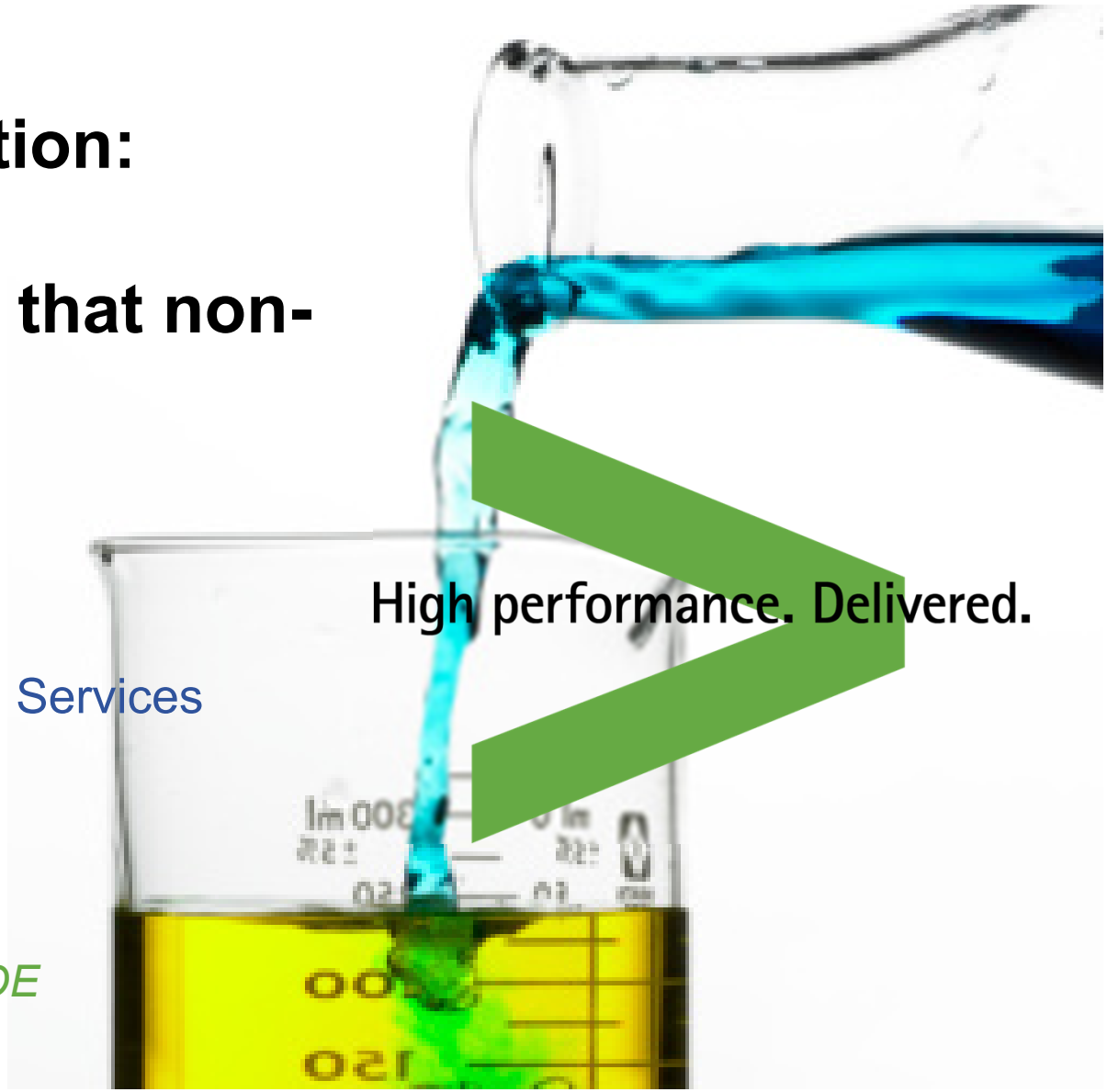
accenture

consulting | technology | outsourcing

SDTM Implementation:

What to do with all that non-standard data?

