

SDTM Trial Summary Domain: Putting Together the TS Puzzle

Kristin Kelly, Accenture Accelerated R&D Services, Berwyn, PA
Jerry Salyers, Accenture Accelerated R&D Services, Berwyn, PA
Fred Wood, Accenture Accelerated R&D Services, Berwyn, PA

ABSTRACT

The SDTM Trial Summary (TS) domain was updated with new variables and implementation strategies with the publication of the SDTM v1.3/SDTMIG v3.1.3. This was an attempt to make the domain more useful to reviewers and to be machine readable to facilitate data warehousing. The FDA has recently considered TS as being essential to include in a study submission. They have developed tools that check that the TS domain is populated according to the guidance in the SDTMIG and the FDA's Study Data Technical Conformance Guide (TCG). Despite the advice for implementation, constructing an informative and complete TS domain can be challenging for sponsors in terms of interpreting the guidance and using the correct controlled terminologies. This paper will discuss some of the challenges associated with populating specific TS parameters, and provide some solutions for addressing these challenges.

INTRODUCTION

The SDTM Trial Design domains represent the plan for what will be done to subjects, and what data will be collected about them in the course of a clinical trial, to address the trial's objectives. The Trial Design domains consist of the following: Trial Arms (TA), Trial Disease Assessments (TD), Trial Elements (TE), Trial Inclusion/Exclusion (TI), Trial Summary (TS), and Trial Visits (TV).

According to the SDTM Implementation Guide (SDTMIG) v3.2, the standard Trial Design Datasets will allow reviewers to:

- clearly and quickly grasp the design of a clinical trial
- compare the designs of different trials
- search a data warehouse for clinical trials with certain features
- compare planned and actual treatments and visits for subjects in a clinical trial.

For these reasons, it is important that the trial-level domains reflect the study design as accurately as possible. Although each of the trial design domains provide a reviewer with important planned aspects of a clinical trial in dataset format, the focus of this paper will be on the Trial Summary (TS) domain.

The FDA has recently developed their own tools that run automated checks on the SDTM data when it is received as part of a submission, and before it is passed to a reviewer. This is done in order to identify data conformance issues earlier in an effort to save time downstream in the review process. The FDA has published business and conformance rules based on the SDTM standard and FDA data requirements. There is one domain, TS, that has become increasingly useful to FDA because it provides a high-level overview of the study without having to go too deep into the data or refer to the protocol. There are several rules published by FDA that pertain to the TS domain that check the adherence to their expectations of the content of this dataset. With the publication of SDTM v1.3/SDTMIG v3.1.3, several new variables were added to the TS domain that increase its usefulness in terms of machine readability and data warehousing. Because of these reasons, it is important that the TS domain be constructed properly. This paper will discuss the TS domain as a whole and also focus on some of the challenges encountered during development.

TS BASICS – THE PUZZLE OUTLINE

The Trial Summary (TS) domain is a trial-level dataset that allows for the submission of a high-level view of a study in a structured format. Each record contains the value of a parameter or characteristic of the trial. It is used to record basic information about the study such as trial phase, protocol title, and trial objectives. It contains both planned and actual aspects of the trial, such as the number of subjects, and study start/end dates.

The TS domain as a whole can be thought of as a jigsaw puzzle made up of interlocking pieces that need to be put together in just the right way to make a complete picture. Typically, when working on a jigsaw puzzle, the first thing one might do is pick out the pieces with the flat edges that would make up the borders, or outline, of the puzzle. In the case of TS, these outline pieces could be compared to the variables that provide the structure for the dataset. Beginning in SDTMIG v3.1.3, several new variables were added to TS in an effort to make the submission of a

minimum set of parameters consistent across sponsors. These variables include TSVALNF, TSVLCD, TSVCDREF, and TSVCDVER. When specific parameter values cannot be provided, a new variable is used to submit reasons why it is missing (TSVALNF). These new variables can also be populated with reference to terminology used to populate TSVCDREF and the version used (TSVCDREF and TSVCDVER). Because TS has become more standardized across studies, this has facilitated machine readability and storage.

Since TS is not a general-observation-class domain, there is a finite set of variables allowed, and no other variables can be added to the domain. The variables are listed in Table 1 below and will be referenced throughout the sections of this paper. Please refer to the SDTM v1.3/SDTMIG v3.1.3 and later versions for more information.

Variable Name	Variable Label	Type	Controlled Terms, Codelist or Format	Role	CDISC Notes	Core
STUDYID	Study Identifier	Char		Identifier	Unique identifier for a study.	Req
DOMAIN	Domain Abbreviation	Char	(DOMAIN)	Identifier	Two-character abbreviation for the domain.	Req
TSSEQ	Sequence Number	Num		Identifier	Sequence number given to ensure uniqueness within a dataset. Allows inclusion of multiple records for the same TSPARMCD, and can be used to join related records.	Req
TSGRPID	Group ID	Char		Identifier	Used to tie together a group of related records	Perm
TSPARMCD	Trial Summary Parameter Short Name	Char	TSPARMCD	Topic	TSPARMCD (the companion to TSPARM) is limited to 8 characters and does not have special character restrictions. These values should be short for ease of use in programming, but it is not expected that TSPARMCD will need to serve as variable names. Examples: AGEMIN, AGEMAX	Req
TSPARM	Trial Summary Parameter	Char	TSPARM	Synonym Qualifier	Term for the Trial Summary Parameter. The value in TSPARM cannot be longer than 40 characters. Examples Planned Minimum Age of Subjects, Planned Maximum Age of Subjects	Req
TSVAL	Parameter Value	Char		Result Qualifier	Value of TSPARM. Example: "ASTHMA" when TSPARM value is "Trial Indication". TSVAL can only be null when TSVALNF is populated. Text over 200 characters can be added to additional columns TSVAL1-TSVALn.	Exp
TSVALNF	Parameter Null Flavor	Char	NULLFLAVOR	Result Qualifier	Null flavor for the value of TSPARM, to be populated if and only if TSVAL is null.	Perm
TSVALCD	Parameter Value Code	Char		Result Qualifier	This is the code of the term in TSVAL. For example; 6CW7F3G59X is the code for Gabapentin, C49488 is the code for Y. The length of this variable can be longer than 8 to accommodate the length of the external terminology.	Exp

