



15th-19th November

EU Connect 2021

The Clinical Data
Science Conference

How to be Traceable in ADaM

Angelo Tinazzi, Cytel Inc.

PHUSE EU Connect 2021

15th -19th November 2021

Paper SI08 – Standards Implementation (SI)

Agenda

- (re)Introducing Traceability
- ADaM Traceability Examples
- Verify Traceability Programmatically
- Conclusions

(re)Introducing Traceability

- What is Traceability
- How to Achieve Traceability in ADaM
- How to be «Technically» Traceable in ADaM

What is Traceability



The property that enables the **understanding of the data's lineage** and/or the **relationship between an element and its predecessor(s)**. Traceability is built by clearly **establishing the path between an element and its immediate predecessor**



Establishing traceability is one of the most problematic issues associated with any data conversion. **If the reviewer is unable to trace study data from the data collection of subjects participating in a study to the analysis of the overall study data, then the regulatory review of a submission may be compromised.**



Why? Transparency and Understanding

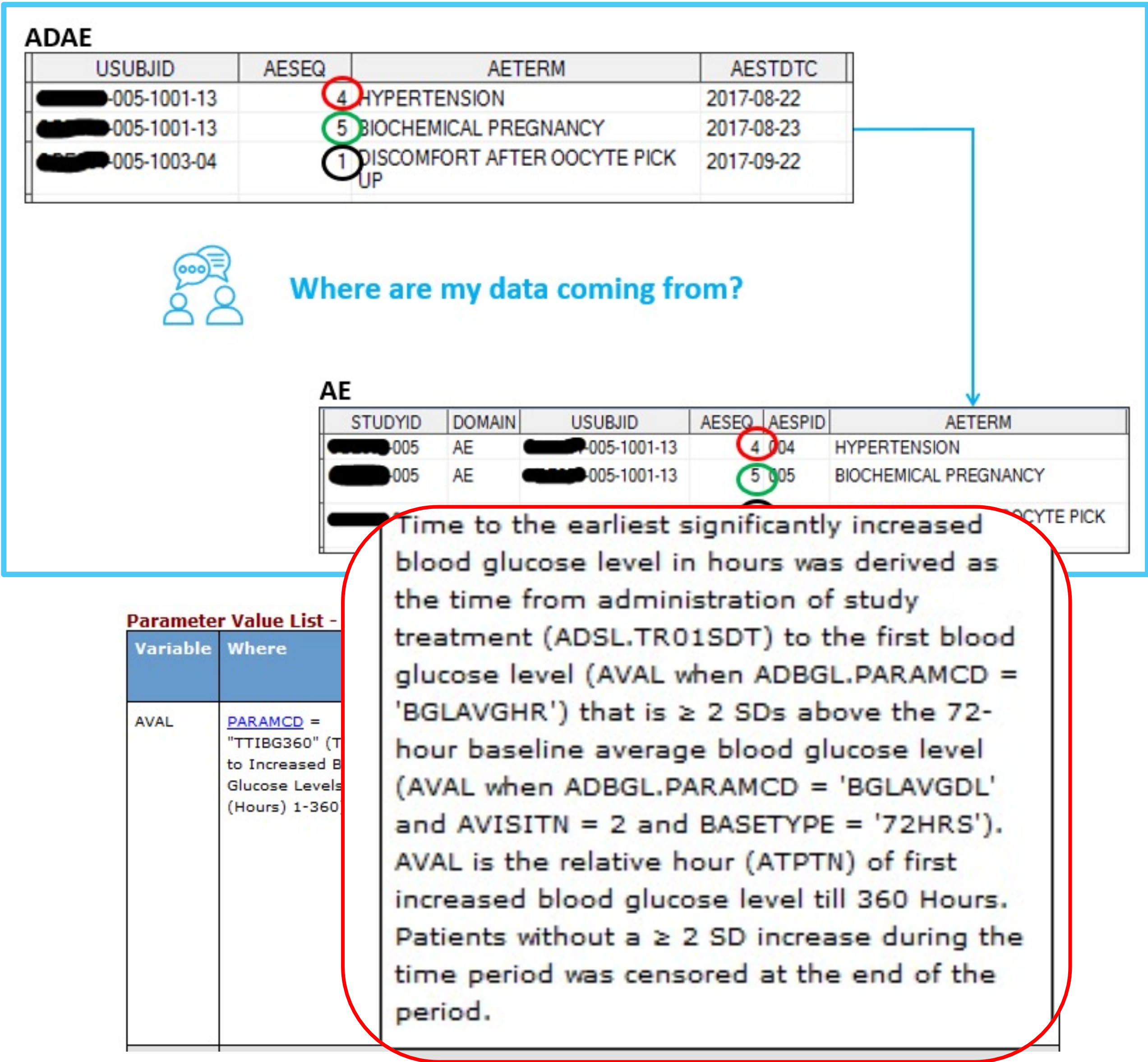
What? Analysis Results → ADaM → SDTM → CRF

