Linking Data in SDTM

Diane Wold
Fred Wood
Kit Howard

October 20, 2020

cdisc

1

Agenda

1. Introduction
2. Linking with Timing Variables
3. Linking with Identifiers
   1. Study-adaptable linking variables
   2. Best practice for choosing IDVAR in SQ and RELREC
4. Linking Comments
5. Conclusions

2

Introduction

Why link data?

Kinds of linking

Linking variables

3

Why link data?

- SDTM-based datasets divide data into domains and into records within domains, but individual records are of very limited use in understanding what's happening to an individual subject, much less understanding the results of a study.
- Linking reconnects data in different records and domains, so that reviewers can make sense of data tabulations and so that meaningful analyses can be performed.

4

Linking data for an individual subject

- Physicians are trained to treat individual patients; medical reviewers sometimes want to understand what happened for a particular subject.
- This arises often with adverse events
  - Associating adverse events with the concomitant medications used to treat the adverse event
  - Associating a hospitalization with an adverse event

5

Grouping data for analysis

- Linking and grouping are closely related; most analyses are based on a group of records identified by a set of linking variables.
- ARMCD and ACTARMCD are extremely important grouping variables, although they appear only in the DM dataset, where they have the role "Record Qualifier"
  - SETCD, when used, is similarly important. Although it is used most often in non-clinical studies, it can be used in human clinical trials.
- The records in an analysis dataset are very often (almost always?) created by linking data across domains, e.g.,
  - By including arm data from the ADSL dataset
  - By combining data from multiple domains for a composite endpoint

6

If records have the same value for a variable, what does it mean?

- Linking in SDTM uses “linking variables”, variables that have the same value in the linked records.
  - Records with the same variable value have some kind of relationship to each other.
- Relationships have a scope, usually described by another variable or variables.
  - Compare the relationships between records with the same value of a VISIT
    - Among all records with the same STUDYID
    - Among all records with the same STUDYID and DOMAIN
    - Among all records with the same STUDYID, DOMAIN, and USUBJID
    - Among all records with the same STUDYID, DOMAIN, and TESTCD
- Not all variables are equally useful for linking
  - Records with the same --ORRES value could have different units for different tests for different subjects in different studies.

7

Traceability for linking variables

- Start with the end in mind!
  - Know what kinds of grouping will be needed for analysis, so that you can be sure that needed linking variables will be collected
- Some needed variables are obvious and routine, such as STUDYID, USUBJID, the arm variables and the visit variables.
- When data collected together are split across SDTM domains, extra care may be needed to assure that they can be reassembled for analysis.
  - More on this later

8

Linking with Timing Variables

Visits
Timepoints
Disease Milestones
Epochs

9

Visits and Time Points

Visits

- VISITNUM, VISIT, and VISITDY, don't have domain prefixes.
- Used to link data across domains.
- Records with the same VISITNUM value are data collected at that visit. They may record events, interventions, or even findings that occurred before the visit, particularly for initial study visits.

Time Points

- --TPTNUM, --TPT, and –ELTM have domain prefixes, but link data within domain
- May be defined differently for different domains.
- Often linked to a Reference Time Point

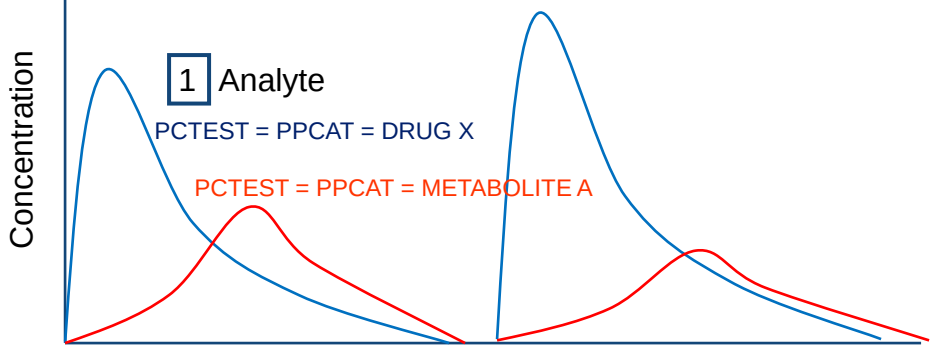
Reference Time Points

- --TPTREF, --RFTDTC are anchors
- Often represent a dose, after which a series of measurements is made
- PCTPTREF/PCRFTDTC linked to PPTPTREF/PPRFTDTC

10

Linking Specific Parameters to Specific Time-Concentration Curves

PC and PP can be related by two variables:



- Day 1
- (PCRFTDTC = PPRFTDTC = 2009-12-10T08:00)

- Day 11
- (PCRFTDTC = PPRFTDTC = 2009-12-21T08:00)

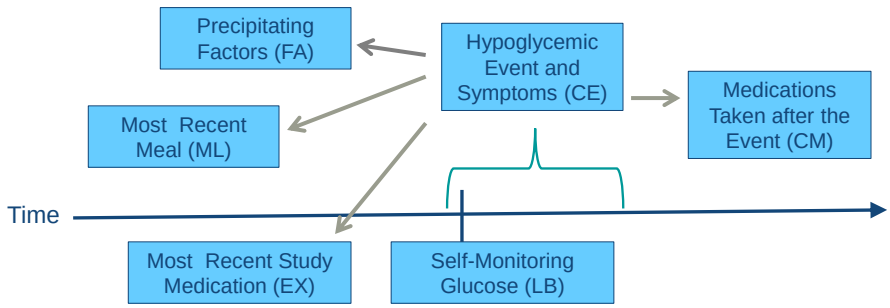
2 Reference Time Point Date/Time

cdisc

11

11

Disease Milestones Example: Hypoglycemic Events (1)

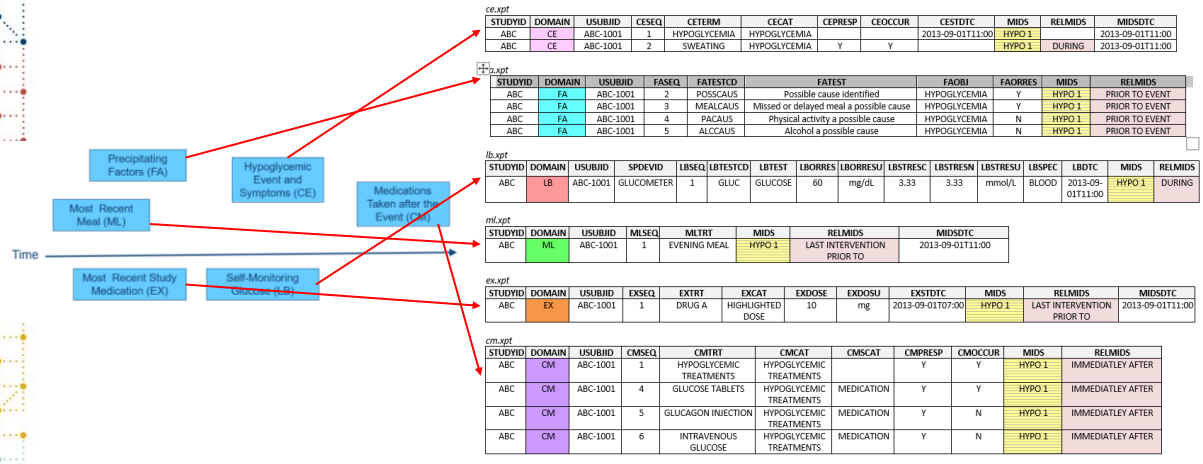


cdisc

12

12

Disease Milestones Example: Hypoglycemic Events (2)



