more comprehensively capture immune response assessments, while the **LB domain** includes additional variables to provide more detailed laboratory test information. Several biospecimen-related domains have been incorporated from the provisional pharmacogenomics SDTMIG to streamline related data collection. Additionally, corrections to core variable values and updated controlled terminology across domains improve overall data accuracy and standardization.

Importantly, SDTM-IG v3.4 encourages replacing **SUPPQUAL** with standardized variables to improve data consistency and downstream usability. It also provides clearer guidance on inter-domain relationships to strengthen traceability across related datasets and introduces stronger metadata requirements to support more efficient regulatory review.
This presentation, *"Evolving with SDTM-IG v3.4: What Clinical Data Professionals Need to Know,"* aims to equip clinical data managers, programmers, and regulatory teams with practical insights and implementation considerations to help them adapt to these changes and future-proof their clinical data standards and submissions.

INTRODUCTION TO NEW DOMAINS

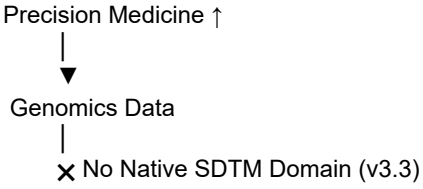
1. Genomics Findings (GF):

Why Genomics Broke SDTM-IG v3.3?

SDTM-IG v3.3 was created before genomics became a routine part of clinical trials. As precision medicine evolved, studies began collecting complex genomic data such as gene mutations and molecular biomarkers. However, SDTM-

IG v3.3 did not include a dedicated domain or standardized variables to represent this information, leading to inconsistent and non-scalable solutions.

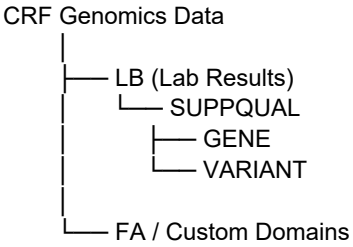
Diagram:



How We Mapped Genomics Before GF?

In the absence of a genomics-specific domain, sponsors relied on workarounds. Genomics results were most mapped to the LB domain, with critical details such as gene and variant stored in SUPPQUAL. In some cases, FA or custom domains were used. These approaches increased complexity and reduced data clarity.

Diagram:



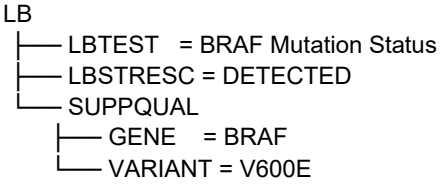
What Is the GF Domain?

The GF domain in SDTM IG 3.4 is designed to store genomic data collected from clinical trials, including information on genetic variations, sequencing results, and biomarkers. A findings domain that contains data related to the structure, function, evolution, mapping, and editing of subject and non-host organism genomic material of interest.
Genomic data helps in personalized medicine (e.g., identifying patient-specific drug responses), used for biomarker discovery in clinical trials, essential for pharmacogenomics and required for regulatory submissions to agencies like FDA and PMDA.

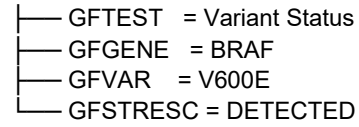
EXAMPLE: BRAF Mutation Status

This example highlights how a BRAF V600E mutation result was captured before and after the introduction of the GF domain.

SDTM-IG v3.3 (LB + SUPPQUAL)



SDTM-IG v3.4 (GF)



2. Cell Phenotype Findings (CP):

Why Cell Phenotype Data Was Challenging in SDTM-IG v3.3?

SDTM-IG v3.3 did not include a dedicated domain for cell phenotype data. As immunology and cell-based therapies became more common, complex data such as immune cell subsets and marker expression could not be represented cleanly.

How Cell Phenotype Data Was Mapped in SDTM-IG v3.3?

Without a CP domain, sponsors used LB, FA, and SUPPQUAL to store cell phenotype results. Critical attributes like cell type and markers were often captured in SUPPQUAL.

Diagram (v3.3 workaround):

CRF Cell Phenotype Data
- LB (Results)
  - SUPPQUAL
    - CELL TYPE
    - MARKER
- FA / Custom Domains

What Is the CP Domain in SDTM-IG v3.4?

