Trials and Tribulations of SDTM Trial Design

Fred Wood, Octagon Research Solutions, Wayne, PA
Mary Lenzen, Octagon Research Solutions, Wayne, PA

ABSTRACT

An increasing number of sponsors are submitting clinical trials data to the FDA in the format of the CDISC Study Data Tabulation Model (SDTM). The Trial Design Model (TDM) datasets can be especially challenging because, in most cases, they are being created retrospectively from the protocol, and cannot be created from electronic data.

This presentation will discuss some of the considerations and challenges in creating the TDM datasets, using case studies of both relatively simple and more-complex trials. Included will be practical advice on which datasets should be created first, which datasets drive the creation of others, and how the trial-level and subject-level datasets relate to each other. The presentation will conclude with a list of resources for TDM dataset creation.

INTRODUCTION

The Trial Design Model, described in the SDTM and the SDTM Implementation Guide (SDTMIG), provides a standardized way to describe those aspects of the planned conduct of a clinical trial. The purpose of standardizing the plan from the protocol is to allow reviewers (and others) to do the following:

- clearly and quickly grasp the design of a clinical trial
- understand the rules for how subjects transition through various periods or phases of the trial
- compare the designs of different trials
- search a data warehouse for clinical trials with certain features
- compare planned and actual treatment paths and visits for subjects in a clinical trial.

Ideally, TDM data should be created before the trial begins; however, that's not usually the case for most studies. There are two major reasons for this:
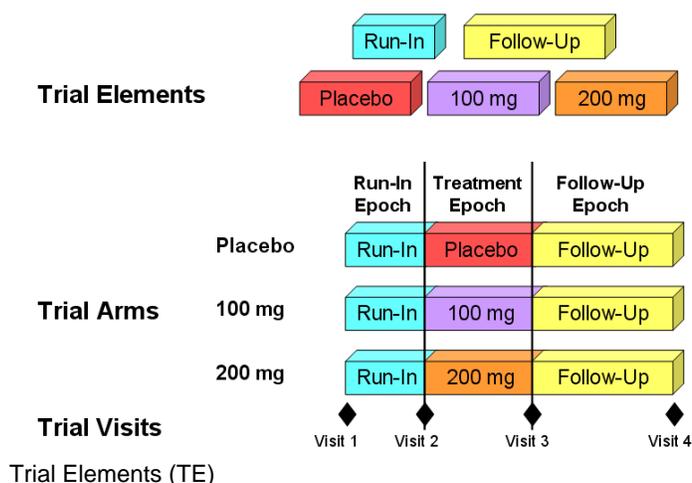
- Most sponsors still rely on paper representations of the protocol, and don't have an electronic format that can be used through all the phases of the data lifecycle
- The fact that an electronic representation of the trial design, such as that described in the TDM, would be part of submission standard was not known before many trials began.

Creating the tables retrospectively presents a number of challenges, which will be discussed later in this paper. First, some TDM basics will be discussed.

TRIAL DESIGN BASICS

The TDM is built upon the concepts of Elements, Arms, Epochs, and Visits. They are shown in Figure 1, and are described in the following sections.

Figure 1. Key Concepts of Trial Design

Trial Elements (TE)

An Element is the basic building block that's related to the administration of planned interventions. Periods of time over which there are no interventions (such as screening or washout) are also considered Elements. Planned Elements are described in the Trial Elements (TE) table. An example for the trial above is shown in the following table.

STUDYID	DOMAIN	ETCD	ELEMENT	TESTRL	TEENRL	TEDUR
1999001	TE	RUNIN	Run-in	First dose of run-in drug	First dose of treatment	P1W
1999001	TE	PLAC	Placebo	First dose of placebo	2 weeks after element start	P2W
1999001	TE	100MG	100 mg	First dose of 100 mg	2 weeks after element start	P2W
1999001	TE	200MG	200 mg	First dose of 200 mg	2 weeks after element start	P2W
1999001	TE	FOLLOW	Follow Up	Last dose of treatment	3 wks after element start	P3W

Trial Arms (TA)
An Arm is a planned sequence of Elements, 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 and Transitions. The Demographics domain contains the short and long names of the Arm to which each subject was randomized. Rows of the TA dataset for the first Arm are shown in the following table.

STUDYID	DOMAIN	ARMCD	ARM	TAETORD	ETCD	TABRANCH	EPOCH
1999001	TA	PLAC	Placebo	1	RUNIN		Run-In
1999001	TA	PLAC	Placebo	2	PLAC		Treatment
1999001	TA	PLAC	Placebo	3	FOLLOW		Follow-Up

Trial Visits (TV)
A Visit is defined as a clinical encounter that encompasses planned and unplanned trial interventions, procedures, and assessments that may be performed on a subject. Planned visits are described in the Trial Visits (TV) table. For the trial above, it would look like the table below. VISITDY and ARM are Permissible, and are not shown here. ARMCD is Expected, but only needs to be populated if the visit schedule varies by Arm.

