The Standard for the Exchange of Nonclinical Data (SEND): History, Basics, and Comparisons with Clinical Data

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

ABSTRACT

The CDISC Standard for the Exchange of Nonclinical Data (SEND) Implementation Guide (SENDIG) contains domains for general toxicology, pharmacology, and carcinogenicity studies. A separate implementation guide (SEND-DART) contains domains for reproductive toxicology studies. The first SEND Model was developed in 2002, utilizing domains described in the CDER 1999 Guidance. In 2007, an effort began to completely align the SENDIG with the SDTM Implementation Guide (SDTMIG), with the first such version (v3.0) published in 2011. Since that time, the SEND Team has been working to create more examples, clarify existing text and examples, and add new domains. Version 3.1, which actually underwent two public reviews (in 2014 and 2015), is expected to be posted in Q2 of this year (2016). This paper provides an overview of the history of SEND and its close ties with the development of the SDTM and the SDTMIG. It also covers some of the basics of the SEND model and how the nonclinical implementation of the SDTM compares with the clinical implementation.

INTRODUCTION

NOTE TO READERS: In an attempt to make wading through the alphabet soup of acronyms easier, some of the more frequently used ones are listed at the end of this paper.

SEND BACKGROUND

This section presents a brief background and history of the Standard for the Exchange of Nonclinical Data (SEND). For a more-detailed historical background of SEND, see Wood and Kramer (1).

Relationship to the Study Data Tabulation Model (SDTM)

The SDTM is the underlying model for five implementation guides (IGs), of which the SEND Implementation Guide (SENDIG) is one. For background and history on the SDTM and SDTMIG, see Wood (2) and Wood and Guinter (3). The development of the SEND standard began in late 2002, and paralleled the development of human-clinical data standards, known as the Submission Data Standards (SDS). The SDS (v1.x and v2.x) was the precursor to the SDTM and the SDTM Implementation Guide (SDTMIG), which began as a v3.0 standard. The initial SEND work product (v1.x through v2.3, 2002-2007) was known simply as SEND. While the name of the principal document changed to the SENDIG (v.3.0, 2011 and its drafts) to parallel what occurred with the SDTMIG, the terms are often used synonymously when referring to the standard.

The Development of SEND

The SEND Team was formed in the first quarter of 2003, starting as the "SEND Consortium", and did not become a CDISC team until several years later. The submission of datasets for the first SEND pilot was essentially completed by the end of 2004. Input from the pilot, as well as efforts to more closely align this implementation with that for human clinical trials, resulted in the Version 2.x standards. The last of these was v2.3, posted in November 2005, at which time the SEND initiative began to lose momentum. As was the case with the SDTM/SDTMIG, many companies were unwilling to commit resources without some type of public statement from the FDA regarding the future of SEND.

There was a resurgence of SEND activity in April and May of 2007. This came from two fronts: 1) there was an industry effort, led largely by people from Lilly and Covance, to develop and improve the SEND model for use as a data-transfer standard (vendor to sponsor), and 2) there was renewed interest from the FDA, not only from CDER but also from the Office of the Commissioner, for its use in submissions. The word spread, and a re-formed SEND Team held its first full-team face-to-face meeting in Washington, DC in September of 2007. This meeting had a number of purposes, including the following:

- Get SEND moving again.
- Improve consistency with the SDTM and the SDTMIG. This was at a time when work was being done by the SDS Team to improve upon SDTM v1.1 and SDTMIG v3.1.1. SEND also needed to be able to address more

complicated trial designs and create a Demographics domain, the data for which had been previously represented only in Subject Characteristics.

- Begin the development of specifications for a regulatory (live submission) pilot, the announcement of which came in October. The pilot was to involve up to eight sponsor companies who would be submitting nonclinical data to an IND (Investigational New Drug) or NDA (New Drug Application) in both the current PDF format as well as the electronic SEND format.
- Continue the development of domains for safety/pharmacology studies.
- Continue the efforts begun in early 2005 to develop domains for reproductive toxicity studies. These studies can be complex due to 1) the staggered timing of the phases of gestation and weaning within treatment groups, resulting in event-based data collection, and 2) the relationships that need to be maintained between mating partners and between parents and offspring, possibly through multiple generations.
- Continue momentum for the development of comprehensive sets of standard terminology, leveraging the work from the CDISC Controlled Terminology Team, and the National Cancer Institute's Enterprise Vocabulary Services (NCI EVS).