ce.xpt

STUDYID	DOMAIN	USUBJID	CESEQ	CETERM	CECAT	CEPRESP	CEOCCUR	CESTDTC	MIDS	RELMIDS	MIDSDTC
ABC	CE	ABC-1001	1	HYPOGLYCEMIA	HYPOGLYCEMIA			2013-09-01T11:00	HYPO 1		2013-09-01T11:00
ABC	CE	ABC-1001	2	SWEATING	HYPOGLYCEMIA	Y	Y		HYPO 1	DURING	2013-09-01T11:00

fa.xpt

STUDYID	DOMAIN	USUBJID	FASEQ	FATESTCD	FATEST	FAOBJ	FAORRES	MIDS	RELMIDS
ABC	FA	ABC-1001	2	POSSCAUS	Possible cause identified	HYPOGLYCEMIA	Y	HYPO 1	PRIOR TO EVENT
ABC	FA	ABC-1001	3	MEALCAUS	Missed or delayed meal a possible cause	HYPOGLYCEMIA	Y	HYPO 1	PRIOR TO EVENT
ABC	FA	ABC-1001	4	PACAUS	Physical activity a possible cause	HYPOGLYCEMIA	N	HYPO 1	PRIOR TO EVENT
ABC	FA	ABC-1001	5	ALCCAUS	Alcohol a possible cause	HYPOGLYCEMIA	N	HYPO 1	PRIOR TO EVENT

lb.xpt

STUDYID	DOMAIN	USUBJID	SPDEVID	LBSEQ	LBTESTCD	LBTEST	LBORRES	LBORRESU	LBSTRESC	LBSTRESN	LBSTRESU	LBSPEC	LBDTC	MIDS	RELMIDS
ABC	LB	ABC-1001	GLUCOMETER	1	GLUC	GLUCOSE	60	mg/dL	3.33	3.33	mmol/L	BLOOD	2013-09-01T11:00	HYPO 1	DURING

ml.xpt

STUDYID	DOMAIN	USUBJID	MLSEQ	MLTRT	MIDS	RELMIDS	MIDSDTC
ABC	ML	ABC-1001	1	EVENING MEAL	HYPO 1	LAST INTERVENTION PRIOR TO	2013-09-01T11:00

ex.xpt

STUDYID	DOMAIN	USUBJID	EXSEQ	EXTRT	EXCAT	EXDOSE	EXDOSU	EXSTDTC	MIDS	RELMIDS	MIDSDTC
ABC	EX	ABC-1001	1	DRUG A	HIGHLIGHTED DOSE	10	mg	2013-09-01T07:00	HYPO 1	LAST INTERVENTION PRIOR TO	2013-09-01T11:00

cm.xpt

STUDYID	DOMAIN	USUBJID	CMSEQ	CMTRT	CMCAT	CMSCAT	CMPRESP	CMOCCUR	MIDS	RELMIDS
ABC	CM	ABC-1001	1	HYPOGLYCEMIC TREATMENTS	HYPOGLYCEMIC TREATMENTS		Y	Y	HYPO 1	IMMEDIATLEY AFTER
ABC	CM	ABC-1001	4	GLUCOSE TABLETS	HYPOGLYCEMIC TREATMENTS	MEDICATION	Y	Y	HYPO 1	IMMEDIATLEY AFTER
ABC	CM	ABC-1001	5	GLUCAGON INJECTION	HYPOGLYCEMIC TREATMENTS	MEDICATION	Y	N	HYPO 1	IMMEDIATLEY AFTER
ABC	CM	ABC-1001	6	INTRAVENOUS GLUCOSE	HYPOGLYCEMIC TREATMENTS	MEDICATION	Y	N	HYPO 1	IMMEDIATLEY AFTER

13

ce.xpt

STUDYID	DOMAIN	USUBJID	CESEQ	CETERM	CECAT	CEPRESP	CEOCCUR	CESTDTC	MIDS	RELMIDS	MIDSDTC
ABC	CE	ABC-1001	1	HYPOGLYCEMIA	HYPOGLYCEMIA			2013-09-01T11:00	HYPO 1		2013-09-01T11:00
ABC	CE	ABC-1001	2	SWEATING	HYPOGLYCEMIA	Y	Y		HYPO 1	DURING	2013-09-01T11:00

fa.xpt

STUDYID	DOMAIN	USUBJID	FASEQ	FATESTCD	FATEST	FAOBJ	FAORRES	MIDS	RELMIDS
ABC	FA	ABC-1001	2	POSSCAUS	Possible cause identified	HYPOGLYCEMIA	Y	HYPO 1	PRIOR TO EVENT
ABC	FA	ABC-1001	3	MEALCAUS	Missed or delayed meal a possible cause	HYPOGLYCEMIA	Y	HYPO 1	PRIOR TO EVENT
ABC	FA	ABC-1001	4	PACAUS	Physical activity a possible cause	HYPOGLYCEMIA	N	HYPO 1	PRIOR TO EVENT
ABC	FA	ABC-1001	5	ALCCAUS	Alcohol a possible cause	HYPOGLYCEMIA	N	HYPO 1	PRIOR TO EVENT

lb.xpt

STUDYID	DOMAIN	USUBJID	SPDEVID	LBSEQ	LBTESTCD	LBTEST	LBORRES	LBORRESU	LBSTRESC	LBSTRESN	LBSTRESU	LBSPEC	LBDTC	MIDS	RELMIDS
ABC	LB	ABC-1001	GLUCOMETER	1	GLUC	GLUCOSE	60	mg/dL	3.33	3.33	mmol/L	BLOOD	2013-09-01T11:00	HYPO 1	DURING

ml.xpt

STUDYID	DOMAIN	USUBJID	MLSEQ	MLTRT	MIDS	RELMIDS	MIDSDTC
ABC	ML	ABC-1001	1	EVENING MEAL	HYPO 1	LAST INTERVENTION PRIOR TO	2013-09-01T11:00

