

ADaMs – Why Do We Produce Them?

DS08



Disclaimer

Views/opinions expressed in this presentation are my own and do not reflect Phastar or CDISC

Target audience are those with minimal SDTM and ADaM knowledge

Agenda

- Introduction to ADaMs
- How to make TFL production easier
- Challenges with ADaM production
- ADaM Define.xml



- **Introduction to ADaMs**
- How to make TFL production easier
- Challenges with ADaM production
- ADaM Define.xml



Introduction to ADaMs

- ADaM = **A**nalysis **D**ata **M**odel
- Datasets for analysis purposes
- Structurally similar to SDTM

Introduction to ADaMs

- Fundamental principles – traceability
 - Raw -> SDTM -> Intermediate datasets (if applicable) -> ADaM
 - Important to provide clear traceability to be able to pinpoint values back to raw level
 - Metadata traceability
 - Include clear derivations in specs
 - Datapoint traceability
 - Include variables in ADaM that can link to SDTM, e.g. xxSEQ

Introduction to ADaMs

- Fundamental principles – analysis ready
 - Minimal programming required in TFL programs
 - ADaMs set up to include all variables and parameters you may need

Introduction to ADaMs

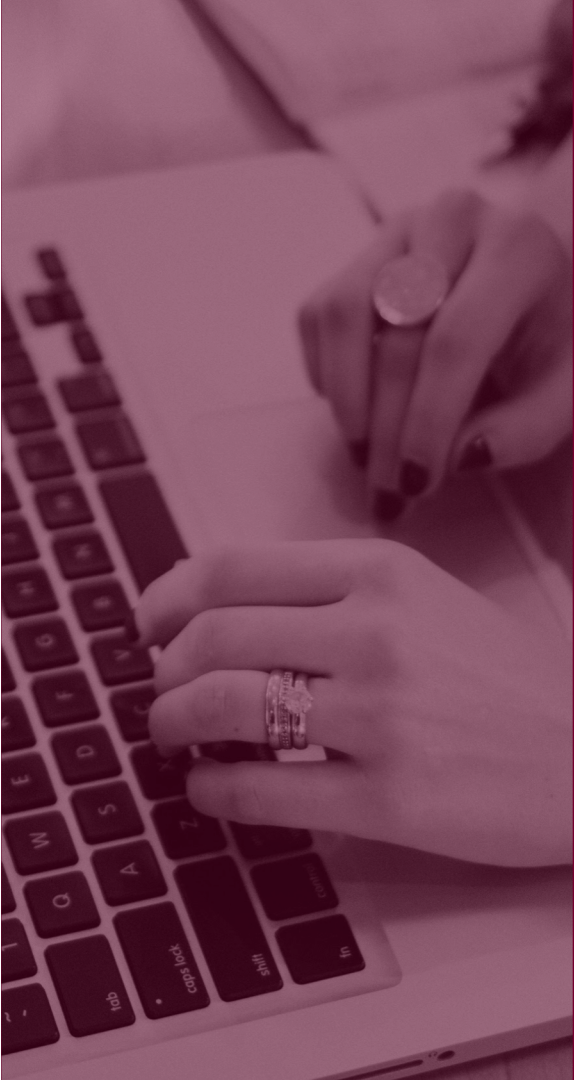
- ADaM structure
 - ADSL
 - One dataset: ADSL, Subject-Level Analysis Dataset
 - One record per subject
 - Contains population flags, treatments, demographic information, disposition information, etc
 - BDS (Basic Data Structure)
 - One record per subject per parameter (per timepoint)
 - Contains parameters and units, timepoints (if applicable), results

Introduction to ADaMs

- ADaM structure
 - OCCDS (Occurrence Data Structure)
 - One record per subject per event
 - Occurrence based

