PharmaSUG 2026 - Paper DS-265

SDTMIG v4.0: Are You Ready For It?

Kristin Kelly, Pinnacle 21 by Certara, Radnor, PA, USA

ABSTRACT

The imminent release of the CDISC SDTMIG v4.0 / SDTM v3.0 represents a pivotal moment for SDTM standards. As a major version update, it introduces significant changes that sponsors should begin preparing for now. Key updates include support for Multiple Subject Instances (MSI), the transition from SUPPQUAL to Non-Standard Variables (NSVs), and metadata restructuring aligned with SDTM Model v3.0. The –BLFL variable has been removed and Sections 1–4, which provide guidance across SDTM domains, have been reorganized for improved clarity and usability. Variables are now organized into Variable Groups, enhancing structure and interpretation. New domains such as Event Adjudication (EA) and Gastrointestinal Findings (GI) further expand the ability to handle more types of data. Also noteworthy are updates to the Protocol Deviations (DV) domain and specimen-based Findings domains. This paper will explore what's new in SDTMIG v4.0 and provide practical guidance to help sponsors navigate this era of transition.

INTRODUCTION

The release of SDTMIG v4.0 / SDTM v3.0 marks a defining moment in SDTM. For many years, the SDTMIG has carried forward familiar structures, assumptions, and guidance. But as clinical research has grown more complex, there is a need for SDTM to evolve to handle new types of data. SDTMIG v4.0 ushers in a new chapter with updated metadata foundations, modernized domain structures, and clearer guidance to support today's study designs.

This paper walks through the major changes in SDTMIG v4.0 from the debut of Multiple Subject Instances (MSI) and NSVs to the restructuring of Sections 1–4 and introduction of new domains and variables. The goal is to help sponsors understand how we got here, what's changing, and how to prepare for where we're going.

STUDY DATA TABULATION MODEL (SDTM) V3.0

Before forging ahead to SDTMIG v4.0, it is essential to discuss the foundation on which it is built, SDTM Model v3.0. The SDTM Model contains the variable and domain metadata that is used for tabulation of clinical and nonclinical data. This metadata is further described in the SDTMIG (clinical trials) and the SENDIG (nonclinical trials). Each implementation guide is paired with a specific version of the SDTM Model that supports it, in this case, SDTMIG v4.0 goes with SDTM v3.0. Significant changes and additions to this version are discussed in the following sections.

NEW DOMAINS

New domains added to SDTM v3.0 that are applicable to human clinical trials are the following:

- Non-standard Variables (NS--) Dataset (Relationship) - replaces SUPPQUAL
- Demographics for Multiple Participations (DC) (Special-purpose) – supports MSI
- Related Devices Dataset (RELDEV) (Study Reference) – supports upcoming SDTMIG-Medical Devices v2.0

Further information for DC and NS—will be discussed in subsequent sections.

NEW VARIABLES

The following table lists the new variables added in SDTM v3.0 that are applicable for human clinical trials.

Class	Variable Name	Variable Label	Root Variable Definition	Usage Restrictions
Interventions	--TRTCD	Standardized Intervention Code	A standardized or dictionary-derived short sequence of characters used to represent the intervention.	
Events	--CLASI	Classification of Protocol Deviation	A classification of protocol deviations based on the potential impact to the completeness, accuracy, and/or reliability of the study data, or to a subject's rights, safety, or well-being. (ICH E3 Q&As (R1))	DV domain only
Findings	--RSCNT	Result Count	The number of occurrences of a result.	EG domain only
Findings	--CBRFL	Conditionally Branched Item Flag	An indication that this instance of an item within a QRS instrument was conditionally branched. An item is conditionally branched when the instrument does not solicit a response from the respondent based on a given condition, although for certain instruments, the instrument may provide a result for the item. Allowed value: Y	QS, FT, RS (clinical classification use case) only
Special-purpose	AGERLO	Age Range Lower Limit	A numeric representation of the lowest possible value for the age of the subject at a specific point in time defined for the trial. Replaces AGETXT.	DM and DC only
Special-purpose	AGERHI	Age Range Upper Limit	A numeric representation of the highest possible value for the age of the subject at a specific point in time defined for the trial. Replaces AGETXT.	DM and DC only
Special-purpose	CRACE	Collected Race	A classification of race as specified by the sponsor in the data collection field.	DM and DC only
Special-purpose	CETHNIC	Collected Ethnicity	A classification of ethnicity as specified by the sponsor in the data collection field.	DM and DC only
Non-Standard Variables	IDVARVLN	Identifying Variable Numeric Value	The numeric value of the variable named in IDVAR used to identify related records.	