ex.xpt

STUDYID	DOMAIN	USUBJID	EXSEQ	EXTRT	EXCAT	EXDOSE	EXDOSU	EXSTDTC	MIDS	RELMIDS	MIDSDTC
ABC	EX	ABC-1001	1	DRUG A	HIGHLIGHTED DOSE	10	mg	2013-09-01T07:00	HYPO 1	LAST INTERVENTION PRIOR TO	2013-09-01T11:00

cm.xpt

STUDYID	DOMAIN	USUBJID	CMSEQ	CMTRT	CMCAT	CMSCAT	CMPRESP	CMOCCUR	MIDS	RELMIDS
ABC	CM	ABC-1001	1	HYPOGLYCEMIC TREATMENTS	HYPOGLYCEMIC TREATMENTS		Y	Y	HYPO 1	IMMEDIATLEY AFTER
ABC	CM	ABC-1001	4	GLUCOSE TABLETS	HYPOGLYCEMIC TREATMENTS	MEDICATION	Y	Y	HYPO 1	IMMEDIATLEY AFTER
ABC	CM	ABC-1001	5	GLUCAGON INJECTION	HYPOGLYCEMIC TREATMENTS	MEDICATION	Y	N	HYPO 1	IMMEDIATLEY AFTER
ABC	CM	ABC-1001	6	INTRAVENOUS GLUCOSE	HYPOGLYCEMIC TREATMENTS	MEDICATION	Y	N	HYPO 1	IMMEDIATLEY AFTER

14

Disease Milestones Variables and Related Domains: TM and SM

Trial Milestones (tm.xpt)

STUDYID	DOMAIN	MIDSTYPE	TMDEF	TMRPT
ABC	TM	HYPOGLYCEMIC EVENT	Hypoglycemic Event, the occurrence of a blood glucose concentration below the specified (by study) level of hypoglycemia	Y

Subject Milestones (sm.xpt)

STUDYID	DOMAIN	USUBJID	SMSEQ	MIDS	MIDSTYPE	SMSTDTC	SMENDTC	SMSTDY	SMENDY
ABC	SM	ABC-1001	2	HYPO 1	HYPOGLYCEMIC EVENT	2013-09-01T11:00	2013-09-01T11:00	25	25
ABC	SM	ABC-1001	3	HYPO 2	HYPOGLYCEMIC EVENT	2013-09-24T08:48	2013-09-24T08:48	50	50

cdisc

Epoch

- In most domains (EX is the usual exception), EPOCH is not collected, but derived.
- Although the SDTMIG does not specify EPOCH as required or expected, the FDA TCG says it should be included in all datasets.
- EPOCH allows a reviewer to quickly see whether events, interventions, and tests occurring before or during treatment.
  - For studies with multiple treatment periods, this includes which treatment period.
- Other timing variables (visit variables, study day variables) are not reliable for this purpose, since the timing of a subject's exposure to study treatment may deviate from the protocol plan.

cdisc

Linking with Identifiers

- Identifier Groups
- Alternatives and Additions to USUBJID
- Study-adaptable Identifiers
- FOCID
- Identifiers for Supplemental Qualifiers and RELREC

17

Identifier Groups

Always available STUDYID, DOMAIN, USUBJID, --SEQ	Replace or Supplement USUBJID POOLID, SPDEVID, NHOID, FOCID, APID
Adaptable at Study Level --GRPID, --REFID, --SPID, --LINKID, --LNKGRP	Others --BEATNO (EG only) FETUSID (SEND only) --RECID (see SENDIG v3.1)

18

The most basic identifiers

- STUDYID, DOMAIN and USUBJID are almost self-explanatory.
- --SEQ is always required in general observation class domains
- The variable name --SEQ and its label (Sequence Number) are somewhat misleading.
  - --SEQ is a record identifier, whose purpose is to distinguish between records
  - If –SEQ is assigned after data have been sorted, --SEQ values may reflect that sorting order, but –SEQ need not indicate any ordering.

Identifiers that can replace USUBJID: POOLID

- For data about a group of subjects that cannot be assigned to a specific subject.
- Commonly used in non-clinical studies, but allowable in human clinical trials

STUDYID	POOLID	USUBJID
CVD-3	SITEDEV2	301
CVD-3	SITEDEV2	306
CVD-3	SITEDEV2	307
CVD-3	SITEDEV2	309
CVD-3	SITEDEV2	312

STUDYID	POOLID	DVTERM	DVCAT
CVD-3	SITEDEV2	SITE SHUT DOWN	SITE DEVIATION

Identifiers that can replace USUBJID: APID

- For data about an associated person
- Used with RSUBJID and SREL.

qs.xpt

STUDYID	APID	RSUBJID	SREL	QSTEST	QSORRES
CT123	CT457	CT123-56	CAREGIVER	CTU01-PERFORMED BATHING/SHOWERING	Y
CT123	CT457	CT123-56	CAREGIVER	CTU02-PERFORMED DRESSING	Y
CT123	CT457	CT123-56	CAREGIVER	CTU03-PERFORMED FEEDING	Y
CT123	CT457	CT123-56	CAREGIVER	CTU04-PERFORMED GIVING MEDICATION	Y
CT123	CT457	CT123-56	CAREGIVER	CTU05-PERFORMED HOUSEKEEPING	Y
CT123	CT457	CT123-56	CAREGIVER	CTU06-PERFORMED TOILETING	Y
CT123	CT457	CT123-56	CAREGIVER	CTU01-PERFORMED MEAL PREPARATION	Y

21

SPDEVID: Can replace or supplement USBUJID

- In a study where the device, not a person, is the object of study, USUBJID may be absent.

de.xpt

STUDYID	USUBJID	SPDEVID	DETERM
ABC-123	2223	334-XRS-01	ALIGNMENT FAILURE
ABC-123		15033	DATA LOSS

- In studies with human study subjects, used to identify the device involved in an observation.

tr.xpt

STUDYID	USUBJID	SPDEVID	TRTEST	TRORRES	TRORRESU	TRMETHOD
XYZ	101	R6VZB30	LONGEST DIAMETER	14	mm	X-RAY
XYZ	101	R6VZB30	CALCIFICATION INDICATOR	Y		X-RAY

22

Identifiers that Supplement USUBJID

- NHOID indicates the organism on which the test is performed
  - USUBJID is still present, to identify the subject from whom the specimen was taken.

ms.xpt

STUDYID	USUBJID	NHOID	MSTEST	MSAGENT	MSCONC	MSCONCU	MSORRES
WXYZ	101	MYCOBACTERIUM TUBERCULOSIS	MICROBIAL SUSCEPIBILITY	RIFAMPICIN	1	ug/mL	RESISTANT
WXYZ	101	MYCOBACTERIUM TUBERCULOSIS	MICROBIAL SUSCEPIBILITY	ISONIAZID	0.2	ug/mL	SUSCEPTIBLE

