SI04 - Analysis Results Metadata 1.0 for Define-XML 2.0 - Benefits to Statistical Analysis

Lionel Debecq, 10th October 2017

Agenda

Analysis Results Metadata for Define-XML?

Purpose of Analysis Results Metadata (ARM)

Generation Method

Conclusions

Q&A

Analysis Results Metadata for Define-XML?

- **Metadata = definition of data**
- **CDISC Define-XML**
  - Study Metadata
  - Machine Readable (XML)
  - Review by Authorities

Analysis Results Metadata for Define-XML?

- **Analysis Results Metadata (ARM) ?**

  = Define-XML Extension

- **Results Description**
  - What ?
  - How ?
  - Where ?

Advantages

- **Multiple Advantages**
  - Reusable (similar studies)
  - Added Value: Improve Review, Add info
  - Traceability

Agenda

Analysis Results Metadata for Define-XML?

Purpose of Analysis Results Metadata (ARM)

Generation Method

Conclusions

Q&A

Data Transparency

*« **Traceability** is the ability to track an element, a piece of data, to its origin. »*

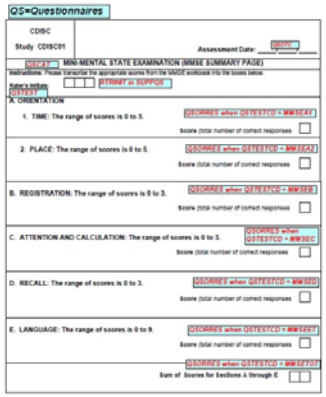
Data Understanding

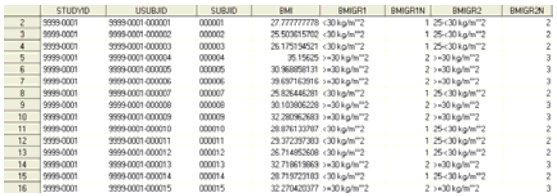
Traceability

- **Relationship : analysis results - analysis datasets - SDTM domains**
- **Path : element - immediate predecessor**
- **Metadata traceability**
  - Analysis Result – Dataset
  - Analysis Variable – Source
- **Data point traceability**
  - Predecessor

Table 1 Demographic Data - Per-Protocol

	Treatment 1	Treatment 2
Baseline body mass index (BMI) [kg/m**2]		
N	167	167
Mean	29.08	29.04
SD	4.84	4.80
Min	20.3	16.0
Median	28.69	28.47
Max	40.1	41.2
Baseline BMI (categorical) [N (%)]		
<25 kg/m**2	41 (24.6%)	71 (21.1%)
25-<30 kg/m**2	60 (35.9%)	130 (38.7%)
>=30 kg/m**2	66 (39.5%)	135 (40.2%)



