

WITH STANDARDS – UNLOCK THE POWER OF DATA

cdisc

2022
US
INTERCHANGE
26-27 OCTOBER | AUSTIN



Making an ADaM Dataset Analysis-Ready

Presented by Sandra Minjoe,
Senior Principal Clinical Data Standards Consultant
Global Data Standards
ICON PLC



Meet the Speaker

Sandra Minjoe

Title: Senior Principal Clinical Data Standards Consultant

Organization: ICON PLC

Sandra Minjoe started programming in the pharma/biotech industry in 1993. She is a Senior Principal Clinical Data Standards Consultant at ICON PLC.

Sandra is a former CDISC ADaM Team Lead. She's been part of the ADaM team since 2001, proposed structures that became ADSL and OCCDS, and continues to work on sub-teams.

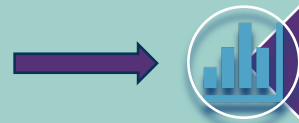


Disclaimer and Disclosures

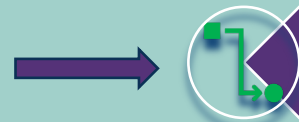
- *The views and opinions expressed in this presentation are those of the author and do not necessarily reflect the official policy or position of CDISC.*
- *The author has no real or apparent conflicts of interest to report.*



Agenda: Analysis-Ready Questions



What does “Analysis-Ready” actually mean?



What types of data manipulation are allowed when producing statistical output?



How do I choose the best dataset structure?



What variables will help me efficiently create different types of outputs?



When does “Analysis-Ready” not apply?



What does “Analysis-Ready” mean?

- Quotes from ADaM model document (v2.1) Section 3.1:
 - *analysis datasets ... have a structure and content that allows statistical analysis to be performed with minimal programming*
 - *No complex data manipulations such as transformations or transpositions are required to perform the supported analysis*
- Quote from ADaMIG v1.3 Section 2.1:
 - *ADaM datasets contain the data needed for the review and re-creation of specific statistical analyses*





What does “Analysis-Ready” mean?

- Quotes from ADaM model document (v2.1) Section 3.1:
 - *analysis datasets ... have a structure and content that allows statistical analysis to be performed with minimal programming*
 - *No complex data manipulations such as transformations or transpositions are required to perform the supported analysis*
- Quote from ADaMIG v1.3 Section 2.1:
 - *ADaM datasets contain the data needed for the review and re-creation of specific statistical analyses*





Analysis-Ready Dataset = Minimal Programming to Produce Results

- Example manipulations allowed when producing results
 - Sorting
 - Subsetting
- Example manipulations not allowed when producing results
 - Transposing
 - Merging/joining datasets
 - Adding variables

➤ ***Instead, do this programming within the analysis dataset***





Agenda: Analysis-Ready Questions



What does “Analysis-Ready” actually mean?



What types of data manipulation are allowed when producing statistical output?



How do I choose the best dataset structure?



What variables will help me efficiently create different types of outputs?



When does “Analysis-Ready” not apply?

Choosing the Dataset Structure to make the Dataset Analysis-Ready

- Most of the time, an ADaM standard structure can be used

ADSL is analysis-ready for summary tables of

- standard demography
- baseline characteristics
- disposition

BDS is analysis-ready for many analyses, including

- change from baseline
- shift
- time-to-event
- ANCOVA and ANOVA
- chi-squared
- CMH
- Fisher's exact
- log rank
- regression
- t-tests

OCCDS is analysis-ready for occurrence analyses, where subjects within a hierarchy are counted, such as tables for standard analysis of

- adverse events
- concomitant meds
- medical history



Common ADaM Structure Choice to make the Dataset Analysis-Ready

SDTM Structure	Dataset Content
Events	AE data
	Disposition data
Findings	Lab data
Interventions	Con Med data
	Exposure data





Common ADaM Structure Choice to make the Dataset Analysis-Ready

SDTM Structure	Dataset Content	Common ADaM Structure
Events	AE data	
	Disposition data	
Findings	Lab data	
Interventions	Con Med data	
	Exposure data	



Common ADaM Structure Choice to make the Dataset Analysis-Ready

SDTM Structure	Dataset Content	Common ADaM Structure
Events	AE data	Occurrence
	Disposition data	ADSL or BDS
Findings	Lab data	
Interventions	Con Med data	
	Exposure data	