- FOCID – covered in a later section

Linking and Grouping in Collection and SDTM

Within or across domains, across records / subjects	• **--CAT, --SCAT**
Within a single subject, within a domain/dataset	• --GRPID
Across Domains, within Subjects	• --LNKID/--LNKGRP • --REFID • --SPID / CMAENO/MHAENO/ DSAENO/PRAENO et al
Traceability	• CDASH to SDTM

--CAT and --SCAT

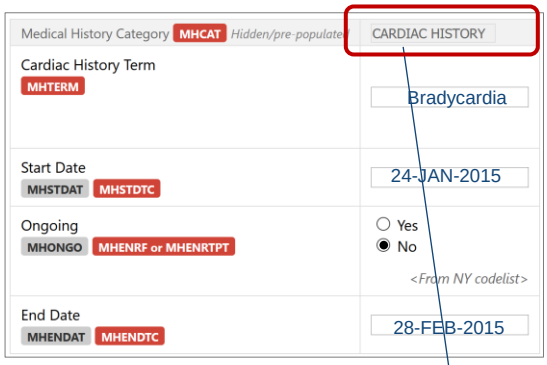
Grouping Qualifiers used to group records

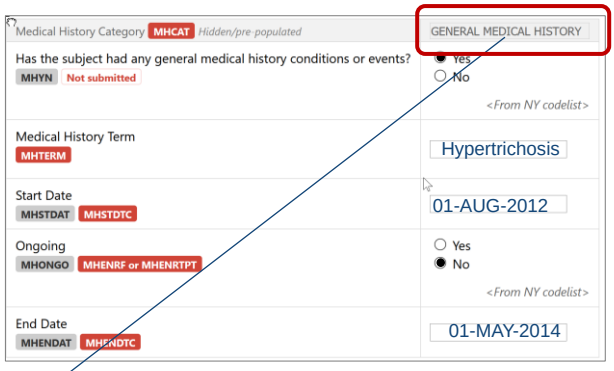
Generally sponsor-defined

- Do not replicate values existing elsewhere, e.g.,
  - Domain name
  - Dictionary classification provided by --DECOD and --BODSYS.

A _few_ domains specify uses and values

Using --CAT for CRF Sections



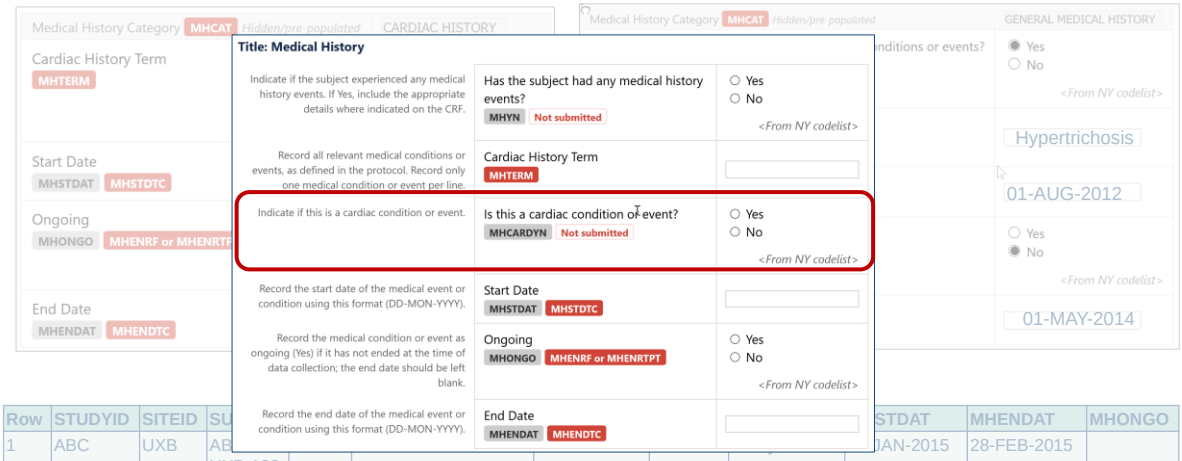


mh

Row	STUDYID	SITEID	SUBJID	MHYN	MHCAT	MHDAT	MHSPID	MHTERM	MHSTDAT	MHENDAT	MHONGO
1	ABC	UXB	ABC-UXB-123	Y	CARDIAC HISTORY	12-JUN-2019	1	Bradycardia	24-JAN-2015	28-FEB-2015	
2	ABC	UXB	ABC-UXB-123	Y	GENERAL MEDICAL HISTORY	12-JUN-2019	2	Hypertrichosis	01-AUG-2012	01-MAY-2014	

Have meaning within the domain

Using --CAT for CRF Sections

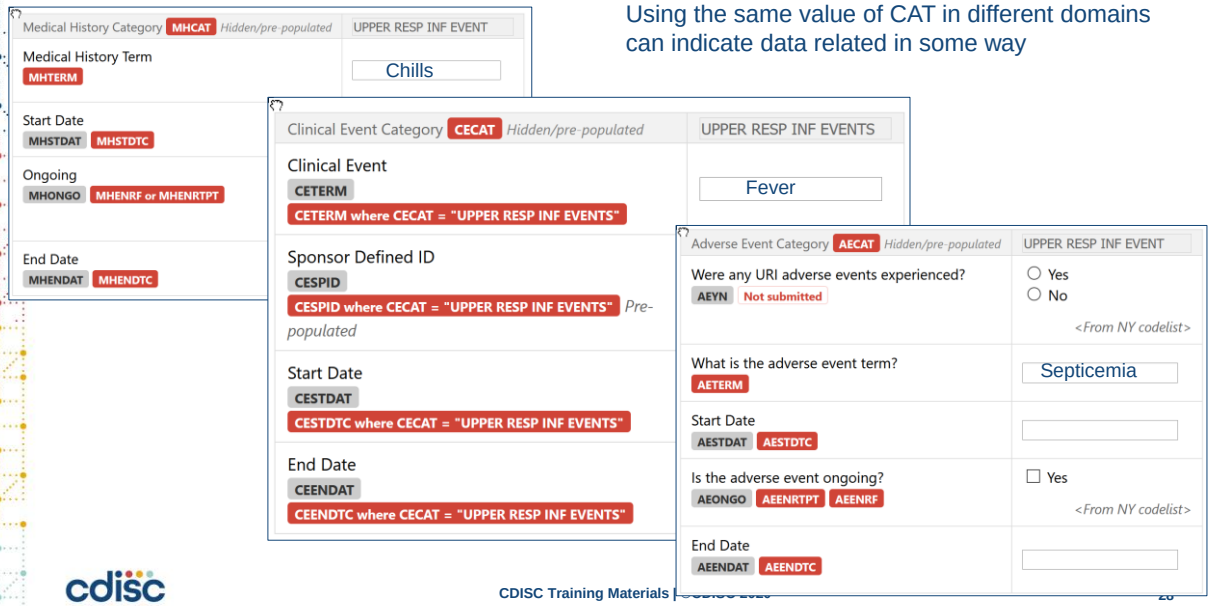


Row	STUDYID	SITEID	SU						STDAT	MHENDAT	MHONGO
1	ABC	UXB	AB UXB-123						JAN-2015	28-FEB-2015	
2	ABC	UXB	ABC-UXB-123	Y	GENERAL MEDICAL HISTORY	12-JUN-2019	2	Hypertrichosis	01-AUG-2012	01-MAY-2014	

