PharmaSUG 2019 - Paper DS-304

Considerations and Updates in the Use of Timing Variables in Submitting SDTM-Compliant Datasets

Jerry Salyers, Data Standards Consulting Group, TalentMine

ABSTRACT

Often, the appropriate use and population of Timing variables can present many challenges for sponsors when converting their operational database or legacy data to an SDTM-compliant format. This paper will discuss many of the common scenarios encountered when converting operational data to SDTM.

One of the common scenarios involve those occasions where the case report form (CRF) allows for checking an "ongoing" or "continuing" box in lieu of providing an end date. In such cases, the SDTM-based datasets require that these data points be represented by the correct use of the Relative-Timing variables (i.e., ---STRF, ---ENRF, --STTPT, --STRTPT, --ENTPT, and --ENRTPT). When doing so, sponsors must address questions such as these: 1) ongoing as of what point in time and 2) is the comparison to the study reference period more appropriate or is there an alternative anchor or reference time point that would be better suited? From controlled terminology, what are the best assigned choices for these different scenarios where an end date might be blank?

We will also explore the appropriate use of other Timing variables such as when a sponsor may need to express an evaluation interval in plain text rather than in ISO 8601 format (--EVINTX). We will also introduce the new Timing variables that will support the new Trial Milestone and Subject Milestone domains.

And finally, we will update areas where we have seen continued issues where data require the use of variables to define sample-collection time points (i.e., --TPT, TPTNUM, and –ELTM), along with the anchors that identify the "reference" or baseline for these collections (i.e., --TPTREF and –RFTDTC). .

INTRODUCTION

The valid SDTM timing variables can be found in Table 2.2.5.1 of Version 1.7 of the model. We can say that any of the timing variables can be used in any domain based on the three general observation classes, except where their use is discouraged (or prohibited) as part of the "Assumptions for Domain Models" as outlined in the SDTM Implementation Guide. An example here is where in a Findings observation class domain, the Timing variable --STDTC should not be used. Also, in this section of the SDTMIG, we're reminded that all domains (or datasets) based on the three general observation classes must contain at least one Timing variable.

As with the other tables in the model, any timing variable used in a SDTM data domain (SDTMIG v 3.1.2 or later) must be in the order shown in Table 2.2.5.1. For instance, any domain where data is "visit-based", the first timing variable encountered should always be VISITNUM.

Many timing variables represent data collected on the CRF (e.g. start date of an Adverse Event or the date of a blood pressure collection) while other timing variables (such as an actual study day of an observation) may be calculated or derived from the date of the observation and the start date of the subject's study reference period as contained in the DM domain. Still other timing variables are plain text that describe when, according to the protocol, an observation should be performed or a blood sample should be taken. There are also timing variables that represent this kind of text in standard format.

An issue for many sponsors is how to accurately express, using controlled terminology, the relative timing variables that are used to indicate, for example, that an event had not resolved at the time the event was captured on the CRF. While the CDISC standard for data collection (CDASH) has data entry fields that collect an "ongoing" check box in many CDASH domains, these ongoing flags need to be translated to

relative timing variables within SDTM. Thus it becomes necessary to define the question "relative to what" point in time (or "anchor") and then use the correct Timing variables and values from controlled terminology to express that relationship.

We also find that sponsors often aren't sure when it's necessary to use the "timepoint" collection variables and then how to correctly define the "reference" for these timepoints, which should be based on collected data in other domains.

As part of SDTM v1.7 and SDTMIG v3.3, there is a new Trial Level domain, TM, which can used to "describe observations or activities expected to occur during the course of the disease under study" and where the timing of these occurrences could be identified at the subject-level in the Subject Milestones (SM) dataset.

REPRESENTING RELATIVE TIMING DATA IN SDTM

--STRTPT, --STTPT, --ENRTPT, --ENTPT, --STRF, --ENRF

In terms of data collection on the CRF, in lieu of collecting a date, we may see values of "PRIOR", "BEFORE", "CONTINUING", or "ONGOING". It's common for a sponsor's operational database to translate these CDASH data points to a value of "Y". However, when mapping this data to the submission standard, a way is needed to answer the all-important question of "relative to what".