VARIABLE DEFINITIONS, GROUPS, AND RELATIONSHIPS

In SDTM v2.0, not all the variables had definitions. In SDTM v3.0, all variables now have Variable Definitions that were reviewed and approved by the CDISC CT, SDTM and appropriate foundational teams. These definitions have been harmonized with the NCI Thesaurus and consistent with the Variable Definitions in the SDTMIG.

All variables have been organized into Variable Groups based on their purpose. A ‘Variable Group’ column has been added to all specification tables in both the SDTM and SDTMIG. Each group is defined in a new ‘Variable Groups Definitions’ table (Section 2.5).

SDTM v2.0 contained a ‘Variables Qualified’ column in each SDTM table. In SDTM v3.0, the column was removed and Variable Relationships introduced to describe relationships between variables in an observation. The tables are included in each section of the SDTM and SDTMIG after the specification tables. A row in this table can be read by concatenating the subject variable, the linking phrase, and the object variable (e.g., "–TEST decodes the value in --TESTCD").

Subject Root Variable	Subject Root Variable Label	Linking Phrase	Predicate Term	Object Root Variable	Object Root Variable Label
--TEST	Name of Measurement, Test, or Exam	decodes the value in	DECODES	--TESTCD	Short Name of Measurement, Test, or Exam
--TESTCD	Short Name of Measurement, Test, or Exam	Is the code for the value in	IS_DECODED_BY	--TEST	Name of Measurement, Test, or Exam

STUDY DATA TABULATION MODEL IMPLEMENTATION GUIDE (SDTMIG) V4.0

SDTMIG v4.0 is considered a major release due to significant revisions to existing content and the introduction of new data structures.

High-level changes in SDTMIG v4.0 include:

- Reorganization/Re-write of Sections 1 – 4
- Metadata Restructuring
- --BLFL variable removal
- Protocol Deviations (DV) updates
- Event Adjudication (EA)
- Gastrointestinal Findings (GI)
- Multiple Subject Instances (MSI)
- SUPPQUAL (SUPP--) to NSV (NS--)

These items will be discussed in the following sections. A comprehensive list of all changes can be found in the Revision History (Appendix E).

METADATA RESTRUCTURING AND REORGANIZATION/RE-WRITE OF SECTIONS 1 - 4

The domain specifications have been restructured to be consistent with the metadata structure in SDTM v3.0:

- 'Variable Name', 'Variable Label', 'Type' and 'Core' columns are unchanged.
- 'Controlled Terms, Codelist or Format' column has been split into 'Codelist', 'Allowed Controlled Terms', and 'Format' columns.
- 'Role' column uses fewer roles; subtypes of Qualifier roles have been eliminated.
- New 'Variable Group' column added. Variable groups collect variables that serve a similar purpose. Variable groups are taken directly from, and defined in, the SDTM v3.0.
- 'Root Variable C-Code' and 'Root Variable Definition' columns added (copied over from SDTM).
- 'CDISC Notes' column was removed. Content redundant with the variable definition was removed.
    - Content describing 'Allowed Controlled Terms' was moved to that column.
    - Content that described general assumptions was moved to Sections 1-4 or to domain-specific assumptions and referenced in the new 'Notes' column. For general assumptions, hyperlinks navigate back to the appropriate section.
    - Examples were moved to the new 'Examples' column.

#	Variable Name	Variable Label	Type	Codelist	Allowed Controlled Terms	Format	Role	Variable Group	Root Variable C-code	Root Variable Definition	Notes	Examples	Core
1	STUDYID	Study Identifier	Char				Identifier	Study	C83082	A sequence of characters used to identify or name the study.			Req
2	DOMAIN	Domain Abbreviation	Char	(DOMAIN)	MH		Identifier	Domain	C49558	An abbreviation for a collection of observations, with a topic-specific commonality.	See Datasets and Domains.		Req
3	USUBJID	Unique Subject Identifier	Char				Identifier	Observation Subject	C69256	A sequence of characters used to uniquely identify a subject across all studies for all applications or submissions involving the product.	See Use of "Subject" and USUBJID and SUBJID.		Req
4	SUBJID	Subject Identifier for the Participation	Char				Identifier	Observation Subject		A sequence of characters used to uniquely identify a subject's participation in a study.	See USUBJID and SUBJID.		Perm
5	MHSEQ	Sequence Number	Num				Identifier	Record	C70710	A number used in combination with the identifier of the subject of the observation to uniquely identify a record within a domain.	See Record Identifiers.		Req
6	MHGRPID	Group ID	Char				Identifier	Record Group	C170996	A sequence of characters used to uniquely identify related records for a subject or pool within a domain, or related parameters in trial design and study reference datasets.	See Relating Groups of Records Within a Domain Using the --GRPID Variable and Category Variables (--CAT and --SCAT) and Identifiers that Group Records.		Perm
7	MHREFID	Reference ID	Char				Identifier	Observation Subject	C82531	A sequence of characters used to uniquely identify a source of information.	See --REFID.	Internal or external medical history identifier.	Perm
8	MHSPID	Sponsor-Defined Identifier	Char				Identifier	Record	C82530	A sponsor-defined sequence of characters used to identify an instance of an observation.	See Record Identifiers.		Perm
9	MHTERM	Reported Term for the Medical History	Char				Topic	Event Name	C82571	The collected name for an event observation.			Req