STUDYID	DOMAIN	VISITNUM	VISIT	ARMCD	TVSTRL	TVENRL
1999001	TV	1	Visit 1		Clinic Admission	Completion of dosing
1999001	TV	2	Visit 2		1 week after dosing of run-in drug	
1999001	TV	3	Visit 3		2 weeks after dosing treatment drug	30 minutes after receipt of blinded treatment
1999001	TV	4	Visit 4		3 weeks after last doses of treatment	

The term Epoch is used to describe a period of time that cuts across Arms, and applies to the trial as a whole. Epochs may contain the same number or a different number of Elements for each Arm. Epochs are defined in the TA table. In the study above the Epochs are Run-In, Treatment, and Follow-Up.

The TDM also includes the Trial Inclusion/Exclusion (TI) and Trial Summary (TS) datasets. The TI table is used to describe the protocol-specified inclusion/exclusion criteria used for all subjects. This is in contrast to the IE domain, which contains the subject-specific exceptions to those criteria (i.e., criteria not met) for subjects included in the submission data. TS is used to submit trial-level summary information, examples of which include the blinding schema, drug indication, and trial objectives.

In contrast to TE and TV, which include the planned Elements and planned Visits, there are two subject-level tables that describe the actual Element sequence (Subject Elements) and visit sequence (Subject Visits). These tables include the actual date/times for the start and end of each Element or visit, and are flexible enough to handle both planned and unplanned Elements or Visits.

CHALLENGES IN CREATING TDM TABLES

The prospective creation of the TDM tables presents fewer challenges than the retrospective creation, and this paper focuses primarily on the latter. One of the major challenges of retrospective creation is actually finding the information needed for an accurate representation of the study. In many cases, not all of the necessary information is in one place, requiring an examination of the CRFs and the actual data, in addition what often amounts to an extensive review of the protocol and any amendments. Because the subject data already exist in the retrospective creation, one needs to ensure that the "plans" represented in TA, TE, and TV support, and are consistent with, the data in Subject Elements and Subject Visits. In other words, one needs to ensure that the interpretation of the protocol by the individuals executing the study is consistent with the data. While such inconsistencies may often be the result of protocol violations, when they are systematic across all subjects and multiple domains, they are usually the result of improper implementation of the protocol.

Creating TDM tables can be as much of an art as it is a science. Two or more individuals may read a protocol and design the TA, TE, and, to a lesser extent, the TV datasets differently. For example, one may interpret a dosing period as a series of

Elements, one per dosing cycle, while another may interpret this treatment period as a single Element. Knowledge of how the data will be analyzed can often influence which of these methods should be followed. Sometimes, however, there is more than one valid, defensible interpretation. Sometimes a consensus approach is best in determining which of the correct methods is the best. Incorrect ways to model trials are usually much easier to exclude.

CREATING TRIAL ELEMENTS

The Trial Elements (TE) dataset provides descriptions of the building blocks that make up the study, including what event indicates their start, what event indicates their end, and their duration (if they have a fixed length). This dataset is usually created before TA, since it defines the components used in the Trial Arms dataset. It's also usually created before Trial Visits, because Visit start and end rules are often described in relation to the TE Elements. Some examples of common Elements are "Screening", "Baseline", "Follow-up", as well as study-specific dosing Elements. The values for ELEMENT should be as consistent as possible with the naming in the protocol.

It is important to note that TE is not simply a duplication of data in the Exposure (EX) dataset. For example, a chemotherapy regimen composed of three concurrently administered drugs is not three different Elements but one treatment Element. The details of the dosing for the three individual drugs are in the EX (Exposure) dataset. It also cannot be stressed enough that these blocks are meant to be described independently of their location(s) within the Arm(s) of the study.

Elements cannot have gaps between them, nor can they overlap. A subject must always be in an Element, and cannot be in two at the same time. Therefore, with consecutive Elements, the end of one must always be the same as the start of the next. The Start Rules and End Rules must be developed in accordance with these concepts. Since the start of one Element in an Arm is the signal for the previous Element to end, the Start Rules are much more important than the End Rules. Both types of rules should be developed without any reference to where they appear in any Arm. If an Element appears in more than one Arm, then the rules cannot reference the Arm.

Some Element Start Rules are described as "active" and some are described as "passive." Active Start Rules, (e.g., dosing Elements) are much more clearly defined than passive start rules (e.g., washout Elements). As an illustration, consider that the Start Rule of a dosing Element is simply the date/time of the first dose for the treatment that defines the Element. It becomes much more difficult to describe the Start Rule for a wash-out period: when does it really begin?

