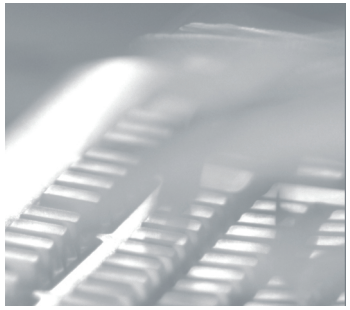
SDTM Amendment 1
and
CDER Data Standards
Common Issues

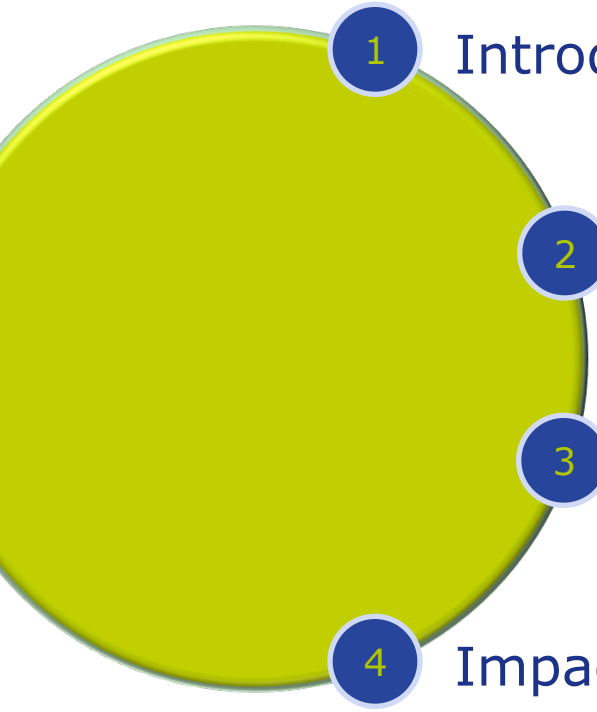

DRIVE YOUR PERFORMANCE

Peter Van Reusel
Business Unit Director Projects
Business & Decision Life Sciences

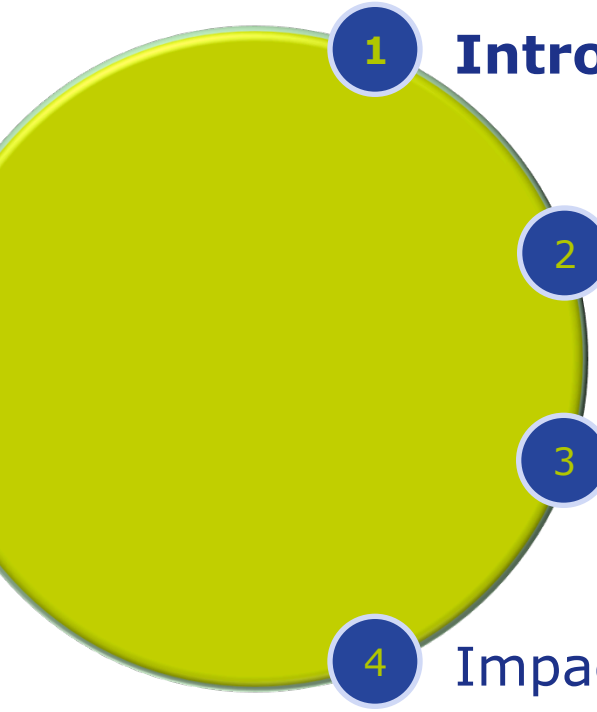
Tel +32 2 774 11 00
Fax +32 2 774 11 99
Mobile +32 476 545 917
peter.vanreusel@businessdecision.com

Rue Saint-Lambert 141
1200 Brussels
Belgium

www.businessdecision-lifesciences.com



1. Introduction
2. CDER Common Data Standards Issues
3. Amendment 1 to the SDTMIG
4. Impact



1. **Introduction**
2. CDER Common Data Standards Issues
3. Amendment 1 to the SDTMIG
4. Impact

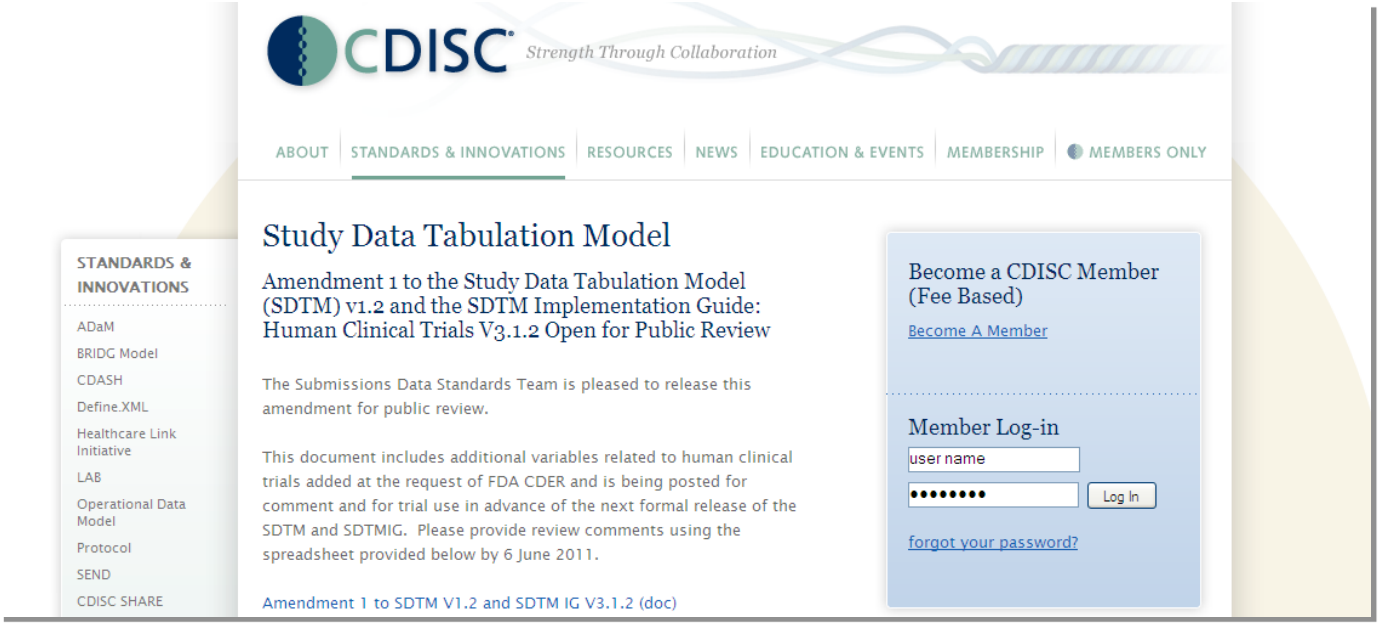
Introduction

- May-2011: CDER published Common Data Standards Issues Document v1.0 on the FDA website
- Dec-2011: v1.1 is published

