

Building Traceability for End Points in Analysis Datasets Using SRCDOM, SRCVAR, and SRCSEQ Triplet

Xiangchen Cui¹, Tathabhai Pakalapati², and Qunming Dong¹

¹Statistical Programming, Vertex Pharmaceuticals, Cambridge, MA 02139

²Clinical Research Services, Cytel Inc., Cambridge, MA 02139

Introduction

To be compliant with ADaM Implementation Guide V1.0, traceability feature should be incorporated to possible extent in study analysis datasets. There are two types of traceability: (1) Metadata Traceability (2) Data Point Traceability. Data Point Traceability provides clear link in the dataset to specific input data values used to derive analysis values. SRCDOM, SRCVAR, and SRCSEQ triplet is one among many ways suggested by CDISC to establish data point traceability in ADaM datasets. This poster provides various examples of applying SRCDOM, SRCVAR, and SRCSEQ triplet to establish traceability in efficacy ADaM datasets from Cystic Fibrosis therapeutic area.

Definition of SRCDOM, SRCVAR, and SRCSEQ Triplet

SRCDOM	Source Domain	The 2-character identifier of the SDTM domain that relates to AVAL or AVALC.
SRCVAR	Source Variable	The name of the column (in the SDTM domain identified by SRCDOM) that relates to AVAL or AVALC.
SRCSEQ	Source Sequence Number	The sequence number SEQ of the row (in the SDTM domain identified by SRCDOM) that relates to AVAL or AVALC.

SRC--- triplet establishes a link from derived value to variables in SDTM that serve as primary candidates to built Data Point Traceability. It aids reviewers (Biostatistician, Agency Reviewer, QC Programmer) to trace back to the specific predecessor record(s) used to derive an analysis value. By definition SRC triplet is feasible to establish traceability only if the derived analysis variable depends only on a single predecessor record and single variable from the source SDTM. Not practical for situations where an analysis value is derived from multiple predecessor records from single variable of source SDTM. This poster provides examples of using SRC--- triplet to handle such situations by slightly modifying the usage of SRC--- triplet without losing the actual definition.

Structure of Sweat Chloride Analysis Dataset (ADSW)

Typically, sweat chloride is collected from both left arm and right arm at every visit. The average of both left arm and right arm is often used as the analyses value at each analysis visit. Three important derivations in ADSW would be

- Baseline value
- Average Sweat Chloride (of both left and right arms) at every analysis visit
- Average on-treatment sweat chloride value

To establish proper traceability, the structure of ADSW will be as follows:

- Inherit all the SDTM SW records to ADSW to establish data point traceability as per ADaM implementation guide V1.0. Values of all analysis variables (such as AVAL, CHG, BASE, ABLFL, AVISIT etc.) will be set to missing for inherited records
- Derive an additional record for every nominal visit in SDTM SW with PARAMTYP="DERIVED" and DTYPE='AVERAGE' whose AVAL is equal to average of sweat chloride at left and right arms
- To ensure uniqueness within each subject for sorting purpose, variable ASWSEQ is created whose value will be same as SW.SWSEQ for records inherited from SDTM and starts from 1001 and incremented by 1 successively for all derived records

