SDTM Variables You Might Forget About

Nadiia Pukhliar, Dariia Tsyhanenko, Iryna Kotenko, Intego Group, LLC.

ABSTRACT

With every new version of the SDTM/SDTMIG more and more examples of raw data mapping are presented, more details on specific variables are described. However, in practice the same mapping rules are transferred from study to study with no changes and everything less common is either mapped to the Supplemental Qualifiers SUPP-- datasets, the Findings About Events or Interventions (FA) domain or is not submitted at all. The paper collects several cases of SDTM mapping providing more coherent and detailed representation of collected data.

Special attention in the paper is given to the Supplemental Qualifiers datasets examining standard supplemental qualifiers name codes per the SDTMIG. Further, we are sharing tricks on using the ADaM IG to get standard qualifier names in SUPP-- domains.

Additional focus of the paper is on using accompanying text in the CRF and the protocol to procure more context in SDTM datasets by creating standard variables from the model that are not described in the implementation guide.

The examples provided represent CRF pages and studies from our practice, they are a great testament to the versatility of the SDTM that covers various study and data collection designs.

INTRODUCTION

The Study Data Tabulation Model Implementation Guide for Human Clinical Trials (SDTMIG) is the most important and used document that you have on hand while working on SDTM creation. It provides general instructions as well as a list of common variables for every standard domain with examples and additional assumptions. In general, an ability to navigate the SDTMIG efficiently allows to map most of collected data for a typical clinical trial. However, the Study Data Tabulation Model (SDTM) is much more than described in the SDTMIG, and using the SDTM as a separate document gives an idea of mapping data that does not fit any variable provided for specific domain in the SDTMIG but still can be mapped to a standard variable from the SDTM.

Talking about study documentation that is actively used during SDTM programming, you think about the Case Report Form (CRF) first. Similarly, a CRF provides needed information for most of the collected data but it is not enough for comprehensive representation of the data in a context of specific study. Study protocol describes a schedule of planned assessments with specifications that can be omitted in the CRF. Also additional details about investigational product, study procedures, schedule of assessments can be mentioned in the protocol.

SDTM datasets mainly include collected data, nevertheless the model allows representation of derived records and/or variables in the main and supplemental datasets. If you need to implement any derivations on the SDTM level it may be a good idea to use the Statistical Analysis Plan (SAP) in addition to the protocol.

Since most subject-level observations collected during the study should be represented according to one of the three SDTM general observation classes, the paper focuses on them. We referred to the SDTM to find the lists of variables allowed to be used in each of the General Observation Classes.

We present several cases of efficient use of study documentation and CDISC standards during the SDTM programming process. Most of them describe non-study specific cases and therefore may be helpful for wide use.

THE INTERVENTIONS OBSERVATION CLASS

Based on the SDTM definition, the Interventions class captures investigational, therapeutic, and other treatments that are administered to the subject (with some actual or expected physiological effect) either as specified by the study protocol (e.g., exposure to study drug), coincident with the study assessment period (e.g., concomitant medications), or self-administered by the subject (such as use of alcohol, tobacco, or caffeine).

We will provide examples of using the following qualifier variables from the class:

Variable Name	Variable Label	Type	Role	Description
--PSTRG	Pharmaceutical Strength	Num	Record Qualifier	Amount of an active ingredient expressed quantitatively per dosage unit, per unit of volume, or per unit of weight, according to the pharmaceutical dose form. Example: "50 mg/TABLET", "300 mg/L".
--PSTRGU	Pharmaceutical Strength Units	Char	Variable Qualifier of --PSTRG	Unit for --PSTRG. Example: "mg/TABLET", "mg/mL".
--TRTV	Treatment Vehicle	Char	Record Qualifier	Vehicle for administration of treatment, such as a liquid in which the treatment drug is dissolved. Example: "SALINE".
--VAMT	Treatment Vehicle Amount	Num	Record Qualifier	Amount of the prepared product (treatment + vehicle) administered or given. Note: should not be diluent amount alone.
--VAMTU	Treatment Vehicle Amount Units	Char	Variable Qualifier of --VAMT	Units for the prepared product (treatment + vehicle). Examples: "mL", "mg".

