Considerations in the Submission of Exposure Data in SDTM-Based Datasets

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

ABSTRACT

The submission of data regarding the subjects' exposure to a study treatment is critical in assessing its safety and efficacy. While more and more sponsors are committing resources to submit SDTM-based datasets, our experience in legacy-data conversion has revealed that many studies don't collect sufficient data to get a reliable assessment of actual exposure. Even when such data have been collected, many sponsors are unsure of how to properly represent that data in the SDTMIG Exposure (EX) and Exposure as Collected (EC; new to SDTMIG v3.2) domains. This paper will discuss methods for representing exposure data, as well as some of the challenges sponsors may face in converting data to be consistent with the SDTMIG.

INTRODUCTION

Based upon FDA input, the SDTMIG states that the EX domain is required for any study where subjects are "exposed" to investigational product. The EX domain is intended for "protocol-specified study treatments." This could include the drug entity under study, any comparator medication, other medications supplied by the sponsor, and/or any placebo dosing.

The goal of EX is to present to a reviewer the complete and accurate picture of a subject's exposure to study medication. The SDTMIG presents five methods that can be used as the basis for the collection and ultimate submission of exposure data, listed as most to least reliable:
1. Actual observation of the administration of drug by the investigator
2. Automated dispensing device which records administrations
3. Subject recall (e.g., via diary)
4. Derived from drug accountability data (e.g., pill counts)
5. Derived from the protocol

The metadata should describe the method by which the exposure data were determined. Our experiences in legacy-data conversion have provided ample evidence that many of these methods have drawbacks and, that in all cases, good data collection and sound data management processes are needed to avoid them.

Exposure data in EX can be represented over the sponsor-defined "constant-dosing interval". The SDTMIG defines it as "any period of time that can be described in terms of a known treatment given at a consistent dose and frequency". The protocol and case report form (CRF) may further define the granularity at which study drug doses are captured across this constant-dosing interval. For example: will each dose be captured on the CRF, will doses be captured within or across protocol "visits", or will a single record with the start and end date across the subjects' study participation suffice? In any event, it is recommended that the data submitted in EX be no less granular than the data that were collected.

A new provisional domain, Procedure Agents (AG), has been used in several Therapeutic-Area User Guides to represent materials administered as part of a procedure. AG is an Interventions domain, created for the submission of compounds that were administered as part of a procedure. Such exposures have generally been difficult to characterize, since they are not really study treatments or concomitant medications. In asthma, for example, it was used to submit dosing information for agents administered as part of an airway responsiveness test, as well as dosing of albuterol for reversibility. It could also be used for contrast agents used as part of an imaging procedure.

CHALLENGES TO THE ACCURATE REPRESENTATION OF EXPOSURE DATA

In the sections below, we will provide examples in which we have had sponsors fail to recognize limitations of the data they intended to submit, and recommendations we have made to remedy some of these. In order to conserve space, the data examples show only the columns relevant to the point being made. There may be other Expected and Permissible variables for which data were collected that would be included in a sponsor's submission dataset.

THE COLLECTION OF DOSING INFORMATION

We have seen many studies in our legacy-data conversion work for which no dosing information has been collected. For these studies, dosing information (e.g., dose levels and frequency) must be derived from the protocol (SDTMIG Method 5). The start and end dates must also be derived, usually from dates of protocol-specified events such as a specific visit at which medication was dispensed. The SDTMIG recognizes that this is the least reliable method for

determining exposure; however, the EX dataset is required, and the derived data at least give the reviewers some approximation of the subjects' exposure in a standard format, and doesn't require a search of the protocol.

THE USE OF DRUG ACCOUNTABILITY DATA

Drug accountability data, which is submitted in the SDTMIG Drug Accountability (DA) domain, is generally collected for use in the calculation of compliance. Although the SDTMIG Method 4 indicates that drug accountability can be used as a source for creating an EX dataset, our experience indicates this is generally not a good practice for the accurate representation of exposure. If, however, it is necessary to do so, it should be done with great care.