Jerry Salyers
Senior Consultant
Data Standards Consulting
Accenture Accelerated R&D Services



San Francisco PharmaSUG SDE
February 10 2015

Introduction and Overview

- Supplemental qualifiers
  - Overview
  - Key principles
  - NSV Challenges
  - Examples and proper use of IDVAR/IDVARVAL
- When should it be FA?
  - Should it just be a “custom” Findings domain?
  - Findings About Overview
  - Difference between a supplemental qualifier and FA
  - Examples of each

Supplemental Qualifiers Overview (1)

MH Dataset

STUDYID	DOMAIN	USUBJID	MHSEQ	MHTERM	MHDTC	MHSTDTC
1999001	MH	0001	1	MULTIPLE MYELOMA	2007-05-11	2004-01-05

SUPPMH Dataset

STUDYID	RDOMAIN	USUBJID	IDVAR	IDVARVAL	QNAM	QLABEL	QVAL	QORIG
1999001	MH	0001	MHSEQ	1	BASIS	Basis for Diagnosis	HISTOPATHOLOGY	CRF
1999001	MH	0001	MHSEQ	1	STAGE	Disease Stage	STAGE 3	CRF
1999001	MH	0001	MHSEQ	1	THRPSTAT	Therapy Status	RELAPSED	CRF

The Relationship

The Data

Supplemental Qualifier Key Principles

- Non-Standard variables in SUPP-- are no less important than the standard Qualifiers.
- One dataset (SUPP--.xpt) for each parent domain that has non-standard variables (NSVs).
- Every record must relate back to at least one valid parent record.
  - If the IDVAR is something other than --SEQ, a single SUPP record may relate to more than one parent record.
- -- QNAMs and QVALs relate only to parent records, but not to each other.
- Data (QVALs) share the timing of the parent record.
- Not for submitting “extra” data such as separate day, month, or year as it may have been collected, or for submitting data that may have been mapped to controlled terminology

Supplemental Qualifiers are Not Always the Best Solution for NSVs

- One can't always conclude that all legacy that doesn't map into a single SDTM-based domain must be a Supplemental Qualifier.
- In some cases, the mapping of non-standard variables into Supplemental Qualifiers resulted in data with different time frames and/or different data structures being shown in merged views of data.
- One must always think of what the resulting merged view will look like in assessing:
  - Which NSVs are submitted as Supplemental Qualifiers
  - The proper selection of IDVAR

Choice of IDVAR

- Caution is needed when choosing the IDVAR, in order to avoid the potential for unintended consequences.
- Consider the following example:

STUDYID	RDOMAIN	USUBJID	IDVAR	IDVARVAL	QNAM	QLABEL	QVAL	QORIG
2008-0001	LB	2008-001	VISITNUM	3	URINTEMP	Urine Temperature within Expected Range	Y	CRF

- The URINTEMP value of 'Y' will appear on all LB records where VISITNUM = 3

Pros	Cons
If the study collected only urine samples, the selection of VISITNUM as the IDVAR could be appropriate.	If the study also collected blood samples, the urine-temperature information would also appear on these records.
It may be desirable to know about the urine temperature when looking at each individual urine result.	The urine temperature will appear on all urine tests and might not be relevant to all of them.

Is the Use of Supplemental Qualifiers the Best Solution? Physical Exam Case Study (1)

- Physical Exam CRF

Body System	Description of Abnormality	
Abdomen		
Extremities/Joints		
General Appearance		
Heart		
HEENT		
Lungs		
Lymph Nodes		
Mental Status		
Neurologic		
Reflexes		
Skin		
Disease Relapse Since Last Visit?	□ Yes □ No	

Sponsors are often tempted to model the last question as a Supplemental Qualifier

PE Dataset (Partial)