Introduction to ADaMs

- Metadata
 - “Data about data”
 - ADaM specifications
 - Shows source, controlled terminology, and derivation of variables/parameters
 - Provides traceability to SDTM
 - Important to set these up correctly when programming ADaMs to ensure nothing is missed and that they are set up with clear information
 - Used in ADaM define.xml

- 
- Introduction to ADaMs
 - **How to make TFL production easier**
 - Challenges with ADaM production
 - ADaM Define.xml



Easy TFL Production

- Some examples of how we can make TFL production easier by using ADaMs:
 - Deriving change/percentage change from baseline
 - Deriving durations (e.g. adverse events, exposure)
 - Deriving parameters/records not collected at CRF level (e.g. corrected calcium, mean, WOCF, LOCF)

Easy TFL Production

- Change and percentage change from baseline
 - Key variables: ABLFL, AVAL, BASE, CHG, PCHG
 - Parameter invariant
 - Populated regardless of whether analysis is needed

	SUBJID	PARAM	AVISIT	ABLFL	AVAL	BASE	CHG	PCHG
1	1	Hemoglobin (g/L)	Baseline	Y	96	96		
2	1	Hemoglobin (g/L)	Week 1		99	96	3	3.125
3	1	Hemoglobin (g/L)	Week 2		106	96	10	10.41666667
4	1	Hemoglobin (g/L)	Week 3		103	96	7	7.291666667
5	1	Hemoglobin (g/L)	Week 4		106	96	10	10.41666667
6	1	Hemoglobin (g/L)	Week 5		103	96	7	7.291666667
7	1	Hemoglobin (g/L)	Week 8		99	96	3	3.125
8	2	Hemoglobin (g/L)	Baseline	Y	132	132		
9	2	Hemoglobin (g/L)	Week 4		121	132	-11	-8.333333333
10	2	Hemoglobin (g/L)	Week 8		122	132	-10	-7.575757576
11	2	Hemoglobin (g/L)	Week 12		134	132	2	1.515151515
12	2	Hemoglobin (g/L)	Week 16		136	132	4	3.030303030
13	2	Hemoglobin (g/L)	Week 20		140	132	8	6.060606061
14	2	Hemoglobin (g/L)	Week 24		143	132	11	8.333333333

Easy TFL Production

- Durations
 - Durations of exposure, adverse events, concomitant medications
 - Key variables: ADURN, ADURU
 - Simple duration derivation: $AENDT - ASTDT$
 - +1 for dates after first dose

	SUBJID	EXTRT	ASTDT	AENDT	ADURN	ADURU
1	1	Treatment 1	17JAN2020	05MAR2020	49 Days	
2	1	Treatment 2	17JAN2020	03JUN2020	139 Days	
3	2	Treatment 1	13FEB2020	17MAR2020	34 Days	
4	2	Treatment 2	13FEB2020	06JUL2020	145 Days	

Easy TFL Production

- Derived records
 - Can derive whole parameters or just records within existing parameters
 - Key variables: DTYPE
 - DTYPE only needed for derived records within a parameter, not needed for wholly derived parameters
 - E.g. corrected calcium (whole parameter), LOCF, WOCF, mean, median

