Considerations for SDTM Implementation in Observational Studies and Real-World Data

Version 1.0 (Final)

Prepared by the
CDISC Observational Studies & Real-World Data Standards Development Team

Notes to Readers

- This document is based on SDTM v2.0 and SDTMIG v3.4.

Revision History

Date	Version
2024-02-28	1.0 Final

See Appendix D for Representations and Warranties, Limitations of Liability, and Disclaimers.

CONTENTS

1 Introduction

Observational studies differ from interventional studies in significant ways regarding study goals, study design, subject populations, treatment (or lack thereof), clinical settings, regulatory oversight requirements, and data collection practices. These differences present challenges with using CDISC standards. This document attempts to present common challenges and suggested strategies or work-arounds for using the Study Data Tabulation Model (SDTM).

This document is not a standard. The content is informative, not normative, and the strategies proposed within should not be interpreted as requirements.

Note that the challenges discussed in this guide are not a complete list of those that may be encountered when attempting to implement CDISC standards for observational research.

Version 1.0 of this document focuses on cases where the modeling of selected observational studies or real-world data (RWD) concepts is not made explicit by or requires deviation from the SDTM and/or the SDTM Implementation Guide (SDTMIG). Use of this guide requires knowledge of the SDTM and the SDTMIG and should be used in conjunction with those standards. This document is based on SDTM v2.0 and SDTMIG v3.4, available at https://www.cdisc.org/standards/foundational. The Clinical Data Acquisition Standards Harmonization (CDASH) model and the Analysis Data Model (ADaM) may be added in a future version.

Highlights to note from the FDA draft guidance *Data Standards for Drug and Biological Product Submissions Containing Real-World Data*[1] that may be helpful to users include:

1. "With adequate documentation of the conformance methods used and their rationale, study data derived from RWD can be transformed to SDTM datasets and submitted to FDA in an applicable drug submission." (p. 4)
2. "Sponsors should document data challenges encountered during transformation to an FDA-supported data standard and a justification of their approach to enable the application of an FDA-supported data standard. Mapping of standards and terminologies can be handled using the Define-XML (see the Appendix) and domain data files. Given that describing the rationale and justification for approaches used to reconcile any challenges in the source data are likely to require free-text description, in addition, a narrative should be presented in the Study Data Reviewer's Guide, either in the body or as an appendix, with appropriate directions for reviewers to the Define-XML and dataset/domains for more detail, if needed." (p. 5) For more information on Define-XML, see https://www.cdisc.org/standards/data-exchange/define-xml.
3. Appendix Table 1: Approach to Using Define-XML to Indicate Decision Involved in Transforming Non-Standardized Data (Race Data) to Standardized Data (i.e., SDTM and ADaM; p. 9)
4. Appendix Table 2: Approach to Using Define-XML to Indicate Decision Involved in Transforming Non-Standardized Data (Drugs Prescribed) to Standardized Data (i.e., SDTM and ADaM; p. 10)

1.1 Purpose

The purpose of Version 1.0 of this document is to address the most commonly encountered issues with using the SDTM for observational studies and RWD and to offer guidance and/or implementation strategies to address them. The focus is on concepts and considerations that are unique to use cases other than typical interventional studies. However, the use case of an external or historical control arm in a clinical trial setting is also included. See Section 1.2, Use Cases/Study Types, for the use cases covered in this guide.

This document assumes users have established fitness of their source data for its intended purpose and seek guidance on how to implement the SDTM for these data.

This document does not address the following topics:

- How to clean source data
- Source-to-target mapping guidance
- How to improve third-party data validation software results

1.2 Use Cases/Study Types

The scope of considerations discussed and examples shown in this document are limited to the use cases listed in the following table. Other use cases exist for observational studies and RWD; these may be explored in a future version. In addition, note that the strategies and examples shown in this guide may be applicable to use cases beyond those listed below.[2,3]

Observational Studies	**Description**
Cohort	A cohort research study compares a particular outcome (e.g., lung cancer) in groups of individuals who are alike in many ways but differ by a certain characteristic (e.g., female nurses who smoke compared with those who do not smoke).
Case control	Case-control studies compare 2 groups of people: those with the disease or condition under study (cases) and a very similar group of people who do not have the disease or condition (controls). Researchers study the medical and lifestyle histories of the people in each group to learn what factors may be associated with the disease or condition. For example, one group may have been exposed to a particular substance that the other was not.
Real World Data	**Description**
External control arm (ECA) or comparator cohort	ECA or comparator cohort studies use RWD as the standard-of-care (control) arm for the comparator cohort to a treatment arm in a clinical trial (phase 2 or 3) of an innovative therapy. This unique study design is typically used in the case of rare diseases when patient recruitment is difficult. For more information on ECA trials, see the FDA guidance *Considerations for the Design and Conduct of Externally Controlled Trials for Drug and Biological Products*.[4]

2 Types of Commonly Encountered Issues

To determine the scope of this guide, a survey was sent to members of the CDISC community to gather information about how the SDTM is being used for observational studies. The survey included the following questions:

1. What types of non-interventional studies have you worked with?
2. What are the biggest challenges you have experienced when using CDISC standards in observational studies?
3. Have you used a model/standard besides SDTM to submit the data to regulatory agencies? In what ways was this easier or more difficult than SDTM?
4. What aspects of the CDISC trial design model have and have not worked well for observational data?
5. Regarding medications data, how have you handled missing dosing information?
6. Have you attempted to create define files for observational study data, and if so, how did you approach define style sheets?
7. Have you encountered any issues with CDISC metadata (e.g., origin of data that was imputed whereas CDISC considers it collected?) If so, how did you handle this?