How? By establishing the path (a link) between an element and its immediate predecessor

How to Achieve Traceability in ADaM

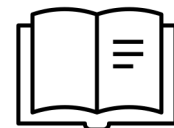
Datapoint Traceability points directly to the specific predecessor record(s) and should be implemented if practical and feasible.

Metadata Traceability facilitates the understanding of the relationship of the analysis variable to its source dataset(s) and variable(s). **This traceability is established by describing (via metadata) the algorithm used or steps taken to derive or populate an analysis variable from its immediate predecessor.** Metadata traceability is also used to establish the relationship between an analysis result and ADaM dataset(s).



How to Achieve Traceability in ADaM more

- When you need to **re-derive** in ADaM a variable collected in SDTM (e.g. AGE) create a new variable (and keep original variables in the ADaM dataset)
- When you need to **derive a variable using variable(s)** from SDTM, keep original SDTM variable(s) used in the derivation
- When you need to **impute a missing or partial value** e.g. date, create a new variable and keep original SDTM variable. For date/time variable makes use of imputation flags variables (e.g. ADTF)
- When applying **windowing** e.g. create **AVISIT** (Analysis Visit) keep “supporting” ADaM windowing variables and keep original SDTM timepoint variable(s) (e.g. VISITNUM)
- When you need to **impute a missing timepoint from other timepoints** e.g. visit derive a new record (use DTYPE) and keep original timepoints records even if not analyzed. Make use of flags (e.g. ANLxxFL) to identify records/visits used in the analysis vs those not used
- When you deal with **complex derivations requiring several steps** make use of Intermediate datasets

- 
- ADaM Intermediate Dataset: how to improve your analysis traceability, A. Tinazzi, T. Curto and A. Aggarwal; PHUSE US Connect, 2018
 - CDISC Breast TAUG Section 5.3.2 “Intermediate Event Dataset”

How to be “Technically” Traceable in ADaM (Summary)

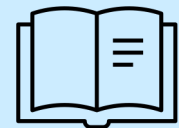
	DATA POINT	METADATA SUPPORTIVE DOCUMENTS
ADaM	• Copy/retain SDTM variables • Copy/retain SDTM records • --SEQ from SDTM • SRCDOM/SRCVAR/SRCSEQ • ADTF • ASEQ • DTYPE • ANLxxFL • Occurrence Flags in OCCDS • Intermediate ADaM Datasets	• define.xml • ADRG • SAP
Analysis Results	«Not applicable»	• define.xml (ARM extension) • ADRG • SAP

ADaM Traceability Examples

- CDISC ADaM Traceability Examples Document
- Example – In your (reviewer) shoes
- Health Authorities care About Traceability

CDISC ADaM Traceability Examples Document

- Completed Public Review
- Planned to be released Q4 2021
- 12 Examples
- Metadata fully explained with outputs and ADaM datasets examples