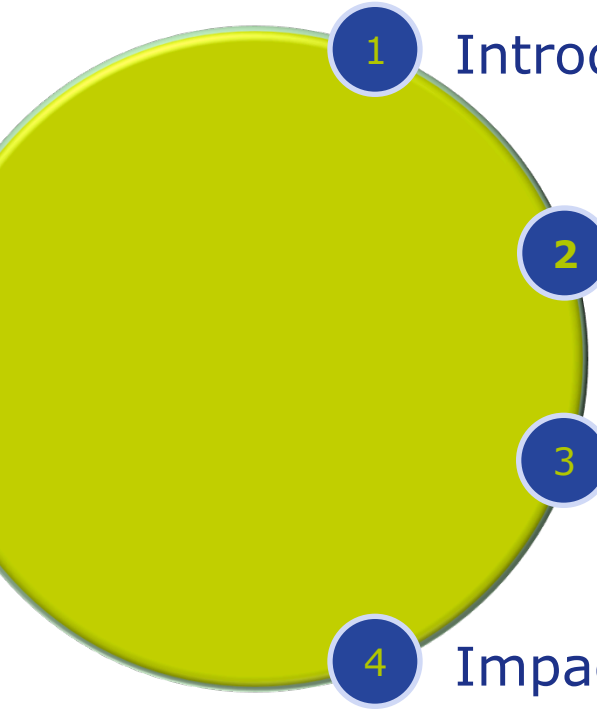
Source:
http://www.fda.gov/Drugs/DevelopmentApprovalProcess/FormsSubmissionRequirements/ElectronicSubmissions/ucm248635.htm

Introduction

- Dec-2011: Amendment 1 to the SDTM V1.2 and SDTMIG V3.1.2 has been posted on the CDISC website



Source: http://www.cdisc.org/sdtm



1. Introduction
2. **CDER Common Data Standards Issues**
3. Amendment 1 to the SDTMIG
4. Impact

General Considerations

- Sponsors should refer to the latest version of SDTMIG
- Sponsors should refer to Amendment 1 to SDTM V1.2
- Sponsors should ensure that every data variable's codelist, origin and derivation is clearly and easily accessible in define file
- Include variables EPOCH for every subject-level observation
- SDTM should be consistent with submitted analysis datasets

Traceability SDTM and ADaM

- Understanding relationship between the analysis results, the analysis datasets and the SDTM domains
- Establishing the path between an element and its immediate predecessor
- Two levels:
  - Metadata traceability
    - Relationship between an analysis result and analysis dataset(s)
    - Relationship of the analysis variable to its source dataset(s) and variable(s)
  - Data point traceability
    - Predecessor record(s)

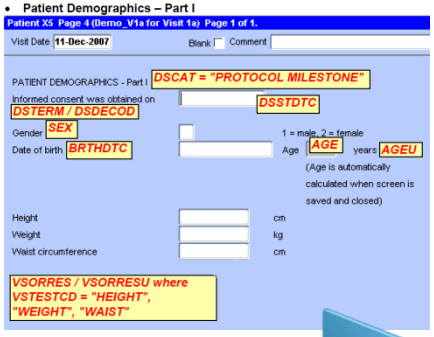
Business & Decision Life Sciences

Traceability SDTM and ADaM

Table 1 Demographic Data - Per-Protocol

	Treatment 1	Treatment 2
Baseline body mass index (BMI) [kg/m**2]		
N	167	167
Mean	29.08	29.04
SD	4.84	4.80
Min	20.3	16.0
Median	28.69	28.47
Max	40.1	41.2
Baseline BMI (categorical) [N (%)]		
<25 kg/m**2	41 (24.6%)	71 (21.1%)
25-<30 kg/m**2	60 (35.9%)	130 (38.7%)
>=30 kg/m**2	66 (39.5%)	135 (40.2%)

- **Patient Demographics – Part I**



	STUDYID	USUBJID	SUBJID	BMI	BMIGR1	BMIGR1N	BMIGR2	BMIGR2N
2	9999-0001	9999-0001-000001	000001	27.777777778	<30 kg/m**2	1	25-<30 kg/m**2	2
3	9999-0001	9999-0001-000002	000002	25.503615702	<30 kg/m**2	1	25-<30 kg/m**2	2
4	9999-0001	9999-0001-000003	000003	26.175194521	<30 kg/m**2	1	25-<30 kg/m**2	2
5	9999-0001	9999-0001-000004	000004	35.15625	>=30 kg/m**2	2	>=30 kg/m**2	3
6	9999-0001	9999-0001-000005	000005	30.968858131	>=30 kg/m**2	2	>=30 kg/m**2	3
7	9999-0001	9999-0001-000006	000006	39.697163916	>=30 kg/m**2	2	>=30 kg/m**2	3
8	9999-0001	9999-0001-000007	000007	25.826446281	<30 kg/m**2	1	25-<30 kg/m**2	2
9	9999-0001	9999-0001-000008	000008	30.103806228	>=30 kg/m**2	2	>=30 kg/m**2	3
10	9999-0001	9999-0001-000009	000009	32.280962683	>=30 kg/m**2	2	>=30 kg/m**2	3
11	9999-0001	9999-0001-000010	000010	28.876133787	<30 kg/m**2	1	25-<30 kg/m**2	2
12	9999-0001	9999-0001-000011	000011	29.372397383	<30 kg/m**2	1	25-<30 kg/m**2	2
13	9999-0001	9999-0001-000012	000012	26.714852608	<30 kg/m**2	1	25-<30 kg/m**2	2
14	9999-0001	9999-0001-000013	000013	32.718619869	>=30 kg/m**2	2	>=30 kg/m**2	3
15	9999-0001	9999-0001-000014	000014	28.715723183	<30 kg/m**2	1	25-<30 kg/m**2	2
16	9999-0001	9999-0001-000015	000015	32.270420377	>=30 kg/m**2	2	>=30 kg/m**2	3