2.1 Using the Adverse Events and Clinical Events Domains

In RWD, medical history, adverse events, and clinical events are all noted as conditions. The distinction between them, when using the SDTM, comes down to study objective and the timing of collection. Conditions occurring prior to the designated study start would be recorded in the Medical History (MH) domain; conditions occurring after the study start reference time point would be recorded in the Adverse Events (AE) and Clinical Events (CE) domains.

In many observational studies, only serious adverse events (SAEs)—as defined by ICH criteria (e.g., death, life threatening, requiring inpatient hospitalization[5])—are collected. Additionally, adverse-event data might not be applicable for certain study types.[5,6,7]

A *clinical event* is an event of interest that would not be classified as an adverse event. It is usually an expected event related to the disease process. For example, in a coronary artery disease study a certain percentage of participants are expected to experience death, myocardial infarction, and so on. In some studies, the CE domain is used to represent the signs and symptoms of interest for a previously seen condition to aid in the review and analysis of data.

Note: Generally, the difference between adverse and clinical events is defined in the protocol, and whether to use the AE or CE domain is ultimately up to the investigator or sponsor.

Summary	Study Type	Challenge Presented	Recommended SDTM Strategy
Using the Adverse Events and Clinical Events domains	Cohort or case-control	• The AE domain was developed for regulatory safety review purposes. • Adverse events might not be relevant for all observational studies.	If an unintended consequence is considered to be an adverse event by the investigator, then the AE domain should be used.
	External control arm	• The regulatory definition of adverse events does not apply to RWD.	When working with RWD, investigators can choose to retroactively apply similar criteria for defining adverse events as defined in the protocol for the randomized controlled trial portion of the study (e.g., after the study reference start date, any new condition found in the data would be noted as an AE). Similarly, clinical events, as defined in the protocol, would be noted as such. Investigators/sponsors should make clear to reviewers how these decisions were made.

2.2 Using the Exposure and Concomitant Medications Domains

The exposure domains are used for protocol-defined treatments that may not be applicable to observational or ECA studies.

Summary	Study Type	Challenge Presented	Recommended SDTM Strategy
Using the Exposure and Concomitant/Prior Medications domains	Cohort or case-control	• No protocol-defined treatment • Distinction between therapies for disease under study vs all other treatments may not be relevant to observational studies	Any medications used to treat the disease under study should be represented using the EC/EX domains, if the investigator deems it appropriate. All other treatments (e.g., prescription, nonprescription, historical treatments) should be represented in the CM domain. Alternatively, the sponsor may choose to represent all medications in the CM domain.
	External control arm	• No protocol-defined treatment • Treatments for the disease under study vs all other treatments are important to distinguish	The medication deemed by the investigator to be the comparator to the experimental drug (i.e., used to treat the disease under study) should be represented using the EC/EX domains. All other treatments (e.g., prescription, nonprescription, historical treatments) should be represented in the CM domain.

2.3 Representing Cohorts with Planned and Actual Arm Variables

Summary	Study Type	Challenge Presented	Recommended SDTM Strategy
Using the ARM/ARMCD and ACTARM/ACTARMCD variables to represent cohorts	Cohort or case-control	• There is no dedicated variable in the Demographics (DM) domain to represent a cohort. • There may not be arms to describe in observational studies. • ARM, ARMCD, ACTARM, and ACTARMCD are expected variables in the SDTMIG.	Use the ARM variable to represent cohorts in the DM domain and NCOHORT in Trial Summary (TS) domain. Populate ARM/ARMCD and ACTARM/ACTARMCD with the same values (there may be cases where the values differ, which makes it necessary to keep both sets of variables). If there is no use case for ARM/ARMCD and ACTARM/ACTARMCD, leave values null and populate ARMNRS with the appropriate reason.
	External control arm	• Traditional arms may not be applicable.	ARM = External control ARMCD = ECA

2.4 Handling Reference Dates and Study Days

Summary	Study Type	Challenge Presented	Recommended SDTM Strategy
Defining Reference Start Date (RFSTDTC) and Reference End Date (RFENDTC)	Cohort	Study reference periods may not be relevant.	• Use registration date as RFSTDTC. • Set the date of occurrence of the evaluated event as RFENDTC. • Document how the RFSTDTC and RFENDTC were defined/populated in the Define.XML or study data reviewer's guide (SDRG).

Summary	Study Type	Challenge Presented	Recommended SDTM Strategy
	Case-control	Study reference periods are not relevant.	• Both RFSTDTC and RFENDTC should be left null. • If validating the SDTM data, explain the error in the SDRG.
	External control arm	Study reference periods may not be relevant.	• The index date[a] or another appropriate milestone date should be used as RFSTDTC. Choose the latest observation in the data as RFENDTC. • Document how RFSTDTC and RFENDTC were defined/populated in the Define.XML.
Handling study day (–DY) variables when complete dates (month, day and year) are not available.	All	• The investigator may have limited information on which to base RFSTDTC (e.g., only birth year, only month and year of participation start). • Dates associated with other observations/records may be incomplete or missing.	Study day variables (--DY) are based on RFSTDTC and --DTC. If either is not a complete date, --DY cannot be used.