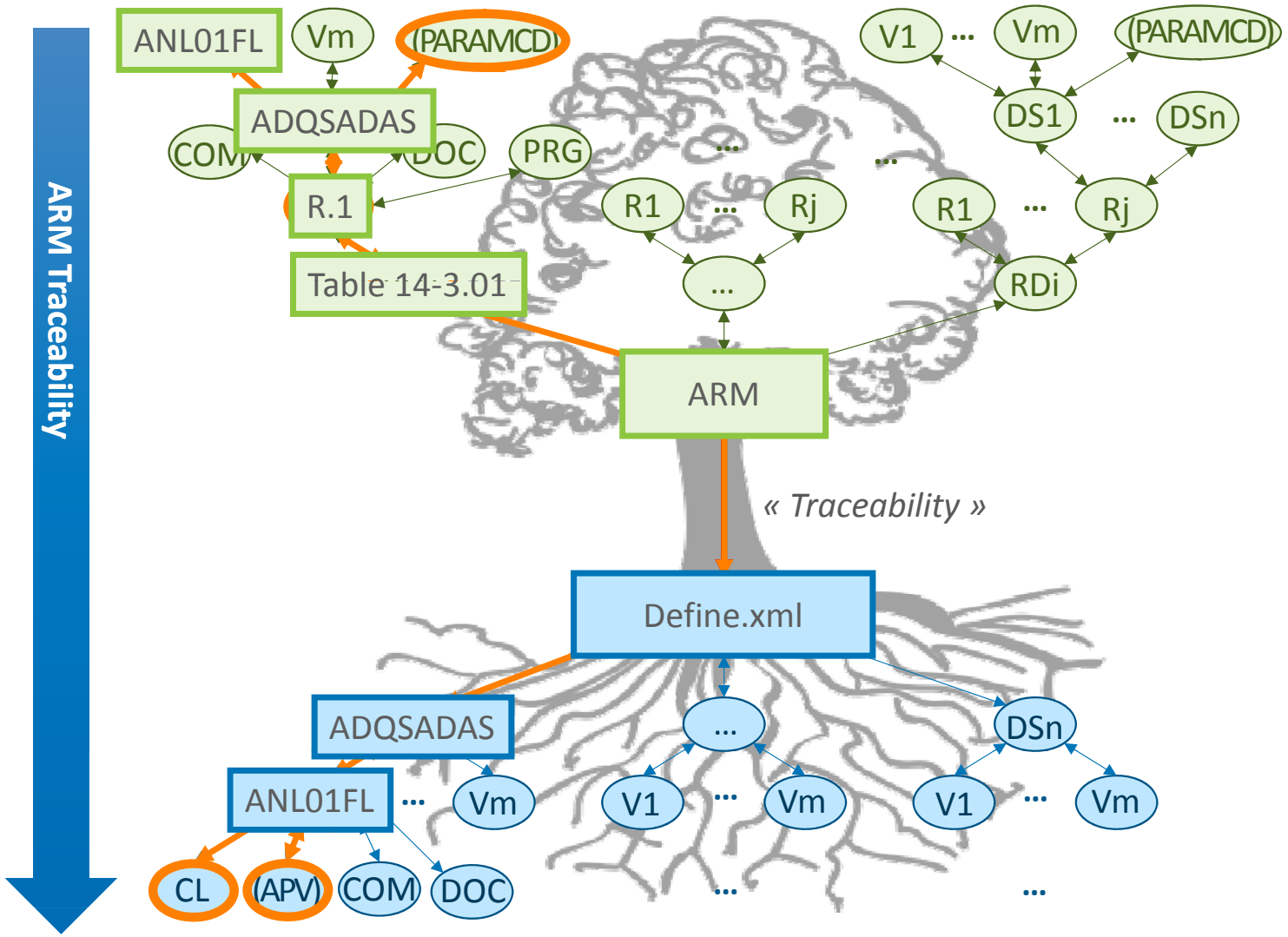


	STUDYID	USUBJID	SUBJID	BMI	BMIGR1	BMIGR1N	BMIGR2	BMIGR2N
2	9999-0001	9999-0001-000001	000001	27.777777778	<30 kg/m**2	1	25-<30 kg/m**2	2
3	9999-0001	9999-0001-000002	000002	25.503615702	<30 kg/m**2	1	25-<30 kg/m**2	2
4	9999-0001	9999-0001-000003	000003	26.175194521	<30 kg/m**2	1	25-<30 kg/m**2	2
5	9999-0001	9999-0001-000004	000004	35.15625	>=30 kg/m**2	2	>=30 kg/m**2	3
6	9999-0001	9999-0001-000005	000005	30.968858131	>=30 kg/m**2	2	>=30 kg/m**2	3
7	9999-0001	9999-0001-000006	000006	39.697163916	>=30 kg/m**2	2	>=30 kg/m**2	3
8	9999-0001	9999-0001-000007	000007	25.826446281	<30 kg/m**2	1	25-<30 kg/m**2	2
9	9999-0001	9999-0001-000008	000008	30.103806228	>=30 kg/m**2	2	>=30 kg/m**2	3
10	9999-0001	9999-0001-000009	000009	32.280962683	>=30 kg/m**2	2	>=30 kg/m**2	3
11	9999-0001	9999-0001-000010	000010	28.876133767	<30 kg/m**2	1	25-<30 kg/m**2	2
12	9999-0001	9999-0001-000011	000011	29.372397383	<30 kg/m**2	1	25-<30 kg/m**2	2
13	9999-0001	9999-0001-000012	000012	26.714852608	<30 kg/m**2	1	25-<30 kg/m**2	2
14	9999-0001	9999-0001-000013	000013	32.718613863	>=30 kg/m**2	2	>=30 kg/m**2	3
15	9999-0001	9999-0001-000014	000014	28.719723183	<30 kg/m**2	1	25-<30 kg/m**2	2
16	9999-0001	9999-0001-000015	000015	32.270420377	>=30 kg/m**2	2	>=30 kg/m**2	3

Branches + Leaves

Trunk

Roots



ARM *(Analysis Result Metadata)*
RD *(Result Display)*
R *(Result)*
DS *(DatatSet)*
V *(Variable)*
COM *(Comments)*
DOC *(Documents)*
APV *(Analysis Parameter Value)*
CL *(Codelist)*
PRG *(Program)*

Agenda

Analysis Results Metadata for Define-XML?