Vital Signs Dataset (VS) — ../Tabulations/vs.xpt

Variable	Label	Type	Controlled Terminology	Origin	Role	Comment
STUDYID	Study Identifier	text		Protocol, CRF Page 1	IDENTIFIER	The STUDYID variable has a fixed format: 'XXXX-YYYY', where 'XXXX' indicates the 4-digit compound code and the 'YYYY' the 4-digit study code
DOMAIN	Domain Abbreviation	text	DOMAIN	Assigned	IDENTIFIER	
USUBJID	Unique Subject Identifier	text		Protocol	IDENTIFIER	The USUBJID variable has a fixed format: 'XXXX-YYYY-ZZZZZZ', where 'XXXX' indicates the 4-digit compound code, 'YYYY' the 4-digit study code and 'ZZZZZZ' the 6-digit patient code
VSSEQ	Sequence Number	integer		Derived	IDENTIFIER	Sequence number (automatically generated) to ensure uniqueness within a dataset for a subject
VSTESTCD	Vital Signs Test Short Name	text		Assigned	TOPIC	
VSTEST	Vital Signs Test Name	text		Derived	SYNONYM QUALIFIER	
VSPOS	Vital Signs Position of Subject	text	POSITION	CRF Page 13	RECORD QUALIFIER	
VSORRES	Result or Finding in Original Units	text		Derived, CRF Page 9, 13	RESULT QUALIFIER	
VSORRESU	Original Units	text	VSRESU	CRF Page 9, 13	VARIABLE QUALIFIER	
VSSTRESC	Character Result/Finding in Std Format	text		Derived	RESULT QUALIFIER	
VSSTRESN	Numeric Result/Finding in Standard Units	float	3.1	Derived	RESULT QUALIFIER	

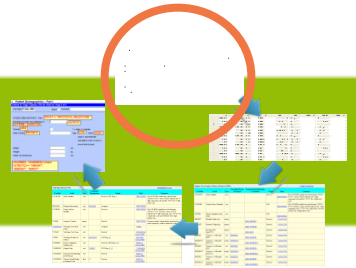
Subject-Level Analysis Dataset Dataset (ADSL) — ../Analysis/adsl.xpt

Variable	Label	Type	Controlled Terms or Format	Computational Algorithm or Method	Origin	Role	Comment
STUDYID	Study Identifier	text			DM	IDENTIFIER	The STUDYID variable has a fixed format: 'XXXX-YYYY', where 'XXXX' indicates the 4-digit compound code and the 'YYYY' the 4-digit study code
USUBJID	Unique Subject Identifier	text			DM	IDENTIFIER	The USUBJID variable has a fixed format: 'XXXX-YYYY-ZZZZZZ', where 'XXXX' indicates the 4-digit compound code, 'YYYY' the 4-digit study code and 'ZZZZZZ' the 5-digit patient code
SUBJID	Subject Identifier for the Study	text			DM	IDENTIFIER	
HEIGHT	Baseline Height (cm)	integer		ADSL.HEIGHT	Derived	ANALYSIS	
WEIGHT	Baseline Weight (kg)	integer		ADSL.WEIGHT	Derived	ANALYSIS	
BMI	Baseline BMI (kg/m**2)	integer		ADSL.BMI	Derived	ANALYSIS	
BMIGR1	Category 1 of Baseline BMI	text	BMIGR1L	ADSL.BMIGR1	Derived	ANALYSIS	
BMIGR1N	Category 1 of Baseline BMI, (N)	integer	BMIGR1N	ADSL.BMIGR1N	Derived	ANALYSIS	
BMIGR2	Category 2 of Baseline BMI	text	BMIGR2L	ADSL.BMIGR2	Derived	ANALYSIS	
BMIGR2N	Category 2 of Baseline BMI, (N)	integer	BMIGR2N	ADSL.BMIGR2N	Derived	ANALYSIS	
BMIGR3	Category 3 of Baseline BMI	text	BMIGR3L	ADSL.BMIGR3	Derived	ANALYSIS	
BMIGR3N	Category 3 of Baseline BMI, (N)	integer	BMIGR3N	ADSL.BMIGR3N	Derived	ANALYSIS	

LIFE SCIENCES

Business & Decision Life Sciences

Traceability SDTM and ADaM



- Analysis Results

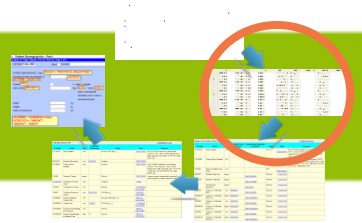
Table 1 Demographic Data - Per-Protocol

	Treatment 1	Treatment 2
Baseline body mass index (BMI) [kg/m**2]		
N	167	167
Mean	29.08	29.04
SD	4.84	4.80
Min	20.3	16.0
Median	28.69	28.47
Max	40.1	41.2
Baseline BMI (categorical) [N (%)]		
<25 kg/m**2	41 (24.6%)	71 (21.1%)
25-<30 kg/m**2	60 (35.9%)	130 (38.7%)
>=30 kg/m**2	66 (39.5%)	135 (40.2%)

LIFE SCIENCES

Business & Decision Life Sciences

Traceability SDTM and ADaM



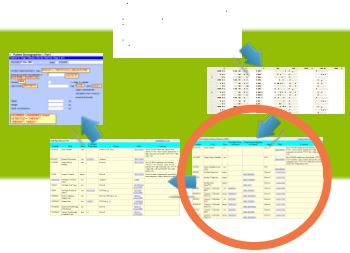
- Analysis Dataset

	STUDYID	USUBJID	SUBJID	BMI	BMIGR1	BMIGR1N	BMIGR2	BMIGR2N
2	9999-0001	9999-0001-000001	000001	27.777777778	<30 kg/m**2	1	25-<30 kg/m**2	2
3	9999-0001	9999-0001-000002	000002	25.503615702	<30 kg/m**2	1	25-<30 kg/m**2	2
4	9999-0001	9999-0001-000003	000003	26.175194521	<30 kg/m**2	1	25-<30 kg/m**2	2
5	9999-0001	9999-0001-000004	000004	35.15625	>=30 kg/m**2	2	>=30 kg/m**2	3
6	9999-0001	9999-0001-000005	000005	30.968858131	>=30 kg/m**2	2	>=30 kg/m**2	3
7	9999-0001	9999-0001-000006	000006	39.697163916	>=30 kg/m**2	2	>=30 kg/m**2	3
8	9999-0001	9999-0001-000007	000007	25.826446281	<30 kg/m**2	1	25-<30 kg/m**2	2
9	9999-0001	9999-0001-000008	000008	30.103806228	>=30 kg/m**2	2	>=30 kg/m**2	3
10	9999-0001	9999-0001-000009	000009	32.280962683	>=30 kg/m**2	2	>=30 kg/m**2	3
11	9999-0001	9999-0001-000010	000010	28.876133787	<30 kg/m**2	1	25-<30 kg/m**2	2
12	9999-0001	9999-0001-000011	000011	29.372397383	<30 kg/m**2	1	25-<30 kg/m**2	2
13	9999-0001	9999-0001-000012	000012	26.714852608	<30 kg/m**2	1	25-<30 kg/m**2	2
14	9999-0001	9999-0001-000013	000013	32.718619869	>=30 kg/m**2	2	>=30 kg/m**2	3
15	9999-0001	9999-0001-000014	000014	28.719723183	<30 kg/m**2	1	25-<30 kg/m**2	2
16	9999-0001	9999-0001-000015	000015	32.270420377	>=30 kg/m**2	2	>=30 kg/m**2	3