[a]Designation of index date (time zero): "A specific and difficult challenge when designing externally controlled trials is specifying the index date (also called time zero or zero time), which is the start of the observation period for assessing endpoints. Given the lack of randomization in externally controlled trials, differences in the way the index date is determined across trial arms may lead to biased effect estimates. The index date for the treatment and control arms in a randomized trial is usually designated as the time when eligibility criteria are determined to have been met and a decision was made regarding the intended treatment strategy for each participant. For an externally controlled trial that relies on RWD, however, the index date for the control arm can be assigned in various ways."[4]

2.5 Trial Summary Issues

Summary	Study Type	Challenge Presented	Recommended SDTM Strategy
How to define study start date	Cohort or case-control	The SDTM defines Study Start Date (TSPARMCD = SSTDTC) as the earliest date of informed consent among any subject (Date/Time of Informed Consent, RFICDTC) that enrolls in the study. Informed consent may not be available for observational studies.	• Sponsors should set the study start date to the earliest reference start date for any subject. • Document how the study start date was defined/populated in the Define.XML or the SDRG if Define.XML is not used.
	External control arm	The SDTM defines Study Start Date (TSPARMCD = SSTDTC) as the earliest date of informed consent among any subject (Date/Time of Informed Consent, RFICDTC) that enrolls in the study. Informed consent may not be available for RWD.	It is recommended to use the earliest start date of any subject's first line of therapy.
Study type	Cohort or case-control	None identified	Use "OBSERVATIONAL" from Study Type Response codelist.
	External control arm	No controlled terminology available at this time. Study Type Response codelist is nonextensible.	Use "EXTERNAL CONTROL ARM" and explain the error in the SDRG.

3 Conformance Rules and Validation Checks

When using the SDTM for observational studies and RWD, it is likely that some conformance rules cannot be followed and validation checks will fail. This section includes examples of rules that may not be applicable to observational studies or RWD, presented at the dataset and variable level.

The CDISC Open Rules Engine (CORE) is open-source software that executes machine-readable CDISC conformance rules. CORE can be used to test study data for conformance to CDISC standards as well as regulatory and sponsor-specific conformance rule sets. In the future, the CORE project may allow users to identify sets of rules that are not applicable to observational studies or RWD, which could be turned off. In the meantime, many of these errors will need to be explained to regulatory authorities in the SDRG.

3.1 Conformance Rules - Dataset Level

Note that if a dataset is not present, subsequent checks for individual records in the required dataset would not be executed. CORE-IDs are in progress for these rules and will be added to the CDISC library.

Conformance Rule Error Message	Challenge Presented
DM dataset is missing	None identified; inclusion of a DM dataset should not present a problem, although populating some DM variables may (see Section 3.2, Conformance Rules - Variable Level). For examples, see Section 4.1, Demographics Examples.
TS dataset is missing	None identified; inclusion of a TS dataset should not present a problem. For examples, see Section 4.2, Trial Summary Examples.
TA dataset is missing	Observational studies may not have planned arms. For examples, see Section 4.3, Trial Arms Examples.
AE dataset is missing	These data may not be available. Adverse events may not be relevant to the study.
LB dataset is missing	These data may not be available.
VS dataset is missing	These data may not be available.
EX dataset is missing	Exposure data may not be relevant to the study.
DS dataset is missing	The concept of protocol-defined milestones may not be relevant to the study.
SE dataset is missing	Trial arms and elements may not be relevant to the study. Therefore, subject-level data on these concepts will not apply.
TE dataset is missing	The concept of trial elements may not be relevant to the study.

3.2 Conformance Rules - Variable Level

There are additional regulatory rules that may not be relevant to observational studies or RWD, such as no Treatment Emergent info for Adverse Event (missing SUPPAE) and inclusion of EPOCH. These errors are not listed in this table but could be explained in the SDRG, as appropriate.

Variable(s)	Domain/Dataset	Core	Conformance Rule Error Message	Rule ID[b]	Challenge Presented
SITEID	DM	Required	SDTM required variable not found in the dataset or is null.	CG0014	This data may not be available.
EPOCH	TA	Required	EPOCH is not in TA.EPOCH	CG0009	Existing controlled terminology was developed for randomized clinical trials. Use cases for observational studies have not been explored.
VISITNUM	Multiple	Expected[a]	SDTM Expected variable not found in the dataset.	CG0016	The concept of “visit” may not be relevant.

