

SDTM TE, TA, and SE Domains: Demystifying the Development of SE

Kristin Kelly, Accenture Accelerated R&D Services, Berwyn, PA

Fred Wood, Accenture Accelerated R&D Services, Berwyn, PA

Jerry Salyers, Accenture Accelerated R&D Services, Berwyn, PA

ABSTRACT

There is a necessary relatedness among the SDTM Trial Design domains, TA (Trial Arms) and TE (Trial Elements), and the special-purpose domain, SE (Subject Elements), that serves to convey both the planned (TE/TA) and actual (SE) treatment paths for a subject within a study. The SE domain is derived from the subject-level general observation class domains based on 1) the Start and End rules in TE, and 2) the Transition and Branching rules (TATRANS and TABRANCH) in the TA domain. The SDTMIG principles state that there are no gaps between Elements, meaning that the start date/time of an Element (SESTDTC) is the same as the end date/time of the previous Element (SEENDTC). The SE domain can be cumbersome to create if the code attempts to account for both the Start Rule and End Rule for each Element. This paper will discuss a streamlined process for deriving SESTDTC and SEENDTC by accounting for only the Start Rule (SESTDTC) for each Element, while including the End Rule (SEENDTC) in the code only if the subject does not fulfill the planned Start Rule of their next Element (e.g., due to study discontinuation, some other unplanned occurrence), or if it is the last Element.

INTRODUCTION

The SDTM Trial Design domains represent the planned treatment paths for a subject in a clinical trial based on the protocol. These trial design domains consist of the following: Trial Arms (TA), 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. Modeling an accurate representation of the treatment paths and rules defined in the protocol is also critical, since these domains serve as the basis for deriving the actual treatment path for a given subject. The Special-Purpose domains, Subject Elements (SE) and Subject Visits (SV), are those domains that represent a subject's actual path through the trial, and are created from the subject-level general observation class (GOC) domains based on the rules in the trial design tables.

There are five SDTM Trial Design domains; however, this paper will focus on TA and TE as well as the Special-Purpose domain, SE. Sponsors and programmers typically find that developing SE can be difficult, especially when the rules in TE do not reference data points that have been collected on the CRF at the subject level. Also, SE can be more work intensive to create when the programmer is accounting for each start and end rule in the code when by definition, the end of one Element is always the start of the next. This paper will discuss a strategy that will make SE creation easier and with fewer errors. First, some basic definitions for TA, TE, and SE will be presented.

TE, TA, AND SE BASICS

Many study protocols provide a study schematic that depicts the design of the trial. These diagrams can be useful in helping to populate the TA and TE domains. The TE and TA domains are built upon the concepts of Elements, Arms, and Epochs. They are shown in Figure 1, and are described in the following sections.

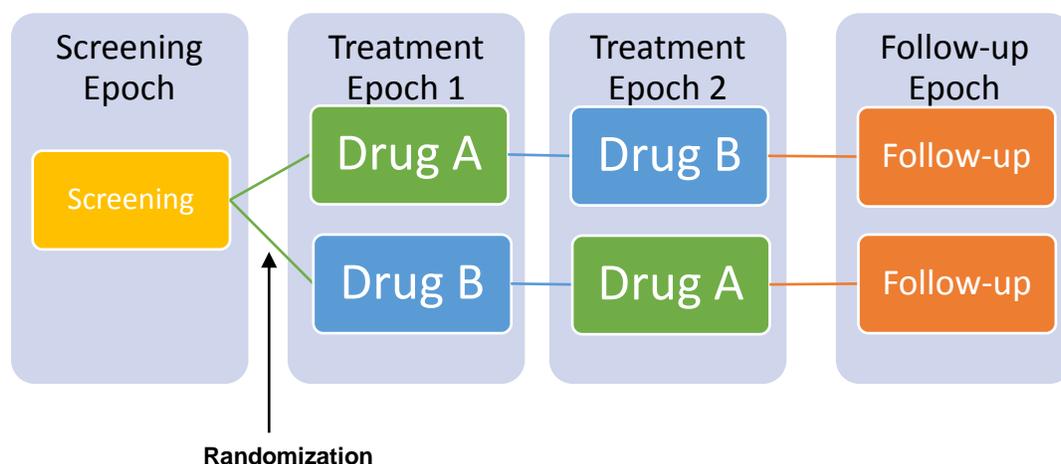


Figure 1. Key Concepts of TE and TA

This example trial shows an open label, crossover study where eligible subjects are randomized to one of two treatment sequences: ‘Drug A – Drug B’ or ‘Drug B – Drug A’. The inner boxes above represent the Elements and the order of the Elements is specific to each treatment sequence, or treatment Arm. The larger boxes depict the Epoch when each Element occurs during the trial. Each treatment period has a planned duration of 7 days in which subjects are dosed on Day 1 of each period with PK sampling at specific time points on the days following up to Day 7. A two-week follow-up Element/Epoch is planned for subjects that complete both treatment phases.