LIFE SCIENCES

Business & Decision Life Sciences

Traceability SDTM and ADaM



- ADaM define.xml

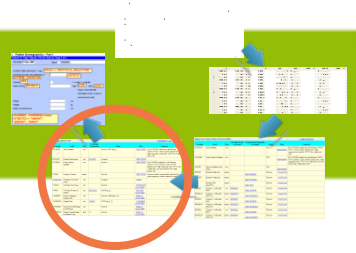
Computational Algorithms (ADSL.BMI)	
Reference Name	**Computation Method**
ADSL.BMI	Continuous variable, calculated using ADSL.WEIGHT/(ADSL.HEIGHT*0,01)**2 value at visit 3 if visit 3 data not available, the last data collected before randomisation
Computational Algorithms (ADSL.HEIGHT)	
Reference Name	**Computation Method**
ADSL.HEIGHT	equal to VS.VSSTRESN when VS.VSTESTCD="HEIGHT"
Computational Algorithms (ADSL.WEIGHT)	
Reference Name	**Computation Method**
ADSL.WEIGHT	equal to VS.VSSTRESN when VS.VSTESTCD="WEIGHT" and VS.VISITNUM=30 if visit 3 data not available, the last data collected before randomisation

				ADSL.HEIGHT			
WEIGHT	Baseline Weight (kg)	integer		ADSL.WEIGHT	Derived	ANALYSIS	
BMI	Baseline BMI (kg/m**2)	integer		ADSL.BMI	Derived	ANALYSIS	
BMIGR1	Category 1 of Baseline BMI	text	BMIGR1L	ADSL.BMIGR1	Derived	ANALYSIS	
BMIGR1N	Category 1 of Baseline BMI, (N)	integer	BMIGR1N	ADSL.BMIGR1N	Derived	ANALYSIS	
BMIGR2	Category 2 of Baseline BMI	text	BMIGR2L	ADSL.BMIGR2	Derived	ANALYSIS	
BMIGR2N	Category 2 of Baseline BMI, (N)	integer	BMIGR2N	ADSL.BMIGR2N	Derived	ANALYSIS	
BMIGR3	Category 3 of Baseline BMI	text	BMIGR3L	ADSL.BMIGR3	Derived	ANALYSIS	
BMIGR3N	Category 3 of Baseline BMI, (N)	integer	BMIGR3N	ADSL.BMIGR3N	Derived	ANALYSIS	

LIFE SCIENCES

Business & Decision Life Sciences

Traceability SDTM and ADaM



- SDTM define.xml and aCRF

Value Level Metadata (ValueList.VS.VSTESTCD)

Source Variable	Value	Label	Type	Controlled Terminology	Origin	Role	Comment
VSTESTCD	DIABP	DIASTOLIC BLOOD PRESSURE	text		CRF Page 13		
VSTESTCD	HEIGHT	HEIGHT	text		CRF Page 9		
VSTESTCD	PULSE	PULSE RATE	text		CRF Page 13		
VSTESTCD	SYSBP	SYSTOLIC BLOOD PRESSURE	text		CRF Page 13		
VSTESTCD	WAIST	WAIST CIRCUMFERENCE	text		CRF Page 9		
VSTESTCD	WEIGHT	WEIGHT	text		CRF Page 9		

							patient code
VSSEQ	Sequence Number						Sequence number (automatically generated) to ensure uniqueness within a dataset for a subject
VSTESTCD	Vital Signs Test Short Name						
VSTEST	Vital Signs Test Name						
VSPOS	Vital Signs Position of Subject						
VSORRES	Result or Finding in Original Units						
VSORRESU	Original Units						
VSSTRESC	Character Result/Finding in Std Format						
VSSTRESN	Numeric Result/Finding in Standard Units						

- **Patient Demographics – Part I**

Patient X5 Page 4 (Demo_V1a for Visit 1a) Page 1 of 1.

Visit Date 11-Dec-2007 Blank ☐ Comment

PATIENT DEMOGRAPHICS - Part I — *DSCAT = "PROTOCOL MILESTONE"*

Informed consent was obtained on — *DSSTDTC*

DSTERM / DSDECOD

Gender — *SEX* — 1 = male, 2 = female

Date of birth — *BRTHDTC* — Age — *AGE* years *AGEU*

(Age is automatically calculated when screen is saved and closed)

Height cm

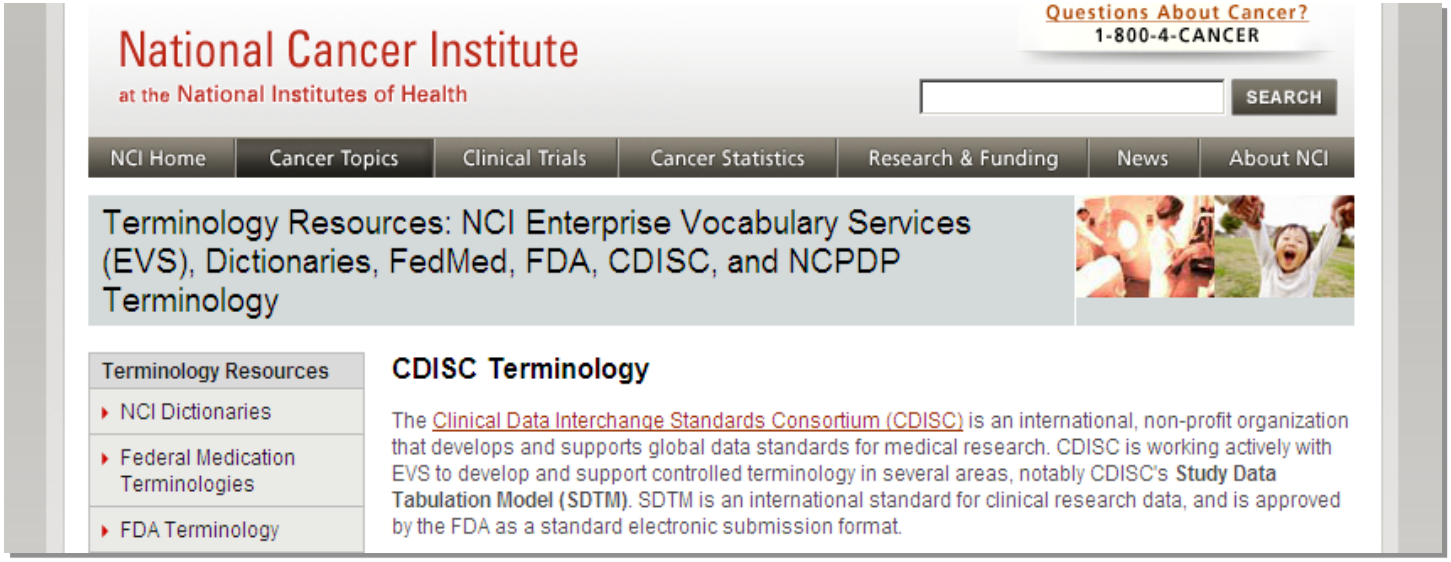
Weight kg

Waist circumference cm

VSORRES / VSORRESU where VSTESTCD = "HEIGHT", "WEIGHT", "WAIST"