Sections 1-4 were reorganized and some content rewritten for further clarity. Across earlier versions of the SDTMIG, these sections were largely untouched, and any new content was added to the end of Section 4. This section contains General Assumptions that apply across all domains. Many SDTM conformance rules are based on this content and since redundancy has been removed from the domain specifications, it is highly recommended to review these sections in SDTMIG v4.0. The list of changes and updates by section can be found in the Revision History (Appendix E).

REMOVAL OF BASELINE FLAG (--BLFL) VARIABLE

Baseline Flag (--BLFL) was deprecated in SDTMIG v3.3 and the Last Observation Before Exposure Flag (--LOBXFL) variable was added. This variable provides a more defined algorithm than –BLFL. In SDTMIG v3.3 and v3.4, 'Core' for –BLFL was updated from 'Expected' to 'Permissible' and –LOBXFL was 'Expected', meaning that it needed to be included in the dataset whether it was populated or not. In SDTM v3.0/SDTMIG v4.0, the –BLFL has been removed from all specification tables. This means that its use is no longer supported and cannot be used in SDTM as a parent variable.

EVENT ADJUDICATION (EA) DOMAIN

With the publication of SDTMIG v4.0, there is now clear guidance for handling adjudication data. The Event Adjudication (EA) domain has been added and can be used to represent data collected about the adjudication of events using the Findings About specialization of the Findings Class. The EAOBJ (Object of the Observation) variable should be populated with the event being adjudicated. The event being adjudicated may reflect terms found in Events class domains, specific events identified by the adjudicators, and/or more generalized terms that appear in the protocol. The EA structure provides the ability to capture multiple judgements by multiple evaluators and aspects of the event. This domain should only contain information about the adjudication process itself. The events or results data collected to support the adjudication should still reside in the appropriate domains, e.g. adverse events data in AE, lab results data in LB, etc. Further, no data should be changed in these domains due to the outcome of the adjudication. RELREC should be used to link information across domains and records to the EA data.

In the EA example dataset below, EAOBJ = 'ACUTE CORONARY SYNDROME' is the event being adjudicated for USUBJID = '01-001' by two evaluators (EAEVALID = 'ADJUDICATOR 1' and 'ADJUDICATOR 2'). The EAACPTFL (Accepted Record Flag) = 'Y' (Yes) for records evaluated by the second adjudicator (rows 4 – 6) to indicate that these records are those accepted by the committee.

USUBJID	EASEQ	EATESTCD	EATEST	EAOBJ	EASTRESC	EAEVAL	EAEVALID	EAACPTFL	EADTC
01-001	1	OCCUR	Occurrence Indicator	ACUTE CORONARY SYNDROME	Y	ADJUDICATOR	ADJUDICATOR 1		2021-06-30
01-001	2	EVTYPE	Event Type	ACUTE CORONARY SYNDROME	Acute myocardial infarction	ADJUDICATOR	ADJUDICATOR 1		2021-06-30
01-001	3	EVSTDTC	Start Date/Time of Event	ACUTE CORONARY SYNDROME	2021-06-21	ADJUDICATOR	ADJUDICATOR 1		2021-06-30
01-001	4	OCCUR	Occurrence Indicator	ACUTE CORONARY SYNDROME	Y	ADJUDICATOR	ADJUDICATOR 2	Y	2021-06-30
01-001	5	EVTYPE	Event Type	ACUTE CORONARY SYNDROME	Acute myocardial infarction	ADJUDICATOR	ADJUDICATOR 2	Y	2021-06-30
01-001	6	EVSTDTC	Start Date/Time of Event	ACUTE CORONARY SYNDROME	2021-06-21	ADJUDICATOR	ADJUDICATOR 2	Y	2021-06-30

For more information on the adjudication process beyond the SDTMIG, the Phuse whitepaper, 'Best Practices for Submission of Event Adjudication Data', can be referenced. The EA domain was first proposed in this guidance. This Findings About structured approach was also described in the FDA's 'Technical Specification for Submitting Clinical Trial Data Sets for treatment of Noncirrhotic Nonalcoholic Steatohepatitis (NASH)'.