27

Cross-Domain Use of --CAT

Using the same value of CAT in different domains can indicate data related in some way



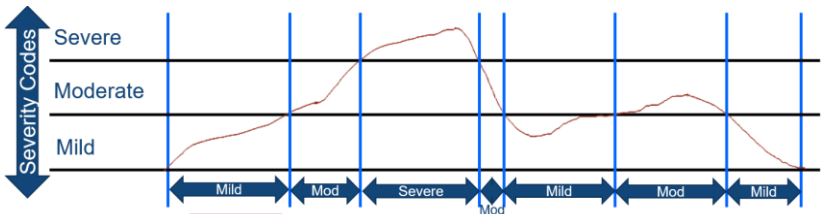
28

Linking and Grouping in Collection and SDTM

Within or across domains, across records / subjects	• --CAT, --SCAT
Within a single subject, within a domain/dataset	• **--GRPID**
Across Domains, within Subjects	• --LNKID/--LNKGRP • --REFID • --SPID / CMAENO/MHAENO/ DSAENO/PRAENO et al
Traceability	• CDASH to SDTM

29

Use of –GRPID - Capturing Each AE Severity Change



ae

Row	DOMAIN	SUBJID	AESPID	AEGRPID	AETERM	AEDECOD	AESEV	AESTDAT	AEENDAT
1	AE	101	1	1	NAUSEA	Nausea	MILD	15-APR-2019	28-APR-2019
2	AE	101	2	1	NAUSEA	Nausea	MODERATE	28-APR-2019	09-MAY-2019
3	AE	101	3	1	NAUSEA	Nausea	SEVERE	09-MAY-2019	29-MAY-2019
4	AE	101	4	1	NAUSEA	Nausea	MODERATE	29-MAY-2019	03-JUN-2019
5	AE	101	5	1	NAUSEA	Nausea	SEVERE	03-JUN-2019	09-JUN-2019
6	AE	101	6	1	NAUSEA	Nausea	MODERATE	09-JUN-2019	27-JUN-2019
7	AE	101	7	1	NAUSEA	Nausea	MILD	27-JUN-2019	10-JUL-2019

30

Combination Therapy Using --GRPID

cm.xpt

Row	DOMAIN	USUBJID	CMSEQ	CMGRPID	CMTRT	CMDECOD	CMDOSE	CMDOSU	CMSTDTC	CMENDTC
1	CM	6526	1	COMBO THPY 1	HYDROCHLOROTHIAZIDE	HYDROCHLOROTHIAZIDE	25	mg	2014-01-21	2014-01-22
2	CM	6526	2	COMBO THPY 1	PROPRANOLOL	PROPRANOLOL	40	mg	2014-01-21	2014-01-22
3	CM	8562	1	COMBO THPY 1	BUPROPION XL	BUPROPION XL	150	mg	2014-03-17	2014-03-25
4	CM	8562	2	COMBO THPY 1	VENLAFAXINE XL	VENLAFAXINE XL	200	mg	2014-03-17	2014-03-25
5	CM	8562	3	COMBO THPY 1	LITHIUM CARBONATE	LITHIUM CARBONATE	300	mg	2014-03-17	2014-03-25

31

Symptoms and Diagnoses

Example- Post Traumatic Stress (PTSD) Medical History CRF

Example CRF Instructions: Indicate any symptoms of Post Traumatic Stress (PTSD) the subject experienced and the date of diagnosis of PTSD,

Category MHCAT *Hidden/pre-populated* — PTSD HISTORY

Subcategory MHSCAT — MHSCAT="PTSD SYMPTOMS" when PTSD symptoms are reported *Hidden/pre-populated* — PTSD SYMPTOMS

Flashbacks FLASHBACKS_MHTERM MHTERM *Hidden/pre-populated* — FLASHBACKS

Indicate if the subject has experienced flashbacks of the traumatic event by checking Yes or No.

Did the subject have flashbacks of the traumatic event ? FLASHBACKS_MHOCCUR MHOCCUR — ○ Yes ○ No

Medical History Event Date Type FLASHBACKS_MHEVDTYP MHEVDTYP *Hidden/pre-populated* — SYMPTOM ONSET

Record the onset date of the specified symptom .

What was the start date of the flashbacks? FLASHBACKS_MHSTDAT MHSTDTC

Record the medical condition or event as ongoing ("Y") i… data collection

Are flashbacks ongoing? — ○ Yes

Post Traumatic Stress Disorder PTSD_MHTERM MHTERM *Hidden/pre-populated* — POST TRAUMATIC STRESS DISORDER

Medical History Event Date Type PTSD_MHEVDTYP MHEVDTYP *Hidden/pre-populated* — DIAGNOSIS

Record the PTSD diagnosis date.

What was the start date of the PTSD diagnosis? PTSD_MHSTDAT MHSTDTC

Record the DSM-5 code for the diagnosis.

What is the DSM-5 code? PTSD_MHDSM5CD NSV.MHDSM5CD

Record the method used to determine the diagnosis.

What is the diagnostic method? PTSD_MHDIAMTH NSV. MHDIAMTH

Symptoms and diagnosis all go into MH

Either hidden in EDC or generated upon export, MHGRPID can tie them together for SDTM

32

Using --GRPID to Group Diagnosis and Symptoms

mh.xpt

Row	MHGRPID	MHTERM	MHDECOD	MHCAT	MHSCAT	MHPRESP	MHOCCUR	MHSTDTC	MHDTC
1	PTSDDIAG	FLASHBACKS	Flashbacks	PTSD HISTORY	PTSD SYMPTOMS	Y	Y		2019-05-15
2	PTSDDIAG	IRRITABILITY	Irritability	PTSD HISTORY	PTSD SYMPTOMS	Y	Y		2019-05-15
3	PTSDDIAG	NIGHTMARES	Nightmares	PTSD HISTORY	PTSD SYMPTOMS	Y	Y		2019-05-15
4	PTSDDIAG	POST TRAUMATIC STRESS DISORDER	Post-traumatic stress disorder	PTSD HISTORY					2019-05-15
5		BROKEN RIGHT FEMUR	Fracture					2017-07-24	2019-05-15