Each of these responsibilities was assigned to a subteam with a designated leader. After the 2007 meeting, the SEND Team saw its active membership increase dramatically, with tremendous commitment from pharmaceutical companies, vendors, CROs, and FDA representatives. Membership took another leap in 2015 after the FDA publication of the guidance documents on electronic submissions (4) and standardized electronic data (5) in December of 2014. The SEND Team has had three or four 4- to 5-day face-to-face meetings each year since 2007.

Throughout 2008, the SEND Team worked on creating what was called Version 3.0 Draft A of the SEND Implementation Guide (SENDIG), intended to be used as a specification for the CDER Pilot described above. In March 2009, it was posted for public awareness. Continuing improvements and new domains led to work on developing Version 3.0 Draft B, to be used primarily as a working draft for team review, in early 2010. In December 2010, the first version intended for production use (Version 3.0) was posted as a draft for public review and comment.

After the final production version was published in May 2011, work then focused on improvements and additions that would be incorporated into v3.1. A draft version for public comment was posted in October 2014. This version included revised domain models to accommodate new variables (Microscopic Findings, ECG Test Results) or modified scope (Vital Signs)) as well as new domains (Cardiovascular Test Results, Respiratory Test Results). The SEND Team spent the early part of 2015 addressing public comments.

As a result of public comments around the use of VISITDY and VISIT, the SEND Team decided to phase out these variables, and replace them with --NOMDY (Nominal Study Day for Tabulations) and --NOMLBL (Label for Nominal Study Day), respectively. The change was made because 1) the SENDIG v3.0 has two slightly different descriptions of the Visit variables, and 2) the concept of "visits" from the SDTM and SDTMIG, developed for clinical-trials data, didn't really apply to nonclinical data. More details are provided in the Special-Purpose Domains section of this paper. These variable changes were considered significant enough to warrant a second public-review period. The additional time allowed a few other additions to be made prior to the second public-review posting in July 2015. That posting resulted in a new set of public comments that needed to be address/resolved, and another series of reviews/approvals from the CDISC Standards Review Council (SRC). The SENDIG v3.1 is expected to be published in Q2 of this year.

The SENDIG-DART (Developmental and Reproductive Toxicology)

As mentioned above, the SEND Reproductive Toxicology Subteam has met regularly since 2005, through the period when work on the SENDIG as we know it today was suspended. Many of the original subteam members continue to work on DART domains. Because of the specialized nature of DART studies, it was decided that any implementation guidelines would be published as a separate document. The SENDIG-DART v1.0 (to be published in Q2 2016) contains domains needed for embryo-fetal development, as well as concepts required for all DART studies. Included in the latter are Trials Stages (TT) and Trial Paths (TP), along with Subject Stages (SJ). These domains have parallels with Trial Elements, Trial Arms, and Subject Elements, except they describe the planned and actual sequence of Reproductive Phases (e.g., mating, gestation, lactation, and weaning), whose timing is often independent of treatment. In addition, new Timing variables were created for the days within a Reproductive Phase, since the timing of observations is often represented with respect to the Phase just as much as with respect to study start.

SIMILARITIES TO THE SDTMIG

- At a high level, the SENDIG is laid out in the same way as the SDTMIG, with introductory sections (Sections 1-3) and sections on assumptions for all domains (Section 4), subject-related special-purpose domains (Section

5), general-observation-class domains (Sections 6.1, 6.2, and 6.3), trial design (Section 7), and relationships (Section 8).

- Within the relevant sections, domain metadata models are displayed in exactly the same format, with columns for variable name, variable label, data type, controlled terminology or format, and role. Likewise, there are the same columns with information to sponsors: CDISC Notes, and Core. Assumptions and examples follow each model.

- Domains the current versions of the SENDIG (v3.0) and SDTMIG (v3.2) have in common include the following:

• Demographics	• ECG Test Results	• Vital Signs
• Comments	• Laboratory Test Results	• Trial Elements
• Subject Elements	• Microscopic Findings	• Trial Arms
• Exposure	• PK Concentrations	• Trial Sets
• Disposition	• PK Parameters	• Trial Summary
• Death Diagnosis	• Subject Characteristics	

  It should be noted, however, that the assumptions are often unique for each domain within the respective IG, and that the examples are always IG specific.