Trial Elements (TE)

An Element is the basic building block that is used to describe the administration of planned interventions. Periods of time when there are no planned interventions (such as screening or washout) are also considered Elements. Planned Elements are described in the Trial Elements (TE) table. Each planned Element will have a rule for when the Element begins (TESTRL) and a rule for when the Element is planned to end (TEENRL). An example for the trial above is shown in Table 1.

| STUDYID | DOMAIN | ETCD | ELEMENT | TESTRL | TEENRL | TEDUR |
|---------|--------|----------|-----------|---|----------------------------|-------|
| XYZ999 | TE | SCREEN | Screening | Informed consent obtained | First dose of treatment | |
| XYZ999 | TE | A | Drug A | First dose of Drug A | 1 week after element start | P7D |
| XYZ999 | TE | B | Drug B | First dose of Drug B | 1 week after element start | P7D |
| XYZ999 | TE | FOLLOWUP | Follow-up | Last PK sample obtained following Period 2 dosing | Last Contact with Subject | P2W |

Table 1. TE Example

Trial Arms (TA)

An Arm is a planned sequence of Elements within Epochs, and is often referred to as a treatment group. The Trial Arms (TA) table describes the sequence of Elements within each Arm, as well as rules for moving through these Elements, referred to as Branches (TABRANCH) and Transitions (TATRANS). An Epoch is a slice of time in the study that typically applies to all subjects independent of their planned Arm or Element path. These Epochs are defined in the TA table and can be derived in the subject-level GOC domains based on TA and TE. Examples include Screening, Treatment, or Follow-up. The Demographics (DM) domain contains the short and long names of both the planned Arm to which each subject was assigned and the actual Arm in which a subject participated. An example for the trial above is shown in Table 2. Please note that only variables relevant to the example are shown due to space restrictions.

| STUDYID | DOMAIN | ARMCD | ARM | TAETORD | ETCD | TABRANCH | EPOCH |
|---------|--------|-------|-----------------|---------|----------|---------------------------|-------------|
| XYZ999 | TA | AB | Drug A - Drug B | 1 | SCREEN | Randomized to sequence AB | SCREENING |
| XYZ999 | TA | AB | Drug A - Drug B | 2 | DRUG A | | TREATMENT 1 |
| XYZ999 | TA | AB | Drug A - Drug B | 3 | DRUG B | | TREATMENT 2 |
| XYZ999 | TA | AB | Drug A - Drug B | 4 | FOLLOWUP | | FOLLOW-UP |
| XYZ999 | TA | BA | Drug B - Drug A | 1 | SCREEN | Randomized to sequence BA | SCREENING |
| XYZ999 | TA | BA | Drug B - Drug A | 2 | DRUG B | | TREATMENT 1 |
| XYZ999 | TA | BA | Drug B - Drug A | 3 | DRUG A | | TREATMENT 2 |
| XYZ999 | TA | BA | Drug B - Drug A | 4 | FOLLOWUP | | FOLLOW-UP |

Table 2. TA Example

Subject Elements (SE)

The Subject Elements domain allows the submission of data on the timing of the Elements a subject actually passed through in their participation in the trial. Using the information defined in the TA and TE domains, the actual start date/time (SESTDTC) and end date/time (SEENDTC) that a subject was in a particular Element can be derived from data in the subject-level GOC domains. According to the SDTMIG, since there are no gaps between Elements, the SEENDTC for one Element will always be the same as the SESTDTC for the next Element. This also means that a subject can only be in one Element at a time.