STUDYID	DOMAIN	USUBJID	PESEQ	PETESTCD	PETEST	PEORRES	PESTRESC
2001-01	PE	2001-01-1008	1	ABDOMEN	Abdomen	NORMAL	NORMAL
2001-01	PE	2001-01-1008	2	EXTRJOIN	Extremities/ Joints	JOINTS SWOLLEN IN FINGERS	JOINTS SWOLLEN IN FINGERS
2001-01	PE	2001-01-1008	3	GENAPP	General Appearance	NORMAL	NORMAL
2001-01	PE	2001-01-1008	4	HEART	Heart	NORMAL	NORMAL
2001-01	PE	2001-01-1008	5	HEENT	HEENT	NORMAL	NORMAL
2001-01	PE	2001-01-1008	6	LUNGS	Lungs	NORMAL	NORMAL
2001-01	PE	2001-01-1008	7	LYMPNODE	Lymph Nodes	NORMAL	NORMAL
2001-01	PE	2001-01-1008	8	MENTSTAT	Mental Status	NORMAL	NORMAL
2001-01	PE	2001-01-1008	9	NEURO	Neurologic	NORMAL	NORMAL
2001-01	PE	2001-01-1008	10	REFLEXES	Reflexes	NORMAL	NORMAL
2001-01	PE	2001-01-1008	11	SKIN	Skin	NORMAL	NORMAL

SUPPPE Representation

Method 1: IDVAR = 'VISIT'

STUDYID	RDOMAIN	USUBJID	IDVAR	IDVARVAL	QNAM	QLABEL	QVAL	QORIG
2001-01	PE	2001-01-1008	VISIT	1	PERELFL	Relapse Since Last Visit	N	CRF

Method 2: IDVAR = 'PESEQ'

STUDYID	RDOMAIN	USUBJID	IDVAR	IDVARVAL	QNAM	QLABEL	QVAL	QORIG
2001-01	PE	2001-01-1008	PESEQ	1	PERELFL	Relapse Since Last Visit	N	CRF
2001-01	PE	2001-01-1008	PESEQ	2	PERELFL	Relapse Since Last Visit	N	CRF
2001-01	PE	2001-01-1008	PESEQ	3	PERELFL	Relapse Since Last Visit	N	CRF
2001-01	PE	2001-01-1008	PESEQ	4	PERELFL	Relapse Since Last Visit	N	CRF
2001-01	PE	2001-01-1008	PESEQ	5	PERELFL	Relapse Since Last Visit	N	CRF
2001-01	PE	2001-01-1008	PESEQ	6	PERELFL	Relapse Since Last Visit	N	CRF
2001-01	PE	2001-01-1008	PESEQ	7	PERELFL	Relapse Since Last Visit	N	CRF
2001-01	PE	2001-01-1008	PESEQ	8	PERELFL	Relapse Since Last Visit	N	CRF
2001-01	PE	2001-01-1008	PESEQ	9	PERELFL	Relapse Since Last Visit	N	CRF
2001-01	PE	2001-01-1008	PESEQ	10	PERELFL	Relapse Since Last Visit	N	CRF
2001-01	PE	2001-01-1008	PESEQ	11	PERELFL	Relapse Since Last Visit	N	CRF

Merged View – Do the different SUPP representations result in a different merged view?

STUDYID	DOMAIN	USUBJID	PESEQ	PETESTCD	PETEST	PEORRES	PESTRESC	PERELFL
2001-01	PE	2001-01-1008	1	ABDOMEN	Abdomen	NORMAL	NORMAL	N
2001-01	PE	2001-01-1008	2	EXTRJOIN	Extremities/ Joints	JOINTS SWOLLEN IN FINGERS	JOINTS SWOLLEN IN FINGERS	N
2001-01	PE	2001-01-1008	3	GENAPP	General Appearance	NORMAL	NORMAL	N
2001-01	PE	2001-01-1008	4	HEART	Heart	NORMAL	NORMAL	N
2001-01	PE	2001-01-1008	5	HEENT	HEENT	NORMAL	NORMAL	N
2001-01	PE	2001-01-1008	6	LUNGS	Lungs	NORMAL	NORMAL	N
2001-01	PE	2001-01-1008	7	LYMPNODE	Lymph Nodes	NORMAL	NORMAL	N
2001-01	PE	2001-01-1008	8	MENTSTAT	Mental Status	NORMAL	NORMAL	N
2001-01	PE	2001-01-1008	9	NEURO	Neurologic	NORMAL	NORMAL	N
2001-01	PE	2001-01-1008	10	REFLEXES	Reflexes	NORMAL	NORMAL	N
2001-01	PE	2001-01-1008	11	SKIN	Skin	NORMAL	NORMAL	N

Regardless of which IDVAR is chosen, the merged view is the same. Method 1 just simply collapses 11 SUPPPE records to a single SUPPPE record (limited usefulness).