Variable Name	Variable Label	Type	Controlled Terms, Codelist or Format	Role	CDISC Notes	Core
TSVCDREF	Name of the Reference Terminology	Char		Result Qualifier	The name of the Reference Terminology from which TSVALCD is taken. For example; CDISC, SNOMED, ISO 8601.	Exp
TSVCDVER	Version of the Reference Terminology	Char		Result Qualifier	The version number of the Reference Terminology, if applicable.	Exp

Table 1: Trial Summary Variable Structure from the SDTMIG¹

TRIAL SUMMARY PARAMETERS (TSPARMCD/TSPARM) – THE INTERIOR PUZZLE PIECES

Within the variable structure of the TS puzzle, are the ‘interior pieces’ or parameters (i.e., the data) that must be fit together to provide a complete picture of a study. In order to make TS useful, a minimum number of trial summary parameters should be provided as shown below in Table 2. Most of the parameters come from www.clinicaltrials.gov, and the controlled terminology shown below is aligned with that source.¹ Definitions of the parameters and controlled terminology values for TSPARMCD, TSPARM, and TSVAL (for CDISC defined codelists) are maintained by the National Cancer Institute's Enterprise Vocabulary Services (NCI EVS)⁶. The TSPARMCD codelist is extensible, meaning more values can be added as needed to describe the trial.

The column titled ‘Inclusion in TS’ in Table 2 indicates whether the parameter should be included in the dataset. If a parameter is listed as ‘Required’, the record should be included in the TS domain and either TSVAL or TSVALNF must be populated. In cases where a parameter is ‘Conditionally Required’ or ‘If Applicable’, the presence and population depend on the study design. For example, if subjects are planned to take a protocol-specified treatment in combination with the investigational drug (ADDON (Added on to Existing Treatments) = ‘Y’), then a record for TSPARMCD = ‘CURTRT’ should be included. If the condition does not apply for a particular study and TSVAL cannot be populated, the corresponding TSPARMCD record need not be included. Please refer to the SDTMIG for specific conditions per parameter for those that are ‘Conditionally Required’. ‘Extensible Addition’ parameters are present in the NCI TSPARMCD codelist and can be added as needed, depending on the protocol and any other information the sponsor wishes to add to the TS domain.

TSPARMCD	TSPARM	TSVCDREF	Inclusion in TS	TSVAL (Codelist or Format)	External Link (If Applicable)
ACTSUB	Actual Number of Subjects		Extensible Addition	Number	
ADAPT	Adaptive Design	CDISC	Required	No Yes Response (C66742)	http://www.cancer.gov/cdisc
ADDON	Added on to Existing Treatments	CDISC	Required	No Yes Response (C66742)	http://www.cancer.gov/cdisc
AGEMAX	Planned Maximum Age of Subjects	ISO 8601	Required	ISO 8601	
AGEMIN	Planned Minimum Age of Subjects	ISO 8601	Required	ISO 8601	
COMPTRT	Comparative Treatment Name	UNII	If Applicable	SRS Preferred Substance Name (or Device Name)	http://fdasis.nlm.nih.gov/srs/srs.jsp
CRMDUR	Confirmed Response Minimum Duration	ISO 8601	If Applicable	ISO 8601	
CURTRT	Current Therapy or Treatment	UNII	Conditionally Required	SRS Preferred Substance Name (or Device Name)	http://fdasis.nlm.nih.gov/srs/srs.jsp
DCUTDESC	Data Cutoff Description		Required	Text	

SDTM Trial Summary Domain: Putting Together the TS Puzzle, continued

TSPARMCD	TSPARM	TSVCDREF	Inclusion in TS	TSVAL (Codelist or Format)	External Link (If Applicable)
DCUTDTC	Data Cutoff Date	ISO 8601	Required	ISO 8601	
DOSE	Dose per Administration		Extensible Addition	Text	
DOSFRQ	Dosing Frequency	CDISC	Extensible Addition	Frequency (C71113)	http://www.cancer.gov/cdisc
DOSU	Dose Units	CDISC	Extensible Addition	Unit (C71620)	http://www.cancer.gov/cdisc
FCNTRY	Planned Country of Investigational Sites	CDISC	Required	ISO 3166-1 alpha-3 Country Code (C66786)	http://www.cancer.gov/cdisc
HLTSUBJI	Healthy Subject Indicator	CDISC	Required	No Yes Response (C66742)	http://www.cancer.gov/cdisc
INDIC	Trial Indication	SNOMED	If Applicable	SNOMED CT	http://ncit.nci.nih.gov/ncitbrowser
INTMODEL	Intervention Model	CDISC	Conditionally Required	Intervention Model (C99076)	http://www.cancer.gov/cdisc
INTTYPE	Intervention Type	CDISC	Conditionally Required	Intervention Type (C99078)	http://www.cancer.gov/cdisc
LENGTH	Trial Length		Required	ISO 8601	
NARMS	Planned Number of Arms		Required	Number	
OBJPRIM	Trial Primary Objective		Required	Text	
OBJSEC	Trial Secondary Objective		If Applicable	Text	
OUTMSEXP	Exploratory Outcome Measure		If Applicable	Text	
OUTMSPRI	Primary Outcome Measure		Required	Text	
OUTMSSEC	Secondary Outcome Measure		If Applicable	Text	
PCLAS	Pharmacological Class of Invest. Therapy	NDF-RT	Conditionally Required	NDF-RT	http://ncit.nci.nih.gov/ncitbrowser
PLANSUB	Planned Number of Subjects		Required	Number	
RANDOM	Trial is Randomized	CDISC	Required	No Yes Response (C66742)	http://www.cancer.gov/cdisc
RANDQT	Randomization Quotient		Conditionally Required	Number	
REGID	Registry Identifier	ClinicalTrials.gov	Required	ClinicalTrials.gov or EUDRAC	http://clinicaltrials.gov/ https://www.clinicaltrialsregister.eu/
ROUTE	Route of Administration	CDISC	Extensible Addition	Route of Administration (C66729)	http://www.cancer.gov/cdisc
SDMDUR	Stable Disease Minimum Duration	ISO 8601	If Applicable	ISO 8601	