In early versions of the SDTM, sponsors were limited to only the two relative timing variables, --STRF and –ENRF. According to controlled terminology, these variables had allowable values of BEFORE, DURING, AFTER and U, with --ENRF having the additional allowable value of DURING/AFTER. These variables were somewhat restrictive in that they could only be used to indicate the start or the end of an event or an intervention in relation to the sponsor-defined reference period, defined in the demographic (DM) dataset as RFSTDTC to RFENDTC. Cases where sponsors may wish to collect this information in relation to other "anchors" or other points in time did not have a solution. For example, there are occasions where sponsors may feel it important to collect concomitant medications by VISIT such that "visit" becomes the anchor, rather than the study reference period. SDTM v1.2 offered these additional possibilities in the form of --STRTPT, --STTPT, --ENRTPT, and –ENTPT. The variables --STTPT and –ENTPT specify the alternative reference timepoints either in ISO 8601 format (if the anchor is a date) or in text (e.g. VISIT 3, STUDY EXIT, etc.). The --STRTPT and --ENRTPT then specify the "relativity" to those anchors according to controlled terminology.

So these "paired" relative Timing variables serve the same purpose as –STRF and –ENRF, that is, identifying the start or end of an event or an intervention (when a date is not collected) "relative" to an anchor. The SDTMIG v3.3, in section 4.4.7, provides detail according to different scenarios as to the valid values for --STRTPT and --ENRTPT. Largely these allowable values are based on whether the reference time point corresponds to the date of collection or assessment or is "prior" to the data of collection or assessment. Generally, --STRTPT has the allowable values of BEFORE, COINCIDENT, AFTER, and U, with –ENRTPT also having the allowable value of ONGOING. The values of COINCIDENT and ONGOING (though in the same codelist), cannot be used in –STRF or –ENRF as these values refer to a specific point in time. Conversely, the values of DURING or DURING/AFTER should not be used in --STRTPT or --ENRTPT as these refer to an "interval" of time, as in the Study Reference Period.

For instance, the SDTMIG, in the modeled CM domain, shows the same codelist STENRF for the variables CMSTRF, CMENRF, CMSTRTPT and CMENRTPT. However, the aforementioned section 4.4.7, under "Assumptions for Domain Models" provides guidance as to which choices are valid for each of these relative Timing variables.

Also, the SDTMIG cautions sponsors not to derive, for example, an "end" relative timing variable in the presence of an end date as this will blur the line between data that is collected and data that is derived. This is more appropriate for "analysis" (ADaM) datasets. In addition, we should emphasize that --STRTPT

must be used with --STTPT, and --ENRTPT must be used with --ENTPT. None of these variables can exist on their own, nor can --STRF be used along with --STTPT/--STRTPT or --ENRF with –ENTPT/--ENRTPT. Within a single domain, it's necessary to determine the answer to the "relative to what" question and then make exclusive use of the appropriate set of variables. We have often seen --STRF and --ENRF misused, but the introduction of the new paired variables, while providing some much needed flexibility, have also introduced new misunderstandings that the current SDTMIG attempts to alleviate.

Example MH record using the "paired" relative timing variables:

MHTERM	MHCAT	MHDTC	MHSTDTC	MHENRTPT	MHENTPT
ASTHMA	GENERAL MEDICAL HISTORY	2019-02-25	2016-10-01	ONGOING	2019-02-25

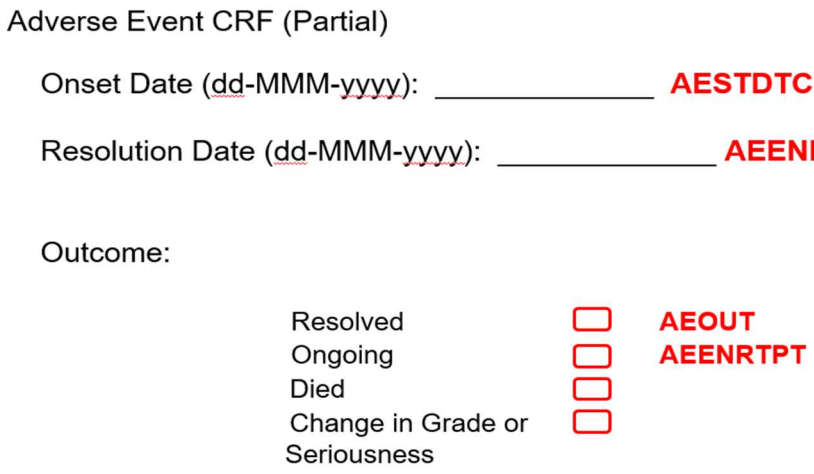
In the above example, the subject's condition of "ASTHMA" was considered "ONGOING" as of the date the assessment was taken. Hence the same value is in both MHDTC and MHENTPT. MHENTPT could also have plain text, such as specifying the anchor as 'SCREENING VISIT", for example. This avoids the redundancy of having MHDTC and MHENTPT being the same value.