Common ADaM Structure Choice to make the Dataset Analysis-Ready

SDTM Structure	Dataset Content	Common ADaM Structure
Events	AE data	Occurrence
	Disposition data	ADSL or BDS
Findings	Lab data	BDS
Interventions	Con Med data	
	Exposure data	

Common ADaM Structure Choice to make the Dataset Analysis-Ready

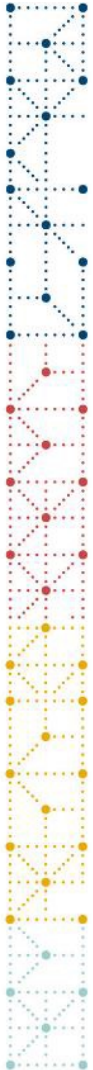
SDTM Structure	Dataset Content	Common ADaM Structure
Events	AE data	Occurrence
	Disposition data	ADSL or BDS
Findings	Lab data	BDS
Interventions	Con Med data	Occurrence
	Exposure data	BDS



Choosing the Dataset Structure to make the Dataset Analysis-Ready

- Full set of ADaM dataset class CT

ADAM OTHER
BASIC DATA STRUCTURE
DEVICE LEVEL ANALYSIS DATASET
MEDICAL DEVICE BASIC DATA STRUCTURE
MEDICAL DEVICE OCCURRENCE DATA STRUCTURE
OCCURRENCE DATA STRUCTURE
SUBJECT LEVEL ANALYSIS DATASET



Choosing the Dataset Structure to make the Dataset Analysis-Ready

- Full set of ADaM dataset class CT

ADAM OTHER
BASIC DATA STRUCTURE
DEVICE LEVEL ANALYSIS DATASET
MEDICAL DEVICE BASIC DATA STRUCTURE
MEDICAL DEVICE OCCURRENCE DATA STRUCTURE
OCCURRENCE DATA STRUCTURE
SUBJECT LEVEL ANALYSIS DATASET

- Make use of ADAM OTHER when no standard structure will work
 - Example: multivariate analysis



Agenda: Analysis-Ready Questions



What does “Analysis-Ready” actually mean?



What types of data manipulation are allowed when producing statistical output?



How do I choose the best dataset structure?



What variables will help me efficiently create different types of outputs?



When does “Analysis-Ready” not apply?



Variables to help Create Outputs

- Examples of what you can do to aid analysis:
 1. Make text in parameters = text displayed on tables and figures
Tip: Use those 200 characters in PARAM so no formatting is needed
 2. Create timing variables that match the categories needed for analysis
Tip: Make the text of timing variables = text displayed on tables and figures
 3. Include variables to arrange content on the tables and figures
Examples: RACEN sorts RACE, AVISITN sorts AVISIT
 4. Include variables that show which rows from the dataset are used for each specific analysis table and figure
Examples: ANL03FL, CRIT2, ACAT4



Variables to help Create Outputs

- Examples of what you can do to aid analysis:
 1. Make text in parameters = text displayed on tables and figures
Tip: Use those 200 characters in PARAM so no formatting is needed
 2. Create timing variables that match the categories needed for analysis
Tip: Make the text of timing variables = text displayed on tables and figures
 3. **Include variables to arrange content on the tables and figures**
Examples: RACEN sorts RACE, AVISITN sorts AVISIT
 4. **Include variables that show which rows from the dataset are used for each specific analysis table and figure**
Examples: ANL03FL, CRIT2, ACAT4





Variables to help Create Outputs

- **Tip:** Use AVAL to sort categorical AVALC

AVALC	AVAL
Low	1
Medium	2
High	3



Variables to help Create Outputs

- **Tip:** Use AVAL for graphical display of AVALC

AVALC	AVAL
0-1	0.5
2-4	3
5-9	7
10-20	15
20-40	30

- *AVAL is the centre of each AVALC range*

- **Tip:** Use AVISTN for graphical display of AVISIT

AVISIT	AVISITN
Baseline	0
Week 1	1
Week 4	4
Month 3	13
Month 6	26

- *Put AVISITN in consistent units*

Variables to Connect Dataset and Output

row	AVISIT	AVISITN	AVAL	BASE	CHG	DTYPE	ANL01FL	ANL02FL
1	Baseline	0	114	114	0		Y	Y
2	Week 2	2	118	114	4		Y	Y
3	Week 2	2	126	114	12			
4	Week 4	4	122	114	8		Y	Y
5	Week 8	8	122	114	8	LOCF	Y	
6	Week 8	8	126	114	12	WOCF		Y
7	Week 12	12	134	114	20		Y	Y

