WITH STANDARDS – UNLOCK THE POWER OF DATA

cdisc

2022
EUROPE
INTERCHANGE
27-28 APRIL | VIRTUAL EVENT

Achieving Traceability in ADaM

Silvia Faini – Principal Statistical Programmer – Cytel Inc.

Meet the Speaker

Silvia Faini

Title: Principal Statistical Programmer

Organization: Cytel Inc.

Silvia Faini has a degree in Statistics, 13 years of experience in clinical trials. During her working experience she acted as Biostatistician and then as Statistical Programmer having the chance to develop standard SAS macros and implement CDISC standards. As Principal Statistical Programmer she gained wide experience in submissions, CDISC ADaM, SDTM and define-xml and in medical device. Active member of Italian CDISC UN, E3C and CDISC Medical Device team.

Disclaimer and Disclosures

- *The views and opinions expressed in this presentation are those of the author(s) and do not necessarily reflect the official policy or position of CDISC.*

- *The author has no real or apparent conflicts of interest to report.*

Agenda

- (re)Introducing Traceability
- ADaM Traceability in Multiple Imputations
- Traceability and Estimands

(re)Introducing Traceability

- What is Traceability
- How to Achieve Traceability in ADaM

What is Traceability

In CDISC Traceability is…

- The property that enables the understanding of the data's lineage and/or the relationship between an element and its predecessor(s).
- A fundamental element of data quality and a requirement for studies submitted to regulatory authorities.
- From data collection to final analysis, traceability plays a crucial role in ensuring the integrity of source data and in reinforcing clinical research results.