Another possibility, using the study reference period as the "anchor":

MHTERM	MHCAT	MHDTC	MHSTDTC	MHENRF
ASTHMA	GENERAL MEDICAL HISTORY	2019-02-25	2016-10-01	DURING/AFTER

In this example, the assessment was again taken on 2019-02-25, however the CRF "ONGOING" flag was mapped to the MHENRF variable and assigned the value of "DURING/AFTER", meaning that the condition of ASTHMA resolved at some future point, either DURING the study reference period or AFTER the study reference period. The condition was not revisited after the initial collection. Note that "ONGOING" is not a valid value for the MHENRF variable.

Consider the following data collection from an AE CRF page where the page "mixes" timing information with the AE "Outcome".



As we know, the "OUT" codelist is "not extensible" such that a value of "ONGOING" in AEOUT would

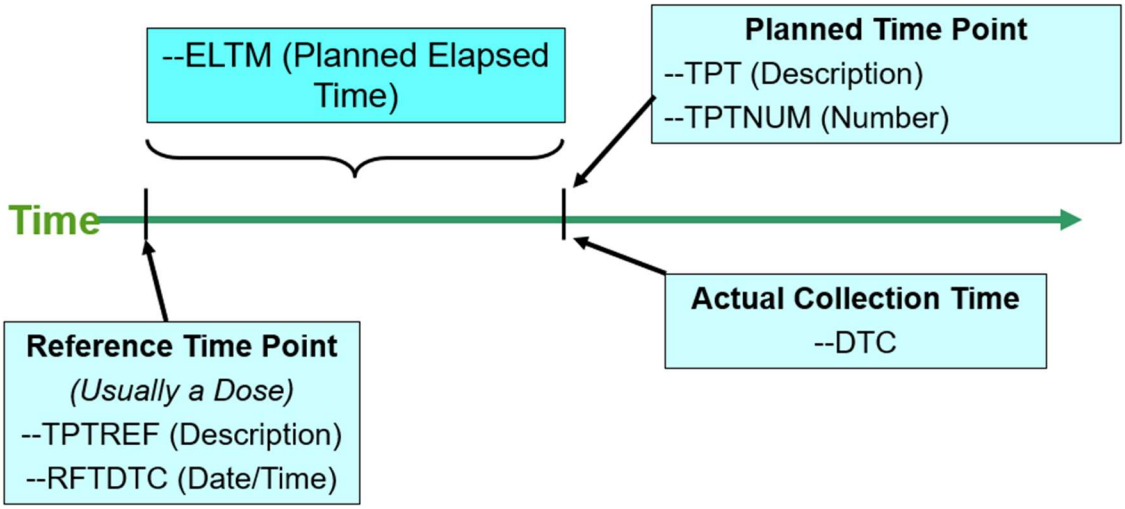
cause an error upon validation. The sponsor actually did a good job of mapping this "non-CDASH conformant" CRF to SDTM as evidenced by the annotation and the below snippet of AE dataset:

AETERM	AEOUT	AESTDTC	AEENRTPT	AEENTPT
SINUS TACHYCARDIA	NOT RECOVERED/NOT RESOLVED	2011-03-22	ONGOING	TRIAL EXIT

The collected value of "Ongoing" (as the "Outcome") is mapped to both AEOUT (according to controlled terminology) as well as to the relative Timing variable AEENRTPT. The reference "time point" is set to "TRIAL EXIT".

OBSERVATION TIMEPOINT VARIABLES

--TPT, --TPTNUM, --ELTM, --TPTREF, --RFTDTC



The above picture shows a continuum along the x-axis where the first vertical line represents some type of a "Reference" time point, most often a dose. That reference is shown as text in the --TPTREF variable while the collected data "behind" the reference (from another domain) is shown in the –RFTDTC variable. The second vertical line represents an actual date/time of an observation collected or performed at a protocol specified "nominal" time. The nominal time is shown as text in the --TPT variable while the actual date/time is, of course, in the --DTC variable. The --ELTM variable is the "Planned Elapsed Time" represented in ISO 8601 format. The resultant data may look something like this (not in prescribed variable order):

