

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

Jerry Salyers, Richard Lewis, and Fred Wood  
Accenture Life Sciences, Wayne, PA

### 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 presentation will discuss some of the issues commonly seen in three general areas.

The first of these is 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 the use of the Relative-Timing variables (i.e., --STRF, --ENRF, --STTPT, --STRTP, --ENTPT, and --ENRTP). When doing so, sponsors must address questions such as these: 1) “ongoing as of what point in time” (in order to ensure the correct values from the controlled terminology codelist), and 2) “is the Study Reference Period more appropriate or is there an alternative study anchor that would be better suited?”

The second area is when data require the use of variables used 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).

The third area involves a misunderstanding of the differences between the --DTC and the --STDTC variables, and the associated derived variables (i.e., --DY and --STDY).

### INTRODUCTION

The standard SDTM Timing variables can be found in Table 2.2.5 of the SDTMI. Any of the Timing variables can be used in any domain based on the three general observation classes, except where their use is discouraged as part of the “assumptions” section shown for each modeled domain in the SDTM Implementation Guide (SDTMIG). As with the other tables in the model, any Timing variable used in an SDTM-based domain must be in the order shown in Table 2.2.5. For instance, any domain where data is collected by Visit, the first Timing variable is always VISITNUM. This is also true for any Timing variable that is added to a domain where it may not exist in the modeled domain within the IG. The added Timing variable still gets slotted in where it appears in Table 2.2.5 and not added to the end of the dataset.

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 defined in the DM domain variable RFSTDTC. 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 as yet resolved at the time the event was captured on the CRF (in the case of Medical History) or had not as yet resolved at the time the study concluded (an Adverse Event). 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 SDTM variables using the relative Timing variables. Thus it becomes necessary to define the question “relative to what” point in time and then use the correct value from controlled terminology to express that relationship.

We also find that sponsors often aren’t sure when it’s necessary to use the “time point” collection variables and then how to correctly define the “reference” for these time points, which should be based on collected data in other domains.

## **CHALLENGES TO THE ACCURATE REPRESENTATION OF RELATIVE TIMING DATA IN SDTM**

### **--STRTPT/--STTPT, --ENRTPT/--ENTPT, --STRF, --ENRF**

In Version 1.1 of the SDTM, sponsors were limited to only the two relative Timing variables --STRF and --ENRF. These variables had allowable values of BEFORE, DURING, AFTER and U, with --ENRF having the additional allowable value of DURING/AFTER. These variables were a little restrictive in that they could only be used to indicate the start and the end of an Event or an Intervention in relation to the sponsor-defined Study Reference Period, defined in the Demographics (DM) dataset, bounded by RFSTDTC and RFENDTC. Cases where sponsors may 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 the visit becomes the anchor, rather than the Study Reference Period. SDTM Version 1.2 offered these additional possibilities in the form of --STRTPT, --STTPT, --ENRTPT, and --ENTPT. The variables --STTPT and --ENTPT were able to specify the additional reference time points either in ISO 8601 format (for instance, if the anchor is a date) or in text (e.g. VISIT 3, FINAL DOSE, etc.). We have seen --STRF and --ENRF misused on a regular basis, but the introduction of the new paired variables, while providing some much needed flexibility, also introduced many new misunderstandings.

First of all, --STRTPT and --ENRTPT serve the same purpose as --STRF and --ENRF, that is identifying the start or end of an Event or an Intervention (in the absence of a collected date) but with one subtle important difference. The newer variables --STRTPT and --ENRTPT have the allowable values of BEFORE, COINCIDENT, AFTER, and U, with --ENRTPT also allowing the value of ONGOING. The new values of COINCIDENT and ONGOING cannot be used in --STRF or --ENRF, nor can the older values of DURING or DURING/AFTER be used as values for the --STRTPT or --ENRTPT variables. Some of this confusion seems to arise from the fact that COINCIDENT and ONGOING were added to the STENRF codelist, the same codelist referenced in the --STRF and --ENRF variables within the SDTMIG. However, it should be emphasized that the SDTMIG specifically indicates in Section 4.1.4.7 which values are allowable for each of these variables. Also, in the SDTMIG, the --STRTPT and --ENRTPT variables in the relevant domain tables do not reference the STENRF codelist, rather they explicitly list the allowable values. The fact that CT utilizes the same codelist for all of the accepted values for any relative Timing variable does not override the explicit values stated in the SDTMIG.