It may be practical, but not without risk, to use the dates dispensed and returned as "anchors" from which to create a "constant-dosing interval." Using drug accountability data for much more than this (e.g., dosing amounts on specific dates, dosing frequencies over the dosing period) could lead to some very misleading information. Consider the case where a subject was given 28 tablets for QD (daily) dosing, and returns 14 of them at the end of the four-week period. The actual exposure could have been one of many scenarios including, but not limited to the following:

- one tablet QOD (every other day)
- one tablet QD for the first two weeks
- one tablet QD for the last two weeks
- one tablet QD every other week

Without the collection of additional information from the subject, it would be impossible to know what their dosing pattern was, and why they didn't take 14 tablets, and creating a meaningful Exposure dataset would be impossible.

REPRESENTING THE COMPLETE DOSING PERIOD

The limitation to SDTMIG Method 1 is that the investigator must be present for each dose given to get the complete dosing information. This is rarely the case, however, in larger, later-phase trials. We have seen a number of cases where some dosing information has been collected, but from which a complete dosing picture cannot be obtained. One of these involves studies that collect only in-clinic doses. The data in the table below resulted from a CRF that collected the date/time of dosing at three visits to the site.

STUDYID	DOMAIN	USUBJID	EXSEQ	EXTRT	EXCAT	EXDOSE	EXDOSFRQ	EXSTDTC	EXENDTC
ABC0001	EX	0001-101	1	DRUG A	AT SITE	150	QD	2012-01-08	
ABC0001	EX	0001-101	2	DRUG A	AT SITE	150	QD	2012-01-15	
ABC0001	EX	0001-101	3	DRUG A	AT SITE	150	QD	2012-01-22	
ABC0001	EX	0001-102	1	DRUG A	AT SITE	150	QD	2012-01-08	
ABC0001	EX	0001-102	2	DRUG A	AT SITE	150	QD	2012-01-15	
ABC0001	EX	0001-102	3	DRUG A	AT SITE	150	QD	2012-01-22	

If these were the only records submitted, the data would lead a reviewer to conclude that the two subjects in this treatment Arm received three doses of "Drug A" over the course of the study. The protocol, however, specified that the subjects were to take the study drug on a daily basis for fifteen days. The sponsor felt that by populating EXDOSFRQ with QD, it would be clear that the dosing was daily; however the absence of a date in EXENDTC does not allow for the establishment of a constant-dosing interval.

While SDTM datasets are expected to represent the raw collected data, this is a case where using only the collected data does not provide an accurate representation of the true exposure. Adding an additional record for each subject that reflects the entire span of dosing (e.g., a "constant-dosing interval" record), is one solution to this problem, as shown in the table below. To clarify that there are two types of dosing records, EXCAT was populated with the value of "AT SITE" for the discrete in-clinic doses and left blank for the "blanket" dosing record. Some implementations have included a value of "INTERVAL" for the added records.

STUDYID	DOMAIN	USUBJID	EXSEQ	EXTRT	EXCAT	EXDOSE	EXDOSFRQ	EXSTDTC	EXENDTC
ABC0001	EX	0001-101	4	DRUG A		150	QD	2012-01-08	2012-01-22
ABC0001	EX	0001-102	4	DRUG A		150	QD	2012-01-08	2012-01-22

Another common scenario is when the CRF captures only the dates of dose adjustments (each dose adjustment record signals the start of a new constant-dosing interval). In such a case, records that "fill in the blanks" between these collected dates must be derived. To avoid this problem, at minimum, the first dose date/time and last dose date/time should be recorded on the CRF. If dosing changes are expected, then an additional CRF page to record these changes would be beneficial. As with the at-site dosing, the EX dataset would include records for the dose adjustments as well as records for the dosing period.