PCTPTNUM	PCTPT	PCELTM	PCTPTREF	PCRFTDTC	PCDTC
Time Point Number	Time Point Name	Planned Elapsed Time	Time Point Reference	Date/Time of Time Point Reference	Sample Date/Time
1	-10 MIN	-PT10M	DAY 1 DOSE	2009-05-17T08:00	2009-05-17T07:50
2	1 HR	PT1H	DAY 1 DOSE	2009-05-17T08:00	2009-05-17T09:00
3	2 HR	PT2H	DAY 1 DOSE	2009-05-17T08:00	2009-05-17T10:02
4	24 HR	PT24H	DAY 1 DOSE	2009-05-17T08:00	2009-05-18T8:02
5	48 HR	PT48H	DAY 1 DOSE	2009-05-17T08:00	2009-05-19T8:10

When time points (--TPTs) are used, --TPTNUMs are expected. As the SDTMIG reminds us, if the scheduling of an observation or sample collection (as for PC) is related to an Intervention, then --TPTREF should also be populated. Section 4.4.10 of the "Assumptions for Domain Models) takes us through the relationships surrounding these timing variables and when they need to be populated in order to confer uniqueness for the record. There are also times that --TPTREF is populated to simply assist in interpreting the data.

The SDTMIG also reminds us that there is always a 1:1 relationship between --TPT and –TPTNUM. Validation tools will identify any instance where this is violated. This 1:1 relationship also extends to --TPT and --ELTM, although it's understood there may be times when a --TPT is populated, however, the --ELTM is NULL. This may be for those occasions where the --ELTM cannot be expressed in ISO 8601 Period format.

To illustrate these relationships and also to identify which variables contribute to a record being unique, the SDTMIG provides a couple of different options as an example. Most often, where time points are involved or when there may be more than 1 "serial collection" within a study or even within a visit, uniqueness is conferred by a combination of VISITNUM/VISIT, --TPTNUM and --TPTREF. Using VS as an example, one option may look like this where VISIT describes the "Period" as a whole while VSTPTREF describes the Dosing Day and the particular dose on that day:

Row	STUDYID	DOMAIN	USUBJID	VSSEQ	VSTESTCD	VSORRES	VSORRESU
1	ABC0001	VS	001-101	1	SYSBP	120	mmHg
2	ABC0001	VS	001-101	2	SYSBP	128	mmHg
3	ABC0001	VS	001-101	3	SYSBP	124	mmHg
4	ABC0001	VS	001-101	4	SYSBP	122	mmHg
5	ABC0001	VS	001-101	5	SYSBP	120	mmHg
6	ABC0001	VS	001-101	6	SYSBP	118	mmHg
7	ABC0001	VS	001-101	7	SYSBP	126	mmHg
8	ABC0001	VS	001-101	8	SYSBP	124	mmHg
9	ABC0001	VS	001-101	9	SYSBP	130	mmHg

Row (Con'd)	VISITNUM	VISIT	VSDTC	VSTPT	VSTPTNUM	VSELTM	VSTPTREF
1	3	PERIOD1	2019-04-01	PREDOSE	1		DAY 1 AM
2	3	PERIOD1	2019-04-01	1 HR POST	2	PT1H	DAY 1 AM

3	3	PERIOD1	2019-04-01	4 HR POST	3	PT4H	DAY 1 AM
4	3	PERIOD1	2019-04-01	PREDOSE	1		DAY 1 PM
5	3	PERIOD1	2019-04-01	1 HR POST	2	PT1H	DAY 1 PM
6	3	PERIOD1	2019-04-01	4 HR POST	3	PT4H	DAY 1 PM
7	3	PERIOD1	2019-04-05	PREDOSE	1		DAY 5 AM
8	3	PERIOD1	2019-04-05	1 HR POST	2	PT1H	DAY 5 AM
9	3	PERIOD1	2019-04-05	4 HR POST	3	PT4H	DAY 5 AM

Of note above, each time the VSTPTREF changes, the specified TPTs, TPTNUMs, and ELTMs are re-used such that we always maintain the 1:1 relationship between TPT and TPTNUM as well as between TPT and ELTM. Also note that VSELTM is not populated for the "PREDOSE" time points.