Table 1. Interventions – Qualifier Variables

CONCOMITANT/PRIOR MEDICATIONS (CM)

Most studies include a separate CRF page for concomitant/prior medications collection. In our experience, the list of variables specified in the SDTMIG for the CM domain covers all data collected at the page for concomitant/prior medications. However, it happens that data collected at other CRF pages has to be mapped to the CM domain. An example of such pages can be Procedural Anesthetics administered to a subject before treatment:

Anesthetic Name	_____
Type	☐ Topical ☐ Local Injection
Concentration	_____ Fixed Unit: %
Volume	_____ Fixed Unit: mL

Table 2. CRF page for Procedural Anesthetics

When we have both concentration and volume provided, it gives an idea that volume represents the amount of substance administered to a subject with active ingredients diluted according to concentration value. Then volume has to be mapped to the CMVAMT variable with CMVAMT = mL. Depending on whether we have additional information in supplemental documentation, we may have or not have the name of the vehicle used that can be mapped to the CMTRTV variable.

EXPOSURE (EX)

The SDTMIG lists several common methods for determining exposure:

1. Derived from actual observation of the administration of drug by the investigator;

2. Derived from automated dispensing device that records administrations;

3. Derived from subject recall;

4. Derived from drug accountability data;

5. Derived from the protocol.

The first method is the most direct and commonly used, however it can be combined with others to get a more comprehensive representation of the data. Let us consider an example of combining data collected in the CRF with the protocol information.

Injectable gel is one of the dose forms widely used in aesthetic treatment in dermatology clinical trials of augmentation therapy, correction of face contour deficiencies, wrinkles treatment. Gels are formed by mixing active ingredients with supportive compounds so that an active pharmaceutical ingredient is released over a period of time. Typical CRF page for treatment administration may look as following:

Treatment Date	_____
Treatment Area	☐ Forehead ☐ Left Cheek ☐ Right Cheek ☐ Chin
Dose Administered	_____ Fixed Unit: mL

Table 3. CRF page for Treatment Administration

Data collected at this page in addition to dose form and name of treatment that we know from the protocol, provides all needed information to be mapped to required and expected variables in the EX domain. However, information obtained from the protocol can be used to indicate more specific details of the treatment in the dataset. It is specified in the protocol for aesthetic treatment that injectable gel contains 10 mg/mL stabilized hyaluronic acid and 2 mg/mL lidocaine hydrochloride in a physiological buffer.

Thus, standard SDTM variables EXPSTRG and EXPSTRGU should be added to the EX dataset to provide full description of the treatment:

USUBJID	EXTRT	EXDOSE	EXDOSU	EXPSTRG	EXPSTRGU	EXSTDTC
123-0001	Drug A	3	mL	10	mg/mL	2023-02-01T08:00
123-0001	Drug A	5	mL	10	mg/mL	2023-02-11T08:00

Table 4. EX dataset

THE EVENTS OBSERVATION CLASS

The Events class captures planned protocol milestones such as randomization and study completion, and occurrences, conditions, or incidents independent of planned study evaluations occurring during the trial (e.g., adverse events) or prior to the trial (e.g., medical history).

We will provide examples and explain the differences among the following qualifier variables from the class presented in Table 2:

Variable Name	Variable Label	Type	Role	Description
--SEV	Severity/Intensity	Char	Record Qualifier	The severity or intensity of the event. Examples: "MILD", "MODERATE", "SEVERE".