Alternative Representation: No SUPPPE

STUDYID	DOMAIN	USUBJID	PESEQ	PETESTCD	PETEST	PEORRES	PESTRESC	PERELFL
2001-01	PE	2001-01-1008	1	ABDOMEN	Abdomen	NORMAL	NORMAL	N
2001-01	PE	2001-01-1008	2	EXTRJOIN	Extremities/ Joints	JOINTS SWOLLEN IN FINGERS	JOINTS SWOLLEN IN FINGERS	N
2001-01	PE	2001-01-1008	3	GENAPP	General Appearance	NORMAL	NORMAL	N
2001-01	PE	2001-01-1008	4	HEART	Heart	NORMAL	NORMAL	N
2001-01	PE	2001-01-1008	5	HEENT	HEENT	NORMAL	NORMAL	N
2001-01	PE	2001-01-1008	6	LUNGS	Lungs	NORMAL	NORMAL	N
2001-01	PE	2001-01-1008	7	LYMPNODE	Lymph Nodes	NORMAL	NORMAL	N
2001-01	PE	2001-01-1008	8	MENTSTAT	Mental Status	NORMAL	NORMAL	N
2001-01	PE	2001-01-1008	9	NEURO	Neurologic	NORMAL	NORMAL	N
2001-01	PE	2001-01-1008	10	REFLEXES	Reflexes	NORMAL	NORMAL	N
2001-01	PE	2001-01-1008	11	SKIN	Skin	NORMAL	NORMAL	N
2001-01	PE	2001-01-1008	12	PERELFL	Relapse Since Last Visit	N	N	

The creation of a new record, rather than the creation of a Supplemental Qualifier, shows that the Relapse Question is related to the physical exam as a whole, and not to every body system.

“Extra” Data – Submitting data “as collected” as well as mapped to Controlled Terminology (1)

Adverse Event CRF (Partial)

Onset Date (dd-MMM-yyyy): _____ **AESTDTC**

Resolution Date (dd-MMM-yyyy): _____ **AEENDTC**

Outcome:

Resolved	☐	**AEOUT**
Ongoing	☐	**AEENRTPT**
Died	☐	
Change in Grade or Seriousness	☐	

"Extra" Data – Submitting data "as collected" as well as mapped to Controlled Terminology (2)

- AE Dataset – Outcome mapped to CT

STUDYID	DOMAIN	USUBJID	AESEQ	AETERM	AEOUT	AESTDTC	AEENRTPT	AEENTPT
ABC15-0806	AE	ABC15-0806-001	7	EJECTION FRACTION DECREASED	NOT RECOVERED/NOT RESOLVED	2011-03-22	ONGOING	TRIAL EXIT
ABC15-0806	AE	ABC15-0806-001	20	SINUS TACHYCARDIA	NOT RECOVERED/NOT RESOLVED	2011-03-22	ONGOING	TRIAL EXIT
ABC15-0806	AE	ABC15-0806-003	1	ALOPECIA	NOT RECOVERED/NOT RESOLVED	2011-03-22	ONGOING	TRIAL EXIT
ABC15-0806	AE	ABC15-0806-003	9	MACULAR UPPER CHEST RASH	NOT RECOVERED/NOT RESOLVED	2011-03-17	ONGOING	TRIAL EXIT
ABC15-0806	AE	ABC15-0806-005	1	ALOPECIA	NOT RECOVERED/NOT RESOLVED	2011-03-29	ONGOING	TRIAL EXIT
ABC15-0806	AE	ABC15-0806-005	6	FATIGUE	NOT RECOVERED/NOT RESOLVED	2011-06-16	ONGOING	TRIAL EXIT
ABC15-0806	AE	ABC15-0806-005	9	RIGHT CHEST WALL PAIN	NOT RECOVERED/NOT RESOLVED	2011-05-10	ONGOING	TRIAL EXIT
ABC15-0806	AE	ABC15-0806=005	11	SINUS TACHYCARDIA	NOT RECOVERED/NOT RESOLVED	2011-03-08	ONGOING	TRIAL EXIT

Collected value of "Ongoing" mapped to Controlled Terminology in AEOUT as well as to an "end" relative timing variable.

