What's New in SDTMIG v4.0

Maria Sekac, Associate Director, Global Clinical Data Standards, Merck & Co., Inc., USA
Michael Wise, Director, Statistical Programming, Edwards LifeSciences

Meet the Speakers

Maria Sekac

Title: Associate Director Clinical Data Management

Organization: Merck & Co., Inc., USA

Maria is a Clinical Data Standards Lead at Merck, bringing over 25 years of experience in the pharmaceutical industry. With a strong foundation in clinical data management and industry standards, she has led initiatives across the clinical trial lifecycle—from protocol development to regulatory submissions—and has developed hundreds of CDASH and SDTM-compliant data collection specifications. Maria also serves on the Leadership Team of the CDISC Submission Data Standards team, contributing to the advancement of industry-wide data standardization efforts.

Michael Wise

Title: Director, Statistical Programming

Organization: Edwards LifeSciences

Michael is the current lead of the CDISC Study Data Submissions (SDS) team and has over 20 years of experience in statistical programming for clinical research including, CROs, pharmaceuticals, biotech, and medical devices. Though the years, he has volunteered for multiple sub-teams within CDISC and had served as president for the Chicago chapter of the American Statistical Association.

Disclaimer and Disclosures

- *The views and opinions expressed in this presentation are those of the author(s) and do not necessarily reflect the official policy or position of CDISC.*

- *The author(s) have no real or apparent conflicts of interest to report.*

Agenda

1. Background
2. Major Scope Changes
   - Metadata Restructuring
   - Reorganization of Sections 1-4
   - New Domains
   - Variable Changes
   - Changing Supplemental Qualifiers to Non-Standard Variables
3. Timelines

Background

- SDTMIG v4.0 will be the next version after SDTMIG version 3.4
  - Major revision
- Scoping of project began in 2020
- Major scope changes have been reviewed for feedback at various intervals by different stakeholder and governance groups, including:
  - CDISC's Global Governance Group (GGG) Meetings
  - CDISC Advisory Council (CAC) Meetings
  - FDA through the CDISC Technical Leadership Committee (TLC) Meetings

- METADATA RESTRUCTURING
- REORGANIZATION AND REWRITE OF SECTIONS 1-4

Metadata Restructuring

- Description:
  - Project to restructure each domain specification table
- Rationale:
  - To match the metadata structure in SDTM v2.0 and to parse out overloaded columns of text
- Benefits:
  - Better organized
  - Include more informative labels
  - Split out metadata that were mixed into one column
  - Move assumptions from CDISC notes to the Assumptions section

Metadata Restructuring

- Old Structure

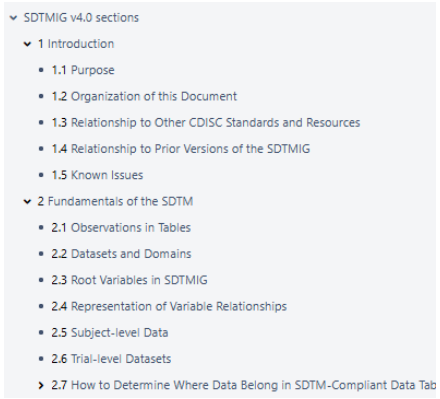
Variable Name	Variable Label	Type	Controlled Terms, Codelist or Format[1]	Role	CDISC Notes	Core
STUDYID	Study Identifier	Char		Identifier	Unique identifier for a study.	Req
DOMAIN	Domain Abbreviation	Char	DD	Identifier	Two-character abbreviation for the domain.	Req
USUBJID	Unique Subject Identifier	Char		Identifier	Identifier used to uniquely identify a subject across all studies for all applications or submissions involving the product.	Req
DDSEQ	Sequence Number	Num		Identifier	Sequence number given to ensure uniqueness of subject records within a domain. May be any valid number.	Req
DDTESTCD	Death Detail Assessment Short Name	Char	(DTHDXCD)	Topic	Short name of the measurement, test, or examination described in DDTEST. It can be used as a column name when converting a dataset from a vertical to a horizontal format. The value in DDTESTCD cannot be longer than 8 characters, nor can it start with a number (e.g., "1TEST" is not valid). DDTESTCD cannot contain characters other than letters, numbers, or underscores. Examples: "PRCDTH", "SECDTH".	Req
DDTEST	Death Detail Assessment Name	Char	(DTHDX)	Synonym Qualifier	Long name for DDTESTCD. The value in DDTEST cannot be longer than 40 characters. Examples: "Primary Cause of Death", "Secondary Cause of Death".	Req