Controlled Terminology

- Use existing CDISC terminology
- If available CDISC terminology is insufficient, sponsors should create their own terminology
- Documentation on sponsor-specific terminology should be included in define.xml



Source: http://www.cancer.gov/cancertopics/cancerlibrary/terminologyresources/cdisc

Business & Decision Life Sciences

MedDRA and Common Dictionaries

- Sponsors should exactly follow spelling and case
- MedDRA version should be consistent across trials within the submission
- Dictionary name and version should be documented in define.xml

Controlled Terminology (External Dictionaries)	
ADVERSE EVENT DICTIONARY, Reference Name (Codelist.AEDICT_F)	
External Dictionary	Dictionary Version
MedDRA	13.1
DRUG DICTIONARY, Reference Name (Codelist.DRUGDICT_F)	
External Dictionary	Dictionary Version
WHO Drug	WHO-DD 10.SEP
ISO 8601, Reference Name (Codelist.ISO 8601)	
External Dictionary	Dictionary Version
ISO 3166, Reference Name (Codelist.ISO 3166)	
External Dictionary	Dictionary Version

SDTM Datasets

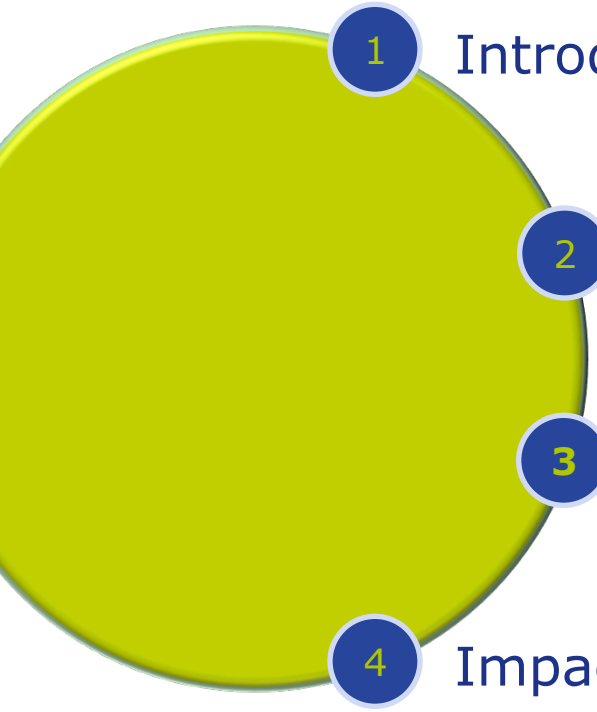
- SUPPQUAL
  - Should not be used as a waste basket
- DM
  - Strongly preferred to use additional variables in Amendment 1
- DS
  - EPOCH should be used to distinguish between multiple disposition events
  - If DEATH occurs, it should be documented in the last record with the associated EPOCH

SDTM Datasets

- AE
  - Provide variables for MedDRA hierarchy
  - Sponsors should include all AEs, not only the one caused by the study treatment
  - AESOC = MedDRA-defined, primary mapped SOC
  - AEBODSYS = SOC used for analysis
- Custom Domains
  - Only to be used for data that does not fit in a published domain
- Filesize
  - Ideal filesize < 1Gb
  - Larger files should be split according to --CAT, --SCAT; Non-split dataset should also be included
  - Discuss with your review division

SDTM Variables

- Permissible variables that CDER expects to see
  - --BLFL (LB, VS, EG, Pharmacokinetics, Microbiology)
  - EPOCH
  - --DY, --STDY, --ENDY (Calculated from RFSTDTC)
- No imputations allowed
- Dates in ISO 8601
  - Missing dates are missing dates
- USUBJID
  - No leading or trailing spaces allowed
  - should match across all datasets (SDTM, ADaM) on a character basis



1. Introduction
2. CDER Common Data Standards Issues
3. **Amendment 1 to the SDTMIG**
4. Impact

Additions to SDTM V1.2/SDTMIG V3.1.2

- New variables in Demographics

- New variables in Events General Observation Class
  - Additional accomodation for MedDRA codings
  - Part of these previously used in SUPPQUAL

Additions to DM

Variable	Label
ACTARMCD	Actual Arm Code
ACTARM	Description of Actual Arm
RFXSTDTC	Date/Time of First Study Drug Exposure
RFXENDTC	Date/Time of Last Study Drug Exposure
RFICDTC	Date/Time of Informed Consent
RFPENDTC	Date/Time of End of Subject Participation
DTHDTC	Date of Death
DTHFL	Subject Died Flag

Additions to DM

- ACTARMCD, ACTARM
  - Actual arm a subject participated in during the trial
  - Randomized subjects that are not treated
    - ACTARMCD/ACTARM= 'NOTTRT '/ 'Not Treated'
- RFXSTDTC, RFXENDTC
  - Date/Time of first/last study treatment exposure
- RFICDTC
  - Date/Time of informed consent
  - Should match entry in DS if this is documented as a protocol milestone

Additions to DM

- RFPENDTC
  - Date/Time of end of participation
  - Last known date of participation FOR DATA
  - NOT the last date of participation in study

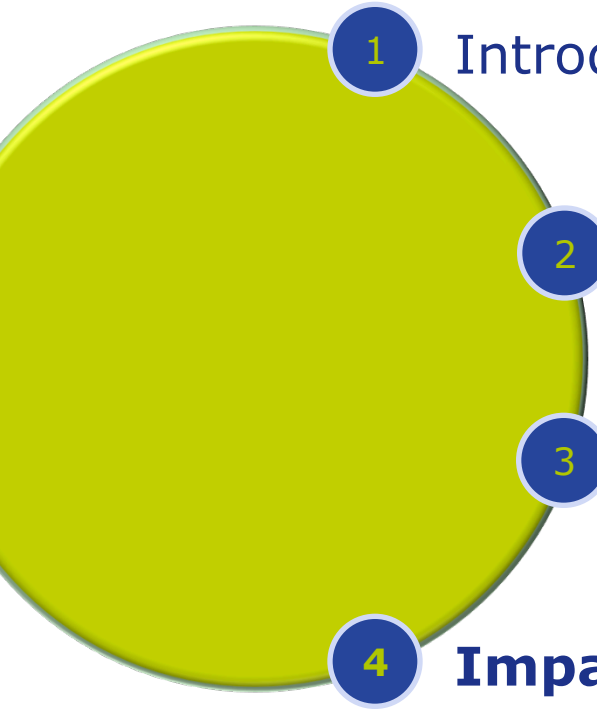
- DTHDTC, DTHFL
  - Date of death, Subject death flag