Only those Elements that a subject actually passed through will be present as records in the SE dataset. For example, a subject that did not pass screening would only have a record for the screening Element in SE. An example for a few subjects from the above trial is provided below in Table 3. Please note that only variables relevant to the example are shown due to space restrictions.

| STUDYID | DOMAIN | USUBJID | SESEQ | ETCD | SESTDTC | SEENDTC | EPOCH |
|---------|--------|------------|-------|----------|------------|------------|-------------|
| XYZ999 | SE | XYZ999-001 | 1 | SCREEN | 2013-02-14 | 2013-02-21 | SCREENING |
| XYZ999 | SE | XYZ999-002 | 1 | SCREEN | 2013-01-27 | 2013-03-02 | SCREENING |
| XYZ999 | SE | XYZ999-002 | 2 | DRUG B | 2013-03-02 | 2013-03-04 | TREATMENT 1 |
| XYZ999 | SE | XYZ999-003 | 1 | SCREEN | 2013-02-27 | 2013-03-22 | SCREENING |
| XYZ999 | SE | XYZ999-003 | 2 | DRUG A | 2013-03-22 | 2013-03-29 | TREATMENT 1 |
| XYZ999 | SE | XYZ999-003 | 3 | DRUG B | 2013-03-29 | 2013-04-05 | TREATMENT 2 |
| XYZ999 | SE | XYZ999-003 | 4 | FOLLOWUP | 2013-04-05 | 2013-04-22 | FOLLOW-UP |

Table 3: SE Example

Subject XYZ999-001 was a screen failure and thus, only a record for the screening Element was created in the SE data set above. Subject XYZ999-002 was assigned to sequence 'BA' (ARMCD = 'BA') but withdrew consent after the period 1 dosing. Only records for screening and the Drug B Elements are present. Lastly, Subject XYZ999-003 was randomized to treatment sequence 'AB' (ARMCD = 'AB') and progressed through the trial as planned.

CHALLENGES IN SE DEVELOPMENT

In order to ensure that an accurate SE domain for a particular trial is created, the Trial Design author will need to take into account where the specific dates would be found in the subject-level SDTM domains that match the TESTRLs and TEENRLs as defined in TE. When a TESTRL for an Element is defined, it is important that it point to a specific date that is planned to be collected, since SESTDTC is the date that a subject actually entered an Element and thus, needs to be present in the data. This is often a challenge when the Trial Design domains are being created prospectively from the protocol before any data is collected for the study and any SDTM domains exist. If it is found during the development of the SE domain that any of the rules in TE do not point to a collected date for the majority of subjects, the rule may need to be revised. For example, if the informed consent date is not collected on the CRF, the TESTRL for ELEMENT = 'Screening' should not be 'Informed consent obtained' since this particular date will not be available in the data for any subject. A more appropriate TESTRL might be 'Earliest date of screening assessments'.

Another challenge in creating SE is when a subject does not progress through the Elements as planned for their treatment arm due to withdrawal of consent, death, lack of efficacy, etc. Since the TESTRL and TEENRL will only include the planned beginning and end of Elements, it is not clear what date to use as the end of a subject's last Element when there is no next Element in SE. Some sponsors try to 'fix' this issue by including every possible unplanned condition that can occur in TEENRL. There are also times when the text in TESTRL and TEENRL is not clear enough to the programmer to be able to determine the domains and specific variables from which to derive values for SESTDTC and SEENDTC. These issues can be solved by adding pseudo-code as further mapping instructions for each TESTRL and TEENRL that would include which specific data point to use. Table 4 below shows the TE table for the example trial with columns added that contain pseudo-code for each TESTRL and TEENRL to aid in SE development. Please note that these columns would only exist in a mapping specification and are dropped in the final TE data set.

| ETCD | TESTRL | <i>Pseudo-code for SESTDTC</i> | TEENRL | <i>Pseudo-code for SEENDTC</i> |
|----------|---|--|----------------------------|---|
| SCREEN | Informed consent obtained | DM.RFICDTC (or DS.DSSTDTC where DSDECOD = 'INFORMED CONSENT OBTAINED') | First dose of treatment | If subject discontinues, then DM.RFPENDTC (or max(SV.SVENDTC)) else SESTDTC of subject's next element |
| A | First dose of Drug A | min(EX.EXSTDTC) where EXTRT = 'DRUG A' | 1 week after element start | If subject discontinues, then DM.RFPENDTC (or max(SV.SVENDTC)) else SESTDTC of subject's next element |
| B | First dose of Drug B | min(EX.EXSTDTC) where EXTRT = 'DRUG B' | 1 week after element start | If subject discontinues, then DM.RFPENDTC (or max(SV.SVENDTC)) else SESTDTC of subject's next element |
| FOLLOWUP | Last PK sample obtained following period 2 dosing | max(PC.PCDTC) per subject where PC.VISIT = 'PERIOD 2' (VISITNUM = 3) | Last Contact with Subject | DM.RFPENDTC |