TSPARMCD	TSPARM	TSVCDREF	Inclusion in TS	TSVAL (Codelist or Format)	External Link (If Applicable)
SENDTC	Study End Date	ISO 8601	Required	ISO 8601	
SEXPOP	Sex of Participants	CDISC	Required	Sex of Participants (C66732)	http://www.cancer.gov/cdisc
SPONSOR	Clinical Study Sponsor	DUNS	Required	Data Universal Numbering System or D-U-N-S (DUNS)	http://mycredit.dnb.com/search-for-duns-number
SSTDTC	Study Start Date	ISO 8601	Required	ISO 8601	
STOPRULE	Study Stop Rules		Required	Text	
STYPE	Study Type	CDISC	Required	Study Type (C99077)	http://www.cancer.gov/cdisc
TBLIND	Trial Blinding Schema	CDISC	Required	Trial Blinding Schema (C66735)	http://www.cancer.gov/cdisc
TCNTRL	Control Type	CDISC	Required	Control Type (C66785)	http://www.cancer.gov/cdisc
TDIGRP	Diagnosis Group	SNOMED	Conditionally Required	SNOMED CT	http://ncit.nci.nih.gov/ncitbrowser
TINDTP	Trial Indication Type	CDISC	Conditionally Required	Trial Indication Type (C66736)	http://www.cancer.gov/cdisc
TITLE	Trial Title		Required	Text	
TPHASE	Trial Phase Classification	CDISC	Required	Trial Phase Classification (C66737)	http://www.cancer.gov/cdisc
TRT	Investigational Therapy or Treatment	UNII	Conditionally Required	SRS Preferred Substance Name (or Device Name)	http://fdasis.nlm.nih.gov/srs/srs_s.jsp
TTYPE	Trial Type	CDISC	Required	Trial Type (C66739)	http://www.cancer.gov/cdisc
STRATFCT	Stratification Factor		Extensible Addition	Any allowable variable name (e.g., AGE, SEX)	

Table 2: Required and Commonly Used TS Parameters

PARAMETER NULL FLAVOR (TSVALNF)

For 'Required' TS parameters where TSVAL cannot be populated, TSVALNF should be used. The controlled terminology associated with this variable is part of the ISO 21090 standard, 'Health Informatics – Harmonized data types for information exchange' which provides the idea of 'null flavor'. A null flavor is data that provides additional information when its primary piece of data is null (has a missing value), in this case, TSVAL. There are 14 terms in the NULLFLAVOR codelist in the ISO 21090 standard. Please refer to the SDTMIG for the complete list. The following are some codes commonly used to populate TSVALNF in SDTM TS:

- PINF (Positive infinity) – Positive infinity of numbers. Typically used when there is no maximum age stipulated in the protocol.
- UNK (Unknown) – A proper value is applicable, but not known.
- NAV (Not available) – Information is not available at this time, but it is expected that it will be available later. Typically used for study end date if the study is ongoing or for pharmacological class if the study drug requires a new class to be added to the applicable dictionary.
- NA (Not applicable) – No proper value is applicable in this context.

PARAMETER VALUE (TSVAL) – CONTROLLED TERMINOLOGY

For TSVAL, there are several dictionaries or codelists associated with specific TS parameters that are noted for each parameter in the 'TSVCDREF' column above in Table 2. Parameters for which there is no codelist mentioned are considered free text, e.g. TSPARMCD = 'TITLE'. Instances where TSVCDREF is 'CDISC' points to NCI EVS Controlled Terminology⁶. The 'TSVAL (Codelist or Format)' column in Table 2 then provides the codelist name and corresponding Concept Code, or C-Code, for that NCI codelist. The following are the terminology and formats used to populate TSVAL:

- CDISC Controlled Terminology – maintained by NCI EVS. Values should come from the ‘CDISC Submission Value’ column in the NCI CT SDTM spreadsheet⁶ (e.g., TSPARMCD = ‘ADDON’, ‘INTMODEL’, ‘TPHASE’)
- UNII (FDA Unique Ingredient Identifier) - used to standardize active ingredients in investigational drugs and comparators (e.g., TSPARMCD = ‘TRT’, ‘CURTRT’, ‘COMPTRT’)
- SNOMED CT (The International Health Terminology Standards Organization’s (IHTSDO) Systematized Nomenclature of Medicine – Clinical Terms) - used to identify the medical condition or problem that the investigational product is intended to affect (e.g. TSPARMCD = ‘INDIC’, ‘TDIGRP’)
- NDF-RT (The Veterans Administration’s National Drug File – Reference Terminology) - used to identify the pharmacologic class of all active investigational drugs (e.g., TSPARMCD = ‘PCLAS’)
- Clinicaltrials.gov/Euradac – used to identify the study from the clinical trial registry (e.g., TSPARMCD = ‘REGID’)
- DUNS (Dun and Bradstreet Data Universal Numbering System) – used to identify a sponsor organization based on the address provided in the protocol (e.g., TSPARMCD = ‘SPONSOR’)
- ISO 8601 Standard – dictates the format for dates and durations (e.g., TSPARMCD = ‘AGEMIN’, ‘LENGTH’, ‘SSTDTC’)
- ISO 3166-1 alpha-3 Standard – provides a 3-letter country code for each country included in the study (e.g., TSPARMCD = ‘FCNTRY’)