- As with the SDTMIG, the SENDIG utilizes an extensive set of controlled terminology that is maintained by CDISC and NCI EVS. While its list is maintained separately from the SDTM terminology, the SEND terminology includes 31 codelists that were initially developed for clinical data. Conversely, the SDTM terminology includes three codelists initially developed by the SEND CT Subteam. Efforts to harmonize wherever possible continue.

DIFFERENCES FROM THE SDTMIG

GENERAL OBSERVATION CLASS DOMAINS

- The Interventions general-observation class has only one modeled domain, Exposure. Domains not modeled in the SENDIG include Concomitant Medications, Exposure as Collected, Procedures, and Substance Use, since these do not apply to toxicology studies. Concomitant Medications and Procedures may be relevant if and when the SEND format is accepted for other types of nonclinical studies, as well as studies supporting veterinary medicines.

- The Events general-observation class has only one modeled domain, Disposition. The Adverse Events, Deviations, Clinical Events, Healthcare Encounters, and Medical History have not been modeled in the SENDIG. It is envisioned that these domains could be used in other types of studies, particularly those for veterinary medicines and those conducted under The Animal Rule. The Animal Rule (for which there is a Guidance document) permits FDA to grant a marketing application based upon studies in animals when 1) studies in humans are not ethical and feasible, and 2) results of the animal studies establish that the drug is reasonably likely to produce clinical benefit in humans. Candidates for this rule include drugs developed to treat or prevent serious or life-threatening conditions caused by exposure to lethal or permanently disabling toxic substances.

- The Findings general-observation class contains a number of SEND-specific domains that do not appear in the SDTMIG, and are unlikely to do so in the future. The SDTMIG has a number of domains that don't appear in the SENDIG, some of which will probably not be used for nonclinical studies.

- The domains specific to each IG are shown in the table below.

SENDIG Only	SDTMIG Only	
• Body Weights • Clinical Observations • Food and Water Consumption • Macroscopic Findings • Palpable Masses • Organ Measurements • Tumor Findings	• Adverse Events • Clinical Events • Concomitant Medications • Deviations • Disease Response • Drug Accountability • Healthcare Encounters • Immunogenicity Specimen • Inclusion Exclusion Exceptions • Medical History	• Microbiology Specimen • Microbiology Susceptibility • Physical Exam • Procedures • Questionnaire • Skin Response • Subject Status • Substance Use • Tumor Identification • Tumor Measurements

FINDINGS DOMAIN CONVENTIONS

With the exception of the differences noted below, the conventions followed by clinical implementations are followed for nonclinical data.

Location Variables

SEND development has been the beneficiary of the development of the BRIDG (Biomedical Research Integrated Domain Group) Model, which was created to help with the alignment of CDISC research concepts with those of HL7 (Health Level Seven). The BRIDG was developed to allow for semantic interoperability (the ability for systems to talk with each other unambiguously). For this to happen, concepts in the standards must defined unambiguously. An example of where there has been noted confusion in the SDTM and SDTMIG is the variable, --LOC (Location). In Findings domains, --LOC can represent either a target location (BRIDG PerformedObservation) or a result location (BRIDG PerformedObservationResult). One looking at the data would, for example, not be able to tell whether a value of STOMACH was the focal point of the exam ("Examine the stomach"), or whether the GI Tract was the focus, and something was found in the stomach ("upon examining the GI tract, lesions were noted in the stomach"). Even more problematic is the situation in which there is a target location (e.g., chest, as in an x-ray), and a result location to report an abnormality in the left lung.

The SENDIG defines the --LOC variable as always being related to the test; the same applies to --LAT (Laterality), and --DIR (Directionality). If more detail is needed for results, additional variables will be created. Names such as --RESLOC (Result Location), --RESLAT (Result Laterality), and --RESDIR (Result Directionality) have been suggested, but no action has been taken for the SENDIG v3.1 or for the SDTM 1.5 upon which it is based.

Standardization of Macroscopic and Microscopic Findings

The SENDIG manages standardization of results in Findings domains according to some specific rules in the MA (Macroscopic Findings) and MI (Microscopic Findings) domains. The overarching rules around standardization of results, described in Section 4.1.5.1 of the SDTMIG and identically in Section 4.5.1.1 of the SENDIG still prevail.