THE REPRESENTATION OF UNBLINDED DATA

The addition of the Exposure as Collected (EC) domain solves the problem encountered by many sponsors whose SOPs required, or data-management functions wanted to submit, blinded data. Some of these sponsors were tempted to submit the blinded data in EX, since that was what they collected. EC now provides a place to submit the

collected, blinded data, if desired or necessary. The example below shows how both the blinded and the unblinded data would be submitted. In this study, one tablet was taken by each subject twice a day for two weeks. The blinded data are shown in the partial EC dataset example below.

ec.xpt (Blinded Data)

STUDYID	DOMAIN	USUBJID	ECSEQ	ECTRT	ECDOSE	ECDOSU	ECDOSFRQ	ECSTDTC	ECENDTC
DEF0001	EX	0001-101	1	BLIND_X	1	TABLET	BID	2012-01-15T08:30	2012-01-29T08:30
DEF0001	EX	0001-102	1	BLIND_Y	1	TABLET	BID	2012-01-17T10:30	2012-01-31T10:30
DEF0001	EX	0001-103	1	BLIND_Z	1	TABLET	BID	2012-01-21T13:00	2012-02-04T13:00

When unblinded, the data reflect that the three subjects had been randomized to the three dose levels planned for the study, with doses represented in the partial EX dataset example below.

ex.xpt (Unblinded Data)

STUDYID	DOMAIN	USUBJID	EXSEQ	EXTRT	EXDOSE	EXDOSU	EXDOSFRQ	EXSTDTC	EXENDTC
DEF0001	EX	0001-101	1	DRUG_A	0	mg	BID	2012-01-15T08:30	2012-01-29T08:30
DEF0001	EX	0001-102	1	DRUG_A	10	mg	BID	2012-01-17T10:30	2012-01-31T10:30
DEF0001	EX	0001-103	1	DRUG_A	20	mg	BID	2012-01-21T13:00	2012-02-04T13:00

THE REPRESENTATION OF DOSES NOT TAKEN

In previous versions of the SDTMIG, there was no recommended method for representing information collected regarding doses not taken. Although it was stated that the --OCCUR variable would not generally be used in the EX dataset, it was not prohibited. Some sponsors used it for this purpose since it represented data that was collected.

In the latest version of the SDTMIG, there is a recommended method that consists of using ECOCCUR for submitting data collected about doses not taken. The reason the dose is not taken would be submitted in SUPPEC, with a QNAM of ECREASOC (QLABEL = Reason for Occur Value). The rationale for this naming is that there could be reasons why ECOCCUR was either Y or N. The guidance for not using EXOCCUR remains.

The following example shows dosing over the period from March 1st through March 29th, 2012. A dose was missed on March 15th, as shown in Row 2, with ECOCCUR = N; ECDOSE is null, and the reason for ECOCCUR being N is in SUPPEC.

ec.xpt

Row	STUDYID	DOMAIN	USUBJID	ECSEQ	ECLNKID	ECTRT	ECPRESP	ECOCCUR
1	XYZ	EC	XYZ-1001	1	1001-01	BOTTLE A	Y	Y
2	XYZ	EC	XYZ-1001	2	1001-02	BOTTLE A	Y	N
3	XYZ	EC	XYZ-1001	3	1001-03	BOTTLE A	Y	Y

Row	ECDOSE	ECDOSU	ECDOSFRQ	ECSTDTC	ECENDTC	ECSTDY	ECENDY
1 (cont.)	2	TABLET	QD	2011-03-01	2011-03-14	1	14
2 (cont.)		TABLET	QD	2011-03-15	2011-03-15	15	15
3 (cont.)	2	TABLET	QD	2011-03-16	2011-03-29	16	29

suppec.xpt

STUDYID	RDOMAIN	USUBJID	IDVAR	IDVARVAL	QNAM	QLABEL	QVAL	QORIG	QEVAL
XYZ	EC	XYZ-1001	ECSEQ	2	ECREASOC	Reason for Occur Value	SUBJECT ERROR	CRF	

The records in EX show the constant-dosing intervals for March 1st through March 14th, and for March 16th through the 29th. The EC and EX records are linked using RELREC, with the IDVAR being --LNKID.

