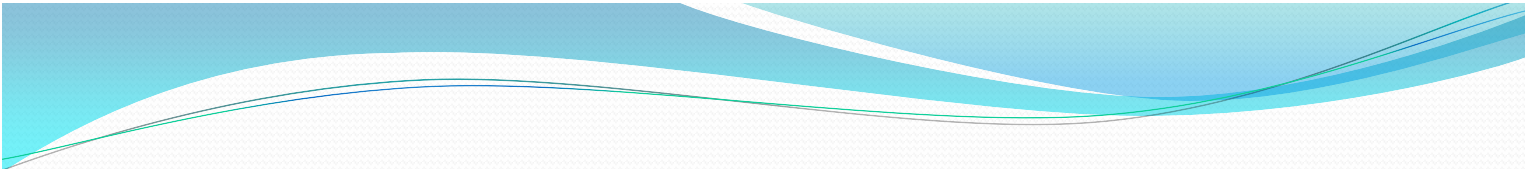


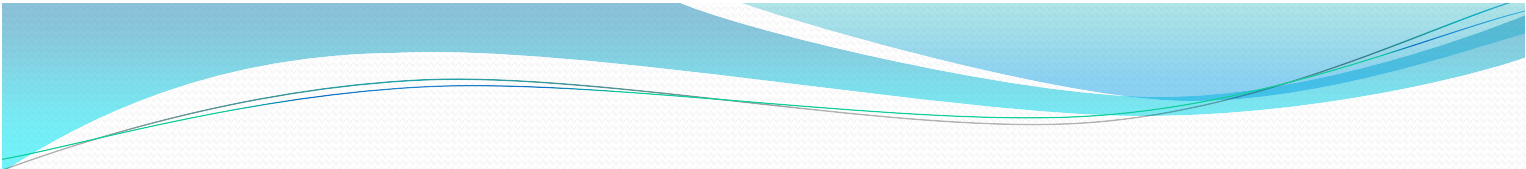
Traceability: Plan Ahead for Future Needs

Sandra Minjoe, Accenture Life Sciences
Tanja Petrowitsch, Bayer Pharma AG



Scenario

- Your company has purchased another company
- You need to create more tables for a study already done
- You have only datasets and tables, no other documentation
- How do you determine, for example:
 - What procedure and data were used to create table values?
 - Which variables in a dataset were derived vs. collected?
- May have to re-program everything to match results
- Lack of traceability documentation makes this a much more difficult chore than it should be