- ▶ Rows 2 & 3 are both windowed to Week 2
 - Row 2 has ANL01FL = 'Y' and ANL02FL = 'Y' (used in both of these analyses)
 - Row 3 has missing values for both (not used in either analyses)

Variables to Connect Dataset and Output

row	AVISIT	AVISITN	AVAL	BASE	CHG	DTYPE	ANL01FL	ANL02FL
1	Baseline	0	114	114	0		Y	Y
2	Week 2	2	118	114	4		Y	Y
3	Week 2	2	126	114	12			
4	Week 4	4	122	114	8		Y	Y
5	Week 8	8	122	114	8	LOCF	Y	
6	Week 8	8	126	114	12	WOCF		Y
7	Week 12	12	134	114	20		Y	Y

- ▶ Rows 5 & 6 are both windowed to Week 8 using different DTYPEs
 - Row 5 (DTYPE=LOCF) is used with ANL01FL analysis
 - Row 6 (DTYPE=WOCF) is used with ANL02FL analysis





Agenda: Analysis-Ready Questions



What does “Analysis-Ready” actually mean?



What types of data manipulation are allowed when producing statistical output?



How do I choose the best dataset structure?



What variables will help me efficiently create different types of outputs?



When does “Analysis-Ready” not apply?



When “Analysis-Ready” does not Apply

- The concept of “Analysis-Ready” applies only to ADaM datasets used for actual analysis
- Datasets not used for analysis do not need to be “Analysis-Ready”
- Common ADaM datasets not used for analysis are:
 - Datasets used solely for listings
 - Intermediate datasets that are not analyzed



When “Analysis-Ready” does not Apply

- The concept of “Analysis-Ready” applies only to ADaM datasets used for actual analysis
- Datasets not used for analysis do not need to be “Analysis-Ready”
- Common ADaM datasets not used for analysis are:
 - Datasets used solely for **listings**
 - **Intermediate datasets** that are not analyzed



When “Analysis-Ready” does not Apply: Listings

- Listings do not need to be “Analysis-Ready”
 - In a listing program you can transpose content and derive new variables
 - A listing can be created from an SDTM dataset merged with ADSL
- It is not necessary to create an ADaM dataset when only listings are needed, but it is not “wrong” to do so
 - Some companies require that all listings be created from ADaM
- **Tip:** watch out for listings that require concatenation of multiple variables into long text strings
 - Must follow SAS XPT file v5 requirements for any submitted dataset
 - May need to concatenate to larger than 200 characters within the listing program



When “Analysis-Ready” does not Apply: Intermediate Datasets

- Intermediate datasets are used to split out some of the work in complex derivations
 - Usually not used for analysis, but useful for review and maybe listings
- Common intermediate datasets
 - Collecting dates, prior to a time-to-event ADaM dataset
 - Doing some prep work, prior to an exposure summary ADaM dataset



When “Analysis-Ready” does not Apply: Intermediate Datasets

- Intermediate datasets are used to split out some of the work in complex derivations
 - Usually not used for analysis, but useful for review and maybe listings
- Common intermediate datasets
 - Collecting dates, prior to a time-to-event ADaM dataset
 - Doing some prep work, prior to an exposure summary ADaM dataset

When “Analysis-Ready” does not Apply: Intermediate Dataset prior to ADTTE

- FDA presentation recommends this approach

ADINTEV - Intermediate Event Dataset				Traceability back to SDTM			Traceability to TTE
USUBJID	PARAM	ADY	ADT	SRCDOM	SRCVAR	SRCSEQ	ASEQ
1001	Date of Randomization	1	5/3/2016	ADSL	RANDDT		1
1001	Date of Progression Per BICR	84	7/25/2016	RS	RELPSDTC	912	2
1001	Date of Last Non-PD Assess per BICR	37	6/8/2016	RS	RSDTC	915	3
1001	Date of Progression Per INV	37	6/8/2016	RS	RSDTC	649	4
1001	Date of Death	195	11/13/2016	DD	DDDTC	672	5
1001	Date of Treatment Discontinuation	151	9/30/2016	ADSL	TR01EDT		6