Easy TFL Production

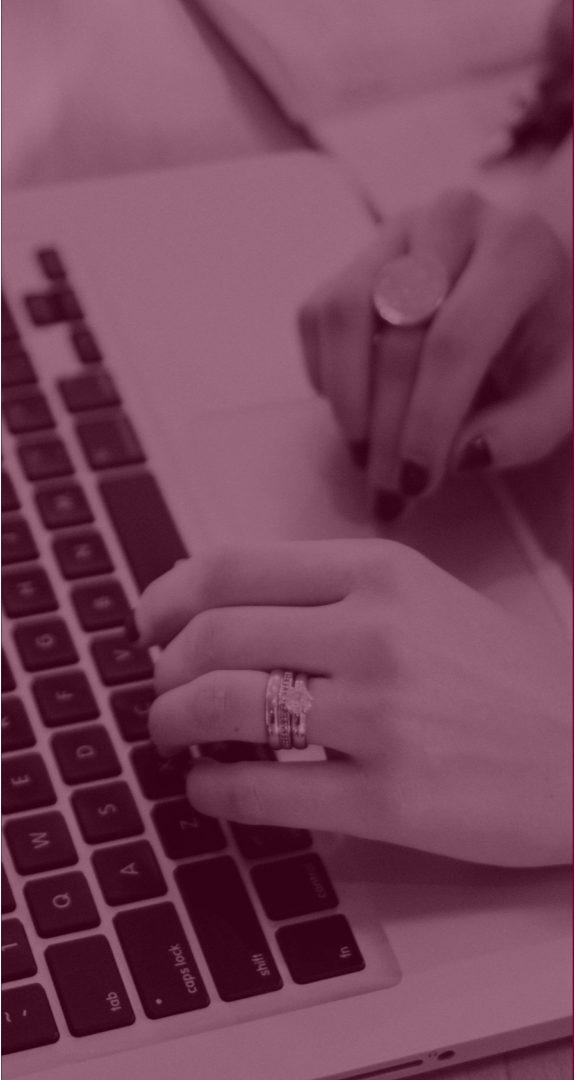
- Derived records
 - Whole parameter derived, no need to populate DTYPE
 - Derivation included in metadata

	SUBJID	PARAM	AVAL
1	1	Calcium (mmol/L)	2.32035
2	1	Calcium (mmol/L)	2.3453
3	1	Albumin (g/L)	44.3
4	1	Albumin (g/L)	43.1
5	1	Calcium, Corrected (mmol/L)	2.23435
6	1	Calcium, Corrected (mmol/L)	2.2833
7	2	Calcium (mmol/L)	2.22055
8	2	Calcium (mmol/L)	2.17065
9	2	Calcium (mmol/L)	2.47005
10	2	Albumin (g/L)	33.5
11	2	Albumin (g/L)	34.3
12	2	Albumin (g/L)	36.4
13	2	Calcium, Corrected (mmol/L)	2.35055
14	2	Calcium, Corrected (mmol/L)	2.28465
15	2	Calcium, Corrected (mmol/L)	2.54205

Easy TFL Production

- Derived records
 - Mean derived within an existing parameter
 - DTYPE added to clearly show derived record, derivation included in metadata
 - DTYPE has a codelist per CDISC controlled terminology, but is extensible

	SUBJID	PARAM	AVISIT	LBDTC	ADT	AVAL	DTYPE
1	1	Prothrombin Complex INR (RATIO)	Screening	2019-11-16	16NOV2019	1.15	
2	1	Prothrombin Complex INR (RATIO)	Screening	2019-11-16	16NOV2019	1.1	
3	1	Prothrombin Complex INR (RATIO)	Baseline		16NOV2019	1.125	AVERAGE

- 
- Introduction to ADaMs
 - How to make TFL production easier
 - **Challenges with ADaM production**
 - ADaM Define.xml



Challenges

- One to one mapping
 - PARAM and PARAMCD
 - PARCATy added complication!

SDTM:

	USUBJID	LBTESTCD	LBTEST	LBCAT
1	1	GLUC	Glucose	CHEMISTRY
2	1	GLUC	Glucose	URINALYSIS

ADaM:

	SUBJID	PARAM	PARAMCD	PARCAT1
1	1	Serum Glucose (mmol/L)	GLUC1	CHEMISTRY
2	1	Urine Glucose (mmol/L)	GLUC2	URINALYSIS