Variable Name	Variable Label	Type	Role	Description
--REL	Causality	Char	Record Qualifier	Records the investigator's opinion as to the causality of the event to the treatment. ICH E2A and E2B examples include "NOT RELATED", "UNLIKELY RELATED", "POSSIBLY RELATED", "RELATED".
--RELNST	Relationship to Non-Study Treatment	Char	Record Qualifier	An opinion as to whether the event may have been due to a treatment other than study drug. Example: "MORE LIKELY RELATED TO ASPIRIN USE".
--TOX	Toxicity	Char	Variable Qualifier of --TOXGR	Description of toxicity quantified by --TOXGR such as NCI CTCAE Short Name. Examples: "HYPERCALCEMIA", "HYPOCALCEMIA". Sponsor should specify which scale and version is used in the Sponsor Comments column of the Define-XML document.
--TOXGR	Toxicity Grade	Char	Record Qualifier	Records toxicity grade using a standard toxicity scale (such as the NCI CTCAE). Sponsor should specify which scale and version is used in the Sponsor Comments column of the Define.XML document.

Table 5. Events – Qualifier Variables

ADVERSE EVENTS (AE)

Severity/Intensity and Toxicity

There are two common examples how severity information could be collected on the CRF:

Severity	☐ Mild ☐ Moderate ☐ Severe

Table 6. CRF field for Severity/Intensity

Severity/ Severity per CTCAE Grade	☐ Mild ☐ Moderate ☐ Severe ☐ Life Threatening or Disabling ☐ Death

Table 7. CRF field for Standard Toxicity Grade

The first case is fairly clear - the raw variable is mapped to the AESEV variable, as it fits the meaning of this SDTM variable and has values completely matched with the AESEV Controlled Terminology codelist, which is non-extensible.

The second case illustrates the CRF field in studies using toxicity grading according to a standard toxicity scale such as Common Terminology Criteria for Adverse Events (CTCAE), so AETOXGR variable should be used for mapping instead of AESEV. If the value is from a numeric scale, AETOXGR represents only the number (e.g., "2" and not "Grade 2"). AETOX describes toxicity quantified by AETOXGR and should not have values "MILD", "MODERATE", etc. since it is already reflected by the grade. For CTCAE v5.0 scale AETOX is a MedDRA Lowest Level Term (LLT) located in the CTCAE Term column. For example, AETOX = "Anemia":

Blood and lymphatic system disorders					
CTCAE Term	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5
Anemia	Hemoglobin (Hgb) <LLN - 10.0 g/dL; <LLN - 6.2 mmol/L; <LLN - 100 g/L	Hgb <10.0 - 8.0 g/dL; <6.2 - 4.9 mmol/L; <100 - 80g/L	Hgb <8.0 g/dL; <4.9 mmol/L; <80 g/L; transfusion indicated	Life-threatening consequences; urgent intervention indicated	Death
Definition: A disorder characterized by a reduction in the amount of hemoglobin in 100 ml of blood. Signs and symptoms of anemia may include pallor of the skin and mucous membranes, shortness of breath, palpitations of the heart, soft systolic murmurs, lethargy, and fatigability. **Navigational Note:** -					

Table 8. CTCAE v5.0 - Anemia

Be aware that according to the FDA validation rule "Toxicity (--TOX) variable should not be present in the domain, when the Toxicity Grade (--TOXGR) variable is missing."

Causality and Relationship

Per ICH E2A, causality assessment is required for clinical investigation cases, which means that we will definitely see the resembling field on the Adverse Events CRF Form:

Relationship to Study Treatment	☐ Not Related ☐ Unlikely ☐ Possibly ☐ Probably ☐ Definitely

Table 9. CRF field for Causality

The SDTM includes standard variables for only two possible causes, AEREL for relationship to study treatment, and AERELNST for relationship to non-study treatments. There is no doubt we have to map the CRF field from Table 5 to the AEREL variable. However, in studies with comparator treatment or drug-drug interaction, the CRF includes similar field for causality information for non-study treatment:

Relationship to Comparator Treatment	☐ Not Related ☐ Unlikely ☐ Possibly ☐ Probably ☐ Definitely ☐ Not Applicable

Table 10. CRF field for Relationship to Non-Study Treatment

As another example, sponsor can define their own pre-specified list of values to check intervention(s) which caused adverse event:

Relationship to Non-Study Treatment	☐ Adjunct Therapy ☐ Concomitant Therapy ☐ Study Device ☐ Study Disease ☐ Study Procedure ☐ Other Medical Condition

Table 11. CRF field for Relationship to Non-Study Treatment with custom codelist

Relationship to Non-Study Treatment may be reported as free text too.