PROTOCOL DEVIATIONS (DV) DOMAIN

The Protocol Deviations (DV) domain was updated to support new regulatory guidance documents including ICH E6 R3. The domain-specific variable, Classification of Protocol Deviation (DVCLASI), was added to provide a standardized way for submitting the severity/classification of the protocol deviation. Terms mentioned in the ICH guidance are 'IMPORTANT'/'NON-IMPORTANT'. Sponsors may also use other term pairs such as 'MAJOR/MINOR' or 'CRITICAL/NONCRITICAL'. Having this new variable allows sponsors to have a specific place for this information rather than using the Category (DVCAT) or Subcategory (DVSCAT) variables, which are not appropriate for this purpose. Because terms can come from different sources, the CDISC Controlled Terminology (CT) and SDS teams chose to not create SDTM CT for DVCLASI at this time.

In the DV example below, protocol deviation data is shown for 3 subjects. The protocol deviation collected is populated in DVTERM and standardized in DVDECOD per sponsor-defined terminology. The category for each is provided in DVCAT. The investigator-assigned classification for the deviation is populated in DVCLASI: 'IMPORTANT' or 'NON-IMPORTANT'.

STUDYID	DOMAIN	USUBJID	DVSEQ	DVTERM	DVDECOD	DVCAT	DVCLASI	EPOCH	DVSTDTC
ABC123	DV	123101	1	IVRS PROCESS DEVIATION - NO DOSE CALL PERFORMED.	TREATMENT DEVIATION	STUDY INTERVENTION	NON-IMPORTANT	TREATMENT	2003-09-21
ABC123	DV	123103	1	DRUG XXX ADMINISTERED DURING STUDY TREATMENT PERIOD	EXCLUDED CONCOMITANT MEDICATION	PROHIBITED CONCOMITANT INTERVENTION	IMPORTANT	TREATMENT	2003-10-30
ABC123	DV	123103	2	VISIT 3 DOSE <15 MG	TREATMENT DEVIATION	STUDY INTERVENTION	IMPORTANT	TREATMENT	2003-10-30
ABC123	DV	123104	1	TOOK ASPIRIN	PROHIBITED MEDS	PROHIBITED CONCOMITANT INTERVENTION	IMPORTANT	TREATMENT	2003-11-30

GASTROINTESTINAL SYSTEM FINDINGS (GI) DOMAIN

The Gastrointestinal System Findings (GI) domain is an additional Morphology/Physiology domain that first appeared in the Crohn's Disease Therapeutic Area User Guide (TAUG) and is now added to the main SDTMIG. This domain is used to represent results and findings related to the gastrointestinal system, including the esophagus, stomach, small and large intestine, anus, liver, biliary tract and pancreas. Findings related to the diagnostic procedure itself should be provided in the Procedures (PR) and/or Findings About Procedures (FAPR) domains and not in GI.

The example GI dataset below shows the results of a diagnostic endoscopy of the stomach.

USUBJID	GISEQ	GITESTCD	GITEST	GIORRES	GIORRESU	GISTRESC	GISTRESN	GISTRESU
6002	1	PAREAULC	Percent Area Covered By Ulcers	11.2	%	11.2	11.2	%

GILOC	GIMETHOD	VISITNUM	VISIT	GIDTC	GIDY
STOMACH WALL	ENDOSCOPY	1	SCREENING	2016-04-18	-10

There are CDISC CT codelists for GI tests, responses and units. The associations between tests and responses are available in a new GI codetable: https://www.cdisc.org/standards/terminology/controlled-terminology

MULTIPLE SUBJECT INSTANCES (MSI)

In clinical trials, it has become more common for a study design to allow subjects to re-enroll in the same study and until now, there was little guidance on how to handle this scenario in SDTM especially since the DM domain can only contain one record per Unique Subject Identifier (USUBJID).

The FDA's Study Data Technical Conformance Guide (sdTCG) provides some details for submitting data for multiple enrollments. It states that if a subject is screened and/or enrolled more than once in a study, then the subject's Subject Identifier for the Participation (SUBJID) should be different for each participation but there should be only one USUBJID assigned within the study and across the entire application. Also, the data collected for each screening or enrollment should be submitted in a custom domain similar to Demographics (DM). Then SUBJID should be included in all subject-level domains to differentiate records across participations. DM should still contain only one record per USUBJID that is considered to be the primary enrollment.

With the publication of SDTMIG v4.0, there is finally clear guidance on submission of data for multiple participations that also incorporates the requirements outlined in the sdTCG. When the protocol allows subjects to screen or enroll more than once in the same study, there are multiple scenarios that typically occur:

1. Subjects that are screened more than once within the same study but are never enrolled. These subjects are identified in DM and DS as screen failures.
2. Subjects that are screened more than once within the same study and eventually enroll. The subjects are identified in DM as having been randomized and/or treated.
3. Subjects that enroll more than once within the same study and are randomized and/or treated more than once. Sponsors must decide how to represent the primary enrollment in DM.
4. Subjects that have multiple foci that enrolled separately, e.g. a subject's left and right eyes in an ophthalmology study. These foci are identified using the Focus of Study-Specific Interest (FOCID) variable.