CP domain that contains data related to the characterization of cell Phenotype, Cell lineage, and function based on the expression of specific markers in single cell or particle suspensions.
The CP domain is modeled for use with disseminated tissue specimens (e.g., blood and other body fluids, bone marrow aspirates) and cell suspensions and is not currently modeled for evaluations of solid tissue specimens. The domain is intended to support tests associated with a cell phenotyping component based on the use of markers and is not intended for tests that are not associated with marker-based phenotyping, which are more appropriate to include in another domain (e.g., Immunogenicity Specimen Assessments (IS), Laboratory Test Results (LB), Microscopic Findings (MI)). The CP domain is not intended to supplant use of the LB domain for routine lab hematology (e.g., blood cell differentials), nor is it intended for findings originating from microscopic assessment of cells, including those employing immunohistochemical (IHC) techniques.

EXAMPLE: CD4+ T-Cell Percentage

SDTM-IG v3.3 (LB + SUPPQUAL)

Cell phenotype results were forced into LB, with cell and marker information stored separately.
Diagram:
LB
- LBTEST = CD4 Positive Cells
- LBSTRESN = 35
- LBSTRESU = %
- SUPPQUAL
  - CELLTYPE = T-Cell
  - MARKER = CD4

SDTM-IG v3.4 (CP Domain)

All cell phenotype attributes are captured natively within CP.
Diagram:
CP
- CPTEST = Cell Percentage
- CPCELL = T-Cell
- CPMARKER= CD4
- CPSTRESN= 35

└── CPSTRESU= %

Replacement of SUPPQUAL with Standardized Variables:

Why SUPPQUAL was used in SDTM-IG v3.3?

In SDTM-IG v3.3, the availability of standard variables was limited, particularly for emerging and complex data types. As clinical trials evolved to include genomics, cell phenotype assessments, and detailed test conditions, many of these data elements did not have a defined place within the parent SDTM domains. To accommodate these gaps, SUPPQUAL was commonly used to store additional attributes that could not be represented using existing standard variables, providing flexibility for sponsors to capture study-specific information.

Challenges faced with SUPPQUAL usage:

1. Critical information stored in SUPPQUAL was fragmented and inconsistent, complicating data interpretation and cross-study comparisons
2. Additional effort and documentation were required for analysis and regulatory review due to non-standard implementations

Why SDTM-IG v3.4 changed the approach?

To address these limitations, SDTM-IG v3.4 introduced new standardized variables and expanded domain specifications to better support commonly collected data elements. The updated guidance encourages mapping key attributes directly within parent domains whenever possible, reducing reliance on SUPPQUAL. This shift reflects CDISC's intent to promote greater standardization and clarity in SDTM datasets while supporting the increasing complexity of modern clinical trials.

Benefits of using standardized variables in SDTM-IG v3.4:

1. Mapping to standardized parent-domain variables improves data clarity, consistency, and traceability from CRF → SDTM → ADaM
2. Facilitates faster, more efficient analysis, validation, and regulatory review
3. Supports automation, data reuse, and submission readiness

EXAMPLE: XXCLSIG and XXREASOC: SUPPQUAL → Parent Domain

In SDTM-IG v3.3, clinically significant (XXCLSIG) and Reason for Occur Value (XXREASOC) were often stored in SUPPQUAL, leading to fragmented data, inconsistent naming, and extra effort for traceability and review.

SDTM-IG v3.4 moves these variables into parent domains (e.g., LB, VS, EG, EC, SV, SU), improving data consistency, traceability, and regulatory readiness while reducing reliance on study-specific mappings.

Improved Clarity in Inter-Domain Relationships:

SDTM-IG v3.4 provides clearer guidance on how domains relate to each other, making it easier to understand and navigate complex clinical datasets. Previously, in v3.3, relationships between findings domains, interventions, and events were sometimes implied or scattered across supplemental documentation, which made data traceability and analysis more challenging.

Benefits:
Links between parent domains and related findings domains are explicitly defined (e.g., linking LB, IS, and GF/CP to Subject or Specimen domains).
Cross-domain relationships, such as between **Biospecimen Events (BE), Biospecimen Findings (BS), and Related Specimens (RELSPEC)**, are now formally specified, enhancing traceability.

EXAMPLE: Immunogenicity Data Modelling – SDTMIG v3.3 vs v3.4

v3.3 Approach:

- **Domain Placement:** Antibody tests against a virus could be in **LB** or **MB**.
- **Linking to Exposure:** RELREC was used to connect tests to **EX**.