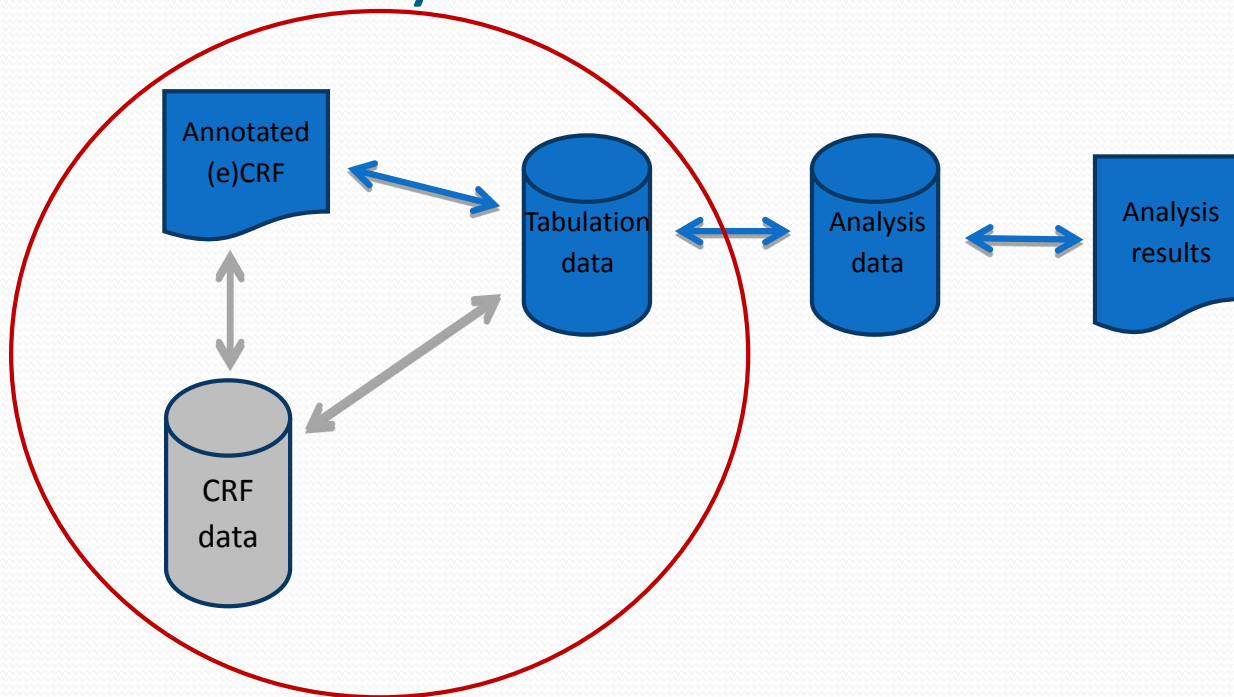


Traceability: Plan Ahead for Future Needs

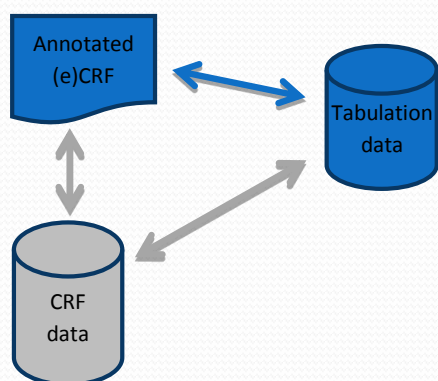
Agenda

- Traceability at Each Step in the Clinical Process
- Metadata as Part of the Process

Traceability in Clinical Trials Data Flow

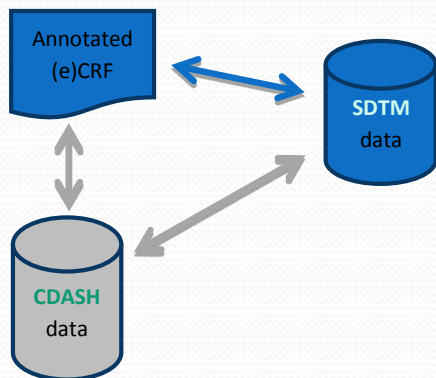


Traceability between Data Collection and Tabulation



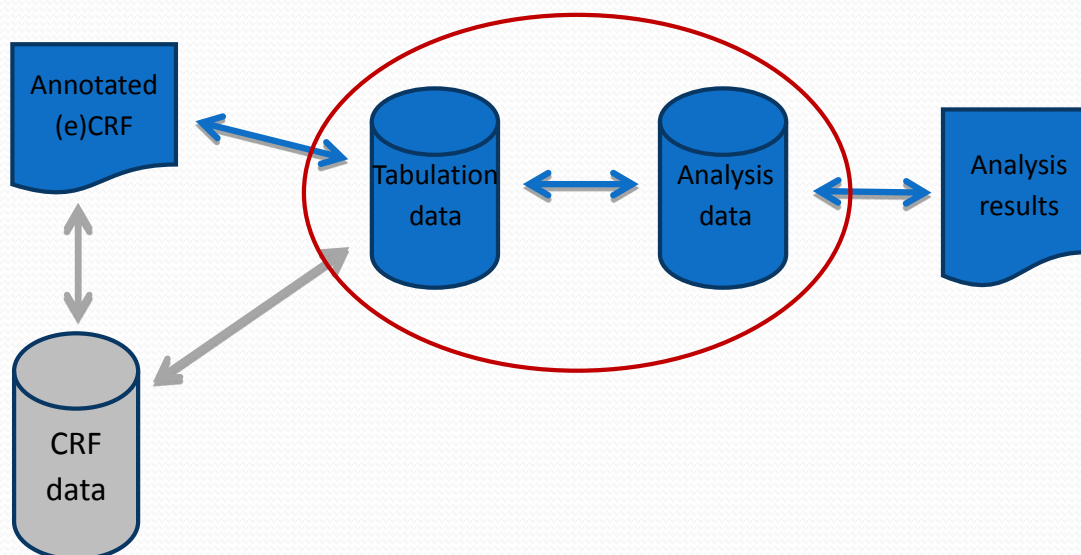
- Two annotated (e)CRFs
 - One for CRF data
 - One for Tabulation data
- Include reference data
 - Row numbers
- Document derivations from CRF to Tabulation data
 - Denote any derived tabulation data
 - Use controlled terminology
 - Minimize transformations
- Create define file and other metadata from specs