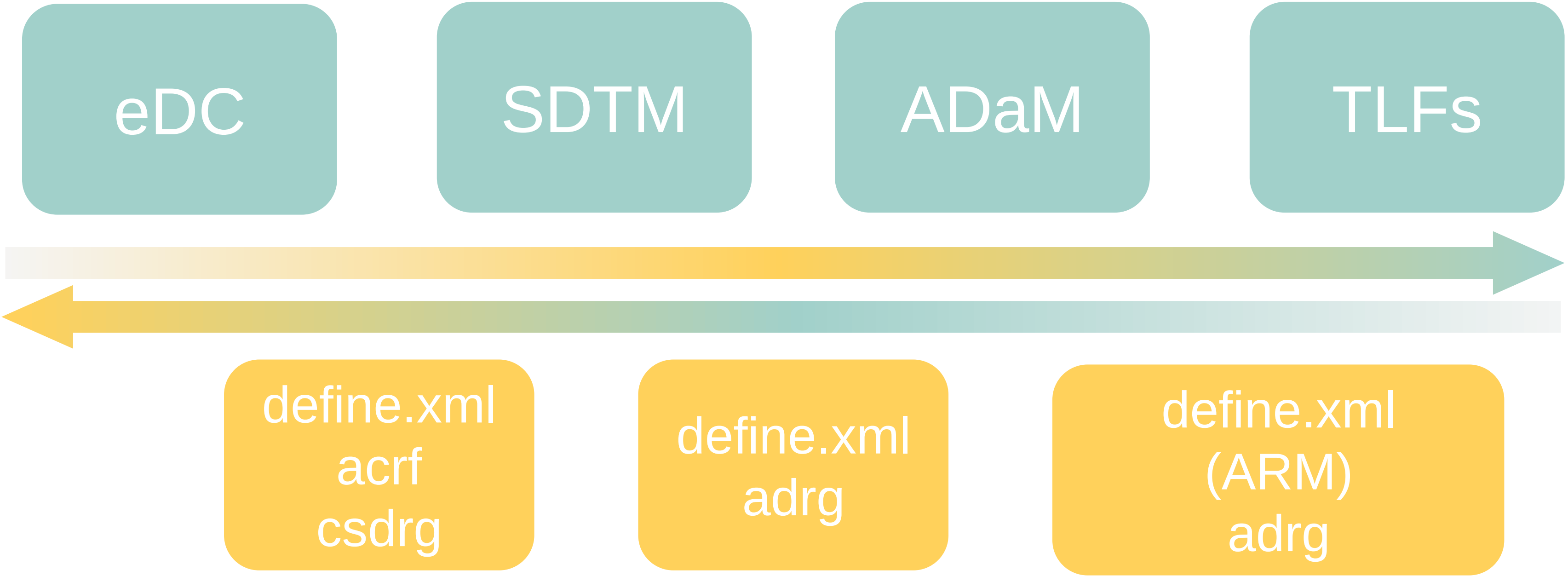
What is Traceability

FDA opinion…

- 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.

What is Traceability

- Traceability Diagram



What is Traceability

Why? What? How?

- **Why?** to facilitate transparency and understanding,
  to boost reliability and integrity

- **What?** Both ADaM and SDTM with support from define.xml provide traceability for data they represent.

- **How?** To have full traceability both SDTM and ADaM must have the appropriate documentation which establish the link between each element and its predecessor.

How to Achieve Traceability in ADaM

Metadata Traceability

- Implemented in define.xml
- Relationship of the analysis variable to other variables within SDTM or ADaM source datasets. 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.
- Relationship between an analysis results and ADaM datasets.

How to Achieve Traceability in ADaM

Datapoint Traceability

- Implemented in ADaM datasets.
- Datapoint traceability can be reached in several ways pointing directly to the specific predecessor records. Typical examples are using SRCDOM, SRCVAR, SRCSEQ variables, or --SEQ from predecessor SDTM.

How to Achieve Traceability in ADaM

	Datapoint	Metadata supportive document
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	N/A	• define.xml (ARM extension) • ADRG • SAP

ADaM Traceability in MI

- Multiple Imputations process
- Multiple Imputations in ADaMIG
- ADaM and MI process in depth

Single and multiple imputations

- Many types of imputation on missing data
- Single value imputation methods: e.g. for continuous data are baseline observation carried forward, last observation carried forward, and worst observation carried forward, for dichotomous endpoints missing values treated as failure/success.
- Multiple imputation (MI), increasing usage in the last years despite it is less easy to implement than the other imputation techniques.

Note: this presentation does not cover the selection of the appropriate multiple imputation method, which is based on Missing Data Pattern, Imputed Variable Type.

Multiple Imputations 3-steps process

Step 1: Imputation

- Each missing value is imputed based on statistical modeling, and this process is repeated several times. The output of interest from PROC MI is a data set containing multiple repetitions of the original data set, along with the newly imputed values. The repetitions are indexed with a variable named _IMPUTATION_.

Step 2: Analysis

- Analysis is done using any SAS statistical procedure the same way we analyze non-imputed data (e.g. FREQ, MEANS, MIXED procedures). However, we need to analyze each MI repetition separately. This is done by adding a BY statement with the _IMPUTATION_ variable.

Step 3: Pooling

- Need to combine all the results obtained in step 2. PROC MIANALYZE combines the results from every MI repetition and provides valid statistical inferences. Regardless of the method used to analyze the data in step 2 considering the variability introduced in step 1.