Table 4. TE Table with Pseudo-Code

STREAMLINING THE PROCESS

Development of SE can sometimes seem overwhelming to a developer, since there can be many conditions that need to be accounted for in order to create an accurate representation of a subject's progression through a trial. The SE domain can take longer to program if the developer chooses to re-derive the specific dates from domains such as EX and DS when these dates have already been derived in the DM domain for RFICDTC, RFSTDTC, RFENDTC, DTHDTC, and so on. For example, DM.RFICDTC is typically derived from DS.DSSTDTC where DSDECOD = 'INFORMED CONSENT OBTAINED'. Rather than doing this same derivation again for the screening Element in SE, it is more efficient to just populate SESTDTC from DM.RFICDTC.

Another instance when SE can take more time to program is when the code includes the conditions to derive each

instance of SESTDTC and SEENDTC for a subject that is going through the trial as planned in TE and TA. Using this strategy could lead to errors that result in the SEENDTC not being the same as the SESTDTC of the next Element, especially in the case of an ongoing trial where the SE code may need to be updated. The developer may update the code properly in SESTDTC for an Element but neglect to update it for the previous Element's SEENDTC.

As mentioned before, since there are no gaps between Elements, and the start of one Element (SESTDTC) is the same as the end of the previous Element (SEENDTC), it would be ideal to be able to apply this definition to the actual development of SE. When writing the code for the SE domain, the programmer can include all the conditions for SESTDTC based on the pseudo-code. Then for SEENDTC, only those conditions that are applicable if a subject discontinues, some other unplanned event occurs, or if it is the last Element can be added. From a programming perspective, this is possible because the SESTDTC derived for one SE record can be used to populate the SEENDTC for the previous record for that subject. If this simple coding strategy is used, this eliminates the possibility of any gaps from the end of one Element to the start of the next. It also reduces the amount of time needed to develop SE because the programmer would only have to account for the conditions for SESTDTC once in the code.

Since there is not one specific tool that is used to create SDTM datasets, the methods used within each specific tool to apply this strategy will vary but the idea remains the same. One example of how this can be done uses SQL language but the same can be done using others.

Sample code for SEENDTC:

```
CASE
  WHEN SEENDTC IS NOT NULL AND SESTDTC IS NOT NULL THEN SEENDTC
  ELSE LEAD(SESTDTC,1) OVER (PARTITION BY USUBJID ORDER BY USUBJID, SESEQ)
END AS SEENDTC
```

In SQL, the LEAD function (in the highlighted text above) will go to the record after the current record in the table that is specified. In this example, it is the next record since LEAD(SESTDTC,1). It is also specified that when SEENDTC is null, then to go to the next record's (or next Element's) SESTDTC to populate SEENDTC for the current record. In the query code for each Element that follows this portion of a header statement, when a subject meets the condition in the pseudo-code for SESTDTC for the next Element, SEENDTC will be null and then the LEAD function will populate the SEENDTC for that Element with the SESTDTC from the next. In instances when a subject discontinues and/or does not enter any new Elements, these are the conditions that should be written into the code for SEENDTC based on the pseudo-code for each Element.

CONCLUSION

Not only is it essential to include SE and SV in addition to the Trial Design domains in a submission to aid in the review, it is also important that they are developed properly so as to provide an accurate representation of the study design and what actually occurred during the trial. Some of the challenges associated with developing some of these domains can be alleviated by looking for solutions that may be a simple application of the principles already outlined in the SDTM.

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. Wood, F., Lenzen, M. (2011) Trials and Tribulations of SDTM Trial Design. PharmaSUG, May 2011.

CONTACT INFORMATION

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

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

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

Name: Jerry Salyers
Title: Senior Consultant, Data Standards Consulting
Enterprise: Accenture Life Sciences
Address: 1160 W. Swedesford Rd, Bldg One
City, State ZIP: Berwyn, PA 19312
E-mail: jerry.j.salyers@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.