Another misunderstood point is that --STRTPT must be used with --STTPT, and --ENRTPT must be used with --ENTPT. None of these variables can be used alone, nor can --STRF be used with --STTPT or --ENRF with --ENTPT. Within a single domain, it is necessary to decide that question “relative to what” and then make exclusive use of the appropriate set of variables. Again, the anchor for the use of a relative Timing variable is either the Study Reference Period or a defined “point in time” (e.g., --STTPT, --ENTPT), but not both. It should also be noted that --STRTPT and --STTPT are meaningless without each other. The same can be said for --ENRTPT and --ENTPT.

In our review of sponsor-supplied SDTM datasets, we have seen an increase in sponsors using the new relative time point variables more so than the older ones that use the Study Reference Period as the anchor. It seems that some sponsors wish to retain their collected “Ongoing” flag, rather than perform a mapping step to the appropriate codelist value that is based on the Study Reference Period. This is perfectly acceptable as long as the sponsor avoids “mixing and matching” the possible combinations of variables along with their allowable values.

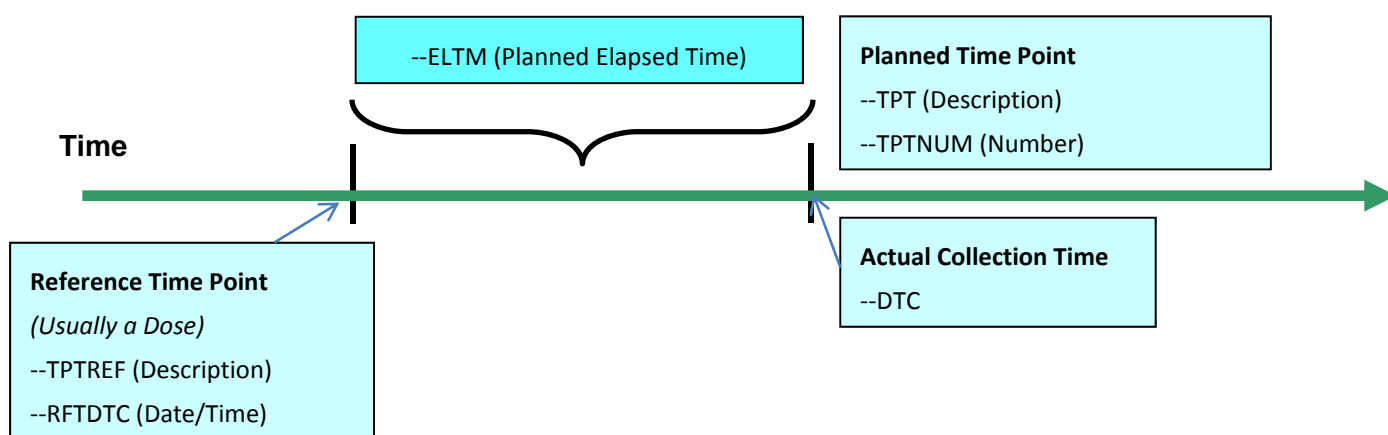
It must be reiterated that any use of a relative Timing variable is always in lieu of a collected date, even if that collected date is only a partial date. The SDTMIG, in Section 4.1.4.7, states that sponsors should withstand the temptation of deriving a relative Timing variable, such as an --ENRF, when an end date is collected. Including both on the same record blurs the line between what is collected on the CRF and what may be derived. Consider the example below where an “Ongoing” box was collected for the records whose CMSEQ values are 2, 4, and 5 and then mapped to the value of ‘AFTER’ from CDISC Controlled Terminology. For records whose CMSEQ values are 1, 3, and 6, the sponsor decided to map CMENRF based on the collected end date and the date in RFENDTC in the DM dataset. The sponsor should not have derived any additional data for those records that contained an end date. This is incorrect and would be flagged by most validation tools.