33

Linking and Grouping in Collection and SDTM

Within or across domains, across records / subjects	• --CAT, --SCAT
Within a single subject, within a domain/dataset	• --GRPID
Across Domains, within Subjects	• --LNKID/--LNKGRP • **--REFID** • --SPID / CMAENO/MHAENO/ DSAENO/PRAENO et al
Traceability	• CDASH to SDTM

34

--REFID: Linking Procedures to Results

nv.xpt

Row	NVREFID	NVTESTCD	NVTEST	NVCAT	NVORRES	NVORRESU	NVMETHOD	NVANMETH	NVDTC	NVENDTC
1	KI3782	SOL	Sleep Latency	SLEEP STUDY	7	min	POLYSOMNOGRAPHY	AASM Scoring Manual	2016-10-15T23:00	2016-10-16T06:59
2	KI3782	TST	Total Sleep Time	SLEEP STUDY	301	min	POLYSOMNOGRAPHY	AASM Scoring Manual	2016-10-15T23:00	2016-10-16T06:59
3	KI3782	SE	Sleep Efficiency	SLEEP STUDY	62.71	%	POLYSOMNOGRAPHY	AASM Scoring Manual	2016-10-15T23:00	2016-10-16T06:59
4	KI3782	WASO	Awake After Sleep Onset	SLEEP STUDY	172	min	POLYSOMNOGRAPHY	AASM Scoring Manual	2016-10-15T23:00	2016-10-16T06:59
5	KI3782	STAGEN1	Stage 1 Sleep N1	SLEEP STUDY	61.5	min	POLYSOMNOGRAPHY	AASM Scoring Manual	2016-10-15T23:00	2016-10-16T06:59

pr.xpt

Row	PRREFID	PRSPID	PRTRT	PRCAT	PRSTDTC	PRENDTC
1	KI3782	PSG001	POLYSOMNOGRAPHY	SLEEP STUDY	2016-10-15T23:00	2016-10-16T06:59

relrec

STUDYID	RDOMAIN	USUBJID	IDVAR	IDVARVAL	RELTYPE	RELID
1999001	PR		PRREFID		ONE	PRNV1
1999001	NV		NVREFID		MANY	PRNV1

35

Linking and Grouping in Collection and SDTM

Within or across domains, across records / subjects	• --CAT, --SCAT
Within a single subject, within a domain/dataset	• --GRPID
Across Domains, within Subjects	• --LNKID/--LNKGRP • --REFID • **--SPID / CMAENO/MHAENO/ DSAENO/PRAENO et al**
Traceability	• CDASH to SDTM

36

Relating Individual Records from Separate Domains

ae

STUDYID	DOMAIN	SITEID	SUBJID	AESPID	AETERM
1999001	AE	03	0001	7	NAUSEA
1999001	AE	03	0001	11	HEADACHE

cm

STUDYID	DOMAIN	SITEID	SUBJID	CMSPID	CMTRT	CMAENO
1999001	CM	03	0001	1	MYLANTA	7
1999001	CM	03	0001	2	TYLENOL	11

ae.xpt

STUDYID	DOMAIN	USUBJID	AESEQ	AESPID	AETERM
1999001	AE	03-0001	1	7	NAUSEA
1999001	AE	03-0001	2	11	HEADACHE

cm.xpt

STUDYID	DOMAIN	USUBJID	CMSEQ	CMSPID	CMTRT	CMAENO
1999001	CM	03-0001	5	1	MYLANTA	7
1999001	CM	03-0001	9	2	TYLENOL	11

relrec

STUDYID	RDOMAIN	USUBJID	IDVAR	IDVARVAL	RELTYPE	RELID
1999001	CM	03-0001	CMSEQ	5		AECM1
1999001	AE	03-0001	AESEQ	1		AECM1
1999001	CM	03-0001	CMSEQ	9		AECM2
1999001	AE	03-0001	AESEQ	2		AECM2

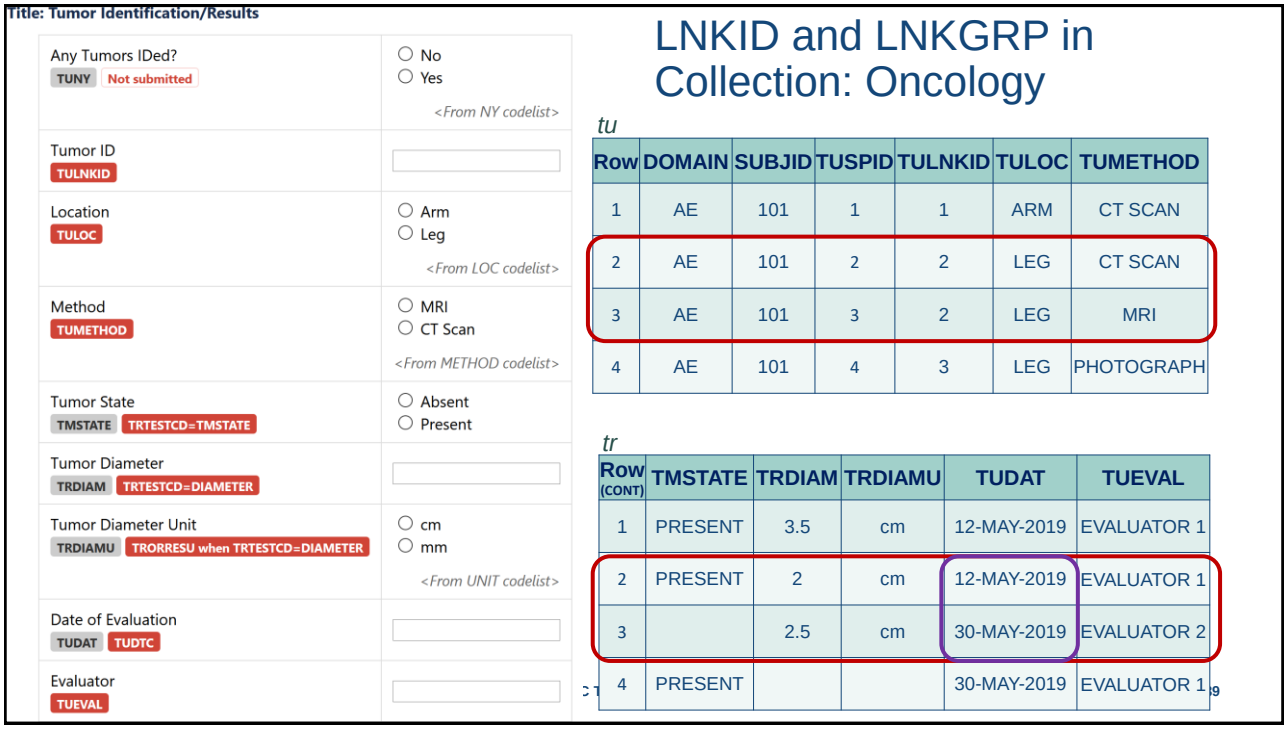
Related records are given same RELID. RELID is only unique within subject