Traceability between Data Collection and Tabulation - **CDISC**

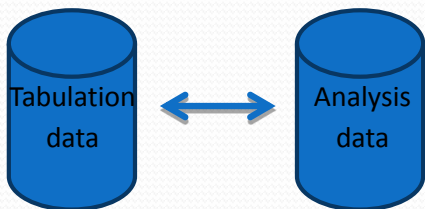


- Two annotated (e)CRFs
 - One for **CDASH** data
 - One for **SDTM** data
- Include reference data
 - Row numbers (**--SPID**)
- Document derivations from **CDASH** to **SDTM** data
 - Denote any derived tabulation data (**--DRVFL**)
 - Use controlled terminology
 - Minimize transformations
- Create define file and other metadata from specs

Traceability in Clinical Trials Data Flow

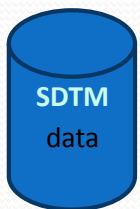


Traceability between Tabulation and Analysis Data



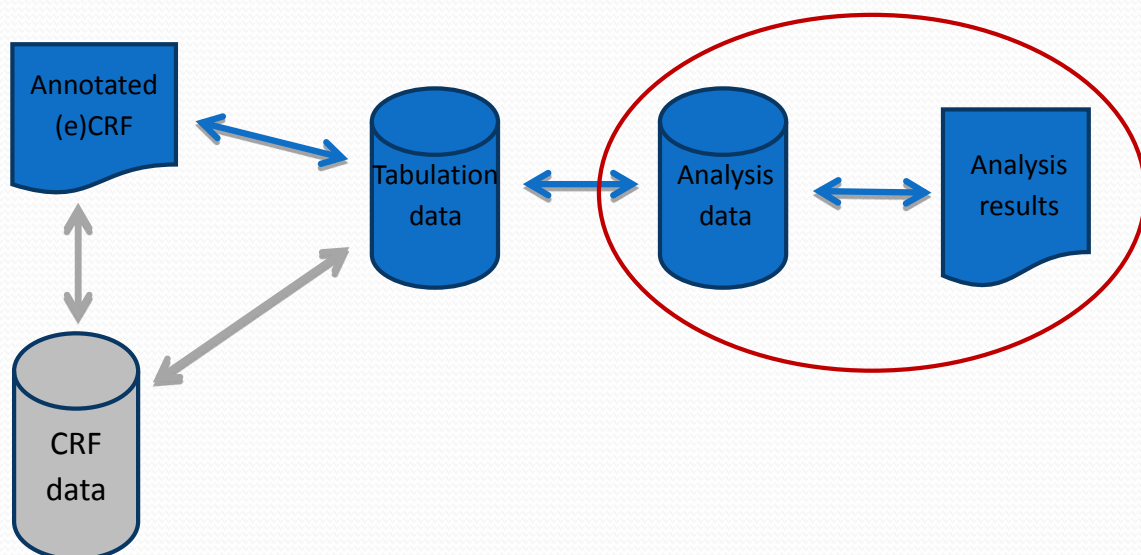
- Use controlled terminology
- Clarify copied vs. derived data
- Use informative labels
- Include intermediate data for complex derivations
- Include reference data
 - Row reference
 - Tabulation data not used for analysis
- Create define file and other metadata from specs

Traceability between Tabulation and Analysis Data - **CDISC**

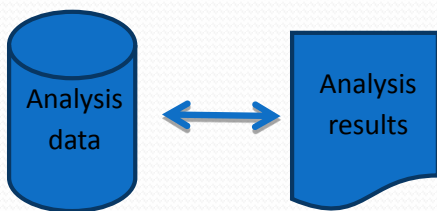


- Use controlled terminology
- Clarify copied vs. derived data (e.g., **DTYPE**, **PARAMTYP**)
- Use informative labels
- Include intermediate data for complex derivations
- Include reference data
 - Row reference (**--SEQ**, **SRCDOM/SRCSEQ**, **ASEQ**)
 - Tabulation data not used for analysis (e.g., ***DTC**)
- Create define file and other metadata from specs