In such instances we should use SDTM standard variable AERELNST rather than map these fields to a separate QNAM in the SUPPAE dataset as you may decide at a glance.

THE FINDINGS OBSERVATION CLASS

The Findings class captures the observations resulting from planned evaluations to address specific tests or questions such as laboratory tests, ECG testing, and questions listed on questionnaires.

On the example of SDTM domain from the findings observation class, we will illustrate importance of the following timing variables presented in Table 9:

Variable Name	Variable Label	Type	Format	Description
--TPTREF	Time Point Reference	Char		Description of the fixed reference point referred to by --ELTM, --TPTNUM, --TPT, --STINT, and --ENINT. Examples: "PREVIOUS DOSE", "PREVIOUS MEAL".
--RFTDTC	Date/Time of Reference Time Point	Char	ISO 8601	Date/time for a fixed reference time point defined by --TPTREF.
--EVLINT	Evaluation Interval	Char	ISO 8601	Duration of interval associated with an observation such as a finding --TESTCD. Usually used with --DTC to describe an interval of this duration that ended at the time represented in --DTC. Example: "-P2M" to represent a period of the past 2 months as the evaluation interval for a question from a questionnaire.
--EVINTX	Evaluation Interval Text	Char		Evaluation interval associated with an observation, where the interval is not able to be represented in ISO 8601 format. Examples: "LIFETIME", "LAST NIGHT", "RECENTLY", "OVER THE LAST FEW WEEKS".

Table 12. All Observation Classes—Timing Variables

PHARMACOKINETICS CONCENTRATIONS (PC)

The SDTMIG provides several variables to represent time points: --TPT, --TPTNUM, --ELTM, and time point anchors: --TPTREF (text description) and --RFTDTC (the date/time).

These variables cover two different types of collected data: repeated measurements and measurements scheduled relative to a reference time point. The second one is typical for pharmacokinetics assessments with results presented in the Pharmacokinetic (PC) domain. Since pharmacokinetics shows how the body interacts with study drug over time, date and time of study treatment should be described in the PCRFTDTC variable with text description in PCTPTREF:

USUBJID	PCTEST	PCTPT	PCTPTNUM	PCTPTREF	PCRFTDTC	PCELTM
123-0001	Drug A Metabolite	PREDOSE	0	Day 1 Dose	2023-02-01T08:00	-PT15M
123-0001	Drug A Metabolite	1H30MIN	1.5	Day 1 Dose	2023-02-01T08:00	PT1H30M
123-0001	Drug A Metabolite	PREDOSE	0	Day 11 Dose	2023-02-11T08:00	-PT15M

USUBJID	PCTEST	PCTPT	PCTPTNUM	PCTPTREF	PCRFTDTC	PCELTM
123-0001	Drug A Metabolite	1H30MIN	1.5	Day 11 Dose	2023-02-11T08:00	PT1H30M

Table 13. PC domain - Reference Time Point representation

The SDTMIG specifies that if there are multiple PK profiles being generated, the zero time for each will be different (e.g., a different dose such as "first dose", "second dose") and, as a result, values for PCTPTREF must be different.

Although PCRFTDTC is permissible in the SDTMIG, it may play an important role in this dataset. It could be used to easily retrieve all needed exposure data from EX dataset, to link the PC records to the respective Pharmacokinetics Parameters (PP) records using PPRFTDTC, or to derive actual elapsed time further in the analysis dataset (PCELTM is the planned elapsed time), if desired, as PCDTC-PCRFTDTC.