Variable(s)	Domain/Dataset	Core	Conformance Rule Error Message	Rule ID[b]	Challenge Presented
AGE	DM	Expected[a]	SDTM Expected variable not found in the dataset.	CG0016	Age data may not be available. Study sponsors submitting such data should explain this validation error accordingly. The source of this data is flexible; origin may be study specific and explained in the Define.XML.
RACE	DM	Expected[a]	SDTM Expected variable not found in the dataset.	CG0016	Race data may not have been collected according to FDA guidance. Sponsors should use the nonstandard variables CRACE (Collected Race) and CETHNIC (Collected Ethnicity) as applicable.
RFSTDTC/RFENDTC	DM	Expected[a]	SDTM Expected variable not found in the dataset.	CG0016	Study reference periods may not be relevant. See Section 2.4, Handling Reference Dates and Study Days.
RFXSTDTC/RFXENDTC	DM	Expected[a]	SDTM Expected variable not found in the dataset.	CG0016	Observational studies do not include regimented exposure to a protocol-defined drug.
RFICDTC	DM	Expected[a]	SDTM Expected variable not found in the dataset.	CG0016	Informed consent/enrollment dates may not be available.
RFPENDTC	DM	Expected[a]	SDTM Expected variable not found in the dataset.	CG0016	Date/time of end of participation may not be available.
ARM/ARMCD	DM	Expected[a]	SDTM Expected variable not found in the dataset.	CG0016	There may not be arms to describe.
ACTARM/ACTARMCD	DM	Expected[a]	SDTM Expected variable not found in the dataset.	CG0016	There may not be arms to describe.
ACTARMUD	DM	Expected[a]	SDTM Expected variable not found in the dataset.	CG0016	Treatment may not be relevant.
ARMCD	DM	Expected[a]	ARMCD is populated for a subject who was not assigned to treatment	CG0523	There may not be arms to describe. See Section 2.3, Representing Cohorts with Planned and Actual Arm Variables.
ACTARMCD	DM	Expected[a]	ACTARMCD is populated for a subject who was not assigned to treatment	CG0524	There may not be arms to describe. See Section 2.3, Representing Cohorts with Planned and Actual Arm Variables.

[a]Per SDTMIG v3.4, Section 4.1.5., "An **Expected** variable is any variable necessary to make a record useful in the context of a specific domain. Expected variables may contain some null values, but in most cases will not contain null values for every record. When the study does not include the data item for an expected variable, however, a null column must still be included in the dataset, and a comment must be included in the Define-XML document to state that the study does not include the data item."

[b]When these rules are published in CORE, they will have a CORE-ID that will differ from the Rule ID, but will be traceable back to the Rule ID via the metadata.

4 Demographics and Study Design Examples

4.1 Demographics Examples

The following example DM datasets illustrate the representation of data from cohort, case-control, and ECA studies.

Example 1: Cohort

This example assumes a prospective cohort study with the purpose of determining the incidence of lung cancer due to smoking, and confirming overall survival.

Row 1: Shows consent was obtained on 10 April 2020 and the subject was registered for participation on 12 April 2020. Lung cancer was diagnosed on 14 April 2022 (see Section 2.4, Handling Reference Dates and Study Days, for a description of RFENDTC in cohort studies). The last observation to confirm survival was completed on 7 July 2022.

Row 2: Shows consent was obtained on 12 April 2020 and observation began on 15 April 2020. Observation ended due to death on 14 June 2022.

dm.xpt

Row	STUDYID	DOMAIN	USUBJID	SUBJID	RFSTDTC	RFENDTC	RFXSTDTC	RFXENDTC	RFICDTC	RFPENDTC	DTHDTC	DTHFL	SITEID	AGE	AGEU	SEX	RACE	ARMCD	ARM	ACTARMCD	ACTARM	ARMNRS	ACTARMUD	COUNTRY
1	COHORT-001	DM	CHORT-001-1	1	2020-04-12	2022-04-14			2020-04-10	2022-07-07			1	30	YEARS	M		SMOKER	Smoker	SMOKER	Smoker			JPN
2	COHORT-001	DM	CHORT-001-2	2	2020-04-15	2022-06-14			2020-04-12	2022-06-14	2022-06-14	Y	1	72	YEARS	M		NONSMOKER	Non-Smoker	NONSMOKER	Non-Smoker			JPN

Example 2: Case Control

This example assumes a case-control study with the purpose of investigating the effects of smoking on the diagnosis of lung cancer. Informed consent was obtained by asking subjects to opt out.

Row 1: Shows there was no date of consent available. The last follow-up was on 11 July 2022.

Row 2: Shows there was no date of consent available. The date of death was 14 June 2022.

dm.xpt

Row	STUDYID	DOMAIN	USUBJID	SUBJID	RFSTDTC	RFENDTC	RFXSTDTC	RFXENDTC	RFICDTC	RFPENDTC	DTHDTC	DTHFL	SITEID	AGE	AGEU	SEX	RACE	ARMCD	ARM	ACTARMCD	ACTARM	ARMNRS	ACTARMUD	COUNTRY
1	CASE-CONTROL-001	DM	CASE-CONTROL-001-1	101						2022-07-11			1			M		OCCUR	Event occur	OCCUR	Event occur			JPN
2	CASE-CONTROL-001	DM	CASE-CONTROL-001-2	102						2022-06-14	2022-06-14	Y	1			M		NOTOCCUR	Event not occur	NOTOCCUR	Event not occur			JPN

Example 3: External Control Arm

This example assumes an ECA study. For this example, RFXSTDTC is the start date of prior line of therapy (the treatment prior to starting the index line of therapy for the same disease). The last known date of contact, death date, date lost to follow-up, and so on are represented in RFPENDTC, which is the same as RFENDTC in this example.

Row 1: Shows the date/time of first exposure to any protocol-specified treatment or therapy for the disease under study (RFXSTDTC).