STUDYID	DOMAIN	USUBJID	CMSEQ	CMTRT	CMSTDTC	CMENDTC	CMENRF
ABC-0001	CM	0012-1001	1	IBUPROFEN	2009-03-24	2009-12-27	DURING
ABC-0001	CM	0012-1001	2	PRILOSEC	2009-03-15		AFTER
ABC-0001	CM	0012-1001	3	LEXAPRO	2008-12-15	2009-07-22	DURING
ABC-0001	CM	0012-1001	4	METFORMIN	2008-12-15		AFTER
ABC-0001	CM	0012-1001	5	AMLODIPINE	2008-11-18		AFTER
ABC-0001	CM	0012-1001	6	HCTZ	2009-01-08	2009-04-28	DURING

## OBSERVATION TIME POINT VARIABLES

### --TPT, --TPTNUM, --ELTM, --TPTREF AND --RFTDTC

By way of review, we can look at how the nominal time point variables can generally be used. Let's look at the following diagram:



This diagram represents an observation being conducted at a specific time point or planned interval from a fixed point of reference. As we see, this diagram accounts for the five nominal time point variables plus the date/time of the observation itself.

The variables --TPT and --TPTNUM are expected to represent planned time points as they are specified in the protocol. Whenever a --TPT is used, the assigning of a corresponding --TPTNUM is expected. There should be a 1:1 relationship between --TPT and --TPTNUM within a reference or "anchor" (--RFTDTC) within a domain. If the domain is comprised of multiple "time over results" curves (more than one --TPTREF), the same --TPT and --TPTNUM combinations can be re-used as required if the nominal times are the same. One of the keys here is to be as specific as possible when populating --TPTREF. All too often we see something as generic as "Dose Administration", even when there is more than a single "time over results" curve. Also, of course, it is imperative that there is data (usually in another domain) to support the reference with which to populate --RFTDTC. A sample of time point data might look like this (variables not necessarily in prescribed order):

*Considerations in the Use of Timing Variables in Submitting SDTM-Compliant Datasets Continued*

--TPT	--TPTNUM	--ELTM	--TPTREF	--RFTDTC	--DTC
Time Point Name	Time point Number	Planned Elapsed Time	Time Point Reference	Date/Time of Time Point Reference	Observation Date/Time
-10 MIN	-1	-PT10M	DAY 1 DOSE	2013-04-01T08:00	2013-04-01T07:50
1 HR POST	1	PT1H	DAY 1 DOSE	2013-04-01T08:00	2013-04-01T09:00
2 HRS POST	2	PT2H	DAY 1 DOSE	2013-04-01T08:00	2013-04-01T10:02
4 HRS POST	4	PT4H	DAY 1 DOSE	2013-04-01T08:00	2013-04-01T12:04
8 HRS POST	8	PT8H	DAY 1 DOSE	2013-04-01T08:00	2013-04-01T16:08

Notice the assigning of the --TPTNUM. Often sponsors will mimic the time point name in the --TPTNUM variable, which is perfectly acceptable, as the primary function of --TPTNUM is as a sort key. Again, this is permissible as long as “planned” times are used. The problem arises when sponsors assign the --TPTNUM based on the actual date/time of the observation rather than the planned time point. We see this fairly frequently in sponsor-supplied SDTM datasets. Based on the above data, we might see something like this for the post-dose time points:

--TPT	--TPTNUM	--ELTM	--DTC
1 HR POST	1	PT1H	2013-04-01T09:00
2.03 HRS POST	2.03	PT2.03H	2013-04-01T10:02
4.07 HRS POST	4.07	PT4.07H	2013-04-01T12:04
8.13 HRS POST	8.13	PT8.13H	2013-04-01T16:08

This representation should be left to analysis. The nominal time point variables are all based on the series of planned observations as outlined in the protocol.

One of the issues that is unique to PC is that the concentrations in samples collected over a protocol-defined interval (sometimes referred to as an “intense” PK collection day or period) are used to calculate PK parameters that are stored in the PP domain. By definition, these parameters are not based on individual samples but on the entire “time vs. concentration” curve produced over the course of the sample collections. If a study has more than one of these “intense” sample collection periods or PK “profiles”, then it becomes necessary to make sure that a given set of PK parameters (in PP) is matched with the correct set of analyte concentrations (in PC). This link is accomplished by making sure that the PCRFTDTC for the set of concentration records matches the PPRFTDTC in the parameter records.