The rules for MA and MI have been developed with pathology experts and with FDA input. The MAORRES and MIORRES variables contain the verbatim description of any findings. MASTRESC and MISTRESC contain only "base processes", without modifiers and/or most adjectives. In addition, there should be only one base process per record, which means that verbatim findings that include two or more base processes must be split into separate records. MISTRESC uses CDISC Controlled Terminology. Any text that was removed from --ORRES in creating the base process in --STRESC appears as standard variables if relevant (--SEV in SENDIG v3.0, with --CHRON [Chronicity] and --DIST [Distribution] added in SENDIG v3.0), or in a Supplemental Qualifier QNAM of --RESMOD if not. The latter was created so reviewers can easily see what text was removed in the standardization process. A flow chart for MI is shown below.
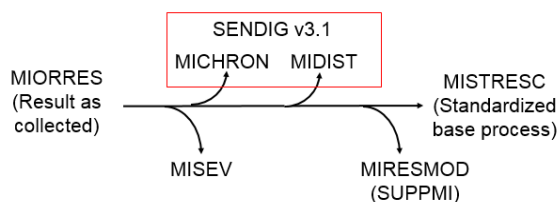
A data example of the standardization process appears in the following table. Note that the oiginal verbatim text contained two base processes that required splitting into two records for standardization. The text, "surrounded by" would appear in as QVAL for the SUPPMI for the QNAM of MIRESMOD.

Row	STUDYID	DOMAIN	USUBJID	MISEQ	MIGRPID	MIREFID	MITESTCD	MITEST
1	2016-1	MI	2016-1-001	12	67	LIV-154	MIEXAM	Microscopic Examination
2	2016-1	MI	2016-1-001	13	67	LIV-154	MIEXAM	Microscopic Examination

Row	MIORRES	MISTRESC	MICHRON	MIDISTR	MISPEC	MISEV
1	Moderate subacute necrosis surrounded by mild diffuse inflammation	NECROSIS	SUBACUTE		LIVER	MODERATE
2	Moderate subacute necrosis surrounded by mild diffuse inflammation	INFLAMMATION		DIFFUSE	LIVER	MILD

SPECIAL-PURPOSE DOMAINS

Subject Visits (SV)

SV is not included in the Special-Purpose Domains, nor is Trial Visits (TV) in the Trial Design section of any version of the SENDIG. Since animals are observed at least daily (often more) in almost all toxicology studies, the concept of documenting these encounters adds little value to the understanding of the data. The Timing variable VISITDY appears in the SENDIG v3.0, but has been replaced in all the domain models of SENDIG v3.1 with what is considered to be a more-relevant variable, --NOMDY (Nominal Study Day), along with a label for this variable, --NOMLBL, as noted previously in the History section above. The implementation of VISITDY is not recommended at this time. Existing implementations of SEND may discontinue or continue use of VISITDY as feasible; conformance to the model is not dependent on whether VISITDY is used or not.

The Pool Definition (POOLDEF) Table

The POOLDEF table appears in the Relationships section of the SENDIG (but not the SDTMIG), and was included in the SDTM v1.3 and later versions. POOLDEF allows for the identification of the individual animals that contributed to data collected for a group or pool of animals. More detail, along with an implementation example for nonclinical data, is provided later in this paper.

The Trial Sets (TX)

TX appears in the Trial Design section of the SENDIG, and was included in the SDTM v1.3 and later versions. TX allows for the subsetting of animals within an Arm (treatment path) as well as for multiple Arms to be "grouped" together. Trial Sets has possible uses in human clinical trials when randomization may occur on factors other than treatment (e.g., subjects who have undergone previous heart surgery vs. those who have not). A clinical example is provided in the next section.

SEND VARIABLES

A number of variables appear in the SDTM as a result of their need in nonclinical studies. These are shown in the table below. The SDTM notes that some of these are not to be used for clinical-trials data. These are designated with an "N" in the last column. Others have no such notation, but their use in human trials is unlikely (designated "U"), while others may have potential applicability to clinical as well as nonclinical data (designated "P"). For more detail on these variables, users should refer to the SDTM v1.5 or the SENDIG v3.1, both of which are expected to be published in Q2 of 2016.