One should approach the creation of TE with the intent of making it as simple as possible. Some simple rules to follow:
- Most studies have a screening period, and hence a Screening Epoch. Since all studies require the signing of an informed consent (which is when a subject is "officially" part of a trial), its date/time can almost always be used as the Start Rule for the Screening Element.
- The start of a treatment Element is the date/time of the first administration of that treatment.
- The end of a treatment Element may not be the last date of the treatment, but rather the beginning of the next Element.
- The end of the last Element usually coincides with the date/time the subject completes the study.

The Trial Elements Start Rules and End Rules play critical roles in the creation of the Subject Elements dataset. Not only should one have a text description for the start rule, as in TESTRL, but one will need some type of pseudo-code that allows a programmer who may not be familiar with the trial to find the actual data point described in the individual Element start rule. For example, if a Start Rule states "First dose of Drug A", pseudo-code might be something such as "Select min(EX.EXSTDTC) where EXTRT = DRUGA".

CREATING TRIAL ARMS
The Trial Arms (TA) dataset describes the planned treatment paths (Arms) and the order of the Elements within each of these. Arms commonly include a screening Element and one or more treatment Elements. The latter may be separated by washout Elements, and sometimes there might be a follow-up Element. Any Elements described in TA must appear in TE.

Trial Arms includes two variables named TABRANCH (Branch) and TATRANS (Transition Rule). The distinction between these variables is not apparent from the variable names and labels. TABRANCH describes a branching rule at the end of an Element that determines where the subject goes next. It might be conditional (e.g., a subject meets certain criteria) or predetermined (e.g., a set time for re-randomization). TATRANS is used to describe the conditions under which subjects may skip one or more Elements, and to which Element would go next.

The values of ARM/ARMCD in TA should be used to populate the corresponding variables in the Demographics (DM) dataset for all randomized subjects. The values in the ARM variable should be as consistent as possible with the naming in the protocol.

CREATING TRIAL VISITS
The Trial Visits (TV) dataset describes the planned schedule of Visits. Visits are defined as "clinical encounters" and are described in the SDTM/SDTMIG using the timing variables, VISITNUM, VISIT, and VISITDY. The concept of visits is widely accepted. The challenge the TDM brings to legacy data conversions is the introduction of Visit Start Rules (TVSTRL) and End

Rules (TVENRL). These may have existed as offsets in some clinical data management systems, but were probably never extracted for submission. Some sponsors may never have thought about what a Visit Start Rule might be. Visits are usually a collection of activities that might include dispensation of study treatment and certain measurements such as vital signs and laboratory tests. Most protocols don't require that these activities are done in a certain order so there is usually no "start" or "end" of a visit. As a result, start rules may often be as general as "Start of assessments on Study Day 14" and end rules may often be as general as "Final assessment on Day 15." End Rules are not needed for single-day visits.

While the Element Start Rules and End Rules are used to create the Subject Elements table, the rules in TV do not drive the creation of Subject Visits. If SDTM data exist, the Subject Visits dataset can be created from the visit-based SDTM datasets. All unique instances of VISITNUM (and corresponding VISIT and VISITDY values, if desired) would be used to populate the corresponding columns of SV. SVSTDTC and SVENDTC would be populated with the minimum and maximum date values for all activities at each visit across all domains.

CREATING TRIAL INCLUSION
Trial Inclusion/Exclusion (TI) describes the inclusion/exclusion criteria used to screen subjects. Its structure is one record per criteria per trial. The biggest challenge is the 200-character limitation of SAS v5 transport files. Truncation to 200 characters could result in duplicate IETEST values if the first 200 characters of multiple criteria are not unique. The SDTMIG recommends that if the text is > 200 characters, put meaningful text in IETEST and describe the full text in the study metadata. It is expected that a reviewer will still need to look at the protocol and/or CRF for the complete text. The values of IETEST/IETESTCD in the subject-level IE (Inclusion/Exclusion Exceptions) dataset must come from those in the TI dataset.

CREATING TRIAL SUMMARY
Trial Summary (TS) includes key details, called Parameters, about the trial that would be useful in clinical trials registry. Examples include the title, objectives, minimum age, maximum age, and indication. CDISC Controlled terminology exists for thirty parameters. The structure is one record per parameter per occurrence. TSSEQ is used as a key to distinguish between multiple occurrences of the same parameter (e.g., each dose would be in a separate record). TSGRPID is used to group multiple parameters together, such as would be needed for Dose, Units, Frequency, and Route in a multiple-dose study.