In support of the guidance in the FDA sdTCG, the Demographics for Multiple Participations (DC) domain was developed to represent subject data for each participation while maintaining the restriction in DM that there can only be one record per subject. This Special-purpose domain is structured similar to DM with the main difference being the addition of Sequence Number (DCSEQ). This is because DC will contain more than one record per subject. Subjects should be assigned only one USUBJID value but SUBJID should be different for each participation instance. Though the Date/Time of Collection (DCDTC) variable can be added, the Study Day (DCDY) variable is not permitted as this would introduce circular logic because the derivation for –DY variables in other domains should still be based on Subject Reference Start Date/Time (RFSTDTC) in DM and DM is essentially derivative of DC.

Prior to SDTMIG v4.0, SUBJID was restricted to the DM domain. In order to differentiate subject records for each participation in other domains, SUBJID can now be added to the domain for this purpose. The SUBJID variable was added to the Identifiers table (Table 3.1.4) in SDTM v3.0 as well as each subject-level domain specification in the SDTMIG v4.0. If DC is used, it should contain records for all subjects in the study regardless of whether they re-screened/re-enrolled. It is important to note that the DC domain and the addition of SUBJID to other domains should only be done for studies that allow multiple participations. If the protocol does not allow for multiple enrollments, then only DM should be submitted.

In this example DC domain, the first three subjects (USUBJID = ABC12301001, ABC12301002, ABC12301003) were screened, randomized and treated within the same participation. All the reference dates and treatment arm variables are populated to reflect this, and they each have only one record in DC. For USUBJID = ABC12301004 (purple rows), the first two records indicate that the subject has two failed screening attempts where RFSTDTC, RFENDTC, RFXSTDTC, RFXENDTC, ARMCD/ARM, ACTARMCD/ACTARM are all null. Further, Reason Arm and/or Actarm is Null (ARMNRS) is set to 'SCREEN FAILURE'. The third record indicates that the subject passed screening and was randomized and treated (ARM/ACTARM = Drug A). RFSTDTC, RFENDTC, RFXSTDTC, and RFXENDTC are also

populated. This subject was assigned a new SUBJID for each screening attempt (SUBJID = 1004, 1004B, and 1004C) and USUBJID is the same for all 3 records.

dc.xpt

STUDYID	DOMAIN	USUBJID	SUBJID	DCSEQ	RFSTDTC	RFENDTC	RFXSTDTC	RFXENDTC	RFICDTC	RFPENDTC
ABC123	DC	ABC12301001	1001	1	2006-01-14	2006-01-28	2006-01-14	2006-01-28	2006-01-05	2006-02-25
ABC123	DC	ABC12301002	1002	1	2006-01-12	2006-01-26	2006-01-12	2006-01-26	2006-01-04	2006-02-23
ABC123	DC	ABC12301003	1003	1	2006-01-08	2006-01-20	2006-01-08	2006-01-20	2006-01-02	2006-02-17
ABC123	DC	ABC12301004	1004	1					2005-09-13	2005-09-21
ABC123	DC	ABC12301004	1004B	2					2005-11-14	2005-11-16
ABC123	DC	ABC12301004	1004C	3	2006-01-16	2006-02-01	2006-01-16	2006-02-01	2006-01-07	2006-03-02

BRTHDTC	AGE	AGEU	SEX	RACE	ETHNIC	ARMCD	ARM	ACTARMCD	ACTARM	ARMNRS
1948-12-13	57	YEARS	M	WHITE	HISPANIC OR LATINO	A	Drug A	A	Drug A	
1955-03-22	50	YEARS	M	WHITE	NOT HISPANIC OR LATINO	P	Placebo	P	Placebo	
1938-01-19	67	YEARS	F	BLACK OR AFRICAN AMERICAN	NOT HISPANIC OR LATINO	P	Placebo	P	Placebo	
1941-07-02			M	ASIAN	NOT HISPANIC OR LATINO					SCREEN FAILURE
										SCREEN FAILURE
	64	YEARS				A	Drug A	A	Drug A	

The DM domain for the same subjects is shown below. For the first 3 USUBJIDs, the information from DC was simply copied over to DM because they only had one record in DC. For USUBJID = ABC12301004, the record indicating that the subject was enrolled is carried forward to the DM domain since it is the primary participation. It is important to note that the SUBJID = 1004C (the SUBJID assigned at the final screening attempt) is populated in DM.SUBJID.