Contrast the above with the following representation where VISITNUM/VISIT is the more granular and VSTPTREF simply denotes the dose within the VISIT (only the time point variables are repeated):

Row (Con'd)	VISITNUM	VISIT	VSDTC	VSTPT	VSTPTNUM	VSELTM	VSTPTREF
1	3	PERIOD 1 DAY 1	2019-04-01	PREDOSE	1		AM DOSE
2	3	PERIOD 1 DAY 1	2019-04-01	1 HR POST	2	PT1H	AM DOSE
3	3	PERIOD 1 DAY 1	2019-04-01	4 HR POST	3	PT4H	AM DOSE
4	3	PERIOD 1 DAY 1	2019-04-01	PREDOSE	1		PM DOSE
5	3	PERIOD 1 DAY 1	2019-04-01	1 HR POST	2	PT1H	PM DOSE
6	3	PERIOD 1 DAY 1	2019-04-01	4 HR POST	3	PT4H	PM DOSE
7	4	PERIOD 1 DAY 5	2019-04-05	PREDOSE	1		AM DOSE
8	4	PERIOD 1 DAY 5	2019-04-05	1 HR POST	2	PT1H	AM DOSE
9	4	PERIOD 1 DAY 5	2019-04-05	4 HR POST	3	PT4H	AM DOSE

Obviously, how the study's VISIT structure is detailed in TV will determine which of the above scenarios is used when conferring uniqueness for these different time points. Again, the 1:1 relationship between TPT and TPTNUM as well as between TPT and ELTM is preserved.

--DTC, --STDTC

As we know, in an Interventions or Events domain, the --DTC is simply the date that the information was recorded. Obviously, of limited value as compared to the far more important –STDTC. The below snippet of dataset depicts an instance where the sponsor chose to use the –DTC for a purpose other than intended (just representative records shown from a CRF that collected Parkinson's Disease symptoms and then a separate CRF that captured "GENERAL" medical history):

mh.xpt

STUDYID	DOMAIN	USUBJID	MHSEQ	MHTERM	MHPRESP	MHOCCUR	MHDTC	MHSTDTC
ABC0001	MH	001-101	1	BRADYKINESIA	Y	Y	2003-05-06	2003-11-21
ABC0001	MH	001-101	2	DYSKINESIA	Y	Y	2003-05-06	2003-05
ABC0001	MH	001-101	3	RIGIDITY	Y	Y	2003-05-06	2003-11-21
ABC0001	MH	001-101	4	TREMOR	Y	Y	2003-05-06	2002-05-06
ABC0001	MH	001-101	5	ANXIETY			2009-03-30	1995-12-05
Abc0001	MH	001-101	6	DEPRESSION			2009-03-30	2005-12-05

The conditions of “ANXIETY” and “DEPRESSION” were spontaneously reported conditions recorded on “2009-03-30”. This date is correctly shown in MHDTC for these two conditions. So what date is recorded in MHDTC for the Parkinson’s Disease symptoms? According to the annotated CRF, this date is the “date of original diagnosis” of Parkinson’s. As is seen, this date then appears on every symptom record. Initially, a reviewer would not have known what this data signified.

What would have been a better and more compliant strategy? We recommended that the sponsor simply create another MH record where MHTERM equals “PARKINSON’S DISEASE” and then the “date of original diagnosis” would go into the MHSTDTC as would be expected. With the advent of SDTMIG v3.3, the domain specific variable MHEVDTYP could also be employed. Obviously, it’s never a good idea to use a variable for another purpose than is intended.

TRIAL MILESTONES AND SUBJECT MILESTONES

In SDTM v1.7 and the SDTMIG v3.3, we see the “trial level” domain “Trial Milestones” (TM) specified as well as the “subject-level” special purpose domain “Subject Milestones” (SM). Along with these new domains, new Timing variables were also created in Table 2.2.5 of the Model and then employed in the specifications of the “Subject Milestones” domain. As with Timing variables in general, these newer Timing variable are eligible to be used in other subject-level data domains where they add needed context to the collected “milestone” information.

In Section 4 of the SDTMIG v3.3 (Assumptions for Domain Models), a Trial Disease Milestone is defined as “an event or activity that is anticipated in the course of the disease, but whose timing is not controlled by the study schedule”. The SDTMIG goes onto say that the identified “milestone” may be something that is anticipated to occur in the subject population and may trigger the collection of other “related” data, outside of the trial visit schedule.

