Considerations in Submitting Non-Standard Variables: Supplemental Qualifiers, Findings About, or a Custom Findings Domain

Jerry Salyers, Accenture Accelerated R&D Services, Berwyn, PA
Richard Lewis, Accenture Accelerated R&D Services, Berwyn, PA
Fred Wood, Accenture Accelerated R&D Services, Berwyn, PA

ABSTRACT

As discussed by one of the authors in 2013 (Wood. F, Considerations in Data Modeling when Creating Supplemental Qualifiers Datasets in SDTM-Based Submissions, PharmaSUG Proceedings, May 2013), the SDTM Implementation Guide (SDTMIG) provides for a standard mechanism and structure for submitting non-standard variables. That paper discussed common issues seen when submitting SUPP-- datasets, from using an inappropriate IDVAR to the practice of submitting data "as collected" (i.e. "coded") though often largely uninterpretable. Examples highlighted some of the unexpected outcomes when the parent domain and data from the supplemental domain are merged together (as during the course of review) based on the identified "merge key" in the IDVAR variable.

With the advent and growing knowledge and use of the Findings About (FA) domain, many sponsors are challenged in determining where non-standard data best fit. It's not uncommon to see sponsors submitting data in FA when supplemental datasets would have been preferred and would not have required the creation of RELREC records. Similarly, we've seen FA used when a custom Findings domain would have sufficed, as the --OBJ variable would not have added any clarity to the data. This paper will highlight several criteria that can be used to determine how best to represent such non-standard data.

INTRODUCTION

The Findings About domain was new to SDTM 3.1.2, and was intended for data that "represent 'findings about' an intervention or event that cannot be represented in an Intervention or Events record or as a Supplemental Qualifier to such a record". The SDTMIG shows an extensive list of conditions when using FA may be the best choice. Not only is it important to understand when FA can and should be used, but it is also important to understand when creating a SUPP-- (where '--' represents the two-letter domain code of the parent dataset) record would still be the simplest and most direct approach. There are also occasions, however, where creating a custom Findings domain would be more appropriate. If the chosen FAOBJ is redundant with the FATESTCD or FACAT, this would be an indicator that a such a domain could be used for the data in question.

FINDINGS ABOUT OVERVIEW

The following list, as outlined in the SDTMIG, shows those instances where a Findings About dataset can be used.

- Data or observations that have different timing than the associated Intervention or Event record as a whole. As when severity of an AE may be captured at various pre-defined junctures along the continuum of the AE. These additional severity data points represent "snapshots" of the AE over time. Thus the timing of each snapshot is different than the timing of the Event record as a whole.
- Data may have "qualifiers" of their own that would need to be represented in other Findings domain variables. Examples might include the size of a rash along with its "unit".
- Data or observation about an Intervention or Event for which no Intervention or Event record is collected.
- Data or information about an Event or Intervention that indicate the occurrence of related symptoms, where the occurrence of the symptom does not represent a separate Event in and of itself.
- Data or information about the occurrence of pre-specified Adverse Events. As the AE dataset cannot contain records for events that did not occur, a Findings About dataset would allow the capture of both the "No" and "Yes" response for each pre-specified AE. With the advent of the therapeutic area user guides (TAUGs), there is growing momentum that this use of FA should be expanded to any pre-specified list of Events or Interventions, not just AEs.

SUPPLEMENTAL QUALIFIER OVERVIEW

Supplemental Qualifiers represent non-standard data that doesn't fit into a standard SDTM variable. This list provides a few key points around the inclusion of supplemental qualifiers and their relationship to parent records in one of the general observation class domains or to Demographics

- One dataset (SUPP--.xpt) for each parent domain that has non-standard variables (pending possible adoption of proposal elevating non-standard variables to the parent domain).

Considerations in the Submission of non-Standard Variables, Continued

- A supplemental qualifier record must relate back to at least one "parent" record. If the IDVAR is anything other than --SEQ, a SUPP record may relate back to multiple parent records.
- QNAMs and QVALs relate only to parent records and not to each other
- QVALs share the timing of the parent record.
- Though data in a supplemental qualifier is no less important than data in the parent domain, it is not for submitting "extra" data or data that may have been mapped to controlled terminology. Also not for submitting ambiguous or "coded" data.

