

DS14

Data Quality in Pooled Studies: Overcoming Challenges and Ensuring Accuracy

Monika Ludwinska

**PHUSE US Connect
Orlando, 2025**

A

INTRODUCTION

- Pooling definition
- Strategies
- Purpose

B

ORGANIZATION

- Planning
- Programming team
- Validation
- Documentation

C

KEY FACTORS AND CHALLENGES

- Mappings of treatments
- Baseline definition
- Dictionaries
- Missing variables
- Ongoing studies

D

METADATA COMPLIANCE AND STRATEGIES

- Data standards
- Programming strategies for validation

WHAT

POOLING

Combining data from multiple sources into a single, separate project to perform robust analysis on derived datasets

HOW

STRATEGY

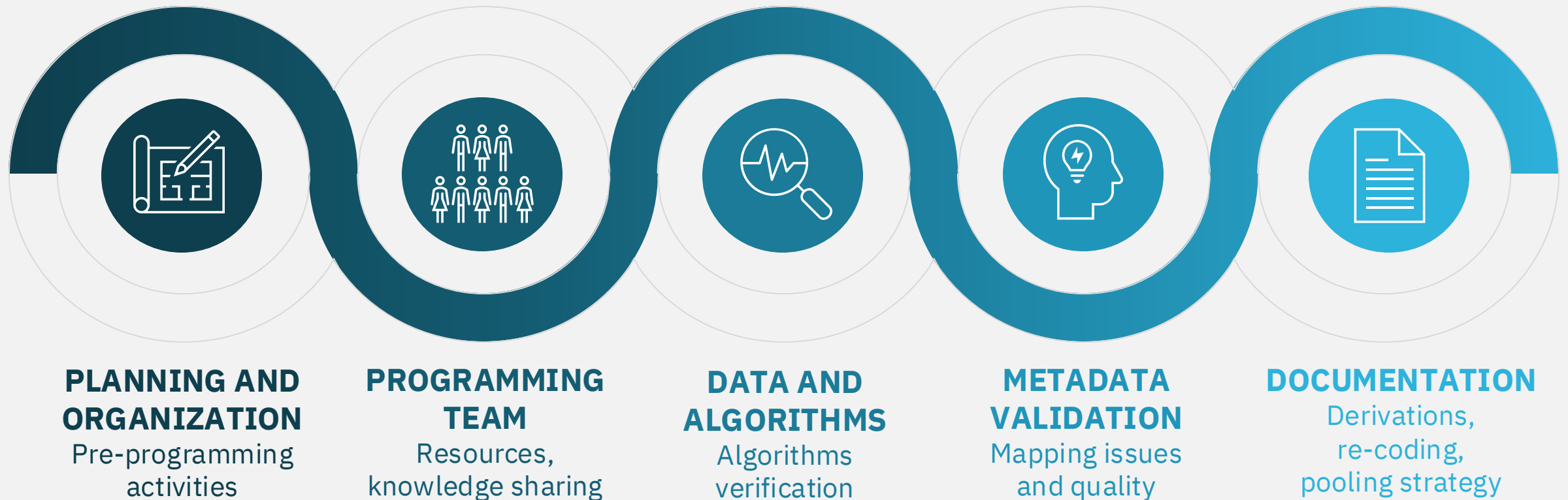
Pooling from ADaMs
Pooling from SDTMs
Pooling from ADaMs + SDTMs

WHY

ANALYSIS

Integrated studies, extension studies, exploratory analysis or publications

ORGANIZATION OF POOLING DATA AND PROGRAMMING



ORGANIZATION OF POOLING DATA AND PROGRAMMING



PLANNING AND ORGANIZATION

- Timelines
- Familiarizing with study designs and documentation
- Data and metadata organization for individual studies
- Programming strategies



PROGRAMMING TEAM

- Smaller teams vs. large teams
- Knowledge accumulation
- Study complexity



DATA AND ALGORITHMS

- Pool quality depends on the quality of individual studies
- Reliable reference
- Individual validation by the programming team



METADATA VALIDATION

- Copy-paste is not pooling
- Review of individual variables
- Codelists harmonization



DOCUMENTATION

- Strategy explanation
- Dictionaries
- Re-coding and re-derivations
- Verification of dictionaries and ADaM IG versions

KEY FACTORS AND POTENTIAL CHALLENGES

TREATMENT ISSUE

Variations in study designs lead to careful treatment mappings.