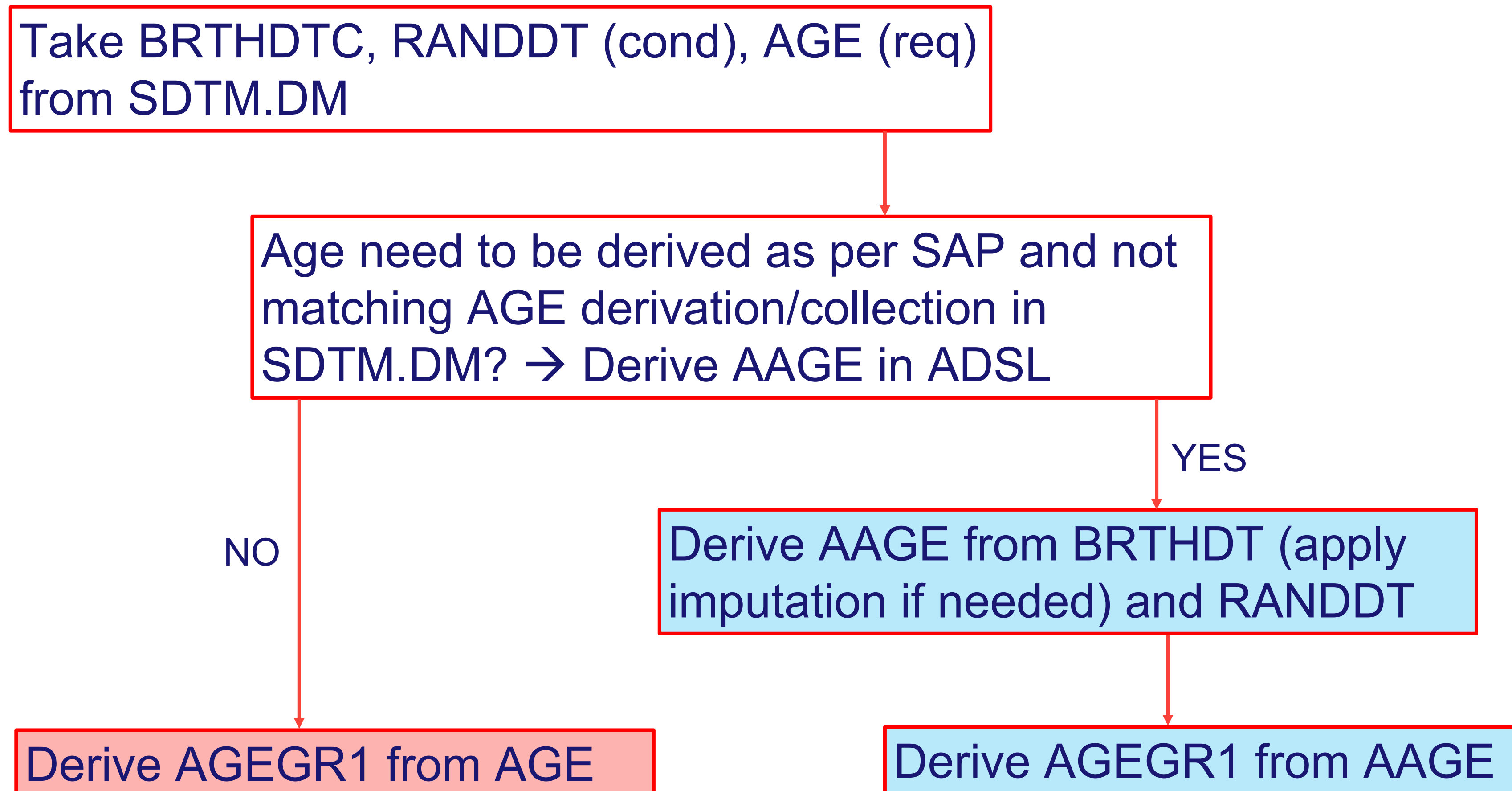
Traceability: Some Thoughts and Examples for ADaM Needs, S. Minjoe, Q. Zhou and R. Watson; PharmaSUG, 2018

Some examples

- General ADSL Traceability
- Traceability with Parameters from Multiple Input Datasets
- Traceability When Multiple Datasets Are Merged
- Traceability When Multiple Input Datasets are Stacked to Create OCCDS
- Traceability When Using a Look-Up Table
- Traceability When Adding a Row to a BDS Dataset
- Traceability When Multiple Analysis Variables are needed on the Same Row
- Using an Intermediate Dataset for BDS Traceability / Using an Intermediate Dataset for ADSL Traceability

CDISC ADaM Traceability Examples Document

General ADSL Traceability A simple example: Derive Age in Categories



CDISC ADaM Traceability Examples Document

BDS Traceability A more complicated example: [Creating Observations](#)

Study Endpoint:

- **Weekly Mean Daily Pain Intensity Score**
- **Observation occurred during the wash-out period should be not considered in the average calculation**