dm.xpt

STUDYID	DOMAIN	USUBJID	SUBJID	RFSTDTC	RFENDTC	RFXSTDTC	RFXENDTC	RFICDTC	RFPENDTC
ABC123	DM	ABC12301001	1001	2006-01-14	2006-01-28	2006-01-14	2006-01-28	2006-01-05	2006-02-25
ABC123	DM	ABC12301002	1002	2006-01-12	2006-01-26	2006-01-12	2006-01-26	2006-01-04	2006-02-23
ABC123	DM	ABC12301003	1003	2006-01-08	2006-01-20	2006-01-08	2006-01-20	2006-01-02	2006-02-17
ABC123	DM	ABC12301004	1004C	2006-01-16	2006-02-01	2006-01-16	2006-02-01	2006-01-07	2006-03-02

BRTHDTC	AGE	AGEU	SEX	RACE	ETHNIC	ARMCD	ARM	ACTARMCD	ACTARM	ARMNRS
1948-12-13	57	YEARS	M	WHITE	HISPANIC OR LATINO	A	Drug A	A	Drug A	
1955-03-22	50	YEARS	M	WHITE	NOT HISPANIC OR LATINO	P	Placebo	P	Placebo	
1938-01-19	67	YEARS	F	BLACK OR AFRICAN AMERICAN	NOT HISPANIC OR LATINO	P	Placebo	P	Placebo	
1941-07-02	64	YEARS	M	ASIAN	NOT HISPANIC OR LATINO	A	Drug A	A	Drug A	

The Disposition (DS) domain for USUBJID = ABC12301004 is shown below. The SUBJID variable was added to the domain and populated with the corresponding value assigned for each participation. The first screening attempt (pink rows), the second (purple rows) and third (green rows) are differentiated by the value in SUBJID. If SUBJID is present in the dataset, then it should be populated for all subjects.

ds.xpt

STUDYID	DOMAIN	USUBJID	SUBJID	DSSEQ	DSTERM	DSDECOD	DSCAT	EPOCH	DSSTDTC
ABC123	DS	ABC12301004	1004	1	INFORMED CONSENT OBTAINED	INFORMED CONSENT OBTAINED	PROTOCOL MILESTONE	SCREENING	2005-09-13
ABC123	DS	ABC12301004	1004	2	SCREEN FAILURE	SCREEN FAILURE	DISPOSITION EVENT	SCREENING	2005-09-21
ABC123	DS	ABC12301004	1004B	3	INFORMED CONSENT OBTAINED	INFORMED CONSENT OBTAINED	PROTOCOL MILESTONE	SCREENING	2005-11-14
ABC123	DS	ABC12301004	1004B	4	SCREEN FAILURE	SCREEN FAILURE	DISPOSITION EVENT	SCREENING	2005-11-16
ABC123	DS	ABC12301004	1004C	5	INFORMED CONSENT OBTAINED	INFORMED CONSENT OBTAINED	PROTOCOL MILESTONE	SCREENING	2006-01-07
ABC123	DS	ABC12301004	1004C	6	COMPLETED	COMPLETED	DISPOSITION EVENT	SCREENING	2006-01-16
ABC123	DS	ABC12301004	1004C	7	COMPLETED	COMPLETED	DISPOSITION EVENT	TREATMENT	2006-02-01
ABC123	DS	ABC12301004	1004C	8	COMPLETED	COMPLETED	DISPOSITION EVENT	FOLLOW-UP	2006-03-02

SUPPLEMENTAL QUALIFIERS (SUPPQUAL) TO NON-STANDARD VARIABLES (NS--)

The most significant change in SDTMIG v4.0 is the vertical structure of SUPPQUAL datasets to horizontal-structured datasets for non-standard qualifier variables (NSVs). Prior to SDTMIG v4.0, SUPPQUAL datasets need to be transposed before they can be appended back onto the parent dataset. With NS—datasets, this step is no longer necessary. Also, many sponsors already store their SDTM datasets in a structure with the NSVs already appended to the parent, sometimes referred to as SDTM+, and then SUPPQUAL datasets are created for submission.

Because the SDTM standard does not allow the addition of non-standard variables to domains, the NS—datasets will still be separate from the parent datasets. The naming of NSV datasets is similar to SUPP-- where the two hyphens in 'NS—' are replaced with the domain abbreviation of the parent domain, e.g. the NS-- dataset for AE is named 'NSAE.xpt'.

The Identifier variables that are used in SUPPQUAL to join back to the parent (STUDYID, RDOMAIN, USUBJID, IDVAR) are the same in NSV datasets. The only difference is that for NSV, the character datatype variable, Identifying Variable Value (IDVARVAL), is replaced by a new numeric datatype variable, Identifying Variable Numeric Value (IDVARVLN). Since it is best practice for SUPP-- to use IDVAR = --SEQ to join back to parent records and –SEQ is a numeric variable, the decision was made to remove IDVARVAL and add a numeric datatype variable, IDVARVLN, eliminating the need to convert the character type to numeric when merging back to the parent. There is an assumption added that states that IDVAR/IDVARVLN must be –SEQ from the parent domain for all NS—datasets except NSDM because STUDYID, RDOMAIN, and USUBJID are sufficient to identify the parent record in DM. In this case, the IDVAR/IDVARVLN variables will be null in NSDM.