- **Challenges:**
    - RELREC implementation was inconsistent.
    - Target of the immune response was unclear.

Example (v3.3):

LBTESTCD	LBORRES	RELREC Link to EX
AVIDITY	High	Yes

Link to intervention relied entirely on RELREC; antigen was not specified.

v3.4 Improvements:

- **Domain Expansion:** IS now includes immune responses to drugs, allergens, or microbes.
- **New Variables:**
    - **ISBDAGNT:** Specifies the antigen/target.
    - **ISCNDAGT:** Captures relevant test agents.
- **Benefits:**
    - Direct, standardized linking to interventions.
    - Clear identification of antibody targets.

Example (v3.4):

ISTESTCD	ISORRES	ISBDAGNT
AVIDITY	High	COVID-19 S-Protein

ISBDAGNT clearly identifies the antigen, removing reliance on RELREC.

Stronger Metadata Guidance: Improvements from SDTM-IG v3.3 to v3.4

SDTM-IG v3.4 introduces enhanced metadata guidance to support the increasing complexity and diversity of clinical data. Compared to v3.3, this version provides more detailed and precise metadata standards, ensuring datasets are consistent, accurate, and easier to interpret across evolving domains.

Key improvements include:

New Variables and Domains: The addition of new variables in domains like IS (Immune Response) and LB (Laboratory Tests) is supported by clear metadata definitions, ensuring consistent capture and interpretation of complex biological data.

Incorporation of Biospecimen-Related Domains: Domains from the provisional pharmacogenomics SDTMIG have been standardized and integrated with strong metadata rules, streamlining biospecimen data collection and linking.

Core Variable Corrections and Controlled Terminology: Corrections to core variables and updates in controlled terminology are carefully documented with precise metadata to improve data consistency and regulatory compliance.

Overall, the stronger metadata guidance in SDTM-IG v3.4 helps ensure that increasingly complex datasets are well-documented, interoperable, and aligned with regulatory expectations, facilitating more efficient data review and analysis compared to the more limited metadata standards of v3.3.

EXAMPLE: Binding Agent: SUPPIS (v3.3) → IS Parent Domain (v3.4)

In **SDTM-IG v3.3**, the binding agent was not a standard IS variable and was captured in **SUPPIS** using study-specific QNAM/QVAL values. This separated assay context from results and reduced consistency.

In **SDTM-IG v3.4**, the binding agent is a **standardized variable in the IS parent domain**, allowing context to be captured directly with immune response results.

Benefits:

Clarity & consistency: Results are self-describing and standardized across studies.

Traceability & review: Eliminates SUPPIS reliance, simplifying CRF → SDTM → ADaM traceability and regulatory review.

SDTM-IG v3.3
SUPPIS
├── QNAM = ISBDAGNT
├── QLABEL = Binding Agent
├── QVAL = NUCLEAR AUTOANTIGENS

SDTM-IG v3.4
IS
├── ISBDAGNT = NUCLEAR AUTOANTIGENS

CONCLUSION

In this paper, we have explained how the transition from SDTMIG v3.3 to v3.4 brings substantial improvements and key differences in the representation of clinical data compared to the prior version. We observe that domains such as Genomics Findings (GF) and Cell Phenotype Findings (CP) are now more clearly defined, providing structured approaches to capture specialized biological information. The replacement of SUPPQUAL with standardized variables enhances data consistency, improves traceability, and reduces ambiguity. Additionally, inter-domain relationships are more clearly articulated, enabling precise linkage between related datasets, while strengthened metadata guidance supports consistent implementation across studies. Overall, we believe these updates significantly enhance clarity, standardization, and traceability, resulting in more robust, interpretable, and regulatory-ready clinical data.

ACKNOWLEDGMENTS

We sincerely thank Efficacy Lifescience Analytics for the opportunity to contribute to this paper. The perspectives presented herein are solely those of the authors and do not necessarily represent the views of Efficacy Life Science Analytics.

CONTACT INFORMATION

(Your comments and questions are valued and encouraged. Contact the author at:
Author Name: Prashanth Jatothu, Nagaraju Salugu
Company: Ephicacy lifescience Analytics
Address: Bengaluru, India
Email: prashanth.jatothu@ephicacy.com, nagaraju.salugu@ephicacy.com
Website: www.ephicacy.com