Row 2: Shows the date/time of first exposure to any protocol-specified treatment or therapy (RFXSTDTC) was not available. The planned study period was 6 months. Data were not available for the reference end date as planned, the treatment start date, and the treatment end date. Observation ended due to death on 14 June 2022.

dm.xpt

Row	STUDYID	DOMAIN	USUBJID	SUBJID	RFSTDTC	RFENDTC	RFXSTDTC	RFXENDTC	RFPENDTC	DTHDTC	DTHFL	SITEID	AGE	AGEU	SEX	RACE	ARMCD	ARM	ACTARMCD	ACTARM	ARMNRS	ACTARMUD	COUNTRY
1	ECA-001	DM	ECA-001-01	1001	2022-01-15	2022-07-16	2010-02-11	2022-07-16	2022-07-16			1	40	YEARS	F		ECA	External Control	ECA	External Control			USA
2	ECA-001	DM	ECA-001-02	1002	2022-02-20	2022-06-14			2022-06-14	2022-06-14	Y	1			M		ECA	External Control	ECA	External Control			USA

4.2 Trial Summary Examples

The following example TS datasets illustrate the representation of data from cohort and case-control studies. A list of values for TSPARM and TSPARMCD can be found in CDISC Controlled Terminology, available at https://www.cancer.gov/research/resources/terminology/cdisc.

Example 1: Cohort Study

ts.xpt

Row	STUDYID	DOMAIN	TSSEQ	TSGRPID	TSPARMCD	TSPARM	TSVAL	TSVALCD	TSVCDREF	TSVCDVER
1	CHORT-001	TS	1		ACTSUB	Actual Number of Subjects	200			
2	CHORT-001	TS	1		AGEMAX	Planned Maximum Age of Subjects	P75Y		ISO 8601	
3	CHORT-001	TS	1		AGEMIN	Planned Minimum Age of Subjects	P20Y		ISO 8601	
4	CHORT-001	TS	1		CRMDUR	Confirmed Response Minimum Duration	P5Y		ISO 8601	
5	CHORT-001	TS	1		CTAUG	CDISC Therapeutic Area User Guide	Lung Cancer Therapeutic Area User Guide v1.0	C177940	CDISC	2022-06-24
6	CHORT-001	TS	1	DBL	DCUTDESC	Data Cutoff Description	DATABASE LOCK			
7	CHORT-001	TS	1	DBL	DCUTDTC	Data Cutoff Date	2022-09-26		ISO 8601	
8	CHORT-001	TS	1		EXTTIND	Extension Trial Indicator	N	C49487	CDISC	2022-06-24
9	CHORT-001	TS	1		FCNTRY	Planned Country of Investigational Sites	JPN		ISO 3166	
11	CHORT-001	TS	1		HLTSUBJI	Healthy Subject Indicator	Y	C49488	CDISC	2022-06-24
12	CHORT-001	TS	1		INDIC	Trial Disease/Condition Indication	Primary malignant neoplasm of lung (disorder)	93880001	SNOMED	
13	CHORT-001	TS	1		LENGTH	Trial Length	P10Y		ISO 8601	
14	CHORT-001	TS	1		NCOHORT	Number of Groups/Cohorts	2			
15	CHORT-001	TS	1		OBJPRIM	Trial Primary Objective	To determine incidence rate of lung cancer at 5 years			
16	CHORT-001	TS	1		OBJSEC	Trial Secondary Objective	Overall survival at 5 years			
17	CHORT-001	TS	1		OBSSMO	Observational Study Model	COHORT	C15208	CDISC	2022-06-24
18	CHORT-001	TS	1		OBSSSM	Observational Study Sampling Method	PROBABILITY SAMPLE	C71517	CDISC	2022-06-24
19	CHORT-001	TS	1		OBSSTP	Observational Study Time Perspective	PROSPECTIVE	C15273	CDISC	2022-06-24
20	CHORT-001	TS	1		OUTMSPRI	Primary Outcome Measure	LUNG CANCER FREQUENCY			
21	CHORT-001	TS	1		OUTMSSEC	Secondary Outcome Measure	DURATION			
22	CHORT-001	TS	1		PDPSTIND	Pediatric Postmarket Study Indicator	N	C49487	CDISC	2022-06-24
23	CHORT-001	TS	1		PDSTIND	Pediatric Study Indicator	N	C49487	CDISC	2022-06-24
24	CHORT-001	TS	1		PIPIND	Pediatric Investigation Plan Indicator	N	C49487	CDISC	2022-06-24

Row	STUDYID	DOMAIN	TSSEQ	TSGRPID	TSPARMCD	TSPARM	TSVAL	TSVALCD	TSVCDREF	TSVCDVER
25	CHORT-001	TS	1		PLANSUB	Planned Number of Subjects	200			
26	CHORT-001	TS	1		RDIND	Rare Disease Indicator	N	C49487	CDISC	2022-06-24
27	CHORT-001	TS	1		REGID	Registry Identifier	NCT123456789	NCT123456789	CDISC	2022-06-24
28	CHORT-001	TS	1		RLPSCRIT	Relapse Criteria	RECIST 1.1	C124415	CDISC	2022-06-24
29	CHORT-001	TS	1		SDMDUR	Stable Disease Minimum Duration	P28D		ISO 8601	
30	CHORT-001	TS	1		SENDTC	Study End Date	2022-09-26		ISO 8601	
31	CHORT-001	TS	1		SEXPOP	Sex of Participants	BOTH	C49636	CDISC	2022-06-24
32	CHORT-001	TS	1		SPONSOR	Clinical Study Sponsor	Pharmaco	123456789	D-U-N-S NUMBER	
33	CHORT-001	TS	1		SDTMVER	SDTM Version	2.0			
34	CHORT-001	TS	1		SDTIGVER	SDTM IG Version	3.4			
35	CHORT-001	TS	1		STOPRULE	Study Stop Rules	None			
36	CHORT-001	TS	1		STRATFCT	Stratification Factor	SMOKING STATUS			
37	CHORT-001	TS	1		SSTDTC	Study Start Date	2012-09-26		ISO 8601	
38	CHORT-001	TS	1		STYPE	Study Type	OBSERVATIONAL	C16084	CDISC	2022-06-24
39	CHORT-001	TS	1		TDIGRP	Diagnosis Group	Primary malignant neoplasm of lung (disorder)	93880001	SNOMED	
40	CHORT-001	TS	1		THERAREA	Therapeutic Area	Lung Cancer			
41	CHORT-001	TS	1		TITLE	Trial Title	Observational study on the incidence rate of lung cancer due to smoking			
42	CHORT-001	TS	1		TPHASE	Trial Phase Classification	NOT APPLICABLE	C48660	CDISC	2022-06-24
43	CHORT-001	TS	1		TTYPE	Trial Type	NOT APPLICABLE			