Another benefit of this change is that the non-standard variables added to NS-- can be either character or numeric type as appropriate. In SUPPQUAL, the Data Value (QVAL) variable was a character datatype only. For NS--, the corresponding metadata for each NSV should be clearly enumerated in the define.xml.

The following is an example of a Healthcare Encounters (HO) domain and a Non-standard HO (NSHO) dataset with its associated NSV metadata. In NSHO, each non-standard variable is added as an additional column/variable so there is a one-to-one relationship between the parent record and the non-standard variables associated with that record noted in IDVAR = HOSEQ and IDVARVLN.

ho.xpt

STUDYID	DOMAIN	USUBJID	HOSEQ	HOTERM	HOSTDTC	HOENDTC
1999001	HO	1001	1	HOSPITAL STAY	2004-01-05	2004-01-12
1999001	HO	1001	2	HOSPITAL STAY	2004-01-23	2004-02-07
1999001	HO	1002	1	HOSPITAL STAY	2004-01-21	2004-01-22

nsho.xpt

STUDYID	RDOMAIN	USUBJID	IDVAR	IDVARVLN	HOAERPFL	HOMEDSFL	HOPROCFL	HONAM	HOSPUTY	HOSPUFL	HORLCNDF
1999001	HO	1001	HOSEQ	1	Y	Y	Y	GENERAL HOSP	ICU	Y	Y
1999001	HO	1001	HOSEQ	2	Y	Y	N	UNIV HOSP	CCU	Y	Y
1999001	HO	1002	HOSEQ	1	Y	N	Y	ST. MARY'S	ICU	N	Y

NSV Metadata

Variable	Label	Type	Codelist	Role
HOAERPFL	AE Reported This Episode	Char	(NY)	Non-Standard Qualifier
HOMEDSFL	Meds Prescribed	Char	(NY)	Non-Standard Qualifier
HOPROCFL	Procedures Performed	Char	(NY)	Non-Standard Qualifier
HONAM	Provider Name	Char		Non-Standard Qualifier
HOSPUTY	Specialized Unit Type	Char		Non-Standard Qualifier
HOSPUFL	Any Time in Spec. Unit	Char	(NY)	Non-Standard Qualifier
HORLCNDF	Visit Related to Study Med Cond.	Char	(NY)	Non-Standard Qualifier

In SDTMIG v4.0, SUPPQUAL has been removed and replaced with the NSV solution. Throughout the SDTMIG, all examples and references have been updated in support of Non-standard Variable datasets. Sponsors should prepare now for adopting this new approach to ensure a smooth transition. Once this approach is adopted, creating NS—datasets and merging them back to the parent dataset should make preparing submission deliverables and review easier. Further, though the NSV approach reduces the number of records in an NS—dataset, care should still be taken when determining which non-standard variable data is really necessary to include.

OTHER NOTABLE UPDATES

Globally, the concept of race and race values can differ across regulatory agencies. For the DM domain in earlier SDTMIGs, the RACE codelist in CDISC CT is assigned to the DM.RACE variable. This codelist was originally defined in the FDA Office of Management and Budget (OMB) document, 'Collection of Race and Ethnicity Data in Clinical Trials' (https://www.fda.gov/regulatory-information/search-fda-guidance-documents/collection-race-and-ethnicity-data-clinical-trials). This document lists and the RACE codelist consists of 5 race categories that are primarily US-focused. In an effort to support the requirements of other regulatory agencies across the world, the RACE codelist is no longer assigned to the DM.RACE variable in the SDTMIG. An assumption was added to explain that the specific race requirements for the regulatory agency being submitted to should be referenced and populated in RACE instead. Also, the Clinical Data Acquisition Harmonization (CDASH) variables, Collected Race (CRACE) and Collected Ethnicity (CETHNIC), have been added to the DM and DC domains in SDTM v3.0/SDTMIG v4.0. Both variables have CDISC CT codelists that should be followed.

The Conditionally Branched Item Flag (--CBRFL) variable was added as a parent variable for the Questionnaires, Ratings, and Scales (QRS) domains: Questionnaires (QS), Functional Tests (FT), and Disease Response and Clinical Classification (RS). Prior to SDTMIG v4.0, this information was mapped to SUPPQUAL in the QRS supplements (https://www.cdisc.org/standards/foundational/qrs). This variable is used for instruments that have questions that are logically skipped based on some condition, e.g. female-only items. Allowed values are 'Y' (Yes) or null. This new variable eliminates the need to hijack Derived Flag (--DRVFL) for this purpose. This variable should not be used in QRS domains as it is not appropriate because all data in QRS domains is considered to be collected data.