SDTM Table	Variable *	Variable Label	Example Value(s)	Used in Clinical Trials **
2.2.1, Interventions	--UNSCHFL *	Unscheduled Flag	Y or null	N
2.2.2, Events	--UNSCHFL *	Unscheduled Flag	Y or null	N
2.2.3, Findings	--CHRON *	Chronicity of Finding		P
	--DISTR *	Distribution Pattern of Finding		P
	--REPNUM *	Repetition Number		P
	--EXCLFL *	Exclude from Statistics	Y or null	N
	--REASEX	Reason for Exclusion from Statistics	Text	N
	--UNSCHFL *	Unscheduled Flag	Y or null	N
2.2.4, Identifiers	POOLID	Pool Identifier	Text	U
	FOCID *	Focus of Study Specific Interest	Text	P
	--RECID *	Invariant Record Identifier	Text	P
2.2.5, Timing	--DETECT	Time in Days to Detection of Tumor	Needed for the creation of the tumor.xpt file from SEND datasets from carcinogenicity studies	U
	--NOMDY *	Nominal Study Day for Tabulations	Number	N
	--NOMLBL	Label for Nominal Study Day	WEEK 1	N

2.2.6, Demographics	SPECIES	Species	RAT, MOUSE	U
	STRAIN	Strain/Substrain	FISCHER 344, B6C3F1	U
	SBSTRAIN	Strain/Substrain Details	Text description of additional STRAIN details	U
	AGETXT	Age Text	4-6 weeks	U
	SETCD	Set Code	Used to divide or group animals within Arms	P
4.1.1, Related Records; 4.1.2, Supplemental Qualifiers	POOLID	Pool Identifier	Text	U

* Denotes new variables in SDTM v1.5. Unless noted, variables shown appeared in SDTM v.1.4.
** See text for full explanation. N = No, U = Unlikely, P = Possible

SENDIG IMPLEMENTATION EXAMPLES FOR SPECIAL-PURPOSE DOMAINS

RELATED RECORDS

As with the SDTMIG, the Related Records (RELREC) dataset is used to represent relationships between records across two (or more) datasets. In the SENDIG, RELREC is frequently needed to relate in-life observations to post mortem findings, or to relate microscopic (histopathology) findings to macroscopic findings noted at necropsy. For example, a mass may be first observed as part of clinical observations, then be measured periodically, observed macroscopically at necropsy, be examined histologically, and then be used in tumor analysis. In this case, the mass would have data in five domains: Clinical Observations, Palpable Masses, Macroscopic Findings, Microscopic Findings, and Tumor Findings. If the same mass was given a single identifier (--SPID value, in this case) across all domains, the representation of such a dataset-to-dataset relationship would look like that in the following table.

STUDYID	RDOMAIN	USUBJID	POOLID*	IDVAR	IDVARVAL	RELTYPE	RELID
2010-001	MA			MASPID		ONE	MASS2
2010-001	MI			MISPID		ONE	MASS2
2010-001	PM			PMSPID		ONE	MASS2
2010-001	CL			CLSPID		ONE	MASS2
2010-001	TF			TFSPID		ONE	MASS2

* POOLID is Permissible, and could be removed if there are no relationships for pooled data.

Some sponsors may choose to use RELREC to represent record-to-record relationships for individual masses in individual animals. In such a case, each mass would have multiple records, as shown in the table below:

STUDYID	RDOMAIN	USUBJID	POOLID	IDVAR	IDVARVAL	RELTYPE *	RELID
2010-001	MA	2001-001-1005		MASPID	25		A
2010-001	MI	2001-001-1005		MIGRPID	GRP 7		A
2010-001	PM	2001-001-1005		PMSPID	2		A
2010-001	CL	2001-001-1005		CLSPID	40		A
2010-001	TF	2001-001-1005		TFSPID	10		A

* RELTYPE is Permissible, and could be removed from SEND submissions if there are no dataset-to-dataset relationships represented in RELREC.

Note that this method could result in a very large RELREC table.

The SEND implementation of RELREC includes the variable, POOLID, which is a Permissible variable. Either USUBJID or POOLID must be populated in all cases where there is not a dataset-to-dataset relationship.