Traceability in Clinical Trials Data Flow

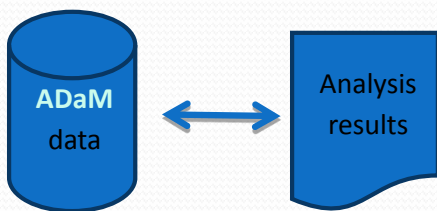


Traceability between Data Analysis and Results

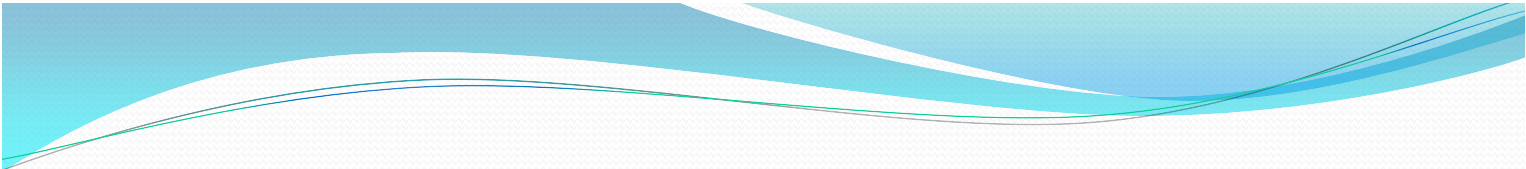


- Ensure analysis datasets are “analysis-ready”
- Use analysis flags to show which rows are included in specific analyses
- Create numeric versions of data that have meaning other than just for sorting
 - Example: numeric visit represents number of weeks

Traceability between Data Analysis and Results - **CDISC**



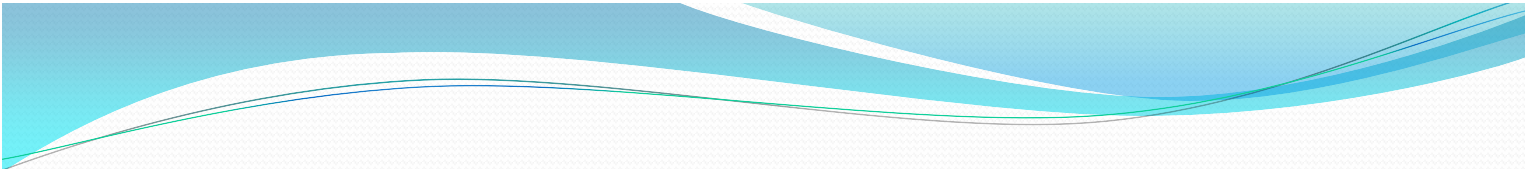
- Ensure analysis datasets are “analysis-ready”
- Use analysis flags to show which rows are included in specific analyses (**ANLzzFL**)
- Create numeric versions of data that have meaning other than just for sorting
 - Example: numeric visit represents number of weeks (**AVISITN**)



Traceability: Plan Ahead for Future Needs

Agenda

- Traceability at Each Step in the Clinical Process
- Metadata as Part of the Process