Example 2: Case-Control Study

ts.xpt

Row	STUDYID	DOMAIN	TSSEQ	TSGRPID	TSPARMCD	TSPARM	TSVAL	TSVALCD	TSVCDREF	TSVCDVER
1	CASE-CONTROL-001	TS	1		ACTSUB	Actual Number of Subjects	200			
2	CASE-CONTROL-001	TS	1		AGEMAX	Planned Maximum Age of Subjects	P75Y		ISO 8601	
3	CASE-CONTROL-001	TS	1		AGEMIN	Planned Minimum Age of Subjects	P20Y		ISO 8601	
4	CASE-CONTROL-001	TS	1		CRMDUR	Confirmed Response Minimum Duration	P5Y		ISO 8601	
5	CASE-CONTROL-001	TS	1		CTAUG	CDISC Therapeutic Area User Guide	Lung Cancer Therapeutic Area User Guide v1.0	C177940	CDISC	2022-06-24
6	CASE-CONTROL-001	TS	1	DBL	DCUTDESC	Data Cutoff Description	DATABASE LOCK			
7	CASE-CONTROL-001	TS	1	DBL	DCUTDTC	Data Cutoff Date	2022-09-26		ISO 8601	

Row	STUDYID	DOMAIN	TSSEQ	TSGRPID	TSPARMCD	TSPARM	TSVAL	TSVALCD	TSVCDREF	TSVCDVER
8	CASE-CONTROL-001	TS	1		EXTTIND	Extension Trial Indicator	N	C49487	CDISC	2022-06-24
9	CASE-CONTROL-001	TS	1		FCNTRY	Planned Country of Investigational Sites	JPN		ISO 3166	
11	CASE-CONTROL-001	TS	1		HLTSUBJI	Healthy Subject Indicator	Y	C49488	CDISC	2022-06-24
12	CASE-CONTROL-001	TS	1		INDIC	Trial Disease/Condition Indication	Primary malignant neoplasm of lung (disorder)	93880001	SNOMED	
13	CASE-CONTROL-001	TS	1		LENGTH	Trial Length	P10Y		ISO 8601	
14	CASE-CONTROL-001	TS	1		NCOHORT	Number of Groups/Cohorts	2			
15	CASE-CONTROL-001	TS	1		OBJPRIM	Trial Primary Objective	Effect of Smoking on Lung Cancer Incidence Rate			
16	CASE-CONTROL-001	TS	1		OBJSEC	Trial Secondary Objective	Overall survival at 5 years			
17	CASE-CONTROL-001	TS	1		OBSSMO	Observational Study Model	CASE CONTROL	C15197	CDISC	2022-06-24
18	CASE-CONTROL-001	TS	1		OBSSSM	Observational Study Sampling Method	PROBABILITY SAMPLE	C71517	CDISC	2022-06-24
19	CASE-CONTROL-001	TS	1		OBSSTP	Observational Study Time Perspective	RETROSPECTIVE	C53312	CDISC	2022-06-24
20	CASE-CONTROL-001	TS	1		OUTMSPRI	Primary Outcome Measure	ODDS RATIO			
21	CASE-CONTROL-001	TS	1		OUTMSSEC	Secondary Outcome Measure	DURATION			
22	CASE-CONTROL-001	TS	1		PDPSTIND	Pediatric Postmarket Study Indicator	N	C49487	CDISC	2022-06-24
23	CASE-CONTROL-001	TS	1		PDSTIND	Pediatric Study Indicator	N	C49487	CDISC	2022-06-24
24	CASE-CONTROL-001	TS	1		PIPIND	Pediatric Investigation Plan Indicator	N	C49487	CDISC	2022-06-24
25	CASE-CONTROL-001	TS	1		PLANSUB	Planned Number of Subjects	200			
26	CASE-CONTROL-001	TS	1		RDIND	Rare Disease Indicator	N	C49487	CDISC	2022-06-24
27	CASE-CONTROL-001	TS	1		REGID	Registry Identifier	NCT123456789	NCT123456789	CDISC	2022-06-24
28	CASE-CONTROL-001	TS	1		RLPSCRIT	Relapse Criteria	RECIST 1.1	C124415	CDISC	2022-06-24
29	CASE-CONTROL-001	TS	1		SDMDUR	Stable Disease Minimum Duration	P28D		ISO 8601	
30	CASE-CONTROL-001	TS	1		SENDTC	Study End Date	2022-09-26		ISO 8601	
31	CASE-CONTROL-001	TS	1		SEXPOP	Sex of Participants	BOTH	C49636	CDISC	2022-06-24
32	CASE-CONTROL-001	TS	1		SPONSOR	Clinical Study Sponsor	Pharmaco	123456789	D-U-N-S NUMBER	
33	CASE-CONTROL-001	TS	1		SDTMVER	SDTM Version	2.0			
34	CASE-CONTROL-001	TS	1		SDTIGVER	SDTM IG Version	3.4			
35	CASE-CONTROL-001	TS	1		STOPRULE	Study Stop Rules	None			
36	CASE-CONTROL-001	TS	1		STRATFCT	Stratification Factor	LUNG CANCER OCCURRENCE			
37	CASE-CONTROL-001	TS	1		SSTDTC	Study Start Date	2012-09-26		ISO 8601	