ex.xpt

Row	STUDYID	DOMAIN	USUBJID	EXSEQ	EXLNKID	EXTRT
1	XYZ	EX	XYZ-1001	1	1001-01	DRUG A
2	XYZ	EX	XYZ-1001	2	1001-03	DRUG A

Row	EXDOSE	EXDOSU	EXDOSFRM	EXDOSFRQ	EXROUTE	EXSTDTC	EXENDTC	EXSTDY	EXENDY
1 (cont.)	500	mg	TABLET	QD	ORAL	2011-03-01	2011-03-14	1	14
2 (cont.)	500	mg	TABLET	QD	ORAL	2011-03-16	2011-03-29	16	29

relrec.xpt

STUDYID	RDOMAIN	USUBJID	IDVAR	IDVARVAL	RELTYPE	RELID
XYZ	EC		ECLNKID		ONE	1
XYZ	EX		EXLNKID		ONE	1

THE REPRESENTATION OF DOSES IN PROTOCOL-SPECIFIED UNITS

In some open-label studies, it's possible that the dosing information could be collected as a weight-adjusted volume of a dosing solution with a known concentration. For example, for a protocol-specified dose of 5 mg/kg, a subject weighing 80 kg would be dosed 10 mL of a 40 mg/mL solution. The CRF may be designed to collect the volume dosed, which could be represented in EC, while EX could contain the dose in mg/kg. This is shown in the following abbreviated example.

ec.xpt

STUDYID	DOMAIN	USUBJID	ECSEQ	ECTRT	ECDOSE	ECDOSU
XYZ	EX	0001-101	1	DRUG_A	10	mL

ex.xpt

STUDYID	DOMAIN	USUBJID	EXSEQ	EXTRT	EXDOSE	EXDOSU
XYZ	EX	0001-101	1	DRUG_A	5	mg/kg

DOSES OF INTEREST

In some studies, the most recent dose taken prior to the start of a PK-sample collection is recorded. This information may be important in assessing the acceptability of PK data from a subject. Such study-drug exposure may be the only dose information collected, or may supplement other exposure data collected for constant-dosing intervals.

In diabetes trials, information around hypoglycemic events is captured. Included is the last dose of study medication prior to the event. This dose has often been taken during a constant-dosing interval, but needs to be represented as a dose of interest. Such doses can be represented as separate records in EX with an appropriate value in EXCAT, as the example below shows. It would be beneficial to explain this use of EXCAT in the Study Data Reviewer's Guide.

STUDYID	DOMAIN	USUBJID	EXSEQ	EXTRT	EXCAT	EXDOSE	EXDOSU	EXDOSFRQ	EXSTDTC	EXENDTC
XYZ	EX	XYZ-001-001	1	DRUG A	DOSING INTERVAL	20	mg	BID	2013-08-20	2013-09-20
XYZ	EX	XYZ-001-001	2	DRUG A	HIGHLIGHTED DOSE	20	mg		2013-09-01T07:00	

NEW EXPOSURE VARIABLES (EX AND EC)

ECMOOD (MOOD)

This variable was added in order to provide a mechanism for representing data collected about 1) what was planned and 2) what actually occurred. It should be recognized that not all studies may need to make this distinction. The name, "mood", comes from BRIDG terminology, which classifies observations as defined (in a global library), planned (in a protocol), scheduled (for a subject), and performed (for a subject). The use in SDTM-based domains is limited to EC, with CDISC Controlled Terminology of SCHEDULED and PERFORMED. Actual dosing in the protocol-specified unit would be submitted in EX.

ECPSTRG and PGSTRGU (Pharmaceutical Strength and Pharmaceutical Units)

These two variables are used to represent the active ingredient expressed quantitatively per dosage unit, per unit of volume, or per unit of weight, depending on the dose form. An example of their use would be to reflect concentrations of active drug in intravenous dosing solutions, when this information is collected. The actual dose, in mg or mg/kg, for example, would be submitted in EX. One use of these variables is shown in Example 7 of the EC/EX section the SDTMIG v3.2.

--FAST (Fasting Status)

The --FAST variable had previously been only in the Findings general observation class. It was added to the EC and EX domains as a Permissible variable. It can also be added to other Interventions domains when needed.

--LAT and --DIR

These variables had previously been only in the Findings general observation class. They were added to the EC and EX domains as Permissible variables, and are Variable Qualifiers of –LOC. They can also be added to other Interventions domains when needed.