SUBJID	ADY	AVISIT	PARAM	AVAL	DTYPE	ANL01FL
010-8008	-7	Baseline	Average Daily Pain Inten. Score	7.00		
010-8008	-6	Baseline	Average Daily Pain Inten. Score	7.00		Y
010-8008	-5	Baseline	Average Daily Pain Inten. Score	6.00		Y
010-8008	-4	Baseline	Average Daily Pain Inten. Score	5.00		Y
010-8008	-3	Baseline	Average Daily Pain Inten. Score	6.00		Y
010-8008	-2	Baseline	Average Daily Pain Inten. Score	7.00		Y
010-8008	.	Baseline	Wk Mean Aver. Daily Pain Inten. Score	6.20	AVERAGE	Y

- Observation occurred during the wash-out period
- The observation is kept in the ADaM dataset but it will not contribute to the AVERAGE derivation (**ANL01FL=Null**)

- Derived Observation
- **DTYPE** flags Derived Observations
- **DTYPE=AVERAGE** specify the method used to derive the observation

CDISC ADaM Traceability Examples Document

BDS Traceability A more complicated example: [Creating Observations](#) (cont)

Study Endpoint:

- Weekly Mean Daily Pain Intensity Score
- Observation occurred during the wash-out period should be not considered in the average calculation

Parameter Value List - ADPAI [AVAL]

Variable	Where	Type	Source/Derivation/Comment
AVAL	PARAMCD = "KNPA0101" (Average Daily Pain Inten. Score)	integer	Derived: QS.QSSTRESN where QS.QSTESTCD=<Pain Original parameter code>. See PARAMCD selection
AVAL	PARAMCD = "KNPA0102" (Worst Pain Intensity Score)	integer	Derived: QS.QSSTRESN where QS.QSTESTCD=<Pain Original parameter code>. See PARAMCD selection
AVAL	PARAMCD = "KNPAWK" (Wk Mean Aver. Daily Pain Inten. Score)	float	Derived: Baseline weekly average daily pain defined as average of daily pain (QS.QSSTRESN where QSTESTCD='KNPA0101'). In the derivation only pain assessment occurred after end of wash-out are considered (QS.QSDTC>DS.DSSTDTC where DS.DSDECOD='END OF WASH-OUT')

Traceability Example – In your (reviewer) shoes

Traceability through the Analysis Results Metadata

ADaM-IG 1.0

Analysis Data Reviewer's Guide
AE by PT Analysis SAS Program
AE by SOC Analysis SAS Program
AE Analysis SAS Program
Alzheimer Analysis
Biomarkers Analysis SAS Program
Demographics Analysis SAS Program
ECG Analysis SAS Program
MRI Analysis SAS Program
Pk Analysis SAS Program
Vital Signs Analysis SAS Program
Analysis Results Metadata
Analysis Datasets
Parameter Value Level Metadata
Controlled Terminology
Analysis Derivations
Comments

Display	Analysis of Biomarker Data by MMRM
Analysis Result	Summary statistics for potential biomarkers and changes from baseline modeled longitudinally using Mixed Model for Repeated Measure (MMRM)
Analysis Parameter(s)	PARAMCD = "TOTSCORE" (Total score of ADAS-Cog items)
Analysis Variable(s)	AVAL (Analysis Value) CHG (Change from Baseline) AVAL (Analysis Value) CHG (Change from Baseline)
Analysis Reason	SPECIFIED IN SAP
Analysis Purpose	SECONDARY OUTCOME MEASURE
Data References (incl. Selection Criteria)	ADAB [PARAMCD = "TOTSCORE"] ADQS [PARAMCD = "TOTSCORE"] ADSL [BMFL = "Y"] Set ADAB and ADQS and Join with ADSL by USUBJID
Documentation	Summary statistics for potential biomarkers and measurements including [REDACTED] ADAS-cog 13 scores were presented by treatment group. The changes from baseline were also modeled longitudinally using Mixed Model for Repeated Measure (MMRM). [SAP section 5.4.2]
Programming Statements	[SAS Version 9.3] <pre>proc mixed data=ADAB order=internal; where BMFL='Y' and PARAMCD ne 'TOTSCORE' and ABLFL ne 'Y'; by PARAMCD; class SUBJID TRTPN VISITNUM; model CHG=BASE VISITNUM TRTPN VISITNUM*TRTPN/ddfm=kr; repeated VISITNUM/type=un subject=SUBJID; lsmeans VISITNUM*TRTPN TRTPN/pdiff cl; ods output diffs=_diff; ods output lsmeans=lsms; run; proc mixed data=ADQS order=internal; where BMFL='Y' and PARAMCD='TOTSCORE' and ABLFL ne 'Y'; class SUBJID TRTPN VISITNUM; model CHG=BASE VISITNUM TRTPN VISITNUM*TRTPN/ddfm=kr; repeated VISITNUM/type=un subject=SUBJID; lsmeans VISITNUM*TRTPN TRTPN/pdiff cl; ods output diffs=_diff; ods output lsmeans=lsms; run;</pre> Biomarkers Analysis SAS Program