The following example from recent development work on the TAUG for COPD shows a possible use of supplemental qualifiers in relation to additional questions being asked surrounding a subject experiencing an exacerbation of COPD. In this example, these additional questions share the timing of the parent record and are appropriately captured in SUPPAE.

ae.xpt

DOMAIN	USUBJID	AESEQ	AESPID	AETERM	AESEV	AESTDTC	AEENDTC
AE	123-1001	1	1	Exacerbation of COPD	Moderate	2013-09-01	2013-09-08

suppae.xpt

RDOMAIN	USUBJID	IDVAR	IDVARVAL	QNAM	QLABEL	QVAL
AE	123-1001	AESEQ	1	PRIMCAUS	Primary cause of Exacerbation	Upper Respiratory Infection
AE	123-1001	AESEQ	1	INTBTION	Intubation Performed?	N
AE	123-1001	AESEQ	1	AEDIS	Lead to Study Discontinuation	N

However, consider the following example that illustrates a fundamental difference between Findings About and Supplemental Qualifiers. In this example, a subject was identified as having a clinical event of "Rash" that was captured as a single clinical event in CE. However, during the course of the event, the rash's severity was assessed at weekly junctures as follows:

CETERM = 'RASH'
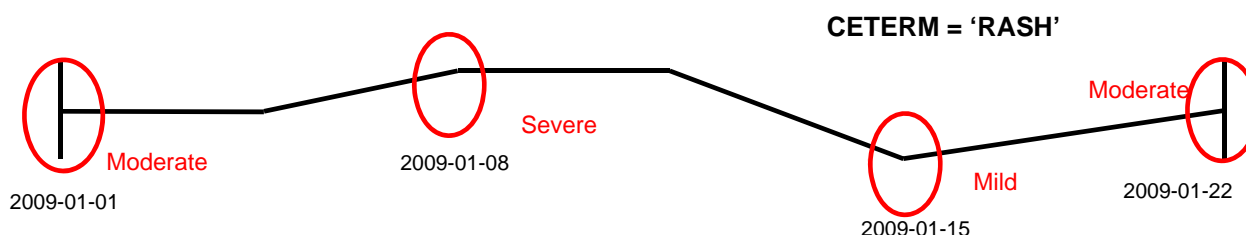
The subject's rash was originally identified with a "moderate" severity on the date shown. A week later, it was assessed as "severe" followed by further weekly assessments of "mild" and then back to "moderate. A look at the partial CE dataset will show a single record with CESEV being set to the highest severity attained over the total course of the record:

DOMAIN	USUBJID	CESEQ	CESPID	CETERM	CESEV	CESTDTC	CEENRF
CE	123-01-01	1	1	RASH	Severe	2009-01-01	AFTER

How should the periodic severity assessments be shown? As we can see by the CE record, the only timing variables we have is the "start date" of the rash and an "ending" relative timing variable. If we used SUPPCE, we might have records such as these:

RDOMAIN	USUBJID	IDVAR	IDVARVAL	QNAM	QLABEL	QVAL
CE	123-01-01	CESEQ	1	CESEV1	Severity 1	Moderate

CE	123-01-01	CESEQ	1	CESEV2	Severity 2	Severe
CE	123-01-01	CESEQ	1	CESEV3	Severity 3	Mild
CE	123-01-01	CESEQ	1	CESEV4	Severity 4	Moderate
CE	123-01-01	CESEQ	1	CESEV1DT	Severity 1 Date	2009-01-01
CE	123-01-01	CESEQ	1	CESEV2DT	Severity 2 Date	2009-01-08
CE	123-01-01	CESEQ	1	CESEV3DT	Severity 3 Date	2009-01-15
CE	123-01-01	CESEQ	1	CESEV4DT	Severity 4 Date	2009-01-22

As we saw above, one of the basic tenets of Supplemental Qualifier records is that they only relate back to the parent record and not to each other. Thus in our example, there is no machine readable, unambiguous relationship between each date and its associated severity. It's clear that the collection of the severity of the rash at weekly junctures represents different "timing" than the original CE record. Thus we need another way to represent this data. The Findings About dataset is the best way to show, on a single record, each severity assessment along with the date of the assessment.