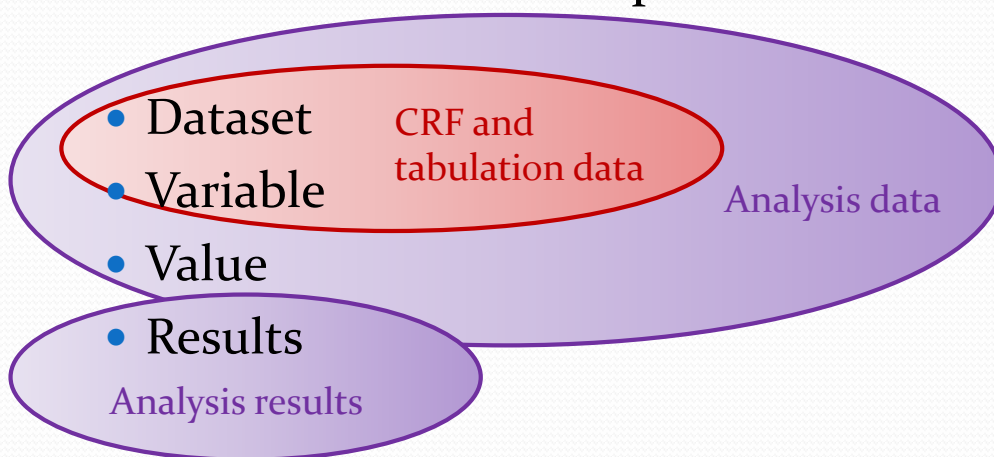


Metadata

- Metadata components for tabulation and analysis data
 - DEFINE.XML
 - Reviewers Guides
 - SDRG = Study Data Reviewers Guide
 - ADRG = Analysis Data Reviewers Guide
- Most of what we need in metadata already exists
 - Re-use material in specifications, programs, tools

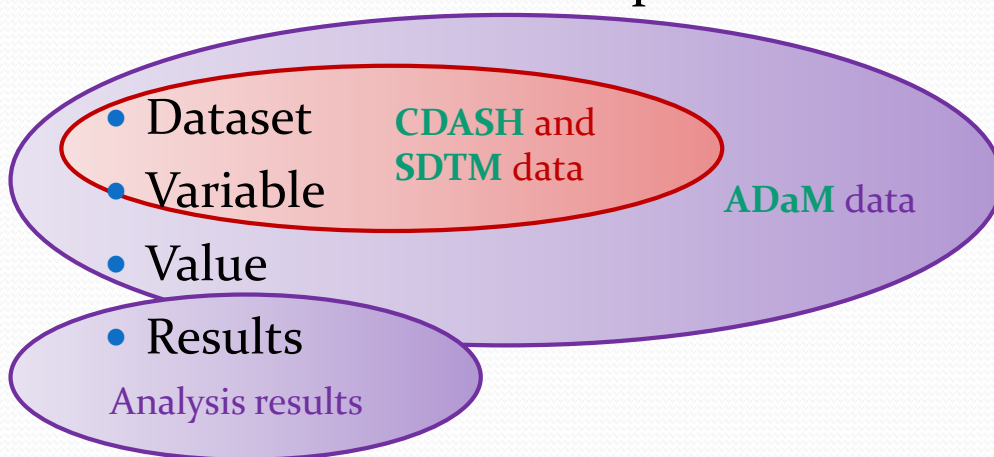
Define and Traceability

- Contains information from specifications
- Each level of metadata provides traceability



Define and Traceability - CDISC

- Contains information from specifications
- Each level of metadata provides traceability



Define: Dataset Metadata Example

Analysis Datasets for Study CDISC-Sample (ADaM-IG 1.0)

Dataset	Description	Class	Structure	Purpose	Keys	Location	Documentation
ADSL	Subject-Level Analysis	SUBJECT LEVEL ANALYSIS DATASET	one record per subject	Analysis	STUDYID, USUBJID	adsl.xpt	Screen Failures are excluded since they are not needed for this study analysis. See Analysis Data Reviewer's Guide, page 6. Analysis Data Reviewer's Guide
ADQSADAS	ADAS-Cog Analysis	BASIC DATA STRUCTURE	One record per subject per parameter per analysis visit per analysis date	Analysis	STUDYID, USUBJID, PARAMCD, AVISIT, ADT	adqsadas.xpt	See referenced dataset creation program and Analysis Data Reviewer's Guide, Section 2.1 adqsadas.sas Analysis Data Reviewer's Guide
ADAE	Adverse Events Analysis Dataset	ADAM OTHER	one record per subject per adverse event	Analysis	STUDYID, USUBJID, AETERM, ASTDT, AESEQ	adae.xpt	See SAS program adae.sas

Define: Variable Metadata Example

ADAS-Cog Analysis (ADQSADAS) [Location: [adqsadas.xpt](#)]

Variable	Label	Key	Type	Length / Display Format	Controlled Terms or Format	Source/Derivation/Comment
STUDYID	Study Identifier	1	text	12		Predecessor: ADSL.STUDYID
SITEID	Study Site Identifier		text	3		Predecessor: ADSL.SITEID
ANL01FL	Analysis Record Flag 01		text	1	["Y" = "Yes"] < No Yes Response - Y subset >	Derived: If multiple visits fall into the same visit window, then the one closest to the target day is chosen for analysis. These are flagged with ANL01FL="Y".
DTYPE	Derivation Type		text	7	["LOCF" = "Last Observation Carried Forward"] < Derivation Type >	
AWRANGE	Analysis Window Valid Relative Range		text	9		Assigned: Window range, specified in the SAP.