37

Linking and Grouping in Collection and SDTM

Within or across domains, across records / subjects	• --CAT, --SCAT
Within a single subject, within a domain/dataset	• --GRPID
Across Domains, within Subjects	• **--LNKID/--LNKGRP** • --REFID • --SPID / CMAENO/MHAENO/ DSAENO/PRAENO et al
Traceability	• CDASH to SDTM

38



LNKID and LNKGRP in Collection: Oncology

tu

Row	DOMAIN	SUBJID	TUSPID	TULNKID	TULOC	TUMETHOD
1	AE	101	1	1	ARM	CT SCAN
2	AE	101	2	2	LEG	CT SCAN
3	AE	101	3	2	LEG	MRI
4	AE	101	4	3	LEG	PHOTOGRAPH

tr

Row (CONT)	TMSTATE	TRDIAM	TRDIAMU	TUDAT	TUEVAL
1	PRESENT	3.5	cm	12-MAY-2019	EVALUATOR 1
2	PRESENT	2	cm	12-MAY-2019	EVALUATOR 1
3		2.5	cm	30-MAY-2019	EVALUATOR 2
4	PRESENT			30-MAY-2019	EVALUATOR 1

39

--LNKID and –LNKGRP in SDTM: Oncology

TU: Tumor Identification
One record per tumor

TULNKID	TUTEST
1	Tumor Identification
2	Tumor Identification
3	Tumor Identification

tu.xpt

TR: Tumor Results
One or many records per tumor

TRLNKGRP	TRLNKID	TRTEST	TRORRES	TRMETHOD	TRDTC
Non-Target	1	Tumor State	PRESENT		2019-05-12
Non-Target	1	Diameter	3.5	CT SCAN	2019-05-12
Target	2	Tumor State	PRESENT		2019-05-12
Target	2	Diameter	2	CT SCAN	2019-05-12
Target	2	Diameter	2.5	MRI	2019-05-30
Non-Target	3	Tumor State	PRESENT		2019-05-30

tr.xpt

STUDYID	RDOMAIN	USUBJID	IDVAR	IDVARVAL	RELTYPE	RELID
20201	TU		TULNKID		ONE	TUTR1
20201	TR		TRLNKID		MANY	TUTR1

relrec

40

--LNKID and --LNKGRP in Other Cases

ec.xpt

STUDYID	DOMAIN	USUBJID	ECSEQ	ECLNKID	ECTRT	ECPRESP	ECOCCUR	ECDOSE	ECDOSU	ECDOSFRM	ECDOSFRQ	ECROUTE	ECLOC	ECLAT
ABC	EC	ABC3001	1	V3	DRUG X	Y	Y	5	mL	INJECTION	ONCE	SUBCUTANEOUS	ABDOMEN	LEFT
ABC	EC	ABC3001	2	V3	DRUG X	Y	Y	5	mL	INJECTION	ONCE	SUBCUTANEOUS	ABDOMEN	CENTER

relrec

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

ex.xpt

STUDYID	DOMAIN	USUBJID	EXSEQ	EXLNKID	EXTRT	EXDOSE	EXDOSU	EXDOSFRM	EXDOSFRQ	EXROUTE	EXLOC
ABC	EX	ABC3001	1	V3	DRUG X	2	mg/kg	INJECTION	ONCE	SUBCUTANEOUS	ABDOMEN

41

Why Not --GRPID or --SEQ or --CAT?

--GRPID

Groups records within a subject *within a domain*

--SEQ

Only has meaning within a subject *within a domain*

--CAT

Groups related topic values; in some domains has a defined role

42

Linking and Grouping in Collection and SDTM

Within or across domains, across records / subjects	• --CAT, --SCAT
Within a single subject, within a domain/dataset	• --GRPID
Across Domains, within Subjects	• --LNKID/--LNKGRP • --REFID • --SPID / CMAENO/MHAENO/ DSAENO/PRAENO et al
Traceability	• **CDASH to SDTM**

43

Supporting Traceability Back to CDASH

Traceability is critically important to ensuring data integrity

Traceability is built into ADaM back into SDTM, and SDTM to the SDTM-annotated CRF but not into CDASH

There are some examples of CDASH variables that can help

- --SPID
- --GRPID

44

--GRPID to Link Horizontal to Vertical

vs

SUBJID	VSGRPID	VSTPT	VSDAT	VSTIM	BPPOS	SYSBP	DIABP	PULSE	TEMP	TEMPU	HEIGHT	HEIGHTU	WEIGHT	WEIGHTU	RESP
031	4	Pre-dose	19-JUN-2017	08:45	Sitting	154	74	72	37.4	C	157	cm	90.5	kg	20

vs.xpt

USUBJID	VSGRPID	VSTESTCD	VSTEST	VSPOS	VSORRES	VSORRESU	VSDTC	VSTPT
C277-031	4	SYSBP	Systolic Blood Pressure	SITTING	154	mmHg	2017-06-19T08:45	Pre-dose
C277-031	4	DIABP	Diastolic Blood Pressure	SITTING	74	mmHg	2017-06-19T08:45	Pre-dose
C277-031	4	PULSE	Pulse		72	beats/min	2017-06-19T08:45	Pre-dose
C277-031	4	TEMP	Temperature		34.5	C	2017-06-19T08:45	Pre-dose
C277-031	4	HEIGHT	Height		157	cm	2017-06-19T08:45	Pre-dose
C277-031	4	WEIGHT	Weight		95.5	kg	2017-06-19T08:45	Pre-dose
C277-031	4	RESP	Respiratory Rate		20	breaths/min	2017-06-19T08:45	Pre-dose