Traceability Example – In your (reviewer) shoes

Traceability through the Analysis Data Reviewer Guide

5.2.7 ADPP – Pharmacokinetic and Pharmacodynamic Parameters

Purpose: XXXXX (PARCAT1= XXXX) Pharmacokinetic and Bioequivalence analysis, and YYYYYY (PARCAT1= XXXXX_YYYYYY) Pharmacodynamics analysis.

Population: all randomized subjects with at least 1 PK or PD measurement after baseline.

Important considerations: The dataset contains the primary and secondary endpoints of the study (PK analysis) and Exploratory Endpoints (PD parameters). For all analyses of the PK and PD parameters, AVAL is the variable analyzed. TRTA was used to identify the treatment a PK/PD parameter corresponds to in each treatment period.

The following table details the program, dataset and selection criteria used for the analysis of the Pk parameters (primary and secondary endpoints of the study). Of note, the geometric mean was calculated in the t-pkparam-sum SAS program.

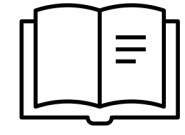
.....

Traceability Example – In your (reviewer) shoes

Traceability through the Analysis Data Reviewer Guide (cont)

Endpoints	Analysis Dataset and Selection Criteria	SAS program name
Primary endpoint: Cmax and AUC _∞	ADPP and PKFL='Y' and and ANL01FL=""Y"" and the following PARAMN/PARAM: 1 / AUC Infinity (h*pg/mL) 7 / Cmax (pg/mL)	<u>t-pk-var1-pkpop-sas.txt</u> <u>t-pk-var1-safpop-sas.txt</u>
Secondary endpoint: Cmax, AUC _∞ and other Pk parameters	ADPP and PKFL='Y' and parcat1='XXXXXX' and ANL01FL=""Y"" and the following PARAMN/PARAM: 2 / AUC0-336 (h*pg/mL) 3 / AUC0-504 (h*pg/mL)	<u>t-pkparam-sum-sas.txt</u> <u>t-pk-var2-pkpop-sas.txt</u> <u>t-pk-var2-safpop-sas.txt</u>

Health Authorities care About Traceability



CDER'S Experience with the ADaM Traceability Assessment and Common Data Quality Issues, J. Anderson; CDISC EU Interchange, 2021

Conclusion



- Traceability is a key area of analysis – an understanding of differences between SDTM, ADaM, and ISS ADaM allows reviewers to confidently use these submitted data packages throughout the review of their application
- The assessment aims to help the reviewer understand the impact of differences observed in the datasets for their application
 - Findings of higher impact, where the differences in datasets are unclear and may call into question the data practices of the applicant, also may result in an Information Request (IR) being filed
- The JumpStart team reports this wide range of observed traceability findings to inform applicants of the need for thorough data and documentation. A comprehensive submission increases the efficiency and timeliness of the review. Two essential components include:
 - Thorough documentation including derivations and other key information submitted in the Reviewer's Guide, SAP, and/or Define.xml File
 - Sequence Numbers submitted in both SDTM and ADaM datasets that map records between comparable findings and events datasets (e.g., AE and ADAE)