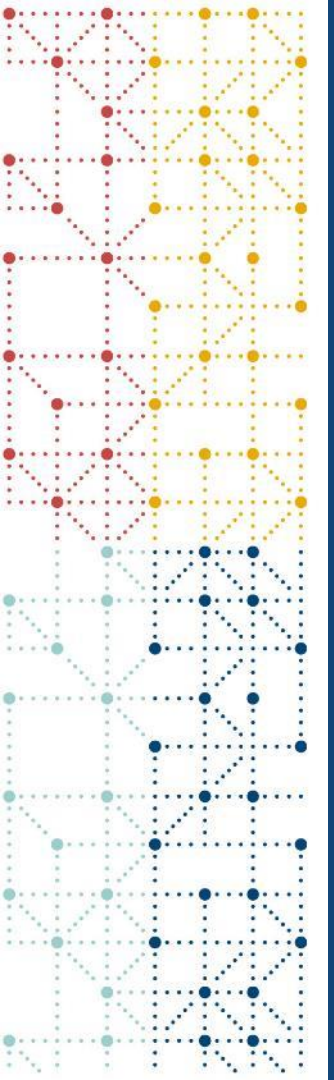
- New Structure

	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 Variable Group	C83082	A sequence of characters used by the sponsor to uniquely identify the study.			Req
2	DOMAIN	Domain Abbreviation	Char	(DOMAIN)	DD		Identifier	Domain Variable Group	C49558	An abbreviation for a collection of observations, with a topic-specific commonality.	See Datasets and Domains.		Req
6	DDTESTCD	Death Detail Assessment Short Name	Char	(DTHDXCD)			Topic	Test Focus Variable Group	C82503	The standardized or dictionary-derived short sequence of characters used to represent the measurement, test, or examination.	See Variable Lengths.	"PRCDTH"; "SECDTH"	Req
7	DDTEST	Death Detail Assessment Name	Char	(DTHDX)			~~Synonym~~ Qualifier	Test Focus Variable Group	C82541	The standardized or dictionary-derived name of the measurement, test, or examination.	See Test Name (--TEST) Greater than 40 Characters.	"Primary Cause of Death"; "Secondary Cause of Death"	Req

Reorganization and Rewrite of Sections 1-4

- Description:
  - Sections 1-4 were rewritten to accommodate restructuring of the metadata tables, to add more clarity, improve the content's organization, and to ensure that general assumptions are consistent, and in one place, rather than dispersed within the CDISC notes for every domain's specification tables.
- Benefits:
  - Clearer and better-organized text within the SDTMIG.





NEW DOMAINS

Demographics for Multiple Participations (DC) Domain

- Description:
  - A special-purpose domain that includes a set of standard variables that describe each participation of a subject in a clinical study (e.g. where subjects can re-screen or participate in the same study more than once)
- Rationale:
  - Clinical trials have historically handled multiple subject instances using different approaches
  - Guidance for how to handle this type of data was provided in the FDA sdTCG in Oct 2018
- Benefits:
  - Facilitates review through a standardized approach for submitting data for multiple subject instances.

Event Adjudication (EA) Domain

- Description:
  - The Event Adjudication (EA) domain is used to represent data about the adjudication of events using the Findings About subclass
- Rationale:
  - Clinical trials have historically handled mapping adjudication data using different approaches and it wasn't addressed in the SDTMIG.
- Benefits:
  - Provides a standard domain that expands upon the PHUSE paper from 2019, which is a similar approach described in the FDA "Technical Specification for Submitting Clinical Trial Data Sets for Treatment of Nonalcoholic Steatohepatitis (NASH), 2022"

EA Domain Use Case

- This example depicts the adjudication of a study endpoint.