Additions to AE

Variable	Label
AELLT	Lowest Level Term
AELLTCD	Lowest Level Term Code
AEPTCD	Preferred Term Code
AEHLT	High Level Term
AEHLTCD	High Level Term Code
AEHLGT	High Level Group Term
AEHLGTCD	High Level Group Term Code
AESOC	Primary System Organ Class
AESOCCD	Primary System Organ Class Code
AEBDSYCD	Body System or Organ Class Code

Additions to AE

- AELLT, AELLTCD, AEPTCD, AEHLT, AEHLTCD, AEHLGT, AEHLGTCD, AESOCCD
  - Promoted from SUPPQUAL (SDTMIG Appendix C5) into the parent domain
- AESOC
  - Primary system organ class
  - AEBODSYS should contain the SOC used in analysis
- AEBDSYCD
  - Body system code



1. Introduction
2. CDER Common Data Standards Issues
3. Amendment 1 to the SDTMIG
4. **Impact**

Impact

1. Additional Variables in AE

   - Low impact
   - The majority of the added variables were previously included in SUPPAE
   - Difference between ‘Primary System Organ Class’ and ‘System Organ Class’ used for the analysis

Impact

2. Additional Value ‘DEATH’ in DS.DSTERM/DSDECOD

- Low impact
- Most sponsors are already using this rule
- Value may be derived from:
  - Death report
  - Survival status
  - Follow up report
  - AE outcome
  - End of study status

Impact

3. Additional Variables in DM

   - Medium impact
   - Clear distinction between planned and actual ARM/ARMCD in their source
   - Additional (complex) derivations from EX, DS, SE

Impact

4. File Size Limitations

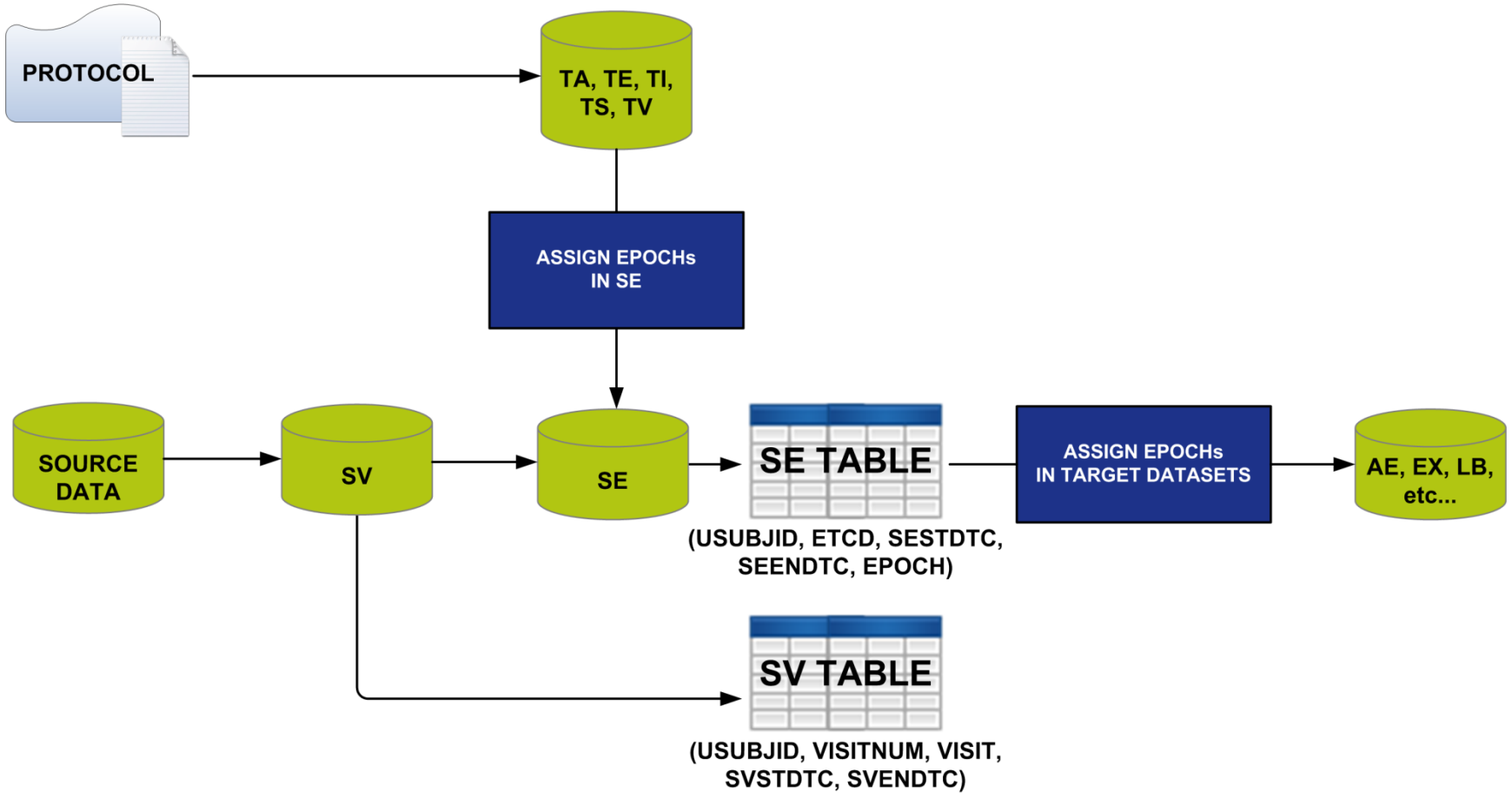
- Medium impact
- Contact FDA Review division, splitting may not be required
- Some additional QC steps are needed to ensure that splitting was done correctly
- Metadata in Define.xml needs to be provided for each split dataset. Includes value level metadata