BEST PRACTICES
The overall best practice for Trial Design is to keep it as simple as possible and present the overall picture of the trial. The selection of Elements and Epochs should take into account how the data will be analyzed. People often tend to over-complicate the TDM datasets, making them so complex that their usefulness is questionable.

Creation of a Spreadsheet Template
One of the most efficient best practices is to have a Trial Design spreadsheet template. The template should include common controlled terminology for Arms, Elements, and Epochs. Sponsors should strive for as much controlled terminology (study independent) as possible. There should be tabs for each TDM dataset, and the template for each should include all the Required and Expected variables. If possible, global decisions on Permissible variables are recommended. The Trial Elements tab should include columns for the start and end rules pseudo-code, which would be dropped from the submitted Trial Elements dataset. The TI tab should include the convention for IETESTCD values such as IN01 or INCL01; or EX01 or EXCL01. It is also useful to include a column which calculates the word count of IETEST to ensure that the 200-character limit is not exceeded. The TS tab can be populated with the TSPARM and TSPARMCD values that would be used in most studies.

Summary Recommendations
It is recommended to begin with the creation of Trial Summary, since doing so helps achieve familiarity and understanding of the protocol, which is essential for creating Trial Arms and Trial Elements. The next step would be to create Trial Elements, followed by Trial Arms and Trial Visits. Sponsors should use or develop controlled terminology wherever possible.

CONCLUSION
TDM datasets are very useful for rapidly understanding the design of a study, even before reading the protocol. The datasets are standard and relatively simple structures. They tend to have relatively small numbers of rows, and are easily comprehended. Not only are they useful for FDA reviewers, but they can be very valuable for internal use. Currently, most protocols are only maintained in the form of documents, and these documents are not centrally accessible or searchable. At a minimum, creation of TDM datasets allows sponsors to have this information in a centrally accessible and searchable format.

RESOURCES

- Study Data Tabulation Model. Clinical Data Interchange Standards Consortium (CDISC) Submission Data Standards (SDS) Team. Version 1.2. November 2008.
- Study Data Tabulation Model Implementation Guide: Human Clinical Trials. Clinical Data Interchange Standards Consortium (CDISC) Submission Data Standards (SDS) Team. Version 3.1.2. November 2008.

In the Members Only section on the CDISC website (http://www.cdisc.org) under "Case Studies" and then "Case Studies" there are several examples posted under the heading "Trial Design Examples":

The Trial Design Team has completed TDM Examples for our members. The following documents are posted there:

- TDM Examples Introduction.doc explains why we're publishing examples on the CDISC website.
- Trial/Study Design Template with Best Practices.xls is the template for Trial/Study Design datasets. The tab "ReadMe" contains information and major changes since the last version was posted. The tabs "Best Practices" and "Hints and Tips" contain advice on creating Trial Design datasets and use of the template.
- TQTc spreadsheet is a Phase 1 "thorough" QTc study. This example is a parallel trial, placebo controlled and positive (moxifloxacin) controlled with therapeutic (2 mg) and supra-therapeutic (20 mg) dosing.
- AGR spreadsheet is a Phase 1 example of an age, gender and race study with a single dose and a multiple dose treatment period.
- RTOG 93-09 spreadsheet and corresponding RT0G 93-09 protocol is an example of a Phase 3 oncology study. This study consists of 2 Arms: 1) CRS includes surgery and 2) CR does not include surgery.
- Antifungal Drug Interaction.xls is a Phase 1, open-label, randomized, pharmacokinetic drug interaction study. The primary analytic objective of this study is a comparison of the PK profile of Drug A alone (Treatment Period 1) against the PK profile of Drug A + either fluconazole or ketoconazole (Treatment Period 2).
- EGF100151 Consolidated.xls is a Phase III Randomized, Open-label Trial for Women with Refractory Advanced or Metastatic Breast Cancer. It is two treatment Arms. Treatment will be administered until disease progression or withdrawal from study drug due to unacceptable toxicity or other reasons (e.g., consent withdrawal, non-compliance, etc.)

ACKNOWLEDGMENTS
The authors wish to thank the following:
- Our colleague, Richard Lewis, for his contributions to this paper.
- The Senior Management Team at Octagon Research Solutions, who have long been advocates for CDISC standards development.

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

Fred Wood
Vice President, Data Standards Consulting
Octagon Research Solutions, Inc.
585 East Swedesford Road, Suite 200
Wayne, PA 19087
Work Phone: 484-881-2297
E-mail: fwood@octagonresearch.com

Mary Lenzen
Principal Consultant
Data Standards Consulting
Octagon Research Solutions, Inc.
585 East Swedesford Road, Suite 200
Wayne, PA 19087
Work Phone: 484-881-2295
E-mail: mlenzen@octagonresearch.com