Multiple Imputations in ADaMIG v1.3

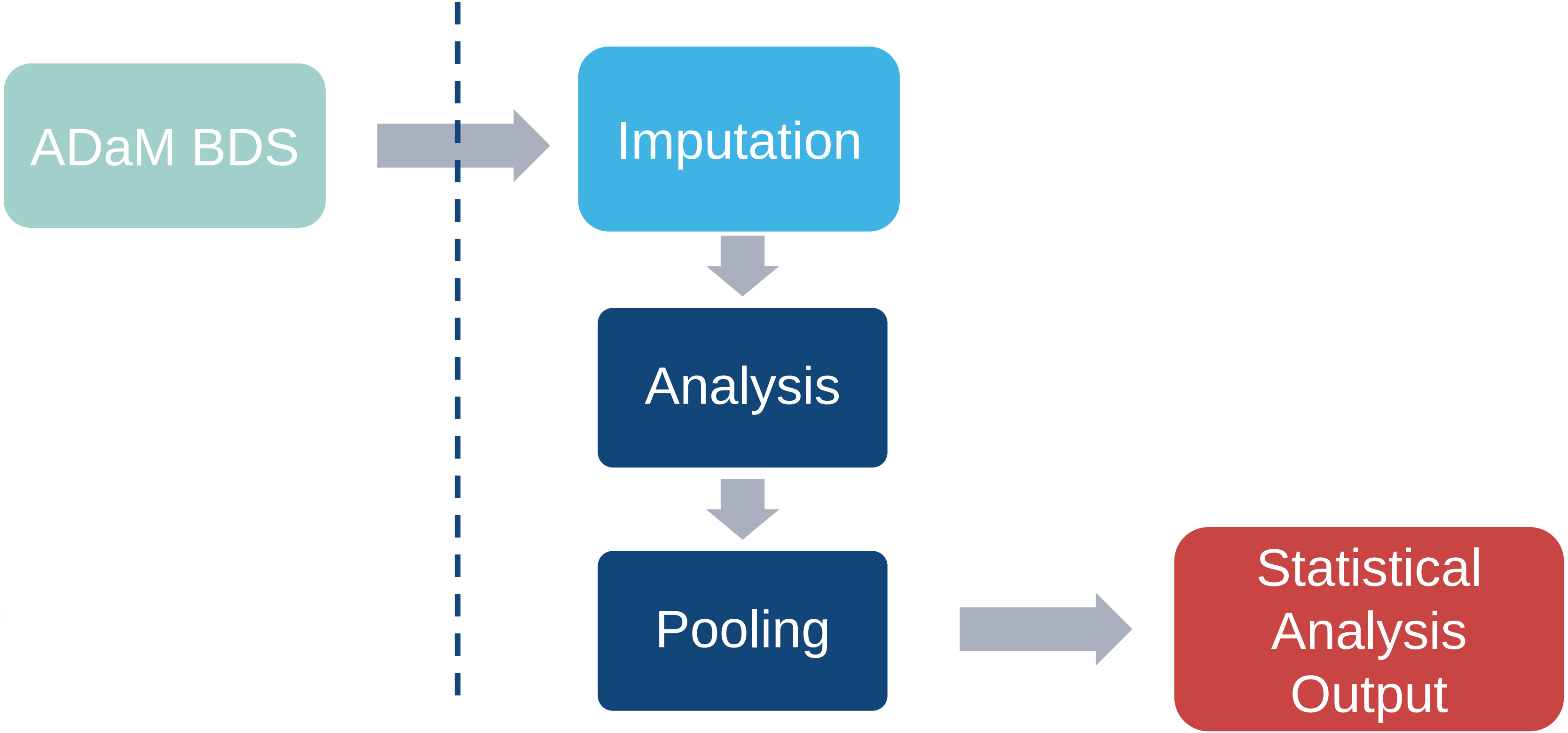
From 4.10.4 Traceability when the Multiple Imputation Method Is Used

"[…] However, documenting the traceability of estimates created via multiple imputation cannot be achieved with these current metadata methods.

In addition, it would not be practical to include all datasets that are created from the PROC MI process as part of a submission.

To address traceability, the ADaM recommendation is to provide the program statements from the three procedures mentioned above as a part of the analysis results metadata."