QUESTIONNAIRES (QS)

There are numerous ways to collect findings in the clinical trials: data may refer to a particular point of time, specific interval, or general period.

For questionnaires it is common to be related to some period of time, because they are often used to quantify feelings, such as pain, which are not simply measured straightaway as laboratory tests.

The SDTM provides two timing variables --EVLINT and --EVLTXT, which can describe evaluation interval or recall period.

The evaluation period of time for the questionnaire is populated in the QSEVLINT variable when it can be represented in ISO 8601 format. Otherwise QSEVINTX with text description is used.

Indication of the interval could be collected on the CRF field directly in a separate variable:

C-SSRS Recall Period	☐ Since Last Study ☐ Past 2 Weeks ☐ Since Last Visit
Date of Administration	_____

Table 14. CRF page for C-SSRS Evaluation Interval

As "Since Last Study" and "Since Last Visit" descriptions cannot be represented in ISO 8601 format we have to use QSEVINTX with the values "SINCE LAST STUDY" and "SINCE LAST VISIT" respectively. Whereas "Past 2 Weeks" has precise ISO 8601 representation - "-P2W", which will be the value of QSEVLINT.

Questionnaires often have supportive text on the CRF pages, and can be considered as a good example when interval is mentioned in the text of question or in subtitle of specific section of the questionnaire:

EORTC QLQ-C30	
During the past week:	
Were you limited in doing either your work or other daily activities?	☐ 1 [Not at All] ☐ 2 [A little] ☐ 3 [Quite a Bit] ☐ 4 [Very Much]
How would you rate your overall health during the past week?	1 2 3 4 5 6 7 Very poor Excellent

Table 15. CRF page for EORTC QLQ-C30 with supportive text

Recall period mentioned in the questionnaire may sometimes seem too indirect. For example, the Epworth Sleepiness Scale (ESS) questionnaire is evaluated by the subject based on "usual way of life in

recent times". Since it still can be precisely described as a period, we can populate the QSEVINTX variable with the value "RECENT TIMES".

Therefore, --EVLINT and --EVLTXT can be assigned to the appropriate value describing the interval even if there is no certain field on the CRF.

SUPPLEMENTAL QUALIFIERS – SUPP-- DATASETS

Since the SDTM does not allow the addition of new variables, the Supplemental Qualifiers special purpose dataset model is used to capture non-standard variables and their association to parent records in general-observation-class datasets (Events, Findings, Interventions) and Demographics. Variable QNAM represents the name of the Qualifier variable being added, and variable QLABEL represents its label.

The SDTMIG provides an initial set of standard name codes to be used in SUPP-- datasets:

QNAM	QLABEL	Applicable Domains
AESOSP	Other Medically Important SAE	AE
AETRTEM	Treatment Emergent Flag	AE
--CLSIG	Clinically Significant	Findings
--REAS	Reason	All general observation classes

Table 16. Supplemental Qualifiers Name Codes

Note: there is the FDA business rule SD1097/SD1321 that states "A treatment-emergent flag should be submitted." Therefore, the SUPPAE dataset has to include a record with QNAM = "AETRTEM" for every adverse event presented in a parent domain.

Additionally, the ADaMIG defines subject-level population flags in relation to SDTM Supplemental Qualifiers, i.e. if you create population flags in the SDTM, they have to follow naming convention as specified:

SDTM QNAM	SDTM QLABEL	ADaM Subject-Level Population Flags
COMPLT	Completers Population Flag	COMPLFL
FULLSET	Full Analysis Set Flag	FASFL
ITT	Intent to Treat Population Flag	ITTFL
PPROT	Per Protocol Set Flag	PPROTFL
SAFETY	Safety Population Flag	SAFFL