MISUSE OF EX VARIABLES

A number of EX variables are sometimes misused. Some of the most common misuses we have seen are as follows:
- In general, it is expected that doses of study drug can be represented in a numeric format, unlike concomitant medications for which dose values such as "200-400" might be collected. Therefore, EXDOSTXT is generally not present in EX datasets. It should not contain a duplication of the value in EXDOSE.
- EXTRT should contain the name of the treatment, and not additional information that can be represented in other variables (e.g., dose, dose units, and dose frequency).
- EXGRPID is often used incorrectly. It is intended to group records within a subject. In any domain, --GRPID has no meaning across subjects. An example of misuse is the case where each subject was supposed to have three injections, and EXGRPID was used instead of EXSPID. We have seen –GRPID used for many other purposes for data that has meaning across subjects, but for which sponsors do not want to create SUPPEX records.
- EXCAT values are intended to be used to identify categories of treatment. Populating EXCAT with a constant value such as "STUDY MEDICATION" would have added no value, as it is redundant to information implied in the domain name itself. EXCAT should also not contain any other information present in other variables.
- EXDOSTOT should be used only when the total daily dose has been collected. It should not be derived (e.g., EXDOSE x EXDOSFRQ) or used for total doses collected over other periods, such as total dose for a week or a dosing period. This information should be submitted in SUPPEX. It has been suggested that, for a future version of the SDTMIG, --DOSTOT may be deprecated, with two new variables created: one for the total dose over an interval, and another to define that interval.
- EXDOSRGM is often confused with EXDOSFRQ. The latter is generally considered to be at a more granular level, and has CDISC Controlled Terminology. EXDOSRGM may be used to represent higher level dosing patterns such as one week on, one week off.

THE IMPORTANCE OF EXPOSURE DATES IN THE DETERMINATION OF SUBJECT ELEMENTS

Trial Element start rules (TESTRL) for any treatment Element are based upon the first date of exposure for that Element. When reviewing trial design specifications for sponsors, we have seen numerous instances where the TESTRL for a planned treatment element references a dose date that is not found or has been represented incorrectly in the EX dataset. The net result is that the Element will be missing the start date in the Subject Elements dataset. The absence of an Element start date will present challenges to reviewers in trying to assess the relationship of other observations, particularly adverse events, to treatment. It is therefore imperative that the design of the CRFs take this into account, and ensure that the appropriate dose dates are collected.

CONCLUSIONS

We have presented a number of challenges sponsors face in providing an accurate representation of a subject's exposure to study medication. Some of these challenges can be addressed by implementing up front, in the protocol and the CRFs, processes that will ensure that one obtains the most accurate picture of study-drug exposure. Other challenges have been lessened by the creation of the Exposure as Collected (EC) domain, new to the SDTMIG v3.2. EC provides mechanisms for representing the following:
- Blinded vs. unblinded data
- Doses not taken/given
- Doses in EX using protocol-specified units when collected in different units
- Scheduled vs. performed doses

These have been discussed in this paper, and implementation examples have been included.

Sponsors are reminded that science and regulation determine the data needed for a submission. The SDTM is simply a standard format for submitting the data that was collected. It will not, in and of itself, remedy deficiencies in data collection and data management. Despite its importance in understanding a drug's safety, our experience indicates that drug exposure data is one of the most vulnerable to poor and incomplete data collection and representation.

ACKNOWLEDGMENTS

CONTACT INFORMATION

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

Fred Wood
Senior Manager, Data Standards Consulting
Accenture Accelerated R&D Services
484-881-2297
f.wood@accenture.com

Jerry Salyers
Senior Consultant, Data Standards Consulting
Accenture Accelerated R&D Services
jerry.j.salyers@accenture.com

Richard Lewis
Senior Consultant, Data Standards Consulting
Accenture Accelerated R&D Services
richard.m.lewis@accenture.com

Kristin Kelly
Consultant, Data Standards Consulting
Accenture Accelerated R&D Services
kristin.c.kelly@accenture.com