Multiple Imputations 3-steps process



Multiple Imputations 3-steps process

Program 1
for ADaM BDS

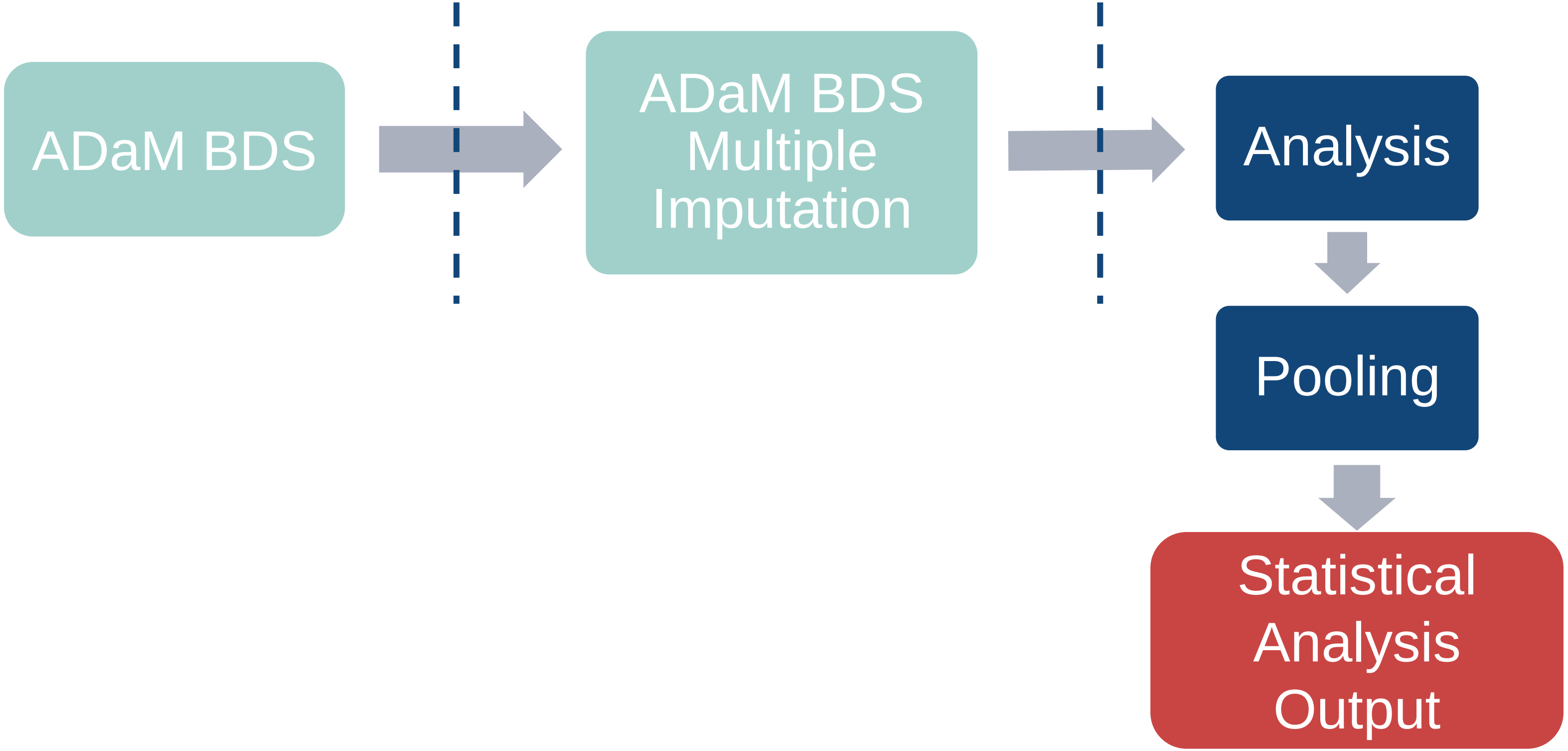
Program 2
for multiple imputations, analysis and pooling steps to produce statistical analysis output

ADaM:
not analysis-ready

Program: complex

Can Traceability be guaranteed?

ADaM and Multiple Imputations



ADaM and Multiple Imputations

Program 1
for ADaM BDS

Program 2
for ADaM BDS with
Multiple Imputations

Program 3
for analysis and
pooling steps to
produce statistical
analysis output

Complexity decrease → Traceability can be guaranteed
ADaM from program 2 is analysis-ready for statistical analysis
based on MI data.