45

Using --SPID to Provide Traceability in Coding Splits

CRF Collection

CMSPID	CMTRT
1	Sudafed and Benadryl

cm

SDTM Dataset

CMSEQ	CMSPID	CMTRT	CMMODIFY	CMDECOD
10	1	Sudafed and Benadryl	Sudafed	pseudoephredrine
11	1	Sudafed and Benadryl	Benadryl	diphenhydramine

cm.xpt

46

Some Data Linking Variables in CDASH and SDTM

--CAT/SCAT*	--GRPID*	--REFID*	--LNKID**	--LNKGRP**	--SPID*	--AENO†	--SEQ‡
Group records within one or more domains/CRFs in the same way; using --SCAT requires --CAT	Groups records within a subject within a domain	Groups within a domain, may be referenced across domains	Links one or more records in one domain to one or more in another; or can link records in a domain	Groups records in or across domains. Does not require the use of –LNKID	Sponsor defined; CDASH uses it as a line number	Holds a variable value representing a link from another domain, often AESPID	With STUDYID, DOMAIN and USUBJID, creates a unique record reference
Values known before the study; usually set in collection, rather than collected; may have CDISC CT	Values often not known prior to study; may have sponsor CT	Values are usually captured as data	Values are often set to make the linking work correctly if a RELREC is used; not usually collected	Other than in Oncology, no restrictions on use	As a line number, values are generated, often by the system, as data are captured	Site enters the AESPID value to show a relationship between the AE and the target domain record	Values set when SDTM dataset generated; good choice for IDVAR in SUPP--; may be less useful for RELREC
Values have meaning across subjects and across domains	Values only have meaning within the domain and subject	Values can group multiple results of one test in the same domain, or link procedures to results	Values have meaning within subjects across domains if used in a RELREC	Values have meaning within subjects across domains if used in a RELREC	May be used during data collection to capture relationships across domains	Values identify specific records. In SDTM, RELREC can be used to represent this relationship.	Values only have meaning within the domain and subject

** Used in CDASH and SDTM; **Used in Gen Obs Classes in SDTM, only modeled in oncology in CDASH; † Used only in CDASH; ‡ Used only in SDTM*

47

47

Notes for Data Linking Variables in CDASH and SDTM

The information grid is intended to provide insight into linking variables that can be used in SDTM and/or CDASH.

- It shows where some variables may be preferred, or less optimal, and indicates where values may originate.

Any variable present in SDTM can be used in CDASH provided it is appropriate. The footnotes on the table show the variables that are *generally* used in each.

Variables referenced in the information cells may be named differently in sponsor systems.

- For example, AESPID in CDASH may be called something else in an EDC system and mapped into AESPID during data export.

Most linking variables have defined use cases in certain domains, for example, TULNKID and TRLNKGRP have a specific use in oncology domains.

- As a result, they are not available for any other uses in those domains.
- Users should check the IGs prior to implementation.

Identical values in corresponding variables in different domains have no innate relationship unless otherwise stated. A RELREC is needed to define links across datasets.

- For example, PRREFID=1 and NVREFID=1 only have a relationship if there is a RELREC record linking them.

48

The FOCID Variable

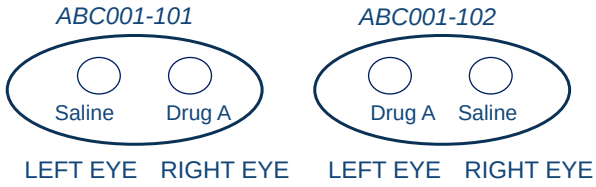
- Added in SDTM v1.7 and SDTMIG v3.3
- An Identifier variable that has no domain prefix.
- Describes a focus of specific interest (e.g., body location) the same way across all domains
  - Example: the right eye might be treated (data in EX) and then evaluated, with results in OE.
- Implementations outside of OE will likely use protocol-defined terminology.
- The Findings variables --LOC (e.g., EYE) and --LAT (e.g., RIGHT), and to a lesser extent, --DIR, and --PORTOT may also be used.

cdisc

49

49

FOCID Example*



ex.xpt

Row	STUDYID	DOMAIN	USUBJID	FOCID	EXSEQ	EXTRT	EXLOC	EXLAT
1	ABC001	EX	ABC001-101	OS	1	Saline	EYE	LEFT
2	ABC001	EX	ABC001-101	OD	2	Drug A	EYE	RIGHT
3	ABC001	EX	ABC001-102	OS	1	Drug A	EYE	LEFT
4	ABC001	EX	ABC001-102	OD	2	Saline	EYE	RIGHT

oe.xpt

Row	STUDYID	DOMAIN	USUBJID	FOCID	OESEQ	OETESTCD	OEORRES	EXLOC	EXLAT
1	ABC001	OE	ABC001-101	OS	1	EDEMA	ABSENT	EYE	LEFT
2	ABC001	OE	ABC001-101	OD	2	EDEMA	PRESENT	EYE	RIGHT
3	ABC001	OE	ABC001-102	OS	1	EDEMA	ABSENT	EYE	LEFT
4	ABC001	OE	ABC001-102	OD	2	EDEMA	ABSENT	EYE	RIGHT

cdisc

** Not all Required and Expected variables shown to conserve space*

50

50

Linking Records, Datasets, and Non-Standard Variables

Considerations in Choosing IDVAR Values for RELREC

- The SDTMIG shows various variables for IDVAR:
  - --SPID
  - --GRPID
  - --REFID
  - --LNKID
  - --LNKGRP
  - --SEQ
- If --SEQ is used and assigned after data collection, then relationships may not be reflected accurately.
- -SEQ does not work well in dataset-to-dataset relationships
- The IDVAR chosen for RELREC should always be in the collected (or database) record.

Linking Records, Datasets, and Non-Standard Variables

Considerations in Choosing IDVAR Values for SUPP--

- The SDTMIG shows various variables for IDVAR:
  - --SEQ
  - --REFID
  - --TSTDTL
  - --GRPID
  - --CAT
  - --SPID
- If --SEQ is assigned systematically after data collection, then it can be used consistently in SUPP-- datasets
- While other IDVAR values may decrease the number of records in a SUPP-- dataset, their use requires that all implementers remember domain-specific processes.

Linking Comments

53

Linking Comments

- Good practice is to limit comment collection.
- However, if you must collect comments, they need to be linked to whatever they are commenting on.
- The Comments domain therefore has a variety of ways to link comments to other data, described in the SDTMIG.

54

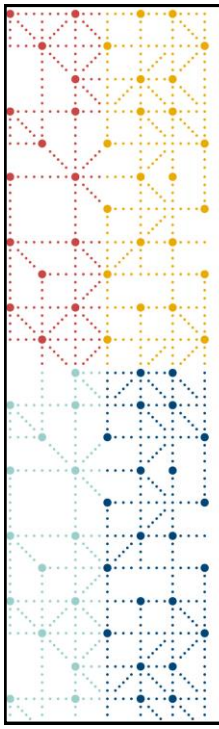
Conclusions

55

Take-aways

- SDTM provides a wide range of ways to link data.
  - These include routine linking by study, subject, and visit
  - Disease Milestones are for non-visit triggered data collection
  - FOCID if multiple parts of the body are treated and assessed
- Many variables are available for study-specific linking needs
  - These have different properties and limitations
- The best choices for IDVAR are different for SUPPQUAL and RELREC
- Linking, especially linking across domains, needs to be planned
  - If a CRF module will be split across SDTM domains, how will you put it back together for analysis?

56



Questions?

Diane Wold diane.wold@cdisc.org

Fred Wood fred.wood@datastandardsconsulting.com

Kit Howard khoward@cdisc.org

Many thanks to Gary Walker for several slides in the data collection section