Row	STUDYID	DOMAIN	TSSEQ	TSGRPID	TSPARMCD	TSPARM	TSVAL	TSVALCD	TSVCDREF	TSVCDVER
38	CASE-CONTROL-001	TS	1		STYPE	Study Type	OBSERVATIONAL	C16084	CDISC	2022-06-24
39	CASE-CONTROL-001	TS	1		TDIGRP	Diagnosis Group	Primary malignant neoplasm of lung (disorder)	93880001	SNOMED	
40	CASE-CONTROL-001	TS	1		THERAREA	Therapeutic Area	Lung Cancer			
41	CASE-CONTROL-001	TS	1		TITLE	Trial Title	Observational study on the effect of smoking of lung cancer incidence rate			
42	CASE-CONTROL-001	TS	1		TPHASE	Trial Phase Classification	NOT APPLICABLE	C48660	CDISC	2022-06-24
43	CASE-CONTROL-001	TS	1		TTYPE	Trial Type	NOT APPLICABLE			

4.3 Trial Arms Examples

The following example TA datasets illustrate the representation of data from cohort and case-control studies. See Section 2.3, Representing Cohorts with Planned and Actual Arm Variables, for details about using ARMCD and ARM.

Example 1: Cohort Study

ta.xpt

Row	STUDYID	DOMAIN	ARMCD	ARM	TAETORD	ETCD	ELEMENT	TABRANCH	TATRANS	EPOCH
1	COHORT-001	TA	SMOKE	Smoking	1	SC	Screen	Assigned to smoking cohort		SCREENING
2	COHORT-001	TA	SMOKE	Smoking	3	OBS	Observation		If lung cancer occurs, go to follow-up epoch	OBSERVATION
3	COHORT-001	TA	SMOKE	Smoking	3	FU	Follow-up			FOLLOW-UP
4	COHORT-001	TA	NONSMOKE	Non-Smoking	1	SC	Screen	Assigned to non-smoking cohort		SCREENING
5	COHORT-001	TA	NONSMOKE	Non-Smoking	2	OBS	Observation		If lung cancer occurs, go to follow-up epoch	OBSERVATION
6	COHORT-001	TA	NONSMOKE	Non-Smoking	2	FU	Follow-up			FOLLOW-UP

Example 2: Case-Control Study

ta.xpt

Row	STUDYID	DOMAIN	ARMCD	ARM	TAETORD	ETCD	ELEMENT	TABRANCH	TATRANS	EPOCH
1	CASE-CONTROL-001	TA	OCCUR	Event occur	1	SS	Smoking status			FOLLOW-UP
2	CASE-CONTROL-001	TA	NOTOCCUR	Event not occur	1	SS	Smoking status			FOLLOW-UP

Appendices

Appendix A: Project Team

Name	Organization
Jon Neville - Team Lead	CDISC
Virginie Aris	Bayer
Yuliia Bahatska	Johnson & Johnson
Rebecca Baker	CDISC
Adrienne Boyance	Amgen
Lorraine Fang	Bristol Myers Squibb
Hidenori Hamano	The University of Tokyo Hospital
Kaori Nagai	NHO Nagoya Medical Center
Toshiki Saito	NHO Headquarters & Nagoya Medical Center
Suneeta Saranu	Bristol Myers Squibb
Alana St. Clair	CDISC
Joe Xi	Bristol Myers Squibb
Matsuo Yamamoto	NHO Nagoya Medical Center

Appendix B: Glossary and Abbreviations

The following abbreviations and terms are used in this document. Additional definitions can be found in the CDISC Glossary, available at https://www.cdisc.org/standards/glossary.

ADaM	Analysis Data Model
CDASH	Clinical Data Acquisition Standards Harmonization (model)
CDISC	Clinical Data Interchange Standards Consortium
Controlled terminology	A finite set of values that represent the only allowed values for a data item. These values may be codes, text, or numeric. A code list is one type of controlled terminology.
CORE	CDISC Open Rules Engine
Domain	A collection of observations with a topic-specific commonality about a subject.
ECA	External control arm
FDA	(US) Food & Drug Administration
ICH	International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use
RWD	Real-world data
SAE	Serious adverse event
SDRG	Study data reviewer's guide
SDTM	Study Data Tabulation Model
SDTMIG	SDTM Implementation Guide (for Human Clinical Trials)
Subject	A participant in a study