SUPPLEMENTAL QUALIFIERS

The main difference from the SDTMIG is the addition of POOLID as a Permissible variable, as was the case for RELREC. Either USUBLID or POOLID must be populated.

POOL DEFINITION

In some animal studies, it's not possible to get certain types of data for individual subjects. Two examples are the following:

- Studies where animals are group housed, and food and water consumption are measured for a cage of animals
- Studies in small animals where blood samples from more than one animal are needed in order to have a quantity sufficient for analysis at a given point in time.

In domains where these measurements are represented, POOLID is populated and USUBJID is null. The POOLDEF table allows for the identification of individual animals within each identified pool. Whenever the composition of the group of animals used for pooling changes, the group must be assigned a new, unique POOLID. The example below shows three animals that were in Cage 1 comprised a pool at some point (Rows 1-3), and that later Animal 2009-008-002 was removed from the pool (Rows 4 and 5). It also shows that some of these same animals were included in another pool, where PK samples were combined for five animals (Rows 6-10). In the example below, the sponsor used POOLID values that convey some type of meaning (cage data and PK samples), although this is not required, since the values used in POOLID are sponsor defined.

Row	STUDYID	POOLID	USUBJID
1	2009-008	CAGE 1	2009-008-001
2	2009-008	CAGE 1	2009-008-002
3	2009-008	CAGE 1	2009-008-003
4	2009-008	CAGE 1-A	2009-008-001
5	2009-008	CAGE 1-A	2009-008-003
6	2009-008	PK1	2009-008-001
7	2009-008	PK1	2009-008-002
8	2009-008	PK1	2009-008-003
9	2009-008	PK1	2009-008-004
10	2009-008	PK1	2009-008-005

THE TRIAL DESIGN MODEL (TDM)

Overview

SEND uses the SDTM Trial Design Model (TDM) tables of Trial Arms (TA), Trial Elements (TE), and Trial Summary (TS) in the same way the SDTMIG does. As mentioned previously, SEND requires the addition of a Trial Sets (TX) table, and does not utilize the Trial Visits (TV) table. SEND also doesn't use Trial Inclusion/Exclusion (TI), since the screening of animals rarely requires the documentation of inclusion/exclusion exceptions. The presentation of the TDM tables is in a slightly different order from that in the SDTMIG, beginning with TE, then TA and TX. Unique to the SENDIG, there are five example studies for which all three of these tables have been created. Having these makes it easier to see how the tables relate to each other, as well as to the experimental design as a whole. These are followed by TS.

The Trial Sets Concept and Splitting Arms

In SEND, a Trial Set represents the most granular subdivision of all the experimental factors, treatment factors, inherent characteristics, and distinct sponsor designations as specified in the design of the study. The author's previous paper (1) presented the design of a nonclinical study in order to illustrate the concept of Trial Sets in conjunction with the better-known concepts of Elements and Arms. The remainder of this section will show an example of how the Trial Sets concept could be used in a clinical trial. Since knowledge of the Elements and Arms is necessary in order to create Sets, these are presented first for the example clinical trial described in the following paragraphs.

In this study, a certain number of smokers and non-smokers are required to be randomized to each of three treatment groups. These consist of a placebo, Drug A, and Drug B. There will be a Screening Epoch and Treatment Epoch. The Elements for this trial, as would be described in TE, include the following:

The Arms, as would be described in the Trial Arms dataset, are illustrated below.
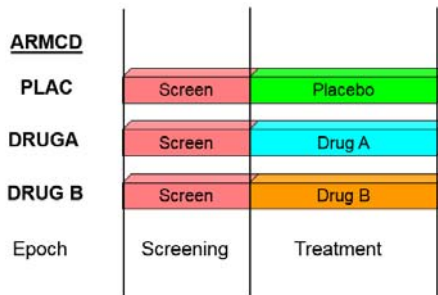
Bringing in the concept of Trial Sets, the experimental design of this study would look like the following diagram.
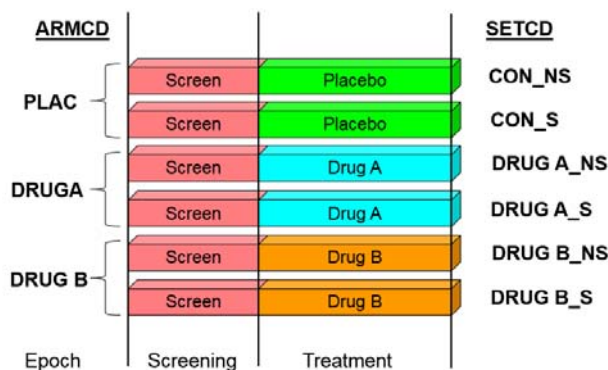
With a Set assignment for each subject (via SETCD in Demographics), it becomes very easy to separate out the smokers from the nonsmokers within each treatment group. Note that since the treatment sequence is not different for smokers and non-smokers, the ARM and ARMCD values are the same regardless of smoking status. How the data are represented in the Trial Sets dataset is shown in the next section.