DOMAIN	USUBJID	FASEQ	FASPID	FATESTCD	FATEST	FAOBJ	FAORRES	FADTC
FA	123-01-01	1	1	SEV	Severity	RASH	Moderate	2009-01-01
FA	123-01-01	2	1	SEV	Severity	RASH	Severe	2009-01-08
FA	123-01-01	3	1	SEV	Severity	RASH	Mild	2009-01-15
FA	123-01-01	4	1	SEV	Severity	RASH	Moderate	2009-01-22

As detailed in the SDTMIG, the value of the --OBJ variable ("RASH") is set to the value of the Topic variable (--TERM or --TRT) of the record in the "parent" domain, in this case, CE. Note the use of an SDTM Qualifier variable "fragment", "SEV", as the FATESTCD.

We should also document in RELREC the relationship between the single "Rash" record in CE with the collected severities.in FA. We can do this in a "dataset to dataset" RELREC as shown below:

RDOMAIN	USUBJID	IDVAR	IDVARVAL	RELTYPE	RELID
CE		CESPID		ONE	CEFA
FA		FASPID		MANY	CEFA

ADDITIONAL USE CASE USING FA

Another one of the instances outlined in the SDTMIG where using FA is acceptable is when collecting the "occurrence" of pre-specified signs or symptoms, where the occurrence ("Yes" collected on the CRF) isn't captured as an event in and of itself. There is some room for interpretation as to when signs and symptoms may best be mapped to a domain such as CE and when they are more appropriately mapped to FA. If the symptom data provides for supportive information for a collected over-arching event record, the data would map nice to FA. However, if the symptom itself could "stand-alone" in an events domain, then collecting this data in CE using CEPRESP and CEOCCUR is easily supported.

Below represents a sample CRF for the collection of a COPD Exacerbation as an Adverse Event, plus the collection of pre-specified signs or symptoms that are often prevalent during an exacerbation:

3

Adverse Event Pre-specified: *Hidden/Pre-specified*	COPD EXACERBATION
Primary cause of exacerbation? QVAL when QNAM= 'PRIMCAUS'	☐ Cold air/Cold weather ☐ Air pollution ☐ Allergy ☐ Tobacco Smoke ☐ Exercise ☐ Stress/Emotions ☐ Upper respiratory infection ☐ Lower respiratory infection ☐ Beta-blockers ☐ Withholding or reducing COPD meds ☐ Lack of efficacy with investigational treatment ☐ Other, specify: _____ ☐ Unknown
Start Date:	AESTDTC _ _ / _ _ _ / _ _ _ _
End Date:	AEENDTC _ _ / _ _ _ / _ _ _ _
Have any of the following symptoms occurred in conjunction with the exacerbation?	
Dyspnea	☐ No ☐ Yes
Sputum Volume	☐ No ☐ Yes
Sputum Purulence	☐ No ☐ Yes
Sore Throat	☐ No ☐ Yes
Nasal Discharge	☐ No ☐ Yes
Cough	☐ No ☐ Yes
Wheezing	☐ No ☐ Yes

The AE dataset would be the same as shown previously. The FA dataset detailing the occurrence of each of the pre-specified symptoms (as noted in FAOBJ), would look like:

DOMAIN	USUBJID	FASEQ	FASPID	FATESTCD	FATEST	FAOBJ	FAORRES
FA	123-1001	1	1	OCCUR	Occurrence	Dyspnea	Y
FA	123-1001	2	1	OCCUR	Occurrence	Sputum Volume	N
FA	123-1001	3	1	OCCUR	Occurrence	Sputum Purulence	N
FA	123-1001	4	1	OCCUR	Occurrence	Sore Throat	Y
FA	123-1001	5	1	OCCUR	Occurrence	Nasal Discharge	Y
FA	123-1001	6	1	OCCUR	Occurrence	Cough	Y
FA	123-1001	7	1	OCCUR	Occurrence	Wheezing	Y