“Extra” Data – Submitting data “as collected” as well as mapped to Controlled Terminology (3)

STUDYID	RDOMAIN	USUBJID	IDVAR	IDVARVAL	QNAM	QLABEL	QVAL	QORIG
ABC15-0806	AE	ABC15-0806-001	AESEQ	7	AEOUT	AE Outcome	Ongoing	CRF
ABC15-0806	AE	ABC15-0806-001	AESEQ	20	AEOUT	AE Outcome	Ongoing	CRF
ABC15-0806	AE	ABC15-0806-003	AESEQ	1	AEOUT	AE Outcome	Ongoing	CRF
ABC15-0806	AE	ABC15-0806-003	AESEQ	9	AEOUT	AE Outcome	Ongoing	CRF
ABC15-0806	AE	ABC15-0806-005	AESEQ	1	AEOUT	AE Outcome	Ongoing	CRF
ABC15-0806	AE	ABC15-0806-005	AESEQ	6	AEOUT	AE Outcome	Ongoing	CRF
ABC15-0806	AE	ABC15-0806-005	AESEQ	9	AEOUT	AE Outcome	Ongoing	CRF
ABC15-0806	AE	ABC15-0806-005	AESEQ	11	AEOUT	AE Outcome	Ongoing	CRF

Collected outcome of “Ongoing” also mapped to SUPPAE with an invalid QNAM (QNAM cannot be the same as a valid SDTM variable name).

Supplemental Qualifiers is not a place to submit ambiguous or “coded” data

STUDYID	RDOMAIN	USUBJID	IDVAR	IDVARVAL	QNAM	QLABEL	QVAL	QORIG
2005-ABC	CM	ABC-0101	CMSEQ	1	CMCONT	Continuing	X	CRF
2005-ABC	CM	ABC-0102	CMSEQ	1	CMCONT	Continuing	X	CRF
2005-ABC	CM	ABC-0102	CMSEQ	2	CMCONT	Continuing	X	CRF
2005-ABC	CM	ABC-0104	CMSEQ	1	CMCONT	Continuing	X	CRF
2005-ABC	CM	ABC-0104	CMSEQ	3	CMCONT	Continuing	X	CRF

This sponsor is attempting to submit the actual act of “checking the box” rather than mapping to a relative timing variable and utilizing controlled terminology

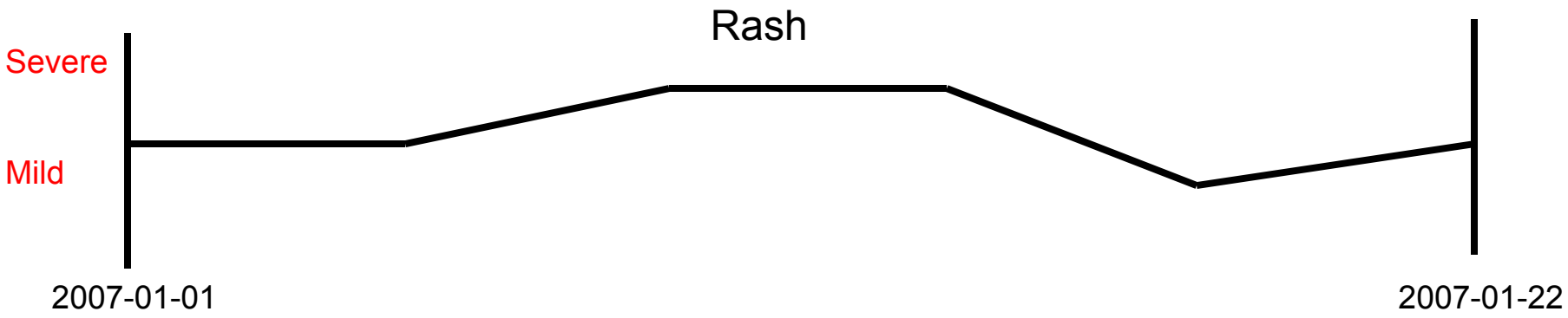
Findings About: Overview

- Specialization of the Findings general observation class
  - --OBJ is needed
  - --OBJ should not be redundant with –TESTCD, --CAT, or domain name
  - --OBJ is used to describe the name of an event or intervention (however a “parent” record is not required for the use of FA)
- Used when data or observations cannot be adequately represented in SUPPQUAL
  - SUPPQUAL requires a parent record
  - SUPPQUAL records inherit the timing of the parent record
  - SUPPQUAL has no way to relate records to each other since all are related only to the parent record
    - Multiple results and units
    - Multiple Qualifiers and encoded values