Purpose of Analysis Results Metadata (ARM)

Generation Method

Conclusions

Q&A

Generation Method:

« How to build ARM ? »

Generation Method: « How to build ARM ? »

- **Two Choices**
  - A. Direct Map to Define-XML v2.0 (1 Document)
    - High structure complexity
    - Time Consuming
  - B. Transform (2 Documents)
    1) « Human Readable » ARM Metadata Specifications + Study Metadata
    2) Transformation to Metadata Compatible with Define-XML v2.0
- **Advantages: « User Friendly »**
  - Lower structure complexity
  - Readable (less columns, less rows per result)
  - Time gain

Pratical Example (1)

Table 14-3.01 — Display Identifier - Document link & Page(s)/Section

Display	Table 14-3.01 **Primary Endpoint Analysis: ADAS-Cog - Summary at Week 24 - LOCF (Efficacy Population)** — Description
Analysis Result	**Dose response analysis for ADAS-Cog changes from baseline** — Result Description (one Display Identifier to many Result Display)
Analysis Parameter(s)	PARAMCD = "ACTOT" (Adas-Cog(11) Subscore) — Parameters
Analysis Variable(s)	CHG (Change from Baseline) — Analysis Variable
Analysis Reason	SPECIFIED IN SAP — Analysis Reason
Analysis Purpose	PRIMARY OUTCOME MEASURE — Analysis Purpose
Data References (incl. Selection Criteria)	ADQSADAS [PARAMCD = "ACTOT" and AVISIT = "Week 24" and EFFFL = "Y" and ANL01FL = "Y"] — Datasets & Selection Criteria Join Description
Documentation	Linear model analysis of CHG for dose response; using randomized dose (0 for placebo; 54 for low dose; 81 for high dose) and site group in model. Used PROC GLM in SAS to produce p-value (from Type III SS for treatment dose). SAP Section 10.1.1 — Documentation Description - Reference & Document link & Page(s)/Selection

Programming Statements	[SAS version 9.2]
	`proc glm data = ADQSADAS;`
	`  where EFFFL='Y' and ANL01FL='Y' and AVISIT='Week 24' and PARAMCD="ACTOT";`
	`  class SITEGR1;`
	`  model CHG = TRTPN SITEGR1;`
	`run;`
	— Programming Statement OR Program File
	+ Programming Language and Version

Practical Example (2)

Table 14-3.01 « one-to-many »

Display	Table 14-3.01 Primary Endpoint Analysis: ADAS-Cog - Summary at Week 24 - LOCF (Efficacy Population)
Analysis Result	Dose response analysis for ADAS-Cog changes from baseline
Analysis Parameter(s)	PARAMCD = "ACTOT" (Adas-Cog(11) Subscore)
Analysis Variable(s)	CHG (Change from Baseline)
Analysis Reason	SPECIFIED IN SAP
Analysis Purpose	PRIMARY OUTCOME MEASURE

In selection Criteria

DisplayIdentifier	xlinkDocument	Page(s)/Section	Description	ResultDescription	Analysis Reason	Analysis Purpose	Dataset(s)	Analysis Variable
Table 14-3.01	csr.pdf	1	Primary Endpoint Analysis: ADAS-Cog - Summary at Week 24 - LOCF (Efficacy Population)	Dose response analysis for ADAS-Cog changes from baseline	SPECIFIED IN SAP	PRIMARY OUTCOME MEASURE	ADQSADAS	CHG

Statements

```
proc glm data = ADQSADAS;
  where EFFFL='Y' and ANL01FL='Y' and AVISIT='Week 24' and PARAMCD="ACTOT";
  class SITEGR1;
  model CHG = TRTPN SITEGR1;
run;
```

Practical Example (3)

Table 14-3.01

Display	Table 14-3.01 Primary Endpoint Analysis: ADAS-Cog - Summary at Week 24 - LOCF (Efficacy Population)						
Analysi	DisplayIdentifier	Join Description	Selection Criteria	Documentation	DOC-REF	Page(s)/Section	xlinkDOC-REF
Analysis							
Analysis				Linear model analysis of CHG for dose response; using randomized dose (0 for placebo; 54 for low dose; 81 for high dose) and site group in model. Used PROC GLM in SAS to produce p-value (from Type III SS for treatment dose).			
Analysis			ADQSADAS [PARAMCD="ACTOT" and AVISIT="Week 24" and EFFFL="Y" and ANL01FL="Y"]		SAP Section 10.1.1	42	sap.pdf
Analysis	Table 14-3.01						

Data References (incl. Selection Criteria)	ADQSADAS [PARAMCD = "ACTOT" and AVISIT = "Week 24" and EFFFL = "Y" and ANL01FL = "Y"]
Documentation	Linear model analysis of CHG for dose response; using randomized dose (0 for placebo; 54 for low dose; 81 for high dose) and site group in model. Used PROC GLM in SAS to produce p-value (from Type III SS for treatment dose). SAP Section 10.1.1