The key challenge is selecting the appropriate approach to treatment variables based on the outputs' purpose and endpoints.

BASELINE DEFINITION

Integrated Studies - separate baseline definition with traceability to individual studies' definitions.

Extension Studies – proper SDTM mapping.

Baseline - the first visit in the extension study

Baseline - the last visit in the parent study

Baseline - the baseline visit in the parent study

DICTIONARIES VERSIONS UPGRADE

Differences in MedDRA and WHO Drug versions across individual studies require upgrading to the latest dictionary versions.

Maintaining traceability to individual studies is crucial - retain original variables and document prior dictionary versions.

MISSING VARIABLES

Missing variables in the ADaM datasets
from individual studies.

Populating from individual SDTM or ADaM datasets
with proper documentation.

ONGOING STUDIES

Pooling during ongoing studies.

Challenges: Effective communication between teams,
up-to-date ADaMs.

Benefits: additional layer of review.

DOCUMENTING ISSUES AND SOLUTIONS

Effective documentation:

Pooling methodology

ADaM IG versions

Inconsistencies, re-derivations, re-codings

Harmonization

DATA STANDARDS AND SOLUTIONS

PROGRAMMING STRATEGIES

ALGORITHMS HARMONIZATION

Differences in algorithms across individual studies are inevitable.

Pooled variables must maintain consistency by using the same algorithm from individual studies.

VARIABLES HARMONIZATION

The same information can be stored in different variables.

Algorithm harmonization needs to be done along with proper naming of the pooled variable in compliance with CDISC standards.

CODELISTS HARMONIZATION

Differences in the order of coded values and variations in decode formats (uppercase vs. title case).

Consistency needs to be achieved in the pooled codelist.

ATTRIBUTES HARMONIZATION

Differences in labeling of the same information between individual studies.

Proper classification of the pooled variable: float vs. integer.

PROGRAMMING SETUP

Gathering individual study ADaMs in separate folders for easy access via libraries.

```
libname studyA "/data/pooled_project/data/studyA" access=readonly;  
libname studyB "/data/pooled_project/data/studyB" access=readonly;  
libname studyC "/data/pooled_project/data/studyC" access=readonly;
```

PROC MEANS

Discrepancies in units and labels across individual studies.

```
proc means data=studyA.ADSL stackods min max;  
run;
```

Variable	Label	Min	Max	Study
XXXBL	Complement C3 (g/L) at Baseline	0.590000	1.940000	STUDY A
XXXBL	Complement C3 at Baseline	0.710000	2.880000	STUDY B
XXXBL	Complement C3 (mg/dL) at Baseline	68.000000	191.000000	STUDY C

PROC FREQ

```
proc freq data=studyA;
```

```
tables &var.N * &var. / out=studyA_CT nocum nocol norow; run;
```

BMIBLG1N	BMIBLG1	COUNT	PERCENT	BMIBLGR1	STUDY
.		267	.		STUDY A
1	< 30	420	92.1053		STUDY A
2	>= 30	36	7.8947		STUDY A
.		45	.		STUDY B
1	< 30	97	89.8148		STUDY B
2	>= 30	11	10.1852		STUDY B
1		904	90.4000	<30 kg/m2	STUDY C
2		96	9.6000	>=30 kg/m2	STUDY C

MAXIMUM LENGTH OF VARIABLES

```
%checkLengthNum(studies=, adams=, variables=);
```

NAME	LENGTH	SIGNIFICANT	NOTE	STUDY
CREATBL	5	2	Float	STUDY A
CREATBL	5	2	Float	STUDY B
CREATBL	12	9	Float	STUDY C

THANK YOU

Monika Ludwinska

Intego Group LLC

Email: monika.ludwinska@intego-group.com

LinkedIn: <https://www.linkedin.com/in/monika-ludwinska/>

QUESTIONS?