Other updates include minor changes to some specimen-based Findings domains guidance and examples. It is recommended to review the Revision History in Appendix E for further details on the updates.

FINAL NOTE: EXPANDED SCOPE OF SDTMIG V4.0

CDISC plans to pause new standards development in favor of Biomedical Concepts (BCs) after this major release of SDTMIG v4.0 (as well as for CDASH, SEND, and ADaM). Because of this, the SDS team has expanded the scope of the SDTMIG v4.0. The additional items are still pending final approval. They will be released for Public Review in a separate cycle from the main SDTMIG and incorporated into the SDTMIG after they have been finalized.

The expanded scope includes the following but is still awaiting CDISC Leadership approval as of March 2026:

- SDTMIG-Associated Persons (SDTMIG-AP) – updates for SUPP to NSV, further clarifications, APCO
- Laboratory Results for Conventional Units (LC) domain – FDA request to submit a separate dataset containing conventional units along with LB as outlined in the sdTCG
- Splitting Disease Response and Clinical Classification (RS) into RS (Disease Response) and CC (Clinical Classifications) domains
- Additional Morphology/Physiology domains

CONCLUSION

SDTMIG v4.0 represents more than a version update—it's a significant era shift for the SDTM standard. With the replacement of SUPPQUAL, clearer metadata structure, expanded domain coverage, and long-needed clarity for MSI, the standard has redefined itself for the complexities of modern clinical trials. Just as each era builds on the last while offering something entirely new, SDTMIG v4.0 honors what worked in earlier versions while letting go of outdated patterns that no longer fit.

Transitioning into this new phase will require planning, updates to internal processes, and time for teams to adapt. But the benefits—greater consistency, better traceability, improved regulatory alignment, and a more intuitive structure—set a strong foundation for the future. As the last major SDTMIG before CDISC shifts toward Biomedical Concepts, v4.0 closes one door and confidently opens the next, inviting sponsors to step forward through it. So now the question is: Are we ready for it?!?

REFERENCES

Study Data Tabulation Model. Clinical Data Interchange Standards Consortium (CDISC) SDTM Team. Version 3.0. 2026. https://wiki.cdisc.org/spaces/SDTM3DOT0/pages/236361197/Study+Data+Tabulation+Model+v3.0+SDTM+v3.0

Study Data Tabulation Model Implementation Guide: Human Clinical Trials. Clinical Data Interchange Standards Consortium (CDISC) Submission Data Standards (SDS) Team. Version 4.0. 2026. https://wiki.cdisc.org/spaces/SDTMIG4DOT0/pages/136085987/SDTMIG+v4.0+Home

FDA. Study Data Technical Conformance Guide. Version 6.1. December, 2025. https://www.fda.gov/media/153632/download

FDA. Collection of Race and Ethnicity Data in Clinical Trials. October 2016. https://www.fda.gov/media/75453/download

FDA. Technical Specifications for Submitting Clinical Trial Data Sets for Treatment of Noncirrhotic Nonalcoholic Steatohepatitis (NASH). December 2024. https://www.fda.gov/media/151870/download

ICH E3: Structure and Content of Clinical Study Reports. ICH; 1996. Accessed September 23, 2024. https://database.ich.org/sites/default/files/E3_Guideline.pdf

ICH E3 Guideline: Structure and Content of Clinical Study Reports Questions & Answers (R1). ICH; 2012. Accessed September 23, 2024. https://database.ich.org/sites/default/files/E3_Q%26As_R1_Q%26As.pdf

PHUSE, Best Practices for Submission of Event Adjudication Data. October 2019. https://phuse.s3.eu-central-1.amazonaws.com/Deliverables/Optimizing+the+Use+of+Data+Standards/Best+Practices+for+Submission+of+Event+Adjudication+Data.pdf

ACKNOWLEDGMENTS

I would like to extend my sincere appreciation to the CDISC SDS and SDTM teams for their exceptional dedication and sustained effort in developing SDTM v3.0 and SDTMIG v4.0 over the past five years. This release reflects countless hours of discussion, collaboration, refinement, and problem-solving across teams. The depth and breadth of updates in this version demonstrate our commitment to strengthening the clarity, consistency, and long-term usability of the SDTM standard. I am grateful to have the privilege of participating on these teams to help continue moving the CDISC standards forward.

CONTACT INFORMATION

Your comments and questions are valued and encouraged.

Contact the author at:
Kristin Kelly
Pinnacle 21 by Certara
kristin.kelly@certara.com