Appendix C: References

1. US Food & Drug Administration. *Data Standards for Drug and Biological Product Submissions Containing Real-World Data*. *Draft Guidance for Industry*. Published October 2021. Accessed August 18, 2023. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/data-standards-drug-and-biological-product-submissions-containing-real-world-data
2. Song JW, Chung KC. Observational studies: cohort and case-control studies. *J Plast Reconstr Aesthet Surg* 2010;126(6):2234-2242. https://doi.org/10.1097/PRS.0b013e3181f44abc
3. Rezigalla AA. Observational study designs: synopsis for selecting an appropriate study design. *Cureus* 2020;12(1):e6692. https://doi.org/10.7759/cureus.6692
4. US Food & Drug Admnistration. *Considerations for the Design and Conduct of Externally Controlled Trials for Drug and Biological Products*. Published February 2023. Accessed August 18, 2023. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/considerations-design-and-conduct-externally-controlled-trials-drug-and-biological-products
5. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. *Post-Approval Safety Data Management: Definitions and Standards for Expedited Reporting E2D*. *Step 4 version*. Published November 2003. Accessed February 9, 2024. https://database.ich.org/sites/default/files/E2D_Guideline.pdf
6. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. *Choice of Control Group and Related Issues in Clinical Trials E10*. Published July 20, 2000. Accessed August 18, 2023. https://database.ich.org/sites/default/files/E10_Guideline.pdf
7. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. *A Selective Approach to Safety Data Collection in Specific Late-Stage Pre-Approval or Post-Approval Clinical Trials E19*. Adopted September 27, 2022. Accessed August 18, 2023. https://database.ich.org/sites/default/files/ICH_E19_Guideline_Step4_2022_0826_0.pdf

Appendix D: Representations and Warranties, Limitations of Liability, and Disclaimers

CDISC Patent Disclaimers

It is possible that implementation of and compliance with this standard may require use of subject matter covered by patent rights. By publication of this standard, no position is taken with respect to the existence or validity of any claim or of any patent rights in connection therewith. CDISC, including the CDISC Board of Directors, shall not be responsible for identifying patent claims for which a license may be required in order to implement this standard or for conducting inquiries into the legal validity or scope of those patents or patent claims that are brought to its attention.

Representations and Warranties

"CDISC grants open public use of this User Guide (or Final Standards) under CDISC's copyright."

Each Participant in the development of this standard shall be deemed to represent, warrant, and covenant, at the time of a Contribution by such Participant (or by its Representative), that to the best of its knowledge and ability: (a) it holds or has the right to grant all relevant licenses to any of its Contributions in all jurisdictions or territories in which it holds relevant intellectual property rights; (b) there are no limits to the Participant's ability to make the grants, acknowledgments, and agreements herein; and (c) the Contribution does not subject any Contribution, Draft Standard, Final Standard, or implementations thereof, in whole or in part, to licensing obligations with additional restrictions or requirements inconsistent with those set forth in this Policy, or that would require any such Contribution, Final Standard, or implementation, in whole or in part, to be either: (i) disclosed or distributed in source code form; (ii) licensed for the purpose of making derivative works (other than as set forth in Section 4.2 of the CDISC Intellectual Property Policy ("the Policy")); or (iii) distributed at no charge, except as set forth in Sections 3, 5.1, and 4.2 of the Policy. If a Participant has knowledge that a Contribution made by any Participant or any other party may subject any Contribution, Draft Standard, Final Standard, or implementation, in whole or in part, to one or more of the licensing obligations listed in Section 9.3, such Participant shall give prompt notice of the same to the CDISC President who shall promptly notify all Participants.

No Other Warranties/Disclaimers. ALL PARTICIPANTS ACKNOWLEDGE THAT, EXCEPT AS PROVIDED UNDER SECTION 9.3 OF THE CDISC INTELLECTUAL PROPERTY POLICY, ALL DRAFT STANDARDS AND FINAL STANDARDS, AND ALL CONTRIBUTIONS TO FINAL STANDARDS AND DRAFT STANDARDS, ARE PROVIDED "AS IS" WITH NO WARRANTIES WHATSOEVER, WHETHER EXPRESS, IMPLIED, STATUTORY, OR OTHERWISE, AND THE PARTICIPANTS, REPRESENTATIVES, THE CDISC PRESIDENT, THE CDISC BOARD OF DIRECTORS, AND CDISC EXPRESSLY DISCLAIM ANY WARRANTY OF MERCHANTABILITY, NONINFRINGEMENT, FITNESS FOR ANY PARTICULAR OR INTENDED PURPOSE, OR ANY OTHER WARRANTY OTHERWISE ARISING OUT OF ANY PROPOSAL, FINAL STANDARDS OR DRAFT STANDARDS, OR CONTRIBUTION.

Limitation of Liability

IN NO EVENT WILL CDISC OR ANY OF ITS CONSTITUENT PARTS (INCLUDING, BUT NOT LIMITED TO, THE CDISC BOARD OF DIRECTORS, THE CDISC PRESIDENT, CDISC STAFF, AND CDISC MEMBERS) BE LIABLE TO ANY OTHER PERSON OR ENTITY FOR ANY LOSS OF PROFITS, LOSS OF USE, DIRECT, INDIRECT, INCIDENTAL, CONSEQUENTIAL, OR SPECIAL DAMAGES, WHETHER UNDER CONTRACT, TORT, WARRANTY, OR OTHERWISE, ARISING IN ANY WAY OUT OF THIS POLICY OR ANY RELATED AGREEMENT, WHETHER OR NOT SUCH PARTY HAD ADVANCE NOTICE OF THE POSSIBILITY OF SUCH DAMAGES.

Note: The CDISC Intellectual Property Policy can be found at
http://www.cdisc.org/system/files/all/article/application/pdf/cdisc_20ip_20policy_final.pdf