Define: Value Metadata Example

Parameter Value List - ADQSADAS [DTYPE]

Variable	Where	Type	Length / Display Format	Controlled Terms or Format	Source/Derivation/Comment
DTYPE	PARAMCD ≠ "ACTOT" (Adas-Cog (11) Subscore)	text	7	["LOCF" = "Last Observation Carried Forward"] < Derivation Type >	Assigned: Value: null
DTYPE	PARAMCD = "ACTOT" (Adas-Cog (11) Subscore)	text	7	["LOCF" = "Last Observation Carried Forward"] < Derivation Type >	Assigned: Value: LOCF denotes that the LOCF imputation method was used to impute the value for the given parameter and analysis visit.

Define: Results Metadata Example

Table 14-5.02

Display	Table 14-5.02 Incidence of Treatment Emergent Serious Adverse Events by Treatment Group
Analysis Result	Incidence of Treatment Emergent Serious Adverse Events by Treatment Group
Analysis Parameter(s)	
Analysis Variable(s)	AEBODSYS (Body System or Organ Class) AEDECOD (Dictionary-Derived Term)
Analysis Reason	SPECIFIED IN SAP
Analysis Purpose	PRIMARY OUTCOME MEASURE
Data References (incl. Selection Criteria)	ADAE [TRTEMFL = "Y" and AESER = "Y"] ADSL [SAFFL = "Y"] Get denominators for percentages from ADSL and counts and numerators from ADAE. Join ADAE with ADSL based on the unique subject identifier (USUBJID) keeping only records in ADAE.
Documentation	Unique count of subjects that experienced an Adverse Event by Preferred Term, System Organ Class, and Treatment Group and percentages based on the number of subjects in the safety population within each treatment group. The total number of times an event occurred was recorded by Preferred Term, System Organ Class, and Treatment Group. Fisher's exact test was used for treatment comparison of event rates. SAP Section 11.2
Programming Statements	[SAS version 9.2] at14-5-02.sas

Define: Controlled Terminology

Example

Age Group [CL.AGEGR1]

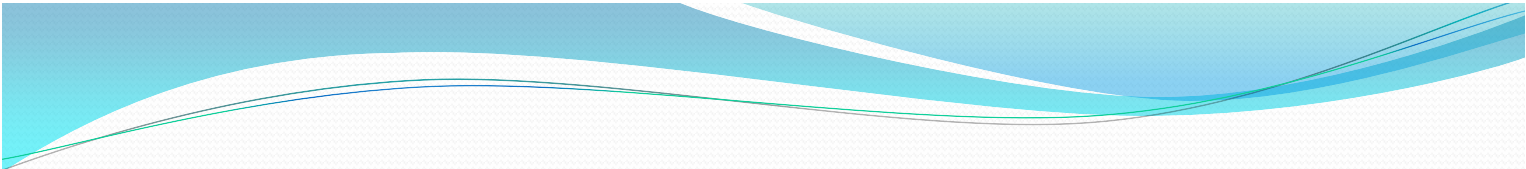
Permitted Value (Code)
<65
65-80
>80

Age Group (N) [CL.AGEGR1N]