Typically, the submission value (e.g., CDISC CT) or the preferred term (e.g., SNOMED) would be placed in TSVAL for the applicable parameter. The corresponding code for that value is placed in TSVALCD and the name of the referenced terminology is populated in TSVCDREF. If there is a version associated with the terminology, as is the case with CDISC CT, this would be noted in TSVCDVER. For records where TSVAL is left null, the applicable code from ISO 21090 (described above) is placed in TSVALNF and TSVCDREF is populated as ‘ISO 21090’.

An example of a TS dataset is provided below in Table 3. Please note that not all Required variables and parameters are included, but only those necessary for demonstration purposes. This table may be referenced for certain parameters throughout other sections of this paper.

TSSEQ	TSGRPID	TSPARMCD	TSPARM	TSVAL	TSVALNF	TSVALCD	TSVCDREF	TSVCDVER
1		ADDON	Added on to Existing Treatments	Y		C49488	CDISC	2014-03-28
1		AGEMAX	Planned Maximum Age of Subjects		PINF			
1		AGEMIN	Planned Minimum Age of Subjects	P18Y			ISO 8601	
1	GEM	CURTRT	Current Therapy or Treatment	GEMCITABINE		B76N6SBZ8R	UNII	
1	CUTOFF	DCUTDESC	Data Cutoff Description	INTERIM ANALYSIS				
1	CUTOFF	DCUTDTC	Data Cutoff Date	2012-11-20			ISO 8601	
1	GEM	DOSE	Dose per Administration	1000				
2	DRUG A	DOSE	Dose per Administration	240				
3	DRUG A	DOSE	Dose per Administration	340				
1		DOSU	Dose Units	mg/m2		C67402	CDISC	2014-03-28
1		HLTSUBJI	Healthy Subject Indicator	N		C49487	CDISC	2014-03-28
1		INDIC	Trial Indication	Adenocarcinoma of pancreas		700423003	SNOMED	

TSSEQ	TSGRPID	TSPARMCD	TSPARM	TSVL	TSVALNF	TSVALCD	TSVCDRE F	TSVCDVE R
1		INTMODEL	Intervention Model	CROSSOVER		C82637	CDISC	2014-03-28
1		INTTYPE	Intervention Type	DRUG		C1909	CDISC	2014-03-28
1		LENGTH NARMS	Trial Length Planned Number of Arms	P30M			ISO 8601	
1			5					
1		PCLAS	Pharmacological Class of Invest. Therapy	Alkylating Drug		N0000175558	NDF-RT	
1		RANDOM	Trial is Randomized	Y		C49488	CDISC	2014-03-28
1		SENDTC	Study End Date		NAV		ISO 21090	
1		SEXPOP	Sex of Participants	BOTH		C49636	CDISC	2014-03-28
1		SPONSOR	Clinical Study Sponsor	ABC Pharma		999999999	DUNS	
1		SSTDTC	Study Start Date	2010-01-01			ISO 8601	
1		STYPE	Study Type	INTERVENTIONAL		C98388	CDISC	2014-03-28
1		TBLIND	Trial Blinding Schema	OPEN LABEL		C49659	CDISC	2014-03-28
1		TCNTRL	Control Type	ACTIVE		C49649	CDISC	2014-03-28
1		TDIGRP	Diagnosis Group	Adenocarcinoma of pancreas		700423003	SNOMED	
1		TINDTP	Trial Indication Type	TREATMENT		C49656	CDISC	2014-03-28
1		TPHASE	Trial Phase Classification	Phase II Trial		C15601	CDISC	2014-03-28
1	DRUG A	TRT	Investigational Therapy or Treatment	DRUG A		9B87ACB9B	UNII	
1		TTYPE	Trial Type	EFFICACY		C98791	CDISC	2014-03-28
2		TTYPE	Trial Type	TOLERABILITY		C49667	CDISC	2014-03-28
3		TTYPE	Trial Type	PHARMACOKINETIC		C49666	CDISC	2014-03-28
4		TTYPE	Trial Type	SAFETY		C49663	CDISC	2014-03-28

Table 3: Example TS Dataset

A NOTE ON SEQUENCE NUMBER AND GROUP ID (TSSEQ AND TSGRPID)

In the example dataset above in Table 3, it's important to note the assignment of TSSEQ and use of TSGRPID. Per the SDTMIG for TS, multiple records of the same parameter should have a unique value of TSSEQ assigned. If there is only one record for a particular parameter (or TSPARMCD/TSPARM) a TSSEQ = 1 is assigned. In the dataset example, there is only one record each for TSPARMCD = 'TPHASE' and 'TRT', thus TSSEQ = 1 for each parameter. There are four records created for TSPARMCD = 'TTYPE' and each is numbered sequentially as TSSEQ = 1, 2, 3, and 4. This is different from --SEQ numbering in the subject-level datasets where --SEQ is assigned based on uniqueness for a subject within a domain, and not at the parameter level.

In the case of TSGRPID, records that need to be tied together across and within parameters are assigned the same value. This is shown where TSGRPID = 'DRUG A' and is populated for records that link TSPARMCD = 'TRT' and two

records for TSPARMCD = 'DOSE' to indicate that for DRUG A, those were the doses administered in the study. The same is done for the other drug administered where TSGRPID = 'GEM' ties together TSPARMCD = 'CURTRT' and the applicable dose for that drug (TSPARMCD = 'DOSE'). Records that apply to all parameters will not have TSGRPID populated as is the case in Table 3 for TSPARMCD = 'DOSU', since TSVAl = 'mg/m2' applies to both treatments administered in the study.