- [Kam, Matilde. 2019. “FDA Review Process: Recommendations for Review-Ready Submissions to CDER”. CDISC Annual Interchange 2019 Proceedings. Available to CDISC members at \[https://www.cdisc.org/system/files/all/event/restricted/2019_US/Session_8/MKamCDISCUSInt2019_04OCT2019_FINAL.pdf\]\(https://www.cdisc.org/system/files/all/event/restricted/2019_US/Session_8/MKamCDISCUSInt2019_04OCT2019_FINAL.pdf\).](https://www.cdisc.org/system/files/all/event/restricted/2019_US/Session_8/MKamCDISCUSInt2019_04OCT2019_FINAL.pdf)



When “Analysis-Ready” does not Apply: Intermediate Dataset prior to ADTTE

- Note: FDA presentation included BDS variable PARAM
 - Since the intermediate dataset is NOT used for analysis, what does “Analysis Parameter” mean in this dataset?
- Intermediate datasets are not used directly in analysis
 - No need to make them “Analysis-Ready” or even use a standard structure
- Instead, make the intermediate dataset “Analysis-Dataset-Ready”
 - Goal is to make it easier to derive the next ADaM dataset
 - **Prostate Cancer TAUG** has a nice pre-ADTTE example
 - Uses some BDS variables, but not PARAM, PARAMCD, or AVAL/AVALC

Agenda: Analysis-Ready Questions



What does “Analysis-Ready” actually mean?

Not allowed:
Transposing,
Merging/joining,
Adding
variables

Allowed:
Sorting,
Subsetting



What types of data manipulation are allowed when producing statistical output?



How do I choose the best dataset structure?

ADSL, BDS, OCCDS, ADAM
OTHER



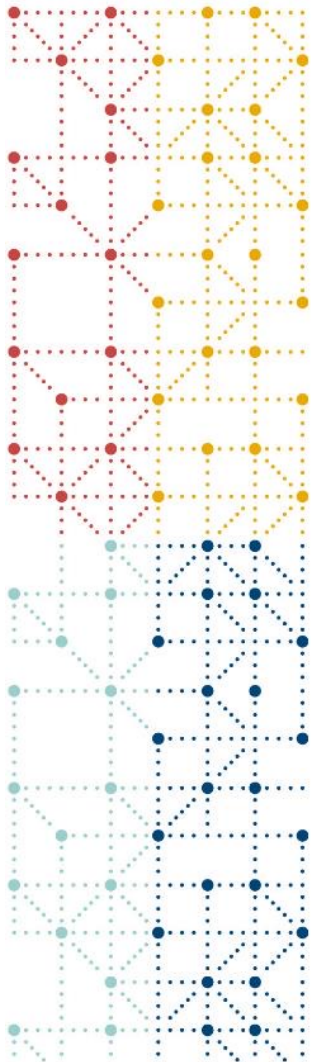
What variables will help me efficiently create different types of outputs?

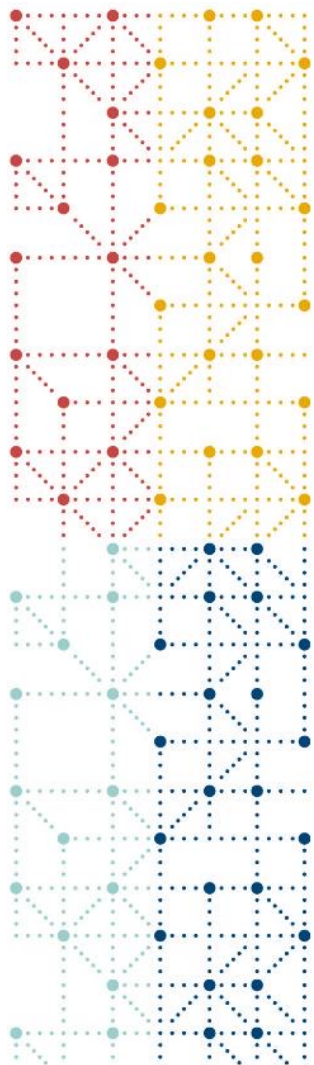
RACEN, AVISITN, AVAL,
ANL01FL



When does “Analysis-Ready” not apply?

Listings, Intermediate
datasets





Thank You!

cdisc