Another point to make regarding assigning of time point variables is there are occasions (most notably in the PC domain for PK concentration data) where collections may not have a “reference” as such. If a subject has a “trough” PK sample collected at a given visit (usually prior to that day’s dose of study medication), the CRF may collect the “most recent” dose or possibly even the two most recent doses. Often a sponsor will indicate the most recent dose as the PCTPTREF and populate PCRFTDTC with the date/time of that dose. However, if the sample is taken to simply confirm the subject has reached “steady state”, the collection of that trough sample has no real reference and the nominal time point variables should remain blank. The “most recent” dose(s) records would be best populated in the Exposure (EX) domain with an appropriate --CAT to signify the dose was the most recent prior to a PK collection.

### **--DTC, --STDTC, --DY AND --STDY**

The variable --STDTC is often misused within Findings domains. For example, even though sponsors know that --DTC and --ENDTC should be used within a Findings domain, when an additional date comes into play, sponsors often feel the need to populate this additional date in a Findings --STDTC. For example, we have seen instances where a Findings domain captures machine-generated data at prescribed intervals. The usual --DTC captures the date/time of the discreet reading while the date/time that the machine was turned on was mapped to a --STDTC variable. Obviously, this is not optimum mapping as the --STDTC variable should not be used in a Findings domain. In the above scenario, the date/time of machine activation should be a record in an appropriate general observation class domain. Only the --DTCs of the discreet values at the prescribed time points should be in the Findings domain.

We've also seen a number of instances where in a single domain (such as EX) both the --DY and --STDY variables are populated. A reviewer would not expect both the --DY and --STDY variables to be populated in the same dataset as it would provide no relevant information. If we use EX as an example, EXDTC and EXDY are not in the modeled EX domain in the implementation guide as EXDTC would rarely be recorded on the CRF. The modeled EX domain only contains the EXSTDY variable which is calculated as the integer days between the subject's RFSTDTC recorded in DM and the EXSTDTC contained in the EX record.

To summarize, --STDY variables are calculated from values contained in --STDTC variables while --DY variables are calculated from values contained in --DTC variables.

## **A WORD ABOUT EPOCH**

There is plenty of confusion regarding this Timing variable. EPOCH is simply a slice of time that cuts across all ARMs of a clinical trial. All ARMs may have a "Screening" epoch, a "Treatment" epoch and a "Follow Up" epoch. These values for epoch first reside in the trial design domain TA. Also, as a Timing variable in Table 2.2.5, epoch is available to be used in any general observation class domain. In one of FDA's most recent "Common Data Standards Issues" documents, FDA has recommended that epoch be added to subject level data domains. Based on this recommendation we most often see sponsors adding epoch to Disposition (DS) and Exposure (EX). Of course, adding epoch to a domain such as Disposition depends on the level of detail captured in the CRF. If the CRF captures only a single end-of-study record, then epoch would not be needed to differentiate multiple disposition events. The first rule of populating epoch, of course, would be that the value must first exist in TA. Care should be taken that epoch is not redundant with "Visit" if visit is added as an additional Timing variable to DS.

## **CONCLUSIONS**

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. Also, they often serve as part of a dataset's natural key structure. So in this way, they are very important. One of the issues that all sponsors face is which Timing variables are most often 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.

## **ACKNOWLEDGMENTS**

The authors would like to acknowledge the management at Accenture Life Sciences who understand the importance of data standards in pharmaceutical research, and who have supported our participation in numerous CDISC teams.

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 registered trademarks or trademarks of their respective companies.

## **CONTACT INFORMATION**

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

Jerry Salyers  
Senior Consultant, Data Standards Consulting  
Accenture Life Sciences  
585 East Swedesford Road  
Wayne, PA 19087  
513.668.8126  
jerry.j.salyers@accenture.com

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

Richard Lewis  
Senior Consultant, Data Standards Consulting  
Accenture Life Sciences  
585 East Swedesford Road  
Wayne, PA 19087  
262.716.4167  
[richard.m.lewis@accenture.com](mailto:richard.m.lewis@accenture.com)

Fred Wood  
Vice President, Data Standards Consulting  
Accenture Life Sciences  
585 East Swedesford Road  
Wayne, PA 19087  
484.881.2297  
[f.wood@accenture.com](mailto:f.wood@accenture.com)