FITTING TOGETHER SOME OF THE TS PUZZLE PIECES

In this section, challenges and solutions typically associated with some specific TS parameters (TSPARMCD/TSPARM) will be discussed. For more information on how to populate other parameters, please refer to the SDTMIG v3.1.3 and later.

STUDY START DATE (TSPARMCD = SSTDTC)

In the FDA Guidance 'Submissions in Electronic Format-Standardized Study Data'⁴, the study start date for a clinical study is described according to the definition provided in the NCI Controlled Terminology (NCI CT)⁶. For the TSPARMCD codelist value, 'SSTDTC' (Study Start Date), the CDISC Definition states that study start date should be 'the earliest date of informed consent among any subject (Date/Time of Informed Consent, RFICDTC) that enrolled in the study. For studies conducted without informed consent (i.e., emergency use) use the date of treatment. Dates for subjects who were screen failures are not included'.⁶ Before there was reference to this parameter in FDA Guidance, the CDISC Definition was typically not referenced by sponsors when developing TS. Instead, the date of 'First Patient In' (FPI) from clinicaltrials.gov has been erroneously used to populate TSVAl where TSPARMCD = 'SSTDTC'. This date does not always correspond to the earliest date for a non-screen failed subject's RFICDTC in DM, and, it is not present in the study data.

Another challenge to populating this TSPARMCD with the correct date is due to only referencing the first part of the definition of 'SSTDTC' from NCI, which would simply be the earliest date of informed consent of any subject in the study, including screen failures. Since the FDA requirement for providing standardized study data is dependent on the study start date, it is important that the correct date according to the full definition is used to populate TSVAl for 'SSTDTC' in TS. Not adhering to the full definition could result in having the study seem to start earlier than it really had per the CDISC definition.

STUDY END DATE (TSPARMCD = SENDTC) VS. DATA CUTOFF DATE (TSPARMCD = DCUTDTC)

Per the NCI CT for the Required TSPARMCD = 'SENDTC' (Study End Date), the date populated in TSVAl for this parameter should be the final date collected for any subject (Date/Time of End of Participation, RFPENDTC), in other words, the maximum SVENDTC for any subject in the Subject Visits (SV) domain. This definition is clear for studies that are completed, but what about when the study is still ongoing? When a study is still ongoing at the time of submission, a date is selected by the sponsor as to when the data will be 'cutoff' for analysis. This date is typically populated in TSVAl for the required TSPARMCD = 'DCUTDTC' (Data Cutoff Date), with the description of what that date is in the TSVAl for TSPARMCD = 'DCUTDESC' (Data Cutoff Description). An example value for an ongoing study might be 'INTERIM ANALYSIS', although there is no controlled terminology associated with this particular parameter. It can be free text that simply describes 'DCUTDTC'. TSGRPID can then be populated to group these two records together.

For TSPARMCD = 'SENDTC', many sponsors will populate TSVAl for an ongoing study with the last RFPENDTC for any subject at the time the data is cut off for submission/analysis. This date may also match the date populated for 'DCUTDTC'. Populating a date in TSVAl for 'Study End Date' in this instance implies that the study is completed. But according to the SDTMIG, even though a TSPARMCD/TSPARM is 'Required', TSVAl may not be able to be populated with a value in some instances, and TSVAlNF should be used to denote the reason why TSVAl is null. For an ongoing study, a recommended practice for handling this in TSPARMCD = 'SENDTC' would be to set TSVAl to null and populate TSVAlNF = 'NAV' ('Not Available') and TSVCDREF = 'ISO 21090'. The TSVAl for TSPARMCD = DCUTDTC would be populated with the data cutoff date with the appropriate data cutoff description in 'DCUTDESC'.

TSSEQ	TSGRPID	TSPARMCD	TSPARM	TSVAL	TSVALNF	TSVALCD	TSVCDREF	TSVCDVER
1	CUTOFF	DCUTDESC	Data Cutoff Description	INTERIM ANALYSIS				
1	CUTOFF	DCUTDTC	Data Cutoff Date	2012-11-20			ISO 8601	
1		SENDTC	Study End Date		NAV		ISO 21090	

Table 4: SENDTC vs DCUTDTC for an ongoing study

Table 4 shows a subset of the TS example in Table 3 of TSPARMCD records for 'SENDTC', 'DCUTDESC', and 'DCUTDTC' for an ongoing study as described previously. After the study is completed, these three TS parameters

should be updated accordingly with the appropriate study end date, data cutoff date and description.

STUDY STOP RULE (TSPARMCD = STOPRULE)

The TSPARMCD = 'STOPRULE' (Study Stop Rule) is a 'Required' parameter in TS for which it is unclear how to populate TSVAl. In NCI CT, 'STOPRULE' is defined as the rule, regulation and/or condition that determines the point in time when a clinical trial will be terminated.⁶ But in the SDTMIG, it is defined simply as the protocol-specified stopping rule. These definitions do not unambiguously state whether this rule applies on the trial level or the subject level. Some sponsors populate TSVAl based on the study discontinuation/withdrawal criteria for a given subject provided in the protocol such as disease progression, lost to follow-up, etc. rather than on the trial stop criteria. Though the NCI definition is much clearer that it is at the study level, it's possible that this definition is not referenced and the SDTMIG is used instead when TS is being constructed. TSVAl for 'STOPRULE' should be populated per the criteria for study termination on the trial level. Examples may include number of events, lack of efficacy, and/or safety across subjects. There is no controlled terminology associated with this TSPARMCD/TSPARM for TSVAl so it should be completed per protocol instructions. If the protocol does not have any study termination rules, then TSVAl can be populated with 'NONE'. In general, it may be helpful to refer to both the SDTMIG as well as the NCI CT definitions when developing TS.