ae.xpt

Row	STUDYID	DOMAIN	USUBJID	AESEQ	AESPID	AETERM	AEDECOD	AEREL	AESTDTC	AEENDTC
1	ABC	AE	ABC-01-101	4	4	DRUG-INDUCED LIVER INJURY	Drug-induced liver injury	RELATED	2015-02-13	2015-03-13

ea.xpt

Row	STUDYID	DOMAIN	USUBJID	EASEQ	EASPID	EATESTCD	EATEST	EAOBJ	EAORRES	EASTRESC	EASTRESN	EAMETHOD	EAEVAL	EAEVALID	EAACPTFL
1	ABC	EA	ABC-01-101	1	4	REL	Relation to Study Treatment or Product	Drug-induced liver injury	PROBABLE	3	3	FONTANA DILIN NUMERIC SCORE 2009	ADJUDICATOR	ADJUDICATOR 1	
2	ABC	EA	ABC-01-101	2	4	REL	Relation to Study Treatment or Product	Drug-induced liver injury	HIGHLY LIKELY	2	2	FONTANA DILIN NUMERIC SCORE 2009	ADJUDICATOR	ADJUDICATOR 2	
3	ABC	EA	ABC-01-101	3	4	REL	Relation to Study Treatment or Product	Drug-induced liver injury	HIGHLY LIKELY	2	2	FONTANA DILIN NUMERIC SCORE 2009	ADJUDICATION COMMITTEE		Y
4	ABC	EA	ABC-01-101	4	4	EVSTDTC	Start Date/Time of Event	Drug-induced liver injury	2015-02-13	2015-02-13			ADJUDICATOR	ADJUDICATOR 1	
5	ABC	EA	ABC-01-101	5	4	EVSTDTC	Start Date/Time of Event	Drug-induced liver injury	2015-02-13	2015-02-13			ADJUDICATOR	ADJUDICATOR 2	Y

Gastrointestinal System Findings (GI) Domain

- Description:
  - A findings domain that contains physiological and morphological findings related to the gastrointestinal system, including the esophagus, stomach, small and large intestine, anus, liver, biliary tract and pancreas.
- Rationale:
  - This domain is used to represent results and findings of gastrointestinal diagnostic procedures.
- Benefits:
  - To support use cases from a variety for Therapeutic Area User Guides

GI Domain Use Case

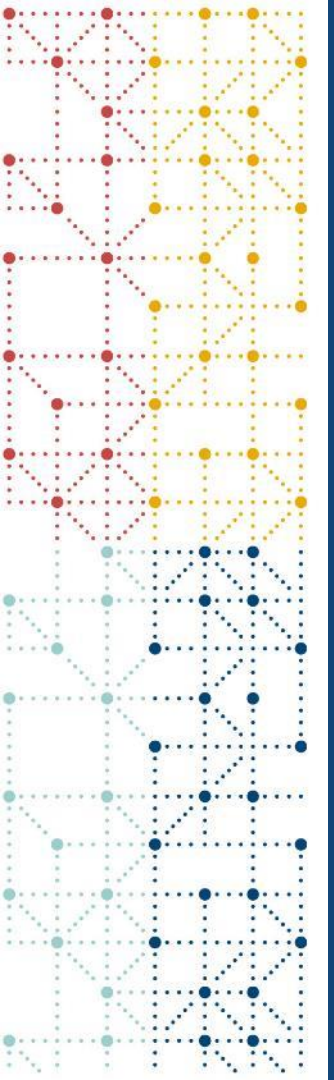
Row 1: Shows the overall image quality of the entire endoscopy scan. Since this scan had poor resolution, no assessments could be measured throughout the entire lower gastrointestinal tract.

Row 2: Shows the ileum could not be assessed because it was resected. No assessments were made for the ileum.