Impact

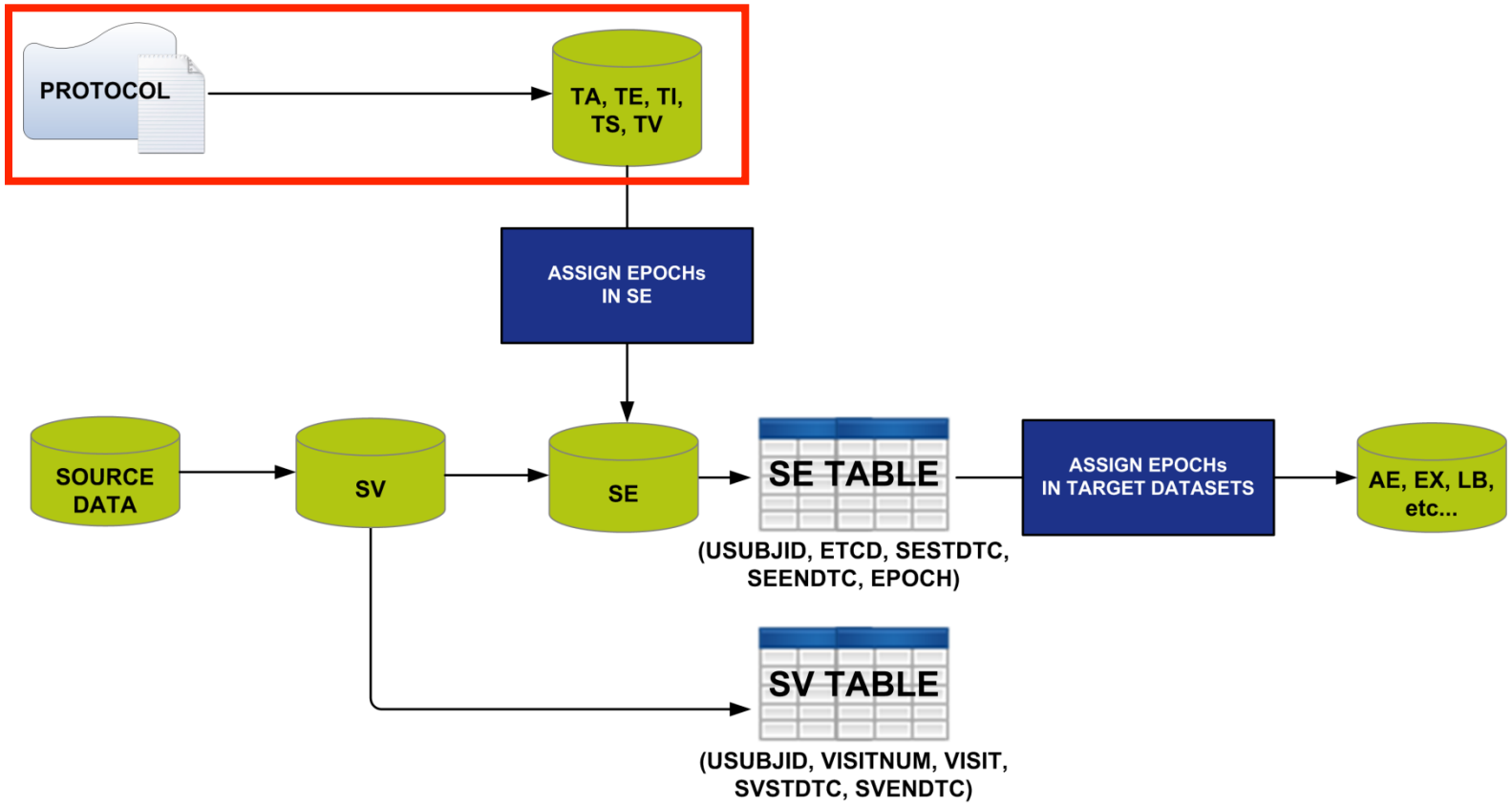
5. Additional variable EPOCH in all domains

- High impact
- Due to the way in which EPOCH is derived (see next slides), generating and/or updating it, is a lengthy process
- Clear business rules for deriving EPOCHs will need to be determined at the beginning of a project

Deriving EPOCHs (1)

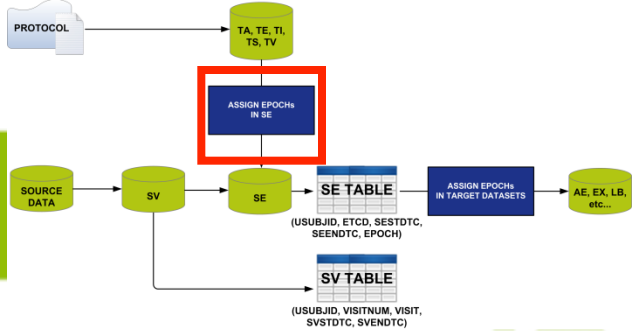


Deriving EPOCHs (2)



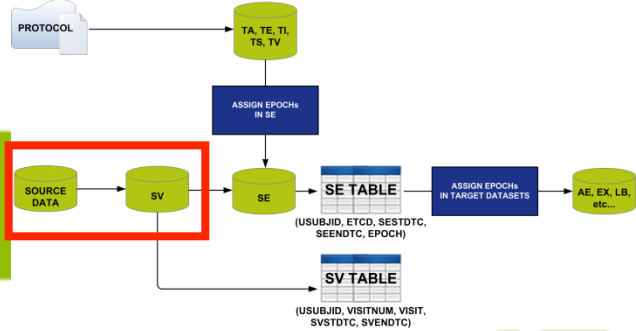
- The Trial Design domains are created based on the protocol

Deriving EPOCHs (3)