SUBMITTING FA

The SDTMIG provides for a mechanism for submitting the FA domain. The major points include:

- As FA doesn't have the "RDOMAIN" variable to easily identify the "parent" domain, sponsors have the option of submitting FA in physically separate datasets based on the parent domain of the value in the --OBJ variable

Considerations in the Submission of non-Standard Variables, Continued

- The dataset name would be "FA" plus the two-letter domain code of the "parent" dataset, as in FAAE, FACE, or FAMH
- The value for the "DOMAIN" variable would remain FA across all of the individual FA datasets
- Also needs to conform to any additional guidance regarding the "splitting" of datasets
    - --SEQ needs to be unique per USUBJID across all of the split datasets
    - Split dataset names can be up to 4 characters in length
    - If the FA domain is split into separate datasets, any Supplemental Qualifier records for FA must also be split into individual SUPP datasets, with "SQ" as the prefix, as in SQFAMH.
- In SDTM 3.2 we have the first Findings About domain without the "FA" 2-letter domain code – SR (Skin Response)

COMPARE AND CONTRAST – SUPPQUAL VS FINDINGS ABOUT

The table below provides a summary as to the differences between mapping data to a Supplemental Qualifier, to a Findings About dataset, or to a custom Findings domain.

CHARACTERISTIC	SUPPQUAL	FINDINGS ABOUT	FINDINGS
CDISC Controlled Terminology	Limited list of QNAMs available in CT	None for –TESTCD and --Test values other than that for tests based upon SDTM Qualifiers (e.g., SEV, DUR, OCCUR)	--TESTCD and --Test values for most modeled domains
Timing	Inherits the timing of the parent record	--DTC for –TESTCD and –TEST	--DTC for –TESTCD and –TEST
Unique record defined by other keys	No	Yes	Yes
Uses and requires --OBJ	No	Yes	No
Ability to relate multiple qualifiers (e.g., results and units) to each other	No	Yes	Yes
Relates to an Intervention or Event record	Always (Must have a "parent" record)	Likely	Possible

CONCLUSIONS

The Findings About domain can be used in many circumstances to submit additional data that does not fit into a standard domain Care should be taken to use it only when required based on a study's needs. The SDS team and data representatives working on Therapeutic Area User Guides continues to explore additional use cases for the valid use of FA for various data submission needs. It should be emphasized that the continued use of Supplemental Qualifier datasets should remain an option when the record shares the timing of the parent. This is still the simpler approach and does not necessitate the creation of RELREC.

RESOURCES

- Study Data Tabulation Model. Clinical Data Interchange Standards Consortium (CDISC) Submission Data Standards (SDS) Team, Version 1.4, December 2013.
- Study Data Tabulation Model Implementation Guide: Human Clinical Trials. Clinical Data Interchange Standards Consortium (CDISC) Submission Data Standards (SDS) Team, Version 3.2, December 2013

ACKNOWLEDGMENTS

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are registered trademarks or trademarks of their respective companies.

CONTACT INFORMATION

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

Jerry Salyers
Senior Consultant, Data Standards Consulting
Accenture Life Sciences
1160 W. Swedesford Road Bldg One
Berwyn, PA 19312
513.668.8126
jerry.j.salyers@accenture.com

Richard Lewis
Senior Consultant, Data Standards Consulting
Accenture Life Sciences
1160 W. Swedesford Road Bldg One
Berwyn, PA 19312
262.716.4167
richard.m.lewis@accenture.com

Fred Wood
Senior Manager, Data Standards Consulting
Accenture Life Sciences
1160 W. Swedesford Road Bldg One
Berwyn, PA 19312
484.881.2297
f.wood@accenture.com