Rows 3-10: Show some examples of data collected by the reader while viewing the endoscopy video. These data were not used as part of a QRS measure for this study.

gi.xpt

Row	STUDYID	DOMAIN	USUBJID	SPDEVID	GISEQ	GILNKID	GITESTCD	GITEST	GIORRES	GIORRESU	GISTRESC	GISTRESN	GISTRESU	GISTAT	GIREASND	GILOC	GIMETHOD	GIEVAL	GIEVALID	GIDTC
1	CR123	GI	CR123-002	245	1	01	GIALL	Gastrointestinal System Findings						NOT DONE	POOR RESOLUTION	GASTROINTESTINAL TRACT, LOWER	ENDOSCOPY			
2	CR123	GI	CR123-001	245	2	01	GIALL	Gastrointestinal System Findings						NOT DONE	RESECTED	ILEUM	ENDOSCOPY			
3	CR123	GI	CR123-001	245	3	01	PAREAULC	Percent Area Covered By Ulcers	30	%	30	30	%			COLON, RIGHT	ENDOSCOPY	READER	READER 1	2018-06-05
4	CR123	GI	CR123-001	245	4	01	STRCTNUM	Number of Strictures	2		2	2				COLON, RIGHT	ENDOSCOPY	READER	READER 1	2018-06-05
5	CR123	GI	CR123-001	245	5	01	PAREAULC	Percent Area Covered By Ulcers	25	%	25	25	%			COLON, TRANSVERSE	ENDOSCOPY	READER	READER 1	2018-06-05
6	CR123	GI	CR123-001	245	6	01	STRCTNUM	Number of Strictures	0		0	0				COLON, TRANSVERSE	ENDOSCOPY	READER	READER 1	2018-06-05
7	CR123	GI	CR123-001	245	7	01	PAREAULC	Percent Area Covered By Ulcers	45	%	45	45	%			COLON, LEFT	ENDOSCOPY	READER	READER 1	2018-06-05
8	CR123	GI	CR123-001	245	8	01	STRCTNUM	Number of Strictures	0		0	0				COLON, LEFT	ENDOSCOPY	READER	READER 1	2018-06-05
9	CR123	GI	CR123-001	245	9	01	PAREAULC	Percent Area Covered By Ulcers	20	%	20	20	%			RECTUM	ENDOSCOPY	READER	READER 1	2018-06-05
10	CR123	GI	CR123-001	245	10	01	STRCTNUM	Number of Strictures	0		0	0				RECTUM	ENDOSCOPY	READER	READER 1	2018-06-05



VARIABLE CHANGES

New Variables

Variable	Label	Observation Class	Definition
--RSCNT	Result Count	Findings, EG domain only	The number of occurrences of a result.
--CBR	Conditionally Branched Item Flag	Findings	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.
--CETHNIC	Collected Ethnicity	Special-purpose	A classification of ethnicity as specified by the sponsor in the data collection field.
--CRACE	Collected Race	Special-purpose	A classification of race as specified by the sponsor in the data collection field.
IDVARVLN	Identifying Variable Numeric Value	Relationship	The numeric value of the variable named in IDVAR used to identify related records.
--TRTCD	Standardized Intervention Code	Interventions	A standardized or dictionary-derived short sequence of characters used to represent the intervention.
--CLASI	Classification of Protocol Deviation	Events, DV domain only	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))

Protocol Deviations (DV) Domain Changes

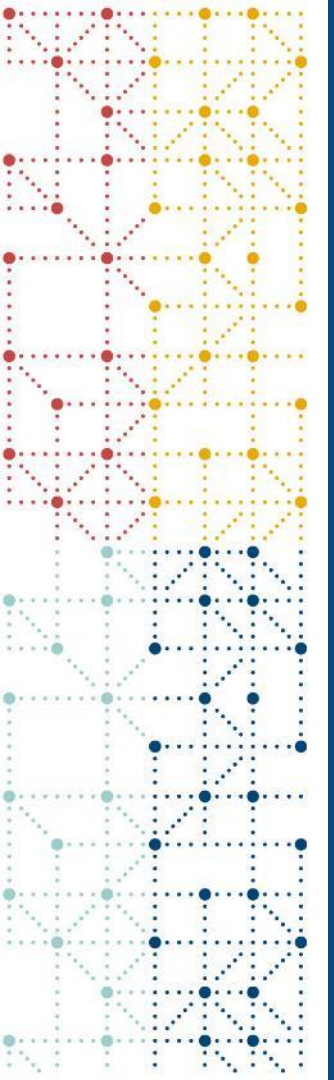
- Description:
  - Further standardization of the DV domain and related terminology in order to align with TransCelerate's Protocol Deviation Data Collection Optimization Initiative. Provides a new variable to capture DV classifications e.g. Important vs Non-Important.
- Impacts:
  - New variable (DVCLASI).
  - New controlled terminology for DVCAT and DVDECOD being evaluated.
- Benefits:
  - Improve the industry standard for submitting protocol deviation data in SDTM for the support of FDA's Bioresearch Monitoring program (BIMO).