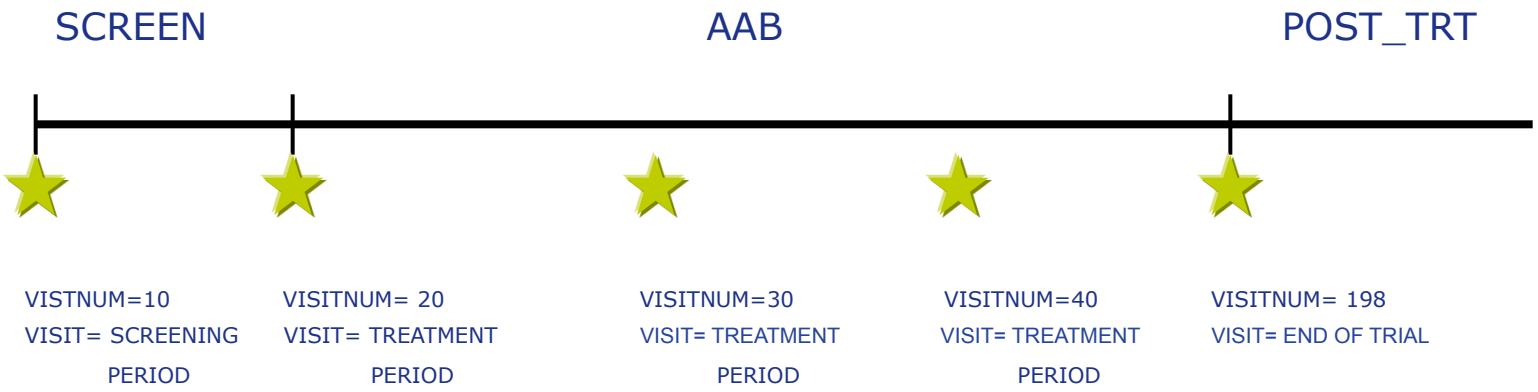


- In the TA domain, EPOCH are assigned per ELEMENT

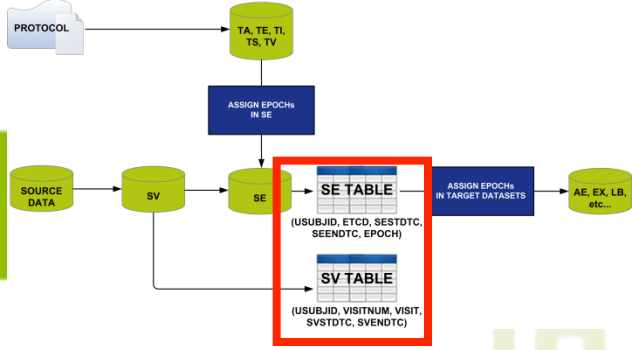
Deriving EPOCHs (4)



- The SE domain is created based on the Trial Design and the source data.

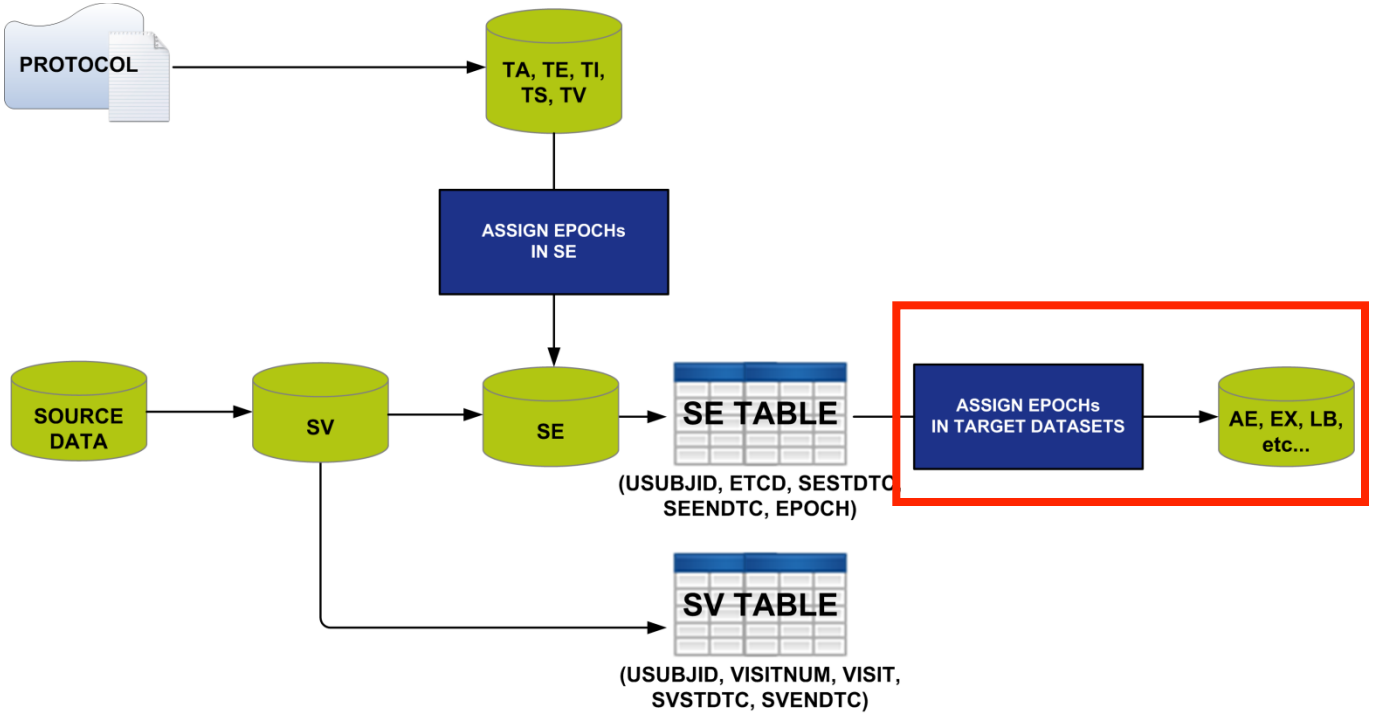


Deriving EPOCHs (5)



- Subject element start and stop are based on VISIT start and stop.
- EPOCHs are assigned to match those in TA.

Deriving EPOCHs (6)



- EPOCH in the target datasets are derived by linking the date variable to the subject element date interval.

Summary

- Adding AE variables has a minor impact.
- Additional DM variables may require complex derivations.
- Deriving EPOCHs is a lengthy and complex process.
- Clear business rules need to be established for the derivations.
- Changing these derivation rules may have a high impact on timing and resources.

Conclusion

- SDTM is for collected data

Conclusion

- SDTM is analysis friendly

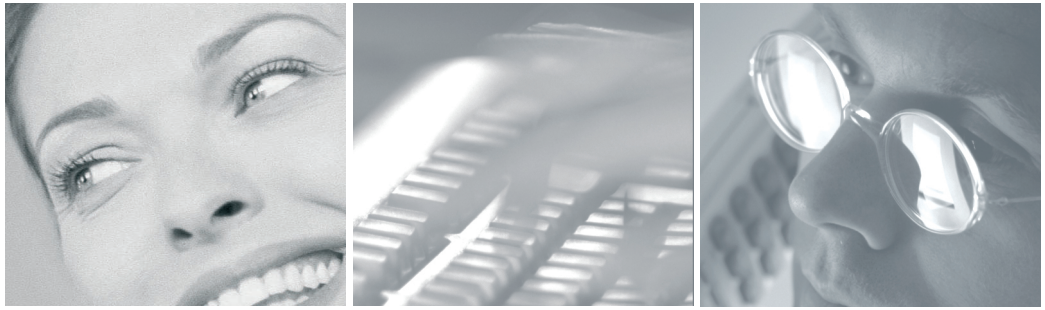
Thank you
for your attention

Peter Van Reusel
Business Unit Director
CRO Services
Business & Decision Life Sciences

Tel +32 2 774 11 00
Fax +32 2 774 11 99
Mobile +32 476 54 59 17
peter.vanreusel@businessdecision.com

Sint-Lambertusstraat 141 Rue Saint-Lambert
1200 Brussels

www.businessdecision-lifesciences.com



DRIVE YOUR PERFORMANCE