HEALTHY SUBJECT INDICATOR (TSPARMCD = HLTSUBJI)

The SDTMIG is descriptive on how to populate TSPARMCD = 'HLTSUBJI' (Healthy Subject Indicator), 'TINDTP' (Trial Indication Type), 'INDIC' (Trial Indication), and 'TDIGRP' (Diagnosis Group) when the study only contains healthy subjects, as is common for Phase I studies. For example, if the TSVAl for TSPARMCD = 'STYPE' (Study Type) is 'INTERVENTIONAL', then a record where TSPARMCD = 'TINDTP' is required to be created. All studies where an investigational substance is administered are considered to be interventional studies, as opposed to an observational study where no drug is given. For a PK study with healthy subjects, TSVAl = 'INTERVENTIONAL' for 'STYPE'. but per the advice in the SDTMIG, there is no trial indication for this type of study with healthy volunteers. The 'Trial Indication Type' codelist includes values such as 'TREATMENT', 'CURE', and 'DIAGNOSIS'. TSVAl for 'HLTSUBJI' would be set 'Y', TSVAl would be null and TSVAlNF for 'TINDTP' would be set to 'NA'. Records where TSPARMCD = 'TDIGRP' and 'INDIC' need not be included, as shown below in Table 5. Given this guidance in the SDTMIG, this would mean that even though the 'TDIGRP' codelist contains a value of 'HEALTHY SUBJECTS', a record indicating this for 'TDIGRP' when 'HLTSUBJI' = 'Y' should not be included.

TSSEQ	TSPARMCD	TSPARM	TSVAL	TSVALNF	TSVALCD	TSVCDREF	TSVCDVER
1	HLTSUBJI	Healthy Subject Indicator	Y		C49488	CDISC	2014-03-28
1	STYPE	Study Type	INTERVENTIONAL		C98388	CDISC	2014-03-28
1	TINDTP	Trial Indication Type		NA		ISO 21090	

Table 5: TSPARMCD = 'HLTSUBJI' when TSVAl = 'Y' for a PK study (Note: Only variables necessary for the example are included due to space restrictions.)

Another example for when TSVAl for 'HLTSUBJI' is 'Y' might be a flu vaccine study that enrolls only healthy subjects. Per the SDTMIG, if the intent of a study is to treat, prevent, or diagnose a condition, then a record for TSPARMCD = 'INDIC' should be included. In this scenario, TSVAl for 'HLTSUBJI' is set to 'Y', TSVAl for 'STYPE' is set to 'INTERVENTIONAL', TSVAl = 'PREVENTION' for 'TINDTP' and TSVAl for 'INDIC' would be 'Influenza' as shown in Table 6. A record for TSPARMCD = 'TDIGRP' does not need to be created when 'HLTSUBJI' is 'Y'

TSSEQ	TSPARMCD	TSPARM	TSVAL	TSVALNF	TSVALCD	TSVCDREF	TSVCDVER
1	HLTSUBJI	Healthy Subject Indicator	Y		C49488	CDISC	2014-03-28
1	INDIC	Trial Indication	Influenza		6142004	SNOMED	
1	STYPE	Study Type	INTERVENTIONAL		C98388	CDISC	2014-03-28
1	TINDTP	Trial Indication Type	PREVENTION		C49657	CDISC	2014-03-28

Table 6: TSPARMCD = 'HLTSUBJI' when TSVAl = 'Y' for a Vaccine study (Note: Only variables necessary for the example are included due to space restrictions.)

There may be some studies that enroll both healthy volunteers and those that are part of the diagnosis group for the drug under study. This type of situation is not specifically handled. In the SDTMIG, it states that when TSVAl for 'HLTSUBJI' is 'N', then 'TDIGRP' is a required parameter and TSVAl cannot be null. And as described above, 'TDIGRP' should not be included when 'HLTSUBJI' is 'Y'. So in this case, a recommended practice would be that TSVAl for 'HLTSUBJI' would be set to 'N' and then to create two records for TSPARMCD = 'TDIGRP'. One record

where TSVAl = 'HEALTHY SUBJECTS' and the other where TSVAl was populated with the condition being studied. Following with the example of a flu vaccine study, this would be 'Influenza'. Since there is more than one record for this parameter, TSSEQ for each would be set as 1 and 2; respectively, as shown in Table 7. Because TSVAl for 'STYPE' is 'INTERVENTIONAL' for this study as well, TSVAl is populated for 'TINDTP' and 'INDIC'.

TSSEQ	TSPARMCD	TSPARM	TSVAL	TSVALNF	TSVALCD	TSVCDREF	TSVCDVER
1	HLTSUBJI	Healthy Subject Indicator	N		C49487	CDISC	2014-03-28
1	INDIC	Trial Indication	Influenza		6142004	SNOMED	
1	STYPE	Study Type	INTERVENTIONAL		C98388	CDISC	2014-03-28
1	TDIGRP	Diagnosis Group	HEALTHY SUBJECTS		C49651	CDISC	2014-03-28
2	TDIGRP	Diagnosis Group	Influenza		6142004	SNOMED	
1	TINDTP	Trial Indication Type	PREVENTION		C49657	CDISC	2014-03-28

Table 7: Example Study with both healthy and ill subjects (Note: Only variables necessary for the example are included due to space restrictions.)