Supplemental Qualifiers vs. Findings About vs. Findings – Compare and Contrast

Characteristic	SUPPQUAL	Findings About	Findings
CDISC Controlled Terminology	Only for a few QNAM values	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 modeled domains
Timing	Relies on parent record	--DTC for --TESTCD/--TEST	--DTC for --TESTCD/--TEST
Uniqueness 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 Event or Intervention record	Always	Likely	Possible

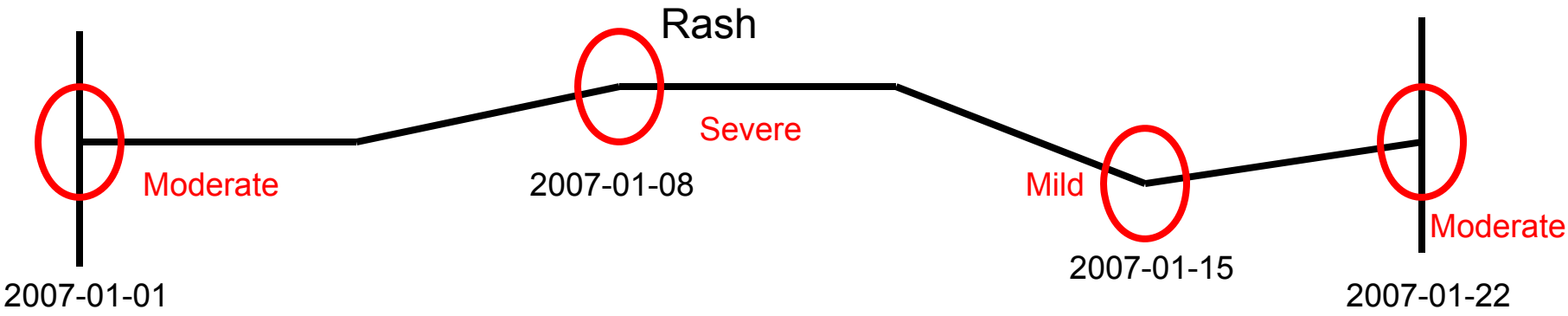
Findings About: Example (1)



DOMAIN	USUBJID	CESEQ	CETERM	CESEV	CESTDTC	CEENRF
CE	123-01-01	1	Rash	Severe	2007-01-01	AFTER

This study collects the rash as an Clinical Event.

Findings About: Example (2)



DOMAIN	USUBJID	CESEQ	CETERM	CESEV	CESTDTC	CEENRF
CE	123-01-01	1	Rash	Severe	2007-01-01	AFTER

This study also collects rash severity at various times.