SWSEQ	ASWSEQ	PARAMCD	PARAMTYP	AVAL	DTYPE	AVISIT	SRCDOM	SRCVAR	SRCSEQ	VISIT	SWSTRESN	SWTPT
1	1	SW_CL_L								SCREENING	82	
3	3	SW_CL_R								SCREENING	66	
5	5	SW_CL_L								DAY 1	53	PREDOSE
7	7	SW_CL_R								DAY 1	38	PREDOSE
9	9	SW_CL_L								DAY 1	84	2-4 HOURS POSTDOSE
11	11	SW_CL_R								DAY 1	94	2-4 HOURS POSTDOSE
13	13	SW_CL_L								DAY 1	87	24 HOURS POSTDOSE
15	15	SW_CL_L								DAY 1	84	24 HOURS POSTDOSE
17	17	SW_CL_R								DAY 7	86	
19	19	SW_CL_L								DAY 7		
21	21	SW_CL_R								DAY 14		
23	23	SW_CL_L								DAY 14	79	
25	25	SW_CL_L								DAY 21	88	
27	27	SW_CL_R								DAY 21		
29	29	SW_CL_L								DAY 28	87	PREDOSE
31	31	SW_CL_R								DAY 28	71	PREDOSE
33	33	SW_CL_L								DAY 28	80	2-4 HOURS POSTDOSE
35	35	SW_CL_R								DAY 28	60	2-4 HOURS POSTDOSE
37	37	SW_CL_L								DAY 28	63	24 HOURS POSTDOSE
39	39	SW_CL_R								DAY 28	88	24 HOURS POSTDOSE
41	41	SW_CL_L								DAY 35	74	
43	43	SW_CL_R								DAY 35	56	
45	45	SW_CL_L								DAY 42	110	
47	47	SW_CL_R								DAY 42	88	
49	49	SW_CL_L								FOLLOW-UP/DAY 56	39	
51	51	SW_CL_L								FOLLOW-UP/DAY 56	69	
1001	1001	SW_CL_M	DERIVED	74	AVERAGE	Screening	SW	SWSTRESN	1\$3	SCREENING	74	
1002	1002	SW_CL_M	DERIVED	45.5	AVERAGE	Day 1	SW	SWSTRESN	5\$7	DAY 1	45.5	PREDOSE
1002.5	1002.5	SW_CL_M	DERIVED	59.8	AVERAGE	Baseline	ADSW	AVAL	1001\$1002		59.8	
1003	1003	SW_CL_M	DERIVED	89	AVERAGE	Day 1, 2-4 hrs post dose	SW	SWSTRESN	9\$11	DAY 1	89	2-4 HOURS POSTDOSE
1004	1004	SW_CL_M	DERIVED	85.5	AVERAGE	Day 1, 24 hrs post dose	SW	SWSTRESN	13\$15	DAY 1	85.5	24 HOURS POSTDOSE
1005	1005	SW_CL_M	DERIVED	86	AVERAGE	Day 7	SW	SWSTRESN	17	DAY 7	86	
1006	1006	SW_CL_M	DERIVED	79	AVERAGE	Day 14	SW	SWSTRESN	23	DAY 14	79	
1007	1007	SW_CL_M	DERIVED	88	AVERAGE	Day 21	SW	SWSTRESN	25	DAY 21	88	
1008	1008	SW_CL_M	DERIVED	79	AVERAGE	Day 28	SW	SWSTRESN	29\$31	DAY 28	79	PREDOSE
1008.5	1008.5	SW_CL_M	DERIVED	83	AVERAGE	Average through Day 28	ADSW	AVAL	DAY 7\$DAY 14\$DAY 21\$DAY 28		83	
1009	1009	SW_CL_M	DERIVED	60	AVERAGE	Day 28, 2-4 hrs post dose	SW	SWSTRESN	35	DAY 28	60	2-4 HOURS POSTDOSE
1010	1010	SW_CL_M	DERIVED	75.5	AVERAGE	Day 28, 24 hrs post dose	SW	SWSTRESN	37\$39	DAY 28	75.5	24 HOURS POSTDOSE
1011	1011	SW_CL_M	DERIVED	65	AVERAGE	Day 35	SW	SWSTRESN	41\$43	DAY 35	65	
1012	1012	SW_CL_M	DERIVED	99	AVERAGE	Day 42	SW	SWSTRESN	45\$47	DAY 42	99	
1013	1013	SW_CL_M	DERIVED	54	AVERAGE	Follow-Up/Day 56	SW	SWSTRESN	49\$51	FOLLOW-UP/DAY 56	54	

Traceability for Average Sweat Chloride at Both Left and Right Arms at Every Analysis Visit

- Analysis value at each analysis visit (AVISIT) is equal to average of sweat chloride at both left and right Arms collected at a nominal visit
- PARAMTYP='DERIVED' tells the reviewer these records are derived in ADaM and DTYPE='AVERAGE' indicates that AVAL was derived by averaging values of records with sequence numbers listed in SRCSEQ
- SRCDOM='SW' and SRCVAR='SWSTRESN' let the reviewer know that the record was derived using values in variable SWSTRESN in SDTM SW domain
- SRCSEQ lists sequence number, separated by '\$', of those records in SDTM SW used in deriving AVAL. SRCSEQ=17 for AVISIT='Day 7' suggests that only sweat chloride at right arm is used to populate AVAL, as sweat chloride at left arm is missing. SRCSEQ='29\$31' for AVISIT='Day 28' suggests that sweat chloride assessments in SDTM SW with sequence numbers 29 and 31, corresponding to left and right arms respectively, are used to derive AVAL.

Traceability for Baseline Analysis Visit

- Analysis value at Baseline analysis visit is equal to average of analysis values from the pre-dose scheduled visits
- Since AVAL at Baseline is average of analysis values corresponding to pre-dose scheduled analysis visits, SRCDOM is equal to 'ADSW' (not SW!) and SRCVAR is equal to 'AVAL' (not SWSTRESN!)
- SRCSEQ='1001\$1002' for AVISIT='Baseline' along with SRCDOM and SRCVAR indicates that baseline analysis value is derived using the AVAL in ADSW, corresponding to sequence numbers 1001 and 1002. PARAMTYP='DERIVED' and DTYPE='AVERAGE' will let the reviewer know that baseline analysis value is the average of AVAL corresponding to records with sequence numbers listed in SRCSEQ.