INVESTIGATIONAL TREATMENT VS. COMPARATOR OR CURRENT TREATMENT

Another set of TS parameters that can get complicated to piece together based on study design consists of the TSPARMCD values of 'TRT' (Investigational Therapy or Treatment), 'CURTRT' (Current Therapy or Treatment), 'COMPTRT' (Comparative Treatment Name) and 'TCNTRL' (Control Type). Referring back to Table 2 that lists the TSPARMCD/TSPARM values, 'TRT' is 'Conditionally Required', and per the SDTMIG, it is required when TSVAl for 'STYPE' (Study Type) is 'INTERVENTIONAL' meaning that a substance or therapy will be administered in the study. TSVAl should be populated with the SRS Preferred Substance Name from the FDA's Unique Ingredient Identifier (UNII) register. The UNII code will be populated in TSVAlCD with 'UNII' in TSVCDREF. Treatments that consist of more than one ingredient (or active moiety) should be split out into separate records. The same convention should be used to populate records for 'CURTRT' and 'COMPTRT' as well.

TSPARMCD = 'CURTRT' is 'Conditionally Required' when TSVAl for TSPARMCD = 'ADDON' ('Added on to Existing Treatments') is 'Y'. A current treatment or therapy is considered to be a drug that is taken in addition to the investigational drug (noted in 'TRT') but is not being compared to it and it is still specified to be taken by subjects per the protocol. An example might be an oncology study where the study drug (SD) and the 'standard of care' (SOC) for that type of cancer are taken together (SD + SOC). The example TS dataset in Table 3 is based on this type of treatment strategy and a snippet from that dataset of these parameters is shown in Table 8.

TSSEQ	TSGRPID	TSPARMCD	TSPARM	TSVAL	TSVALNF	TSVALCD	TSVCDREF	TSVCDVER
1		ADDON	Added on to Existing Treatments	Y		C49488	CDISC	2014-03-28
1	GEM	CURTRT	Current Therapy or Treatment	GEMCITABINE		B76N6SBZ8R	UNII	
1		TCNTRL	Control Type	ACTIVE		C49649	CDISC	2014-03-28
1	DRUG A	TRT	Investigational Therapy or Treatment	DRUG A		9B87ACB9B	UNII	

Table 8: TS Example of 'ADDON', 'CURTRT', 'TCNTRL', and 'TRT'

TSPARMCD = 'TCNTRL' is a 'Required' parameter in TS (Table 8) that indicates the control type in the study (e.g., placebo, active control, or none). TSPARMCD = 'COMPTRT' is included in TS as a record if it is applicable for that particular study and TSVAl should only contain active treatment. This would mean that for studies where the study drug is being compared to placebo, then a record for 'COMPTRT' would not be included in TS. Instead, TSVAl for 'TCNTRL' would be set to 'PLACEBO' from the 'Control Type' CDISC codelist. An active drug that is acting as a control and not the drug under study would have a TSVAl of 'ACTIVE' and would be noted in either 'CURTRT' and/or 'COMPTRT' depending on study design, as described above.

Treatment strategies commonly seen in oncology studies can be difficult to determine how to model in 'TRT', 'CURTRT', and 'COMPTRT' in TS. The following are some example scenarios that are frequently planned treatment paths in studies that do not have placebo as a control type. Some of the recommended practices for handling these three parameters when developing TS can be applied to studies for other indications with similar study design.

1. Example Study 1:
 - Treatment Arms
 - i. SD (Study Drug) + SOC (Standard of Care)

- ii. SD (Study Drug)
 - iii. SOC (Standard of Care)
 - TSVAl for 'TRT' is SD, TSVAl for 'COMPTRT' and 'CURTRT' will be the same, as SOC
- 2. Example Study 2:
 - Treatment Arms
 - i. SD + SOC
 - ii. SOC
 - TSVAl for 'TRT' is SD, TSVAl for 'CURTRT' is SOC, no record created for 'COMPTRT'
- 3. Example Study 3
 - Treatment Arms
 - i. SD + SOC
 - ii. TD (Test Drug, Active Comparator) + SOC
 - TSVAl for 'TRT' is SD, TSVAl for 'CURTRT' is SOC, TSVAl for 'COMPTRT' is TD
- 4. Example Study 4
 - Treatment Arms
 - i. SD + SOC1
 - ii. SD + SOC2
 - TSVAl for 'TRT' is SD, two records for 'CURTRT' where TSVAl is each SOC drug
- 5. Example Study 5
 - Treatment Arms
 - i. SD1 + SOC
 - ii. SD2 + SOC
 - iii. SOC
 - Two records for 'TRT' where TSVAl is each SD, TSVAl for both 'CURTRT' and 'COMPTRT' is SOC

The examples mentioned above represent only some of the complicated scenarios that may be planned for a study. It is recommended that as new study designs are encountered, that the conventions used for constructing TS for that study type be documented so that others can have something to refer back to, and that will ensure consistency within a company across similar studies.

PHARMACOLOGICAL CLASS (TSPARMCD = PCLAS)

Per the SDTMIG, TSPARMCD = 'PCLAS' (Pharmacological Class of Invest. Therapy) is a 'Conditionally Required' parameter when TSVAl for 'STYPE' is 'INTERVENTIONAL' and TSVAl for 'INTTYPE' (Intervention Type) is of a value for which pharmacological class is applicable¹, e.g. 'DRUG' or 'BIOLOGIC'. TSVAl values should be assigned from the Veterans Administration's National Drug File – Reference Terminology (NDF-RT) and the corresponding code should be populated in TSVAlCD with TSVCDREF set to 'NDF-RT'.

Pharmacological class is a complex concept that consists of three components of a substance: mechanism of action (MOA), physiologic effect (PE), and chemical structure (CS).³ Per the FDA's Study Data Technical Conformance Guide, TSVAl for 'PCLAS' should be populated with the Established Pharmacological Class (EPC) term from the NDF-RT based on the unique ingredient (active moiety) of the study treatment.