Protocol Deviations (DV) Domain Example

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

Removed Variables

- --BLFL removed from both SDTM v3.0 and SDTMIG v4.0
  - Previously, identifying a 'baseline' record to be flagged by --BLFL was operationally difficult:
    - Sponsors would either leave the baseline flag value blank or flag a record to approximate the true baseline value.
  - Last Observation Before Exposure Flag (--LOBXFL) provides a more defined algorithm for derivation:
    - --LOBXFL is an indication that the record is the last non-missing baseline assessment prior to the date and time of first protocol specified treatment exposure.
    - First introduced in SDTMIG v3.3, --LOBXFL is present in domains that previously contained --BLFL.
- To be consistent with restrictions reflected in SDTM v3.0:
  - --MODIFY and --BODSYS were removed from Findings
  - --PORTOT was removed from Events and Interventions



CHANGING SUPPLEMENTAL QUALIFIERS TO NON-STANDARD VARIABLES

Non-Standard Variables (NSV) replaces Supplemental Qualifiers

- Prior to SDTMIG 4.0 non-standard variables, variables not defined within a domain structure were mapped to SUPP.
- SUPP datasets have a vertical structure and need to be transposed before appending to the parent dataset.
- The structure of the new NS-- datasets is horizontal which will make it much easier for reviewers to join the data back to the parent.
- NSVs can have defined variable level metadata in addition to value level metadata.
- NSVs can now be a numeric datatype; SUPP only allowed for character values.

AE Example: Comparing SUPPAE to NSAE

ae.xpt

Row	STUDYID	DOMAIN	USUBJID	AESEQ	AETERM	AESEV	AESER	AESMIE	AESTDTC	AEENDTC
1	1996001	AE	99-401	1	UTERINE FIBROIDS	SEVERE	Y	Y	2023-01-05	2023-01-12
2	1996001	AE	99-567	1	FEVER	MILD	N		2023-09-25	2023-09-25

suppae.xpt

Always character

Row	STUDYID	RDOMAIN	USUBJID	IDVAR	IDVARVAL	QNAM	QLABEL	QVAL	QORIG	QEVAL
1	1996001	AE	99-401	AESEQ	1	AESOSP	Other Medically Important SAE	SPONTANEOUS ABORTION	CRF	
2	1996001	AE	99-401	AESEQ	1	AETRTEM	Treatment Emergent Flag	Y	Derived	
3	1996001	AE	99-567	AESEQ	1	AETRTEM	Treatment Emergent Flag	N	Derived	

nsae.xpt

Row	STUDYID	RDOMAIN	USUBJID	IDVAR	IDVARVAL	AESOSP	AETRTEM
1	1996001	AE	99-401	AESEQ	1	SPONTANEOUS ABORTION	Y
2	1996001	AE	99-567	AESEQ	1		N

NSAE metadata

Can be character or numeric

Variable	Label	Type	Codelist	Role	Origin Type	Origin Source
AESOSP	Other Medically Important SAE	Char		Non-Standard Qualifier	Collected	Investigator
AETRTEM	Treatment Emergent Flag	Char	(NY)	Non-Standard Qualifier	Derived	Sponsor

TIMELINES

Progress and Milestones

Month/Year	Tasks
Oct 2021 – Mar 2025	Development of Draft Standards
Apr 2025	Internal Review
May 2025 – present	Internal Review Comment Resolution
Q4 2025	Public Review Targeted to Begin*

* Pending review and resolution of remaining Internal Review comments

- CDISC Standards Timelines & Scope Changes:
  - https://www.cdisc.org/standards/timeline

Questions?