The Trial Sets (TX) Dataset

The Trial Sets dataset allows for a complete description of all the parameters that make a Set unique by allowing as many rows per Set as there are parameters of interest. TX has a normalized structure, modeled similarly to Trial Summary, with TXPARMCD/TXPARM showing the parameters and TXVAL housing the values. The TX dataset for the study described above might look like the following for the Drug A non-smokers.
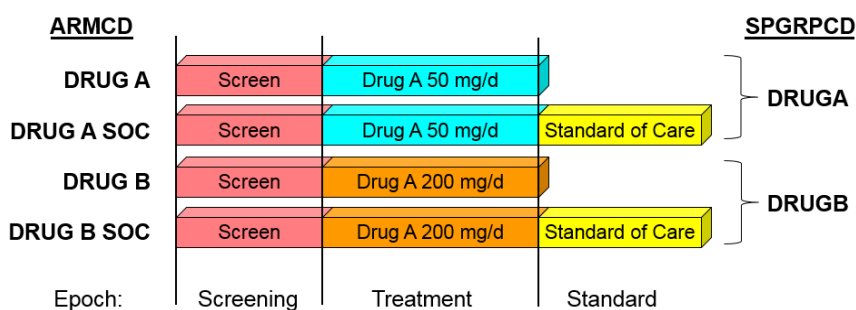
STUDYID	DOMAIN	SETCD	SET	TXSEQ	TXPARMCD	TXPARM	TXVAL
TDM10	TX	DRGA-NS	Drug A Non-Smokers	1	ARMCD	Arm Code	1
TDM10	TX	DRGA-NS	Drug A Non-Smokers	2	GRPLBL	Group Label	Drug A, NS
TDM10	TX	DRGA-NS	Drug A Non-Smokers	3	TRTDOS	Dose Level	200
TDM10	TX	DRGA-NS	Drug A Non-Smokers	4	TRTDOSU	Dose Units	mg/day
TDM10	TX	DRGA-NS	Drug A Non-Smokers	5	POPTYPE	Population Type	Smokers

Using Trial Sets (TX) to Group Arms

While the discussion until now has focused primarily how Arms can be divided into Sets, there are occasions where a sponsor may want, for some analyses, to have certain Arms or Sets grouped together. An example would be the combining of the in-life data from Arms whose only difference is the presence of an additional treatment, follow-up, or other type of period, since prior to the additional period there are no experimental factors that make the Arms unique. In TX, both Arms (TXPARMCD = ARMCD) would be given the same Sponsor Group Code (TXPARMCD = SPGRPCD). Below is an example of what the design and TX rows might look like. Only the relevant rows for TX are

shown. This is only an example, and sponsors 1) may have implemented Trial Arms differently, 2) may decide not to implement Trial Sets, and 3) may decide to do their grouping in the analysis datasets.

STUDYID	DOMAIN	SETCD	SET	TXSEQ	TXPARMCD	TXPARM	TXVAL
2004-1	TX	DRUGA	Drug A	1	ARMCD	Arm Code	DRUG A
2004-1	TX	DRUGA	Drug A	2	SPGRPCD	Sponsor Group Code	A
2004-1	TX	DRUGAREC	Drug A Standard of Care	3	TRTDOS	Arm Code	DRUG A REC
2004-1	TX	DRUGAREC	Drug A Standard of Care	4	TRTDOSU	Sponsor Group Code	A

The Trial Sets: A Final Word

It has been mentioned several times previously in this paper that Trial Sets is designed to represent criteria upon which randomization is based. Trials Sets is not intended to replace any post-randomization groupings or analyses that may be needed to support the safety and efficacy objectives of a study. If there are multiple pre-randomization criteria, then each gets its own Set. If, in the previous example, randomization was not only based upon smoking status, but also upon the naïveté of subjects to treatment(s) for the condition under study, then one could have Sets for 1) Drug A smokers who are naïve, 2) Drug A smokers who had been on previous treatment, 3) Drug A non-smokers who are naïve, and 4) Drug A non-smokers who had been on previous treatment.