Permitted Value (Code)	Display Value (Decode)
1	<65
2	65-80
3	>80

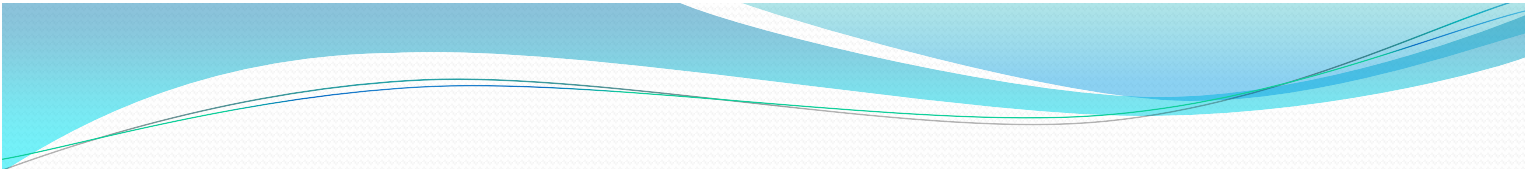
Age Unit [CL.AGEU, C66781]

Permitted Value (Code)
YEARS [C29848]



Reviewers Guides and Traceability

- Provides additional information beyond define
 - Define is basically a set of tables of information
 - Reviewers Guides are text documents
- CSS Working Groups have created standards
- Templates, examples, instructions all available at wiki
 - Study Data Reviewers Guide (SDRG)
 - Analysis Data Reviewers Guide (ADRG)



Reviewers Guide Example

3.5 Use of Visit Windowing, Unscheduled Visits, and Record Selection

- Was windowing used in one or more analysis datasets?

No visit windowing was done in the analysis datasets.

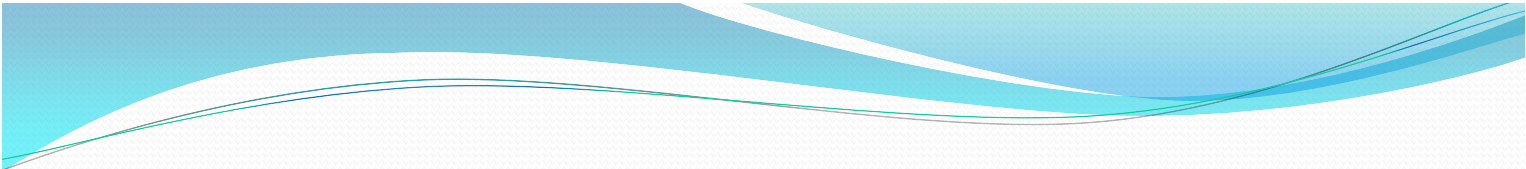
- Were unscheduled visits used in any analyses?

Unscheduled visits were not used in analysis.

3.6 Imputation/Derivation Methods

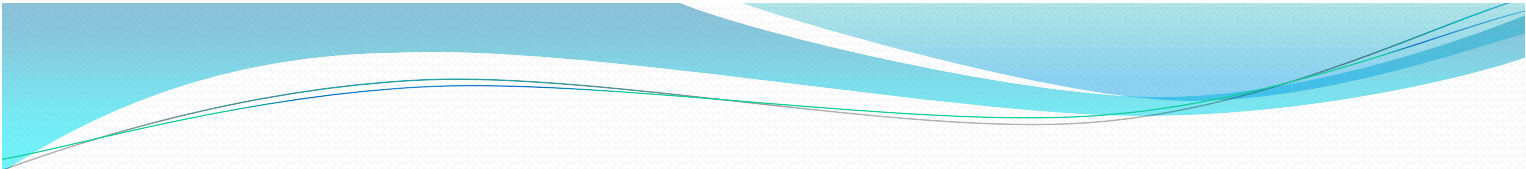
- If date imputation was performed, were there rules that were used in multiple analysis datasets?

For datasets that may contain partial dates, imputation rules are as follows: If day is missing and the year and month are the same as the year and month of the first date of study drug dosing, then the date is imputed as the first dosing date. Otherwise, if day is missing and month and year are present, day is imputed as the first day of the month.



Metadata as Part of the Process

- Most of what we need in metadata already exists
 - Re-use material in specifications, programs, tools
- We can automate the creation of content in define.xml and reviewers guides
 - Apply stringent rules to structure of Excel specs
 - Consider a metadata repository
 - Could even use specs to drive executable code
- Don't wait until after data is complete to think about adding metadata



References

- CSS wiki: phusewiki.org -> CSS Working Groups
 - White paper on Linear Data Flow: -> [Optimizing the Use of Data Standards](#)
 - ADRG and SDRG: -> [CSS Catalog of Deliverables](#)
- CDISC, including define examples: cdisc.org

Questions?