Table 17. ADaM Subject-Level Population Flags Corresponding to SDTM Supplemental Qualifiers

For concomitant medications, supplemental qualifiers are widely used to represent additional coding dictionary information such as a drug's Anatomical Therapeutic Chemical code from WHO Drug dictionary (ATC codes). There are no standard QNAM/QLABEL specified for these variables in the SDTMIG, however you can follow three easy steps that lead you to the only correct result. Let us move backwards from ADaM datasets to SDTM:

1. The ADaM Occurrence Data Structure (OCCDS) version 1.0 indicates the following variables representing WHO Drug terms and codes:

Variable Name	Variable Label	Type	Codelist / Controlled Terms	Core	CDISC Notes
CMDECOD	Standardized Medication Name	Char	WHO Drug	Cond	CM.CMDECOD This is typically one of the primary variables used in CM analysis and would be copied from the SDTM CM domain. Include the dictionary version in the variable metadata. Conditional on whether coded and used for analysis.
ATCy	ATC Level y Text	Char	WHO Drug	Cond	Corresponds to the ATC Level Text for WHO Drug Conditional, based on analysis at multiple levels (y)
ATCyCD	ATC Level y Code	Char	WHO Drug	Cond	Corresponds to the ATC Level Code for WHO Drug Conditional, based on analysis at multiple levels (y)

Table 18. WHO Drug Dictionary Coding Variables

2. Next, in the ADaM OCCDS it is specified that "For traceability when copying variables from SUPPQUAL, it is recommended to use variable names that exactly match the corresponding SUPPQUAL.QNAM values".

3. Thus, to be able to comply with the naming conventions from the item 1 and follow traceability rule from the item 2, we have to define supplemental qualifiers names and labels for ATC coding mapping accordingly:

QNAM	QLABEL	Applicable Domains
ATCy	ATC Level y Text	CM
ATCyCD	ATC Level y Code	CM

Table 19. QNAM/QLABEL naming convention for ATC Coding Variables in SUPPCM

This is a common situation when SUPP-- datasets are treated as a way to map any data out of standard variables list in freestyle, while we have to remember that there are still some restrictions and conventions required to follow for them.

CONCLUSION

To conclude the paper, we would like to list recommendations to follow to take the SDTM creation process to the next level:

- during the specification writing you need to have not only raw data and the CRF on your hand but also the study protocol and statistical analysis plan which describe details for comprehensive representation of the data;

- use all available documentation describing guidance on standardized data such as the SDTMIG, the SDTM, Therapeutic Area User Guides (TAUG) etc. Each of them may include that certain details and examples applicable to your specific study;

- work on your understanding of the big picture of clinical trials - the CDASH establishes a standard way to collect data; acquaintance with the CDASH helps to understand the step that precedes the SDTM creation, while good navigation through the ADaM and ADaMIG provides insight into the next step of collected data analysis and data representation in the outputs.

We hope that our experience will be found useful for mapping less common data into the SDTM. We have

to follow all available standards and constantly improve processes as new versions are developing continuously.

REFERENCES

1. Study Data Tabulation Model Implementation Guide: Human Clinical Trials. Clinical Data Interchange Standards Consortium (CDISC) Submission Data Standards (SDS) Team. Version 3.3. November 2018.

2. Study Data Tabulation Model. CDISC Submission Data Standards Team and CDISC SDTM Governance Committee. Version 1.7. November 2018.

3. Analysis Data Model Implementation Guide. CDISC Analysis Data Model Team. Version 1.1. February 2016.

4. Common Terminology Criteria for Adverse Events (CTCAE). U.S. Department of Health and Human Services. Version 5.0. November 27, 2017.

CONTACT INFORMATION

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

Nadiia Pukhliar
Intego Group, LLC
+38 098 39 29 644
nadiia.pukhliar@intego-group.com

Dariia Tsyhanenko
Intego Group, LLC
+38 050 08 92 214
dariia.tsyhanenko@intego-group.com