ADaM MI: an example

- Example: datapoint traceability
- Example: metadata traceability

ADaM MI example: datapoint traceability

In below screenshot from ADaM BDS a test parameter for which one subject has missing visits 5, 6 and 7.

ID	AVISIT	PARAMCD	AVAL	BASE	CHG	ABLFL
006	Visit 1	[redacted]	4			
006	Visit 2	[redacted]	4	4	0	Y
006	Visit 3	[redacted]	1	4	-3	
006	Visit 4	[redacted]	3	4	-1	
006	Visit 8	[redacted]	4	4	0	

Transpose to apply multiple imputation

	PARAMCD	AVAL2	AVAL3	AVAL4	AVAL5	AVAL6	AVAL7	AVAL8
006	[redacted]	4	1	3				4
008	[redacted]	4	15	3		2	21	20

ADaM MI example: datapoint traceability

Output from PROC Multiple Imputation

	IMPUTATION	PARAMCD	AVAL2	AVAL3	AVAL4	AVAL5	AVAL6	AVAL7	AVAL8
006	1	[redacted]	4	1	3	2	4.3	3.6	4
006	2	[redacted]	4	1	3	3	1.4	6.7	4
006	3	[redacted]	4	1	3	9.7	2.2	2.3	4
006	4	[redacted]	4	1	3	10.1	7.2	9.9	4
006	23	[redacted]	4	1	3	8.4	9.1	6.1	4
006	24	[redacted]	4	1	3	4.8	9.2	3.8	4
006	25	[redacted]	4	1	3	1.4	7	4.3	4

ADaM MI example: datapoint traceability

- Re-transpose to fit BDS structure. Highlighted data from PROC MI and info to ensure traceability in ADaM.

JID	IMPUT	AVISIT	PARAMCD	AVAL	BASE	CHG	DTYPE	SRCDOM	SRCVAR	SRCSEQ
006	1	Visit 3 [redacted]	SCHIRMER	1	4	-3		ADOE	ASEQ	138
006	1	Visit 4 [redacted]	SCHIRMER	3	4	-1		ADOE	ASEQ	139
006	1	Visit 5 [redacted]	SCHIRMER	2	4	-2	MCMC MI			
006	1	Visit 6 [redacted]	SCHIRMER	4.3	4	0.3	MCMC MI			
006	1	Visit 7 [redacted]	SCHIRMER	3.6	4	-0.4	MCMC MI			
006	1	Visit 8 [redacted]	SCHIRMER	4	4	0		ADOE	ASEQ	140
006	2	Visit 3 [redacted]	SCHIRMER	1	4	-3		ADOE	ASEQ	138
006	2	Visit 4 [redacted]	SCHIRMER	3	4	-1		ADOE	ASEQ	139
006	2	Visit 5 [redacted]	SCHIRMER	3	4	-1	MCMC MI			
006	2	Visit 6 [redacted]	SCHIRMER	1.4	4	-2.6	MCMC MI			
006	2	Visit 7 [redacted]	SCHIRMER	6.7	4	2.7	MCMC MI			
006	2	Visit 8 [redacted]	SCHIRMER	4	4	0		ADOE	ASEQ	140

ADaM MI example: metadata traceability

- In define.xml sections: Datasets, Variables, Methods

Datasets

Dataset	Description	Class	Structure	Purpose	Keys	Documentation	Location
ADPRMI1	[redacted] MCMC MI Analysis Dataset	BASIC DATA STRUCTURE	One or more records per subject per eye per analysis parameter per analysis timepoint per	Analysis	STUDYID, USUBJID, FOCID, IMPUT, PARAM, PARAMCD, AVISITN	Include [redacted] data for primary analysis with MCMC Multiple Imputation. Input records for MI are [redacted] data for ITT subjects at scheduled visits.	adprmi1.xpt

ADPRMI1 (Schirmer Test MCMC MI Analysis Dataset) - BASIC DATA STRUCTURE

Location: adprmi1.xpt

Variable	Label / Description	Type	Length or Display Format	Controlled Terms or ISO Format	Origin / Source / Method / Comment
IMPUT	Imputation Number	integer	2		Derived Equal to _IMPUTATION_ variable derived in PROC MI procedure from the setting NIMPUTE.
AVAL	Analysis Value	float	4		Derived Equal to ADOE.AVAL for all records from ADOE where [redacted] with non-missing values. Derived with Markov Chain Monte Carlo (MCMC) Multiple Imputation method within USUBJID and FOCID for all visits for which the parameter was not collected until Visit 8 [redacted].