Programming Statements	[SAS version 9.2]
	proc glm data = ADQSADAS;
	where EFFFL='Y' and ANL01FL='Y' and AVISIT='Week 24' and PARAMCD="ACTOT";
	class SITEGR1;
	model CHG = TRTPN SITEGR1;
	run;

Practical Example (4)

Table 14-3.01

Display	Table 14-3.01 Primary Endpoint Analysis: ADAS-Cog - Summary at Week 24 - LOCF (Efficacy Population)
Analysis Result	**Dose response analysis for ADAS-Cog changes from baseline**
Analysis Parameter(s)	PARAMCD = "ACTOT" (Adas-Cog(11) Subscore)
Analysis Variable(s)	CHG (Change from Baseline)
Analysis Reason	
Analysis Purpose	
Data References / Selection Criteria	
Documentation	

DisplayIdentifier	Programming Statement	Programming Language and Version	Program

Table 14-3.01	proc glm data = ADQSADAS;
	where EFFFL='Y' and ANL01FL='Y' and AVISIT='Week 24' and PARAMCD="ACTOT";
	class SITEGR1;
	model CHG = TRTPN SITEGR1;
	run;
		SAS version 9.2	

Programming Statements	[SAS version 9.2]

```
proc glm data = ADQSADAS;
  where EFFFL='Y' and ANL01FL='Y' and AVISIT='Week 24' and PARAMCD="ACTOT";
  class SITEGR1;
  model CHG = TRTPN SITEGR1;
run;
```

Pratical Example (5)

« Human friendly » ARM Specifications:

1 Result – 1 Row – 1 « group » of Selection Criteria per dataset – 19 Columns: i.e. 1 group of selection criteria, 1 dataset

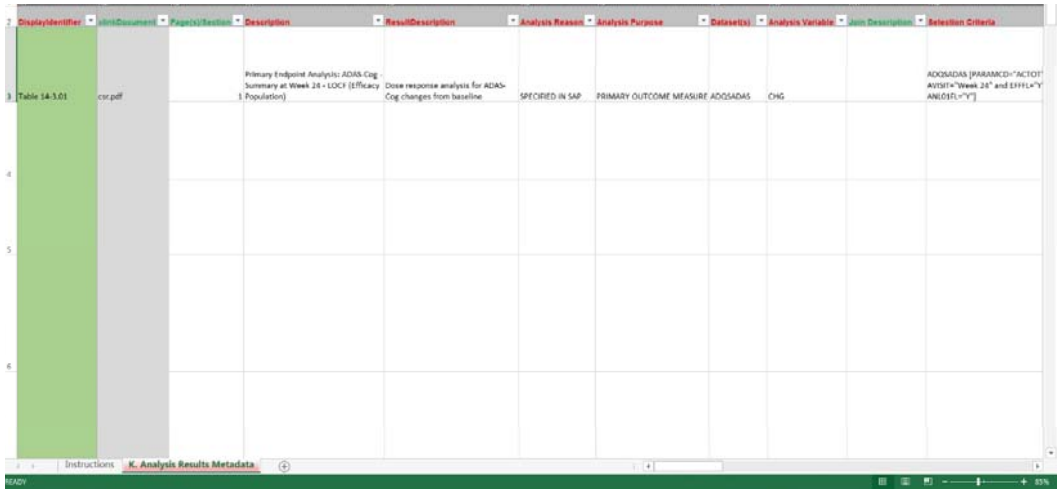
DisplayIdentifier	xlinkDocument	Page(s)/Section	Description	ResultDescription	Analysis Reason	Analysis Purpose	Dataset(s)	Analysis Variable	Join Description	Selection Criteria	Documentation
Table 14-3.01	csr.pdf	1	Primary Endpoint Analysis: ADAS-Cog - Summary at Week 24 - LOCF (Efficacy Population)	Dose response analysis for ADAS-Cog changes from baseline	SPECIFIED IN SAP	PRIMARY OUTCOME MEASURE	ADQSADAS	CHG		ADQSADAS (PARAMCD="ACTOT" and AVISIT="Week 24" and EFFFL="Y" and ANL01FL="Y")	Linear model analysis of CHG fo using randomized dose (0 for pl dose; 81 for high dose) and site Used PROC GLM in SAS to pro (from Type III SS for treatment d

ARM Specifications Define-XML 2.0:

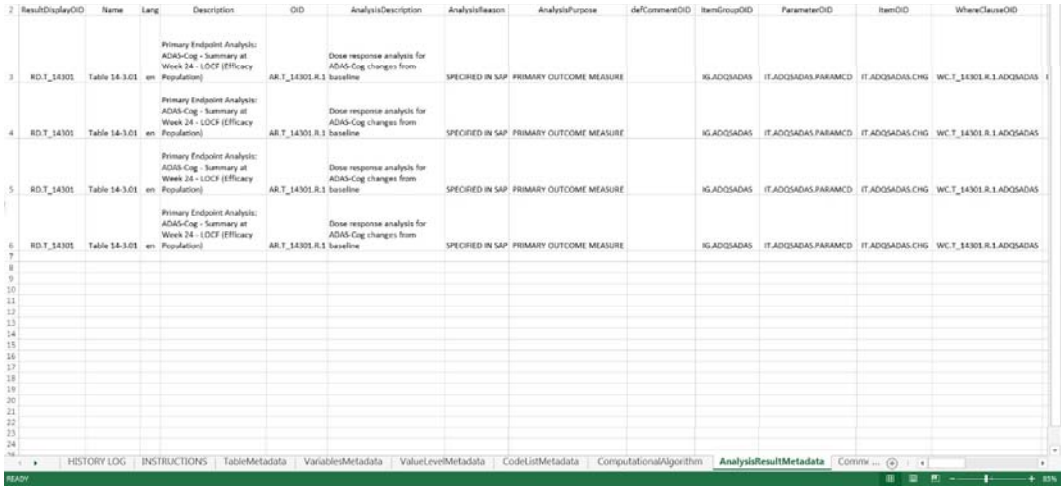
1 Result – n Rows per group of Selection Criteria per dataset – 32 Columns: i.e. 4 rows for 1 group of selection criteria for 1 dataset

ResultDisplayOID	Name	Lang	Description	OID	AnalysisDescription	AnalysisReason	AnalysisPurpose	defItemOID	Comparator	CheckValue	Soft/Hard	ocumentsLeaf	Type	PageRefs	FirstPage	LastPage	Docu
RD.T_14301	Table 14-3.01	en	Primary Endpoint Analysis: ADAS-Cog - Summary at Week 24 - LOCF (Efficacy Population)	AR.T_14301.R.1	Dose response analysis for ADAS-Cog changes from baseline	SPECIFIED IN SAP	PRIMARY OUTCOME MEASURE	IT.ADQSADAS.PARAMCD	EQ	ACTOT	Soft	LF.T14301	PhysicalRef	1			LF.SA
RD.T_14301	Table 14-3.01	en	Primary Endpoint Analysis: ADAS-Cog - Summary at Week 24 - LOCF (Efficacy Population)	AR.T_14301.R.1	Dose response analysis for ADAS-Cog changes from baseline	SPECIFIED IN SAP	PRIMARY OUTCOME MEASURE	IT.ADQSADAS.AVISIT	EQ	Week 24	Soft	LF.T14301	PhysicalRef	1			LF.SA
RD.T_14301	Table 14-3.01	en	Primary Endpoint Analysis: ADAS-Cog - Summary at Week 24 - LOCF (Efficacy Population)	AR.T_14301.R.1	Dose response analysis for ADAS-Cog changes from baseline	SPECIFIED IN SAP	PRIMARY OUTCOME MEASURE	IT.ADQSADAS.EFFFL	EQ	Y	Soft	LF.T14301	PhysicalRef	1			LF.SA
RD.T_14301	Table 14-3.01	en	Primary Endpoint Analysis: ADAS-Cog - Summary at Week 24 - LOCF (Efficacy Population)	AR.T_14301.R.1	Dose response analysis for ADAS-Cog changes from baseline	SPECIFIED IN SAP	PRIMARY OUTCOME MEASURE	IT.ADQSADAS.ANL01FL	EQ	Y	Soft	LF.T14301	PhysicalRef	1			LF.SA

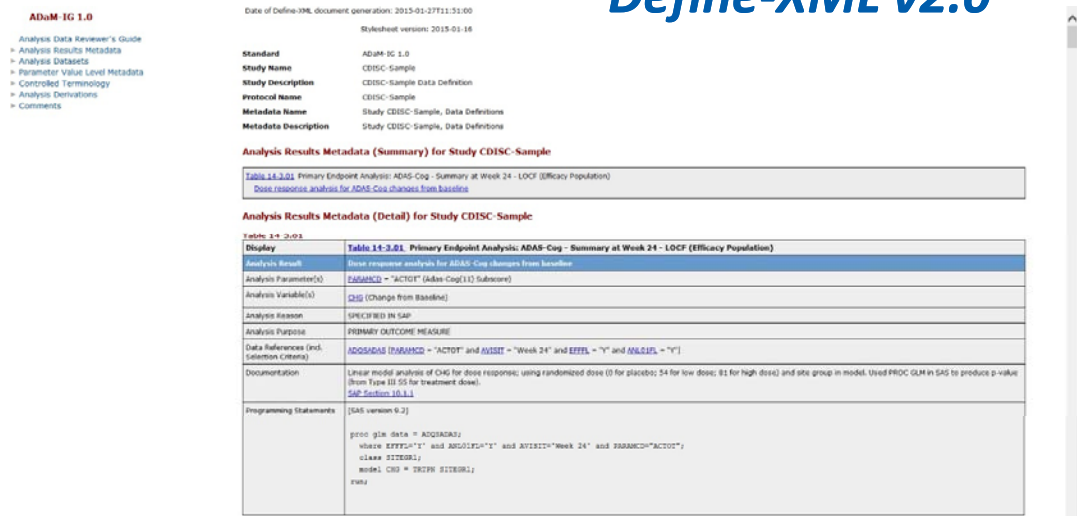
ADaM Metadata Specifications + ARM Specifications