THE FUTURE

As mentioned previously, the FDA is requiring all nonclinical and clinical study data to be submitted in a standardized electronic format. This requirement is described in two binding-guidance documents issued by CDER and CBER (4, 5). The details regarding the specific standards to be followed are described in the Study Data Technical Conformance Guide (6) and the Data Standards Catalog (7). These four documents allow CDER to achieve its goals that include the rapid acquisition, analysis, storage, and reporting of regulatory data. The use of a common standard will enable a more efficient review of the data via use of standard review tools within the FDA's Nonclinical Information Management System (NIMS).

While binding guidance documents have been, and will continue to be, a significant motivating factor in the adoption, implementation, and use of data standards within the pharmaceutical industry, there are other benefits that standards should bring:

- Increased reusability, since information can be reused across multiple documents, databases, and systems
- Better integration of data across studies and submissions
- Fewer requests from FDA for specific data representations
- The ability for increased data sharing by providing a common language
- An increase in the ability to identify safety signals by providing the ability to mine data within and across therapeutic areas
- The ability to use automated tools to create specialized views of data

The adoption effort is expected to be more challenging for nonclinical data than clinical data, since the former has not been typically submitted in any type of usable, standardizable electronic format. Nonclinical studies may have PDFs and/or Excel spreadsheets formatted as summary tables for reports. While these might be considered by some to be "electronic" it can be difficult, if not impossible, to convert them to SEND datasets. Database extracts often face the same problem.

REFERENCES

1. Wood, F., and Kramer, L. (2011) SEND History and Basics. PharmaSUG Proceedings, May 2011. http://www.lexjansen.com/pharmasug/2011/CD/PharmaSUG-2011-CD14.pdf.
2. Wood, F. (2008) The CDISC Study Data Tabulation Model (SDTM): History, Perspective, and Basics. PharmaSUG Proceedings, June 2008.
3. Wood, F., and Guinter, T. (2008) Evolution and Implementation of the CDISC Study Data Tabulation Model (SDTM). Pharmaceutical Programming 1 (1): 20-27.
4. Providing Regulatory Submissions in Electronic Format — Submissions Under Section 745A(a) of the Federal Food, Drug, and Cosmetic Act (December 2014).
http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM384686.pdf
5. Providing Regulatory Submissions In Electronic Format — Standardized Study Data (December 2014).
http://www.fda.gov/ucm/groups/fdagov-public/@fdagov-drugs-gen/documents/document/ucm292334.pdf
6. Study Data Technical Conformance Guide (2016).
http://www.fda.gov/ForIndustry/DataStandards/StudyDataStandards/default.htm
7. FDA Data Standards Catalog (2015).
http://www.fda.gov/downloads/ForIndustry/DataStandards/StudyDataStandards/UCM340684.xlsx

ACKNOWLEDGEMENTS

SOME OF THE ACRONYMS USED IN THIS PAPER

BRIGD	Biomedical Research Integrated Domain Group
CDER	Center for Drug Evaluation and Research (FDA)
CDISC	Clinical Data Interchange Standards Consortium
CRO	Contract Research Organization
DART	Developmental and Reproductive Toxicology
FDA	Food and Drug Administration (U.S.)
HL7	Health Level Seven
IG	Implementation Guide
IND	Investigational New Drug (Application)
NDA	New Drug Application
NIMS	(FDA) Nonclinical Information Management System
SDS	Submission Data Standards
SDTM	Study Data Tabulation Model
SDTMIG	SDTM Implementation Guide: Human Clinical Trials
SEND	Standard for Exchange of Nonclinical Data
SENDIG	SEND Implementation Guide
TDM	Trial Design Model

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at: Fred Wood Senior Manager and Lead The Data Standards Consulting Group Accenture Accelerated R&D Services 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 registered trademarks or trademarks of their respective companies