ADaM MI example: metadata traceability

ADRG section 3.5 Imputation/Derivation Methods

- For efficacy endpoints Markov Chain Monte Carlo (MCMC) multiple imputation, Fully Conditional Specification (FCS) multiple imputation and last observation carried forward (LOCF) methods were used. These are described in SAP [redacted]. Records imputed with one of the above listed method are identified respectively with DTYPE equal to MCMC MI, FCS MI, LOCF.

Analysis Datasets	DTYPE
ADPRMI1	MCMC MI

ADRG section 4.2 Data Dependencies

Datasets	Input ADaM Datasets
ADPRMI1	ADSL, ADEL, ADOE

ADRG specific subsection of 5.2 Analysis Datasets

5.2.6 ADPRMI1 – [redacted] MCMC MI Analysis Dataset

This is a BDS analysis dataset with more records per subject per eye per analysis parameter per analysis timepoint per imputation number. PROC MI repetitions are indexed in a variable named _IMPUTATION_, this is kept in the final dataset and renamed to have a valid ADaM name not exceeding eight characters (IMPUT).

Starting from ADOE for [redacted] records Multiple Imputations based on MCMC method was done only for scheduled visits from AVISIT=Visit 3 [redacted] to AVISIT=Visit 8 [redacted]. Imputed records have DTYPE=MCMC MI. Baseline records have been used in the program, but in the final dataset baseline values have been kept only in BASE variable.

Traceability and Estimands

Some thoughts

Traceability and Estimands

From ICH E9 R1

“An **estimand** is a precise description of the treatment effect reflecting the clinical question posed by a given clinical trial objective”

“The targets of estimation are to be defined in advance of a clinical trial.”

As per GCP

new studies must start with the end in mind

(e.g. trial design and CRF must reflect these needs)

Traceability and Estimands

From ICH E9 R1

“The description of an estimand involves precise specifications of certain attributes, which should be developed **based** not only on clinical considerations but also **on how intercurrent events are reflected in the clinical question of interest**.”

Intercurrent events concept

Use CDISC standards to achieve traceability

→ PHUSE Working Group working on this

Achieving Traceability in ADaM

Conclusion

Conclusion

Achieving Traceability in ADaM

- Metadata and Datapoint levels of traceability
- ADaM classes and ADaM intermediate datasets
- Clear and complete Documentation: define.xml and ADRG

→ achieving traceability is always possible by exploiting the CDISC standards

References

- CDISC ADaM guidelines, https://www.cdisc.org/standards/foundational/adam
- CDISC define.xml guidelines, https://www.cdisc.org/standards/data-exchange/define-xml
- PHUSE EU Connect 2021 - Paper SI08 “How to be Traceable in ADaM”, A.Tinazzi, Cytel Inc.
- PharmaSUG 2017 - Paper SP01 “Multiple Imputation: A Statistical Programming Story”, C.Smith, Cytel Inc., S.Kosten, DataCeutics Inc.
- PharmaSUG 2019 - Paper ST-160 “Experiences in Building CDISC Compliant ADaM Dataset to Support Multiple Imputation Analysis for Clinical Trials”, X.B.Cui, Alkermes Inc.
- “ICH E9 (R1) Addendum on Estimands and sensitivity analysis in clinical trials”, Final version 20Nov2019, ICH

Thank You!

Silvia Faini

Principal Statistical Programmer

Cytel Inc.

silvia.faini@cytel.com

cdisc