Challenges

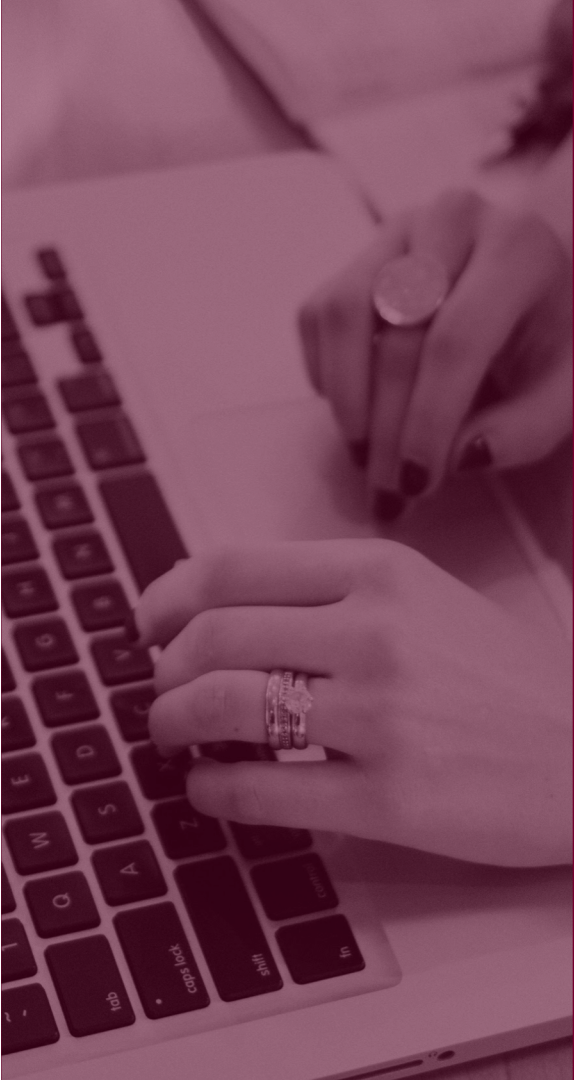
- New variable vs new record
 - New variable:
 - Parameter invariant
 - Derivation includes AVAL and/or BASE, and BASE is not changed in any way
 - Example: CHG, PCHG, R2BASE, $\log_{10}(\text{AVAL})$

Challenges

- New variable vs new record
 - New record:
 - Parameter variance, or AVAL and BASE are needed but BASE needs to be transformed (e.g. $\log_{10}(\text{BASE})$)
 - Multiple timepoints within the same parameter are needed for the derivation -> new record within the parameter derived
 - Multiple parameter results needed for the derivation -> new record derived for the parameter
 - Multiple baselines within a parameter -> new records within the parameter derived for the baselines

Challenges

- Large amounts of data
 - Large amount of data from SDTM + deriving records + adding variables...!
 - Large datasets can be an issue to transfer to reviewers
 - May need to split datasets
 - Datasets greater than 5GB (XPT format) should be split (follow FDA guidance in technical conformance guide)
 - Splits and unsplit versions sent for review
 - Example: Splitting ADLB into chemistry, haematology, urinalysis data

- 
- Introduction to ADaMs
 - How to make TFL production easier
 - Challenges with ADaM production
 - **ADaM Define.xml**

Define.xml

- What is a define.xml?
 - Data definitions document to show datasets, variables, controlled terminology, and derivations for the study
 - One for SDTM, one for ADaM
 - Specifications usually form the basis of the define
 - Regulators review the define package

Define.xml

- What makes up the ADaM define package?
 - Define.xml
 - Analysis Data Reviewers Guide (ADRG), complex algorithms document
 - Datasets (in XPT format)
 - Analysis Results Metadata (ARM)
 - Only required for PMDA at this time, not a requirement for studies submitting to the FDA currently
 - ADaM and table/figure analysis programs (ASCII text format)

Define.xml

- Analysis Data Reviewer's Guide (ADRG)
 - Document containing information on ADaMs included in study
 - Provides understanding of the study design, any important data dependencies, the derivations used in datasets, and validation issues
 - Included in define package as a PDF -> "adrg.pdf"
 - Note, different name required for PMDA submission
 - Also included as a hyperlink in define.xml



Thank you for listening!

Any questions, please email kate.booth@phastar.com