"The established pharmacological class is generally the MOA, PE, or CS term that is considered the most scientifically valid and clinically meaningful. Sponsors should include in TS the established pharmacologic class of all active moieties of investigational products used in a study. FDA maintains a list of established pharmacologic classes of approved moieties. If the established pharmacologic class is not available for an active moiety, then the sponsor should discuss the appropriate MOA, PE, and CS terms with the review division. For unapproved investigational active moieties where the pharmacologic class is unknown, the PCLAS record may not be available."³

In the case where the pharmacological class is not yet known, TSVAl should be left null for 'PCLAS', TSVAlNF should be 'NAV' (Not Available), and TSVCDREF should be 'ISO 21090' (Table 9).

TSSEQ	TSGRPID	TSPARMCD	TSPARM	TSVAL	TSVALNF	TSVALCD	TSVCDREF	TSVCDVER
1		PCLAS	Pharmacological Class of Invest. Therapy		NAV		ISO 21090	
1	DRUG A	TRT	Investigational Therapy or Treatment	DRUG A		9B87ACB9B	UNII	

Table 9: Example of 'Not Available' Pharmacological Class

Once the class has been added to the NDF-RT, the TS domain should be updated to include the EPC and the

corresponding NDF-RT concept code (Table 10).

TSSEQ	TSGRPID	TSPARMCD	TSPARM	TSVAL	TSVALNF	TSVALCD	TSVCDREF	TSVCDVER
1		PCLAS	Pharmacological Class of Invest. Therapy	Alkylating Drug		N0000175558	NDF-RT	
1	DRUG A	TRT	Investigational Therapy or Treatment	DRUG A		9B87ACB9B	UNII	

Table 10: Example of completed record for 'PCLAS'

When referencing the NDF-RT, only the NDF-RT concept terms and codes for the EPC, MOA, CS, and PE are listed. When the FDA creates new established pharmacological classes, the active moiety of the drug is used to create the 'FDA EPC Text Phrase' that is to be used in the 'Highlights, Indications, and Usage' heading that is part of an NDA's Structured Product Labeling (SPL) where the following sentence is required to be included:

"(Drug) is a (name of class) indicated for (indication(s))." ²

The FDA EPC text phrase should be populated in the '(name of class)' field. Once the FDA EPC text phrase is created, it is then added to the NDF-RT as the 'Established Pharmacological Class'. 'PCLAS' in TS can be difficult to populate if sponsors only reference the NDF-RT, since there is a link that is needed between the active moiety and the EPC term/code to assign the correct values in TSVAL and TSVALCD that is not included in the NDF-RT. However, the FDA does maintain a list of active moieties and their corresponding FDA text phrase that matches the text of the EPC term from the NDF-RT called 'FDA EPC Text Phrases for Highlights Indications and Usage heading'.⁵ This provides the connection necessary between the active moiety of a drug and its related NDF-RT EPC concept. This document is updated periodically by the FDA and can be found in the following location on the FDA website: <http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/LawsActsandRules/ucm084159.htm>

One can search this document by the active moiety of the study drug to get the FDA EPC text phrase and then this phrase can then be used to search the NDF-RT list for the correct casing/format of the EPC term and its corresponding code to populate TSVAL and TSVALCD; respectively, for TSPARMCD = 'PCLAS'.

CONCLUSIONS

The Trial Summary domain has increased in usability with the addition of new variables in SDTM v1.3/SDTMIG v3.1.3 that help to better describe the key characteristics of a clinical trial. Providing a complete picture of a study in TS involves fitting together the pieces or 'data' properly, keeping in mind that it is only one small aspect necessary for a successful submission. Requirements for the TS domain now include several inputs such as the SDTMIG, the FDA Study Data Conformance Guide, NCI Controlled Terminology, SNOMED CT, and others. The creation of useful and consistent TS datasets across studies requires proper documentation by sponsors of implementation strategies based on the study design, as well as recording how to utilize the input resources that need to be referenced. The more robust the TS puzzle is, the more valuable the dataset becomes, which aids in accelerating FDA review of submissions.

REFERENCES

1. Study Data Tabulation Model Implementation Guide: Human Clinical Trials. Clinical Data Interchange Standards Consortium (CDISC) Submission Data Standards (SDS) Team. Version 3.2. November 2013.
2. Guidance for Industry and Review Staff: Labeling for Human Prescription Drug and Biological Products — Determining Established Pharmacologic Class for Use in the Highlights of Prescribing Information. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER). October 2009.
3. Study Data Technical Conformance Guide. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER). Version 2.3. October 2015.
4. Providing Regulatory Submissions In Electronic Format — Standardized Study Data. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER). December 2014.
5. FDA EPC Text Phrases for Highlights Indications and Usage heading. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER). October 2015.
6. CDISC SDTM Controlled Terminology. NCI Enterprise Vocabulary Services.2015-12-18. Available at: <http://www.cancer.gov/research/resources/terminology/cdisc>

CONTACT INFORMATION

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

Name: Kristin Kelly
Title: Consultant, Data Standards Consulting
Enterprise: Accenture Accelerated R&D Services
Address: 1160 W. Swedesford Rd, Bldg One
City, State ZIP: Berwyn, PA 19312
E-mail: kristin.c.kelly@accenture.com

Name: Jerry Salyers
Title: Senior Consultant, Data Standards Consulting
Enterprise: Accenture Accelerated R&D Services
Address: 1160 W. Swedesford Rd, Bldg One
City, State ZIP: Berwyn, PA 19312
E-mail: jerry.j.salyers@accenture.com

Name: Fred Wood
Title: Senior Manager and Lead, Data Standards Consulting
Enterprise: Accenture Accelerated R&D Services
Address: 1160 W. Swedesford Rd, Bldg One
City, State ZIP: Berwyn, PA 19312
E-mail: f.wood@accenture.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.