Below is a representative record in the “Trial Milestone” dataset where the TMDEF variable provides the definition of the disease milestone and the TMRPT variable indicates if the milestone can occur only once (TMRPT = ‘N”) or can occur multiple times over the course of the trial (TMRPT = “Y”):

tm.xpt

Row	STUDYID	DOMAIN	MIDSTYPE	TMDEF	TMRPT
1	DEF123	TM	DIAGNOSIS	Initial diagnosis of diabetes	N
2	DEF123	TM	HYPOGLYCEMIC EVENT	Occurrence of a hypoglycemic event defined as a glucose level below a threshold as defined in the protocol	Y

In the "Special Purpose" subject-level domain SM, the defined MIDSTYPE from TM is detailed along with the variable MIDS which either is a shortened version of MIDSTYPE (where TMRPT = "N") or represents unique occurrences of MIDSTYPE (where TMRPT = "Y"). As v3.3 of the SDTMIG says, the MIDS variable "serves to link observations associated with a disease milestone similar to the say that VISITNUM links observation performed at a given VISIT". MIDS is the Topic variable of SM.

STUDYID	DOMAIN	USUBJID	SMSEQ	MIDS	MIDSTYPE	SMSTDTC
DEF123	SM	015-001	1	DIAG	DIAGNOSIS	2010-03-01
DEF123	SM	015-001	2	HYPO1	HYPOGLYCEMIC EVENT	2013-05-01T08:00
DEF123	SM	015-001	3	HYPO2	HYPOGLYCEMIC EVENT	2013-05-17T09:30

Table 2.2.5 of the Model shows three new Timing variables as follows:

MIDS: The name of the specific instance of a Disease Milestone Type (MIDSTYPE)

RELMIDS: The temporal relationship of the observation to the Disease Milestone instance as shown in MIDS

MIDSDTC: The start date/time of the Disease Milestone instance as shown in MIDS

By definition, the SM domain is populated based on data present in other domains. For example, the above "Diagnosis" of diabetes would have been detailed in the MH domain where the new Timing variable MIDS has been added (Note the inclusion of the new domain-specific variable MHEVDTYP):

STUDYID	DOMAIN	USUBJID	MHSEQ	MHTERM	MHEVDTYP	MHSTDTC	MIDS
DEF123	MH	015-001	1	TYPE 2 DIABETES	DIAGNOSIS	2010-03-01	DIAG

As stated above, the occurrence of a disease milestone as represented in SM may trigger the collection of data in other domains, such as the collection (for example, in CE) of each occurrence of a hypoglycemic event:

STUDYID	DOMAIN	USUBJID	CESEQ	CETERM	CESTDTC	MIDS
DEF123	CE	015-001	1	HYPOGLYCEMIC EVENT	2013-05-01T08:00	HYPO1
DEF123	CE	015-001	2	HYPOGLYCEMIC EVENT	2013-05-17T09:30	HYPO2

In addition to records in CE, we may also have records in other domains that are pertinent to the subject milestone. The below represents records that may be in Exposure (EX) that represent the most recent dose of study medication prior to the milestone:

Row	STUDYID	DOMAIN	USUBJID	EXSEQ	EXTRT	EXDOSE	EXDOSU	EXSTDTC
1	DEF123	EX	015-001	1	METFORMIN	500	mg	2013-05-01T06:30
2	DEF123	EX	015-001	2	METFORMIN	500	mg	2013-05-17T06:30

Row	MIDS	RELMIDS	MIDSDTC
1 (Cont)	HYPO1	MOST RECENT DOSE PRIOR TO	2013-05-01T08:00
2(Cont)	HYPO2	MOST RECENT DOSE PRIOR TO	2013-05-17T09:30

CONCLUSION

Timing variables add often-needed context around the SDTM-based tabulation data. The day of a blood pressure measurement, the start date of study medication, the end date of an adverse event are all submitted in Timing variables. Relative Timing variables also are used to translate an "Ongoing" box to actual data as they help to define the all important question of "relative to what". Also, they often serve as part of a dataset's natural key structure, particularly Timing variables such as a --TPT. So in this way, they are very important. One of the issues that all sponsors face is which Timing variables, as well as the most correct codelist value, should be used with each of the general observation classes and thus, with which domains. How Timing variables are reported, the order they are presented in, their structure, and their associated controlled terminology all play a part in their correct use.

REFERENCES

SDTM Version 1.4 and SDTMIG Version 3.2

SDTM Version 1.7 and SDTMIG Version 3.3

CONTACT INFORMATION

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

Jerry Salyers
TalentMine
jerry.salyers@datastandardsconsulting.com