Traceability for Average On-Treatment Sweat Chloride

- Analysis value for Average On-treatment Sweat Chloride is equal to average of analysis values of on-treatment analysis visits
- Since the analysis value is equal to average of analysis values corresponding to on-treatment analysis visits, SRCDOM=ADSW, SRCVAR=AVAL, PARAMTYP='DERIVED' and DTYPE='AVERAGE'

SUBJID	SWSEQ	ASWSEQ	PARAMCD	PARAMTYP	AVAL	DTYPE	AVISIT	ANLOTFL	SRCDOM	SRCVAR	SRCSEQ	VISIT	SWSTRESN	SWTPT
999006	17	17	SW_CL_R									DAY 7	99	
999006	19	19	SW_CL_L									DAY 7		
999006	21	21	SW_CL_R									DAY 14		
999006	23	23	SW_CL_L									DAY 14	89	
999006	25	25	SW_CL_L									DAY 21	98	
999006	27	27	SW_CL_R									DAY 21		
999006	29	29	SW_CL_L									DAY 28	97	PREDOSE
999006	31	31	SW_CL_R									DAY 28	56	PREDOSE
999006	.	1005	SW_CL_M	DERIVED	99		Day 7	Y	SW	SWSTRESN	17	DAY 7	99	
999006	.	1006	SW_CL_M	DERIVED	89		Day 14	Y	SW	SWSTRESN	23	DAY 14	89	
999006	.	1007	SW_CL_M	DERIVED	98		Day 21	Y	SW	SWSTRESN	25	DAY 21	98	
999006	.	1008	SW_CL_M	DERIVED	76.5		Day 28	Y	SW	SWSTRESN	29&31	DAY 28	76.5	PREDOSE
999006	.	1008.5	SW_CL_M	DERIVED	90.6	AVERAGE	Average through Day 28	Y	ADSW	AVAL	DAY 7&DAY 14&DAY 21&DAY 28		90.6	
999007	17	17	SW_CL_R									DAY 7	96	
999007	19	19	SW_CL_L									DAY 7	93	
999007	21	21	SW_CL_R									DAY 14	56	
999007	23	23	SW_CL_L									DAY 14		
999007	25	25	SW_CL_R									DAY 21		
999007	27	27	SW_CL_L									DAY 21		
999007	29	29	SW_CL_R									DAY 28		PREDOSE
999007	31	31	SW_CL_L									DAY 28		PREDOSE
999007	.	1005	SW_CL_M	DERIVED	94.5		Day 7	Y	SW	SWSTRESN	17&19	DAY 7	94.5	
999007	.	1006	SW_CL_M	DERIVED	96		Day 14	Y	SW	SWSTRESN	21	DAY 14	56	
999007	.	1006.5	SW_CL_M	DERIVED	75.3	AVERAGE	Average through Day 28	Y	ADSW	AVAL	DAY 7&DAY 14		75.3	

- Listing visits contributing to average in SRCSEQ instead of sequence numbers will be more readable and easy for reviewer
- For subject 999006, SRCSEQ='DAY 7\$DAY 14\$DAY 21\$DAY 28', SRCDOM='ADSW', SRCVAR='AVAL', PARAMTYP='DERIVED', and DTYPE='AVERAGE' for AVISIT='Average through Day 28' indicates that analysis value is average of ADSW.AVAL corresponding to analysis visits Day 7, Day 14, Day 21, and Day 28
- Importance of traceability will be more evident when all possible on-treatment visits don't contribute to average. For subject 999007, SRCSEQ='DAY 7\$DAY 14' only!

Advantages of Establishing Traceability in ADaMs

- Helps in effective program validation
- Speeds up the review process
- Transparency in submitted analysis data
- Good Relationship with FDA

Conclusion

SRCDOM, SRCVAR, and SRCSEQ definition in ADaM IG V1.0 can be used to establish traceability when an analysis value is derived from a single predecessor record from single variable of source SDTM. This poster introduces a slightly modified approach to apply SRC--- triplet when an analysis value is derived from multiple predecessor records from single variable of source SDTM.

Reference

- [1] CDISC Analysis Data Model Team. "Analysis Data Model (ADaM) Implementation Guide". December 2009. <http://www.cdisc.org/adam>
- [2] Zhu, Songhui and Yan, Lin. "Methods of Building Traceability for ADaM Data". Proceedings of PharmaSUG 2011 Conference.
- [3] Cui, Xiangchen, Liu, Hongyu, and Pakalapati, Tathabhai. "Examples of Building Traceability in CDISC ADaM Datasets for FDA Submission". Proceedings of SAS Global Forum 2012.