Verify Traceability Programmatically

- Pinnacle21 ADaM Traceability Checks
- We need more Programmatic Checks

Pinnacle21 ADaM Traceability Checks

➤ 13 Checks

➤ Need to include DM, AE, EX when validating ADaM datasets packages

- AD0053 (Error) - The combination of STUDYID and USUBJID value does not exist in the SDTM **DM** domain
- AD0061 (Error) - SDTM.**EX** is present but neither ADSL TRTSDT nor TRTSDTM are present
- AD0061A (Error) - SDTM.**EX** is present but neither ADSL TRTEDT nor TRTEDTM are present
- AD0204 (Error) - For the same USUBJID, the ADSL.AGE != **DM**.AGE
- AD0205 (Error) - For the same USUBJID, the ADSL.AGEU != **DM**.AGEU
- AD0206 (Error) - For the same USUBJID, the ADSL.SEX != **DM**.SEX
- AD0207 (Error) - For the same USUBJID, the ADSL.RACE != **DM**.RACE
- AD0208 (Error) - For the same USUBJID, the ADSL.SUBJID != **DM**.SUBJID
- AD0209 (Error) - For the same USUBJID, the ADSL.SITEID != **DM**.SITEID
- AD0210 (Error) - For the same USUBJID, the ADSL.ARM != **DM**.ARM
- AD0253 (Error) - Record key from SDTM **AE** is not traceable to ADaM ADAE (not enough ADAE recs)
- AD0258 (Error) - Record key from ADaM ADAE is not traceable to SDTM.**AE** (extra ADAE recs)
- AD0367 (Error) - For the same USUBJID, the ADSL.ACTARM != **DM**.ACTARM

We need more Programmatic Checks

Check Through Metadata (define-xml) – Check Predecessor – ADaM vs its SDTM Predecessor

ADEG (ECG Analysis Dataset) - BASIC DATA STRUCTURE Location: [adeq.xpt](#)

Variable	Where Condition	Label / Description	Type	Length or Display Format	Controlled Terms or ISO Format	Origin / Source / Method / Comment
EGDTC		Date/Time of ECG	datetime		ISO 8601	Predecessor: EG.EGDTC

*If an SDTM variable appears in an ADaM dataset, **then ALL attributes and content must remain the same.** ADaM adheres to the principle of harmonization known as “**same name, same meaning, same values**”*

- ⓪ Did we change attributes / metadata
- ⓪ Did we change variable content
- ⓪ Does EGDTC exists in EG dataset



Please Take Care of your Metadata, A. Tinazzi; PHUSE EU Connect, 2020

We need more Programmatic Checks for example

Check Through Metadata (define-xml) – Check Derivations

ADEG (ECG Analysis Dataset) - BASIC DATA STRUCTURE Location: [adeq.xpt](#)

Variable	Where Condition	Label / Description	Type	Length or Display Format	Controlled Terms or ISO Format	Origin / Source / Method / Comment
AVISIT		Analysis Visit	text	34	ADEGAVISIT "BASELINE" "DAY 1 - 1 HR POST-DOSE" "DAY 2 - 24 HRS POST-DOSE" "DAY 85" "UNSCHEDULED"	Derived If ABLFL is equal to 'Y', then AVISIT='BASELINE'. Else if ADY<0 but ABLFL is NOT equal to 'Y', then set AVISIT as missing. Else if EG.EGTPT contains "DAY XX", then AVISIT=EG.EGTPT. Else if EG.EGTPT contains "UNSCHEDULED", then AVISIT='UNSCHEDULED';



Complicated to check (Manual Review)

We need more Programmatic Checks for example

Check Through Data – Check Predecessor – ADaM vs ADSL Predecessor

ECG Analysis Dataset (ADEG) [Location: [adeg.xpt](#)]

Variable	Label	Key	Type	Length / Display Format	Controlled Terms or Format	Source/Derivation/Comment
STUDYID	Study Identifier	1	text	8		Predecessor: ADSL.STUDYID
USUBJID	Unique Subject Identifier	2	text	16		Predecessor: ADSL.USUBJID
SUBJID	Subject Identifier for the Study		text	7		Predecessor: ADSL.SUBJID
SAFFL	Safety Population Flag		text	1	["Y"] <No Yes Response (Yes Only)>	Predecessor: ADSL.SAFFL