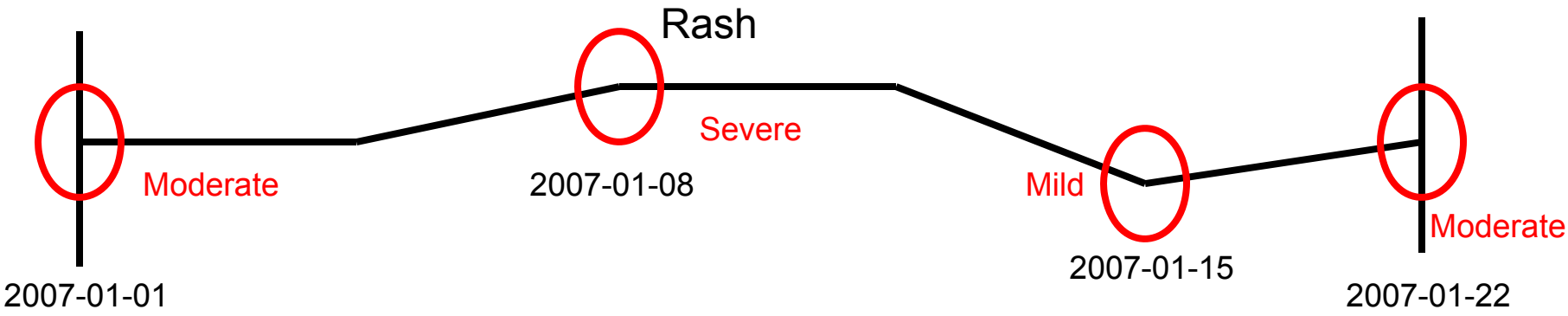
Why SUPPAE Would Not Work

STUDYID	RDOMAIN	USUBJID	IDVAR	IDVARVAL	QNAM	QLABEL	QVAL
1999001	AE	123-01-01	AESEQ	1	AESEV1	Severity 1	Moderate
1999001	AE	123-01-01	AESEQ	1	AESEV2	Severity 2	Severe
1999001	AE	123-01-01	AESEQ	1	AESEV3	Severity 3	Mild
1999001	AE	123-01-01	AESEQ	1	AESEV4	Severity 4	Moderate
1999001	AE	123-01-01	AESEQ	1	AESEV1 DT	Severity 1 Date	2009-01-01
1999001	AE	123-01-01	AESEQ	1	AESEV2 DT	Severity 2 Date	2009-01-08
1999001	AE	123-01-01	AESEQ	1	AESEV3 DT	Severity 3 Date	2009-01-15
1999001	AE	123-01-01	AESEQ	1	AESEV4 DT	Severity 4 Date	2009-01-22

Problems:

- No machine-readable, unambiguous relationship between each date and its severity
- SUPP-- records inherit timing from the parent record

Findings About: Example (3)



DOMAIN	USUBJID	CESPID	CETERM	CESEV	CESTDTC	CEENRF
CE	123-01-01	1	Rash	Severe	2007-01-01	AFTER

DOMAIN	USUBJID	FASEQ	FASPID	FATESTCD	FATEST	FAOBJ	FAORRES	FADTC
FA	123-01-01	1	1	SEV	Severity	Rash	Moderate	2007-01-01
FA	123-01-01	2	1	SEV	Severity	Rash	Severe	2007-01-08
FA	123-01-01	3	1	SEV	Severity	Rash	Mild	2007-01-15
FA	123-01-01	4	1	SEV	Severity	Rash	Moderate	2007-01-22

Findings About: Example (4)

Using RELREC

STUDYID	RDOMAIN	USUBJID	IDVAR	IDVARVAL	RELTYPE	RELID
1999001	CE		CESPID		ONE	CEFA
1999001	FA		FASPID		MANY	CEFA

DOMAIN	USUBJID	CESPID	CETERM	CESEV	CESTDTC	CEENRF
CE	ABC-0001	1	Rash	Severe	2007-01-01	AFTER

DOMAIN	USUBJID	FASEQ	FASPID	FATESTCD	FATEST	FAOBJ	FAORRES	FADTC
FA	ABC-0001	1	1	SEV	Severity	Rash	Moderate	2007-01-01
FA	ABC-0001	2	1	SEV	Severity	Rash	Severe	2007-01-08
FA	ABC-0001	3	1	SEV	Severity	Rash	Mild	2007-01-15
FA	ABC-0001	4	1	SEV	Severity	Rash	Moderate	2007-01-22

Submitting FA – Splitting Domains and Naming of Datasets

- Unlike SUPP datasets, FA doesn't have the "RDOMAIN" variable to easily identify the "parent" domain
- Sponsors may choose to split FA into physically separate datasets based on the parent domain of the value in --OBJ.
- Dataset name would combine 'FA' with the two-letter domain code of the parent record (e.g., FAMH, FAAE, FACM)
- Value of "DOMAIN" would remain 'FA' in all the split datasets.
- Conform to additional guidance in section 4.1.1.7 of the SDTMIG:
  - --SEQ must be *unique* within USUBJID across the split datasets.
  - Split dataset names can be up to four characters in length.
  - SUPP-- datasets for split domains would also be split, and can have names up to eight characters, beginning with "SQ" rather than "SUPP" (e.g., SQFAMH).
- With SDTMIG v3.2, the first Findings About domain without the two-letter FA prefix was introduced – Skin Response (SR)

The End