ADaM Metadata Specifications Reworked With Metadata Specifications (included)



Define-XML v2.0



Business & Decision Life Sciences

Date of Define-XML document generation: 2015-01-27T11:51:00

Stylesheet version: 2015-01-16

Standard	ADaM-IG 1.0
Study Name	CDISC-Sample
Study Description	CDISC-Sample Data Definition
Protocol Name	CDISC-Sample
Metadata Name	Study CDISC-Sample, Data Definitions
Metadata Description	Study CDISC-Sample, Data Definitions

Analysis Results Metadata (Summary) for Study CDISC-Sample

Table 14-3.01 Primary Endpoint Analysis: ADAS-Cog - Summary at Week 24 - LOCF (Efficacy Population)
Dose response analysis for ADAS-Cog changes from baseline

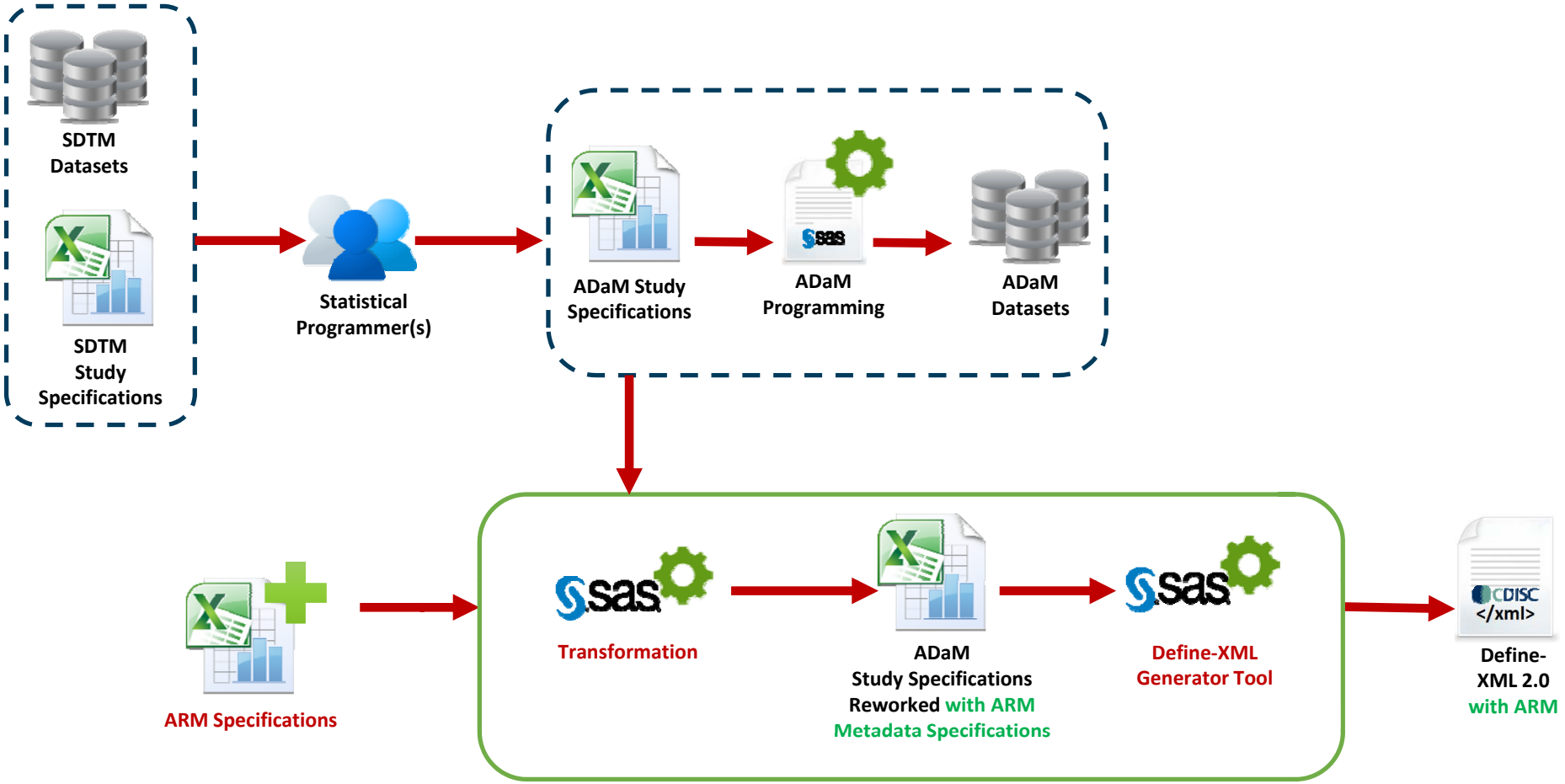
Analysis Results Metadata (Detail) for Study CDISC-Sample