- Variables Attributes Can be not Changed
- Variable Content Can be not changed → Safety Population (SAFFL) can be not “re-defined” in each ADaM dataset

```
%*--- Check Variable in ADSL same in other data;
%MACRO ADSL_CHK (INDS=, SOURCE=, SOURCEWH=);

  title1 "Checking obs in ADSL vs ADaMs";
  title2 "&INDS";

  /*Variable in ADSL also in &INDS*/
  proc sql noprint;
    select distinct a.name into :KEEP separated by ' '
    from metadata a, metadata b
    where a.name=b.name and a.memname='ADSL' and b.memname=upcase("&INDS");
  quit;

  /*Sort ADSL*/
  proc sort data=idad.ADSL(keep=&KEEP) out=ADSL_compare;
    by USUBJID;
  run;

  /*Nodupkey and sort &INDS*/
  proc sort data=idad.&INDS.(keep=&KEEP) out=&INDS._compare nodupkey;
    by &KEEP;
  run;
  proc sort data=&INDS._compare nodupkey;
    by USUBJID;
  run;

  /*Compare*/
  title3 "Compare Content";
  proc compare data=ADSL_compare(drop=ADSDTM) compare=&INDS._compare(drop=ADSDTM) listall MAXPRINT=30000;
    id STUDYID USUBJID;
  run;

%MEND ADSL_CHK;
```

Conclusions

Conclusions

- Good ADaM Traceability could avoid misunderstanding and questions from the Reviewer
- Traceability might facilitate automation of processes

***Traceability is the art of making
complex things simple, clear and
transparent***

References

- **« Lost » in Traceability, from SDTM to ADaM finally Analysis Results Metadata**, A. Tinazzi; CDISC EU Interchange, 2016
- **ADaM Intermediate Dataset: how to improve your analysis traceability**, A. Tinazzi, T. Curto and A. Aggarwal; PHUSE US Connect, 2018
- **Lost in Traceability- From SDTM to ADaM**, A. Tinazzi, Cytel Blog; <https://www.cytel.com/blog/lost-in-traceability>
- **Please Take Care of your Metadata**, A. Tinazzi; PHUSE EU Connect, 2020
- **CDER'S Experience with the ADaM Traceability Assessment and Common Data Quality Issues**, J. Anderson; CDISC EU Interchange, 2021
- **CDISC ADaM Example of Traceability Examples**, CDISC, To be released Q4 2021
- **Traceability: Some Thoughts and Examples for ADaM Needs**, S. Minjoe, Q. Zhou and R. Watson; PharmaSUG, 2018
- **More Traceability: Clarity in ADaM Metadata and Beyond**, W. Zhong, R. Watson, D. Ewing and J. Zhang; PharmaSUG, 2019
- **Leveraging Intermediate Data Sets to Achieve ADaM Traceability**, Y. Zhuo; PharmaSUG, 2019

Abstract

Traceability is crucial in all steps of clinical data handling, from data collection to final analysis. For analysis, good traceability starts with a clever design of ADaM datasets, where both data and metadata traceability should be guaranteed.

The CDISC ADaM Team should release in 2021 a document where additional examples of traceability will be provided; furthermore, the importance of traceability in ADaM in data submission, has been recently highlighted by the FDA and PMDA at CDISC EU 2021 (“CDER’s experience with the ADaM traceability assessment and common data quality issues”, Jesse Anderson).

Retaining --SEQ variable from SDTM source dataset is a starting point, together with other techniques recommended in the CDISC ADaM IG i.e. keeping original variables (and records) used in the derivation.

The purpose of this presentation is to re-enforce the importance of traceability in ADaM through some real examples; moreover, some programming techniques will be illustrated on how to electronically check that data traceability is satisfied, from both SDTM to ADaM and within ADaM i.e. variables copied from ADSL onto other ADaM datasets.

Thank you.

Angelo Tinazzi, Senior Director
Cytel Inc.

Standards, Systems, CDISC Consulting, Statistical Programming
Clinical Research Services

Route de Pré-Bois 20

C.P 1839, 1215 Geneva, SWITZERLAND

email: angelo.tinazzi@cytel.com

Subscribe to Cytel Blog to receive update on my
Good Data Submission Doctor Blog Series
<https://www.cytel.com/blog/topic/cdisc>

Backup Slides

How to be Traceable in ADaM (Summary)

When traceability is successfully implemented, it is possible to identify:

- Information that **exists in the submitted SDTM data**
- Information that is **derived or imputed** within the ADaM dataset
- The method used to create **derived or imputed** data
- **Information used for analyses**, in contrast to **information that is not used for analyses** yet is included to support traceability or future analysis



CDISC ADAM IG 1.2