Table 14-3.01

Display	**Table 14-3.01 Primary Endpoint Analysis: ADAS-Cog - Summary at Week 24 - LOCF (Efficacy Population)**
Analysis Result	**Dose response analysis for ADAS-Cog changes from baseline**
Analysis Parameter(s)	PARAMCD = "ACTOT" (Adas-Cog(11) Subscore)
Analysis Variable(s)	CHG (Change from Baseline)
Analysis Reason	SPECIFIED IN SAP
Analysis Purpose	PRIMARY OUTCOME MEASURE
Data References (incl. Selection Criteria)	ADQSADAS [PARAMCD = "ACTOT" and AVISIT = "Week 24" and EFFFL = "Y" and ANL01FL = "Y"]
Documentation	Linear model analysis of CHG for dose response; using randomized dose (0 for placebo; 54 for low dose; 81 for high dose) and site group in model. Used PROC GLM in SAS to produce p-value (from Type III SS for treatment dose). SAP Section 10.1.1

Programming Statements	[SAS version 9.2]
	
	`proc glm data = ADQSADAS;`
	`  where EFFFL='Y' and ANL01FL='Y' and AVISIT='Week 24' and PARAMCD="ACTOT";`
	`  class SITEGR1;`
	`  model CHG = TRTPN SITEGR1;`
	`run;`

Process Overview



Generation Method

- **Challenges**
- Traceability
- Completion
- Display

Loss of Traceability (1)

- **Root missing !**
  - Selection Criteria
  - Analysis Variable(s)
  - Analysis Parameter(s)

i.e. ADQSADAS [PARAMCD="ACTOT" and AVISIT="Week 24" and EFFFL="Y" and ANL01FL="Y"]

ADAS-Cog Analysis (ADQSADAS) [Location: adqsadas.xpt]

Variable	Label	Key	Type	Length / Display Format	Controlled Terms or Format	Source/Derivation/Comment
STUDYID	Study Identifier	1	text	12		Predecessor: ADSL.STUDYID
SITEID	Study Site Identifier		text	3		Predecessor: ADSL.SITEID
USUBJID	Unique Subject Identifier	2	text	11		Predecessor: ADSL.USUBJID
TRTP	Planned Treatment		text	20	["Placebo", "Xanomeline Low Dose", "Xanomeline High Dose"] <Actual Treatment>	Predecessor: ADSL.TRT01P
TRTPN	Planned Treatment (N)		integer	8	[0 = "Placebo", 54 = "Xanomeline Low Dose", 81 = "Xanomeline High Dose"] <Actual Treatment (N)>	Predecessor: ADSL.TRT01PN
AGE	Age		integer	8		Predecessor: ADSL.AGE
RACE	Race		text	32	["WHITE", "BLACK OR AFRICAN AMERICAN", "ASIAN", "AMERICAN INDIAN OR ALASKA NATIVE"] <Race>	Predecessor: ADSL.RACE
RACEN	Race (N)		integer	8	[1 = "WHITE", 2 = "BLACK OR AFRICAN AMERICAN", 6 = "AMERICAN INDIAN OR ALASKA NATIVE", 7 = "ASIAN"] <Race (N)>	Predecessor: ADSL.RACEN
SEX	Sex		text	1	["F" = "Female", "M" = "Male", "U" = "Unknown"] <Sex>	Predecessor: ADSL.SEX
EFFFL	Efficacy Population Flag		text	1	["N" = "No", "Y" = "Yes"] <No Yes Response>	Predecessor: ADSL.FASFL
AVISIT	Analysis Visit	4	text	16	["Baseline", "Week 8", "Week 16", "Week 24"] <Analysis Visit>	Derived:
AVISITN	Analysis Visit (N)		integer	8	[0 = "Baseline", 8 = "Week 8", 16 = "Week 16", 24 = "Week 24"] <Analysis Visit (N)>	Assigned: Numeric code for AVISIT
VISIT	Visit Name		text	19	Visit	Predecessor: QS.VISIT
VISITNUM	Visit Number		float	8	Visit Number	Predecessor: QS.VISITNUM
ADY	Analysis Relative Day		integer	8		Derived:
ADT	Analysis Date	5	integer	date9.		Derived:
PARAM	Parameter		text	100	ADAS-Cog Parameter	Assigned:
PARAMCD	Parameter Code	3	text	8	ADAS-Cog Parameter Code	Assigned:
AVAL	Analysis Value		integer	8		
BASE	Baseline Value		integer	8		Derived:
CHG	Change from Baseline		integer	8		Derived:
PCHG	Percent Change from Baseline		integer	8		Derived:
ABLFL	Baseline Record Flag		text	1	["Y" = "Yes"] <No Yes Response - Y subset>	Predecessor: QS.QSBLFL

Loss of Traceability (2)

Analysis Results Metadata (Detail) for Study CDISC-Sample

Table 14-3.01

Display	**Table 14-3.01 Primary Endpoint Analysis: ADAS-Cog - Summary at Week 24 - LOCF (Efficacy Population)**
Analysis Result	**Dose response analysis for ADAS-Cog changes from baseline**
Analysis Parameter(s)	PARAMCD = "ACTOT" (Adas-Cog(11) Subscore)
Analysis Variable(s)	CHG (Change from Baseline)
Analysis Reason	SPECIFIED IN SAP
Analysis Purpose	PRIMARY OUTCOME MEASURE
Data References (incl. Selection Criteria)	ADQSADAS [PARAMCD = "ACTOT" and AVISIT = "Week 24" and EFFFL = "Y" and =] **= Missing Root**
Documentation	Linear model analysis of CHG for dose response; using randomized dose (0 for placebo; 54 for low dose; 81 for high dose) and site group in model. Used PROC GLM in SAS to produce p-value (from Type III SS for treatment dose). SAP Section 10.1.1

Programming Statements	[SAS version 9.2]
	
	`proc glm data = ADQSADAS;`
	`  where EFFFL='Y' and ANL01FL='Y' and AVISIT='Week 24' and PARAMCD="ACTOT";`
	`  class SITEGR1;`
	`  model CHG = TRTPN SITEGR1;`
	`run;`

Completion

- ***Selection Criteria: Expressions not allowed***
  - *i.e. 0 < AVAL < 25*

    ⇒AVAL > 0 and AVAL < 25

  - *i.e. (AVISIT IN ('Week 12' 'Week 24')) as AVISIT*

    ⇒AVISIT IN ('Week 12' 'Week 24')

⇒ Must remain simple

- ***Selection Criteria: Not Existing variables or datasets***
- ***Programming statements: Length, indentation, spaces, etc.***

Display (1)

Programming Statements 1

[SAS version 9.2]

```
proc glm data = ADQSADAS;
  where EFFFL='Y' and ANL01FL='Y' and AVISIT='Week 24' and PARAMCD="ACTOT";
  class SITEGR1;
  model CHG = TRTPN SITEGR1;
run;
```

Programming Statements 2

[SAS version 9.2]

```
proc glm data = ADQSADAS;              where EFFFL='Y' and ANL01FL='Y' and AVISIT='Week 24' and PARAMCD="ACTOT";                    class SITEGR1;
```

Display (2)

Solution in Stylesheet (xsl): Blank Space

Before

```
.code{
 font-family:"Courier New", monospace, serif;
 font-size:1.2em;
 line-height:150%;
 white-space:pre;
 display:block;
 vertical-align:top;
 padding:5px;
}
```

After

```
.code{
 font-family:"Courier New", monospace, serif;
 font-size:1.2em;
 line-height:100%;
 white-space:pre-wrap;
 display:block;
 vertical-align:top;
 padding:5px;
}
```

Display (2)

Solution in Stylesheet (xsl): Blank Space

Programming Statements	[SAS version 9.2]

```
proc glm data = ADQSADAS;                where EFFFL='Y' and ANL01FL='Y' and AVISIT='Week 24' and PARAMCD="ACTOT";
        class SITEGR1;                          model CHG = TRTPN SITEGR1; run;
```

Agenda

Analysis Results Metadata for Define-XML?

Purpose of Analysis Results Metadata (ARM)

Generation Method

Conclusions

Q&A

Conclusions

- ✓ **Study Traceability** Improvement
  - ✓ Data Transparency and Understanding
- ✓ Added Value **for Sponsors**
  - ✓ Easier Review
- ✓ « Reusable » **in similar studies**
- ✓ **Method:** Time gain

Questions...

ROI²
return on investment
through
return on information

Business & Decision

Life Sciences

Lionel Debecq

Statistical Programmer
lionel.